



Ingenieurfacultät Bau Geo Umwelt
Signalverarbeitung in der Erdbeobachtung

From TomoSAR Point Clouds To Objects

Muhammad Shahzad, M.Sc.

Vollständiger Abdruck
der von der Ingenieurfacultät Bau Geo Umwelt
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigten Dissertation.

Vorsitzender:

Univ.-Prof. Dr.-Ing. Uwe Stilla

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. habil. Xiao Xiang Zhu
2. Univ.-Prof. Dr.-Ing. habil. Richard Bamler
3. Univ.-Prof. Dr. Bernhard Höfle
Ruprecht-Karls-Universität Heidelberg

Die Dissertation wurde am 16.12.2015 bei der Technischen Universität München eingereicht und durch die Ingenieurfacultät Bau Geo Umwelt am 08.02.2016 angenommen.

Abstract

Synthetic aperture radar (SAR) projects a 3-D scene onto two native coordinates i.e., “range” and “azimuth”. In order to fully localize a point in 3-D, advanced interferometric SAR (InSAR) techniques are required that process stack(s) of complex-valued SAR images to retrieve information in the third dimension “elevation”. Among advanced InSAR methods, SAR tomography (TomoSAR) is the ultimate way of 3-D SAR imaging. By exploiting stack(s) of SAR images taken from slightly different positions, it builds up a synthetic aperture in elevation that enables retrieval of 3-D position of dominant scatterers within each azimuth-range SAR pixel. Moreover, considering the fact that acquisitions are taken at different times, the synthetic aperture principle can also be extended to the temporal domain which enables 4-D (space-time) SAR imaging. The technique is thus capable of retrieving elevation and deformation information (linear, seasonal, etc.) even of multiple scatterers inside a single SAR pixel. Along with very high resolution (VHR) data provided by modern SAR satellites (including TerraSAR-X, TanDEM-X and CosmoSkyMED), TomoSAR enables the generation of 4-D (space-time) point clouds of illuminated objects with a point (scatterer) density comparable to LiDAR. E.g., experiments using TerraSAR-X high-resolution spotlight data stacks show that the scatterer density retrieved using TomoSAR is on the order of 1 million points/km². Object reconstruction from these TomoSAR point clouds can greatly support the reconstruction of dynamic city models that could potentially be used to monitor and visualize the dynamics (i.e., long-term deformation in the mm- and cm-range, e.g. subsidence/uplift caused by earthquakes, bad construction, seasonal changes etc.) of urban infrastructures in very high level of details. Motivated by these opportunities, the goal of this thesis is to explore for the first time the potential of object reconstruction from this class of point cloud data. To this end, the work presented in this thesis contributes to the field of SAR based object reconstruction by addressing the following four new aspects:

Reconstruction of building façades: Complex multiple scattering and different microwave scattering properties of the objects in the scene which possess different geometrical and material features render TomoSAR point clouds to exhibit some special characteristics such as low positioning accuracy, high number of outliers, gaps in the data and rich façade information due to the side looking geometry. Taking into consideration these associated properties, in particular the rich façade information, a robust procedure has been developed and presented in this thesis that exploits advanced robust estimation methods to reconstruct 2-D/3-D building shape models (or footprints with extruded heights) via façade point analysis. The basic idea is to extract façade points via point density analysis, segment them into clusters of points belonging to individual façades, and finally reconstruct 3-D building models via independent modeling of each individual façade to reconstruct the overall 2-D building footprints, followed by its representation in 3-D using height information. The developed approach is illustrated and validated by examples using TomoSAR point clouds generated from a stack of TerraSAR-X high-resolution spotlight images from ascending orbit covering an approximately 2 km² high-rise area in the city of Las Vegas.

Reconstruction of building footprints from roof points: The aforementioned *façade-only* based approach is a good option to detect the shapes of high-rise buildings because many points on their façades are available. However there are cases occurring especially for lower buildings where no or only very few scatterers are available in the data. Also, since the satellite orbits are bound to pass close to the poles of the Earth, such an approach may fail to reconstruct buildings whose façades face North or South, due to the missing measurements. Moreover, due to the side looking geometry, multiple views are required to reconstruct the complete building shapes. To cope with these issues, a novel framework is developed that encompasses the façade reconstruction approach and additionally incorporates the available roof points to determine the complete shape/footprint of buildings. The developed framework

is modular and completely data driven with no imposition of any restrictions on building shape (i.e., any arbitrarily shaped footprint could be reconstructed). This as a consequent allows for a robust reconstruction of both tall and low buildings within the area of interest, and is therefore well suited for urban monitoring of larger areas from space. The developed approach is illustrated and validated over different large area scenes containing taller and moderate sized buildings in the cities of Las Vegas, USA and Berlin, Germany.

Innovative applications: To explore the possible applications of the reconstructed 2-D/3-D building shape models, three innovative applications have been demonstrated in this thesis: 1) A first 4-D building model is presented; 2) Model based elevation refinement is carried out to refine the elevation position of each raw TomoSAR point by using its more accurate azimuth and range coordinates together with the corresponding reconstructed model; and 3) A concept that utilizes the reconstructed model to improve the TomoSAR inversion in a feedback iterative manner (i.e., to use reconstructed footprints as iterative priors) is presented.

Reconstruction of individual trees: Besides man-made structures in particular buildings, trees (or vegetation) are also of key interest when it comes to city modeling. Spaceborne repeat-pass TomoSAR data however lacks the ability to reconstruct temporarily incoherent objects (i.e., no points are available over vegetation or trees). In this regard, a model based individual tree reconstruction methodology has been developed and validated using a 3-D TomoSAR point cloud generated from multi-baseline millimeterwave InSAR data acquired from multi-aspects with an airborne single-pass MEMPHIS SAR system. This leads to the first reconstruction of individual trees using millimeterwave SAR image stacks.

Since these point clouds are very recent and have not been exploited yet, I hope that the presented methods may not only take SAR based object reconstruction a step further but also substantially contribute to the vision of developing dynamic city models. Such models may help in monitoring and visualization of the dynamics of urban objects in very high level of details.

Zusammenfassung

Synthetisches Apertur Radar (SAR) projiziert eine 3-D-Szene auf die zwei nativen Radarkoordinaten "Range" und "Azimuth". Um einen Punkt in 3-D zu bestimmen sind fortgeschrittene interferometrische SAR (InSAR) Techniken vonnöten, die Stapel von komplexwertigen SAR-Bildern verarbeiten um Information über die dritte Dimension "Elevation" zu gewinnen. Unter den fortgeschrittenen InSAR-Methoden ist SAR-Tomographie (TomoSAR) der aktuelle Stand der Technik für 3-D SAR-Bildgebung. Stapel von SAR-Bildern, die von geringfügig unterschiedlichen Positionen aufgenommen wurden, können dazu benutzt werden eine synthetische Apertur in Elevationsrichtung zu bilden, die es ermöglicht die 3-D-Position von dominierenden Streuern in jedem Azimuth-Range-Pixel zu bestimmen. Des Weiteren, unter der Berücksichtigung der Tatsache, dass die Aufnahmen zu unterschiedlichen Zeitpunkten gemacht wurden, kann das synthetische Aperturprinzip auch um die zeitliche Domäne erweitert werden und ermöglicht dadurch 4-D (Raum-Zeit) SAR-Bildgebung. Die Technik ist dadurch in der Lage die Elevation- und Deformationsinformation (linear, saisonal, etc.) von sogar mehreren Streuern in einem einzigen SAR-Pixel zu bestimmen. Zusammen mit Daten mit sehr hoher Auflösung (VHR für very high resolution), welche von modernen SAR-Satelliten (unter anderem TerraSAR-X, TanDEM-X und CosmoSkyMED) geliefert werden, ermöglicht TomoSAR die Erstellung von 4-D (Raum-Zeit) Punktwolken der ausgeleuchteten Objekte mit einer Punkt- bzw. Streuerdichte, die mit der von LiDAR vergleichbar ist. Zum Beispiel zeigen Experimente mit hochaufgelösten TerraSAR-X spotlight Datenstapeln, dass die mittels TomoSAR erzeugte Streuerdichte in der Größenordnung von 1 Millionen Punkte/km² liegt. Die Objektrekonstruktion von diesen TomoSAR-Punktwolken kann die Rekonstruktion von dynamischen Stadtmodellen beträchtlich unterstützen, welche dann zur Überwachung und Visualisierung dynamischer Prozesse (z.B. langfristige Deformation in der Größenordnung von mm und cm, wie etwa die durch Erdbeben, mangelhafte Baustellen, Jahreszeiten, etc. verursachten Absenk- oder Anhebungen.) urbaner Infrastruktur mit einem hohen Maß an Details genutzt werden könnten. Motiviert durch diese Möglichkeiten ist das Ziel dieser Dissertation zum ersten Mal das Potential zu untersuchen, das die Objektrekonstruktion für diese Klasse von Punktwolkendaten bietet. Zu diesem Zweck trägt die Arbeit, die in dieser Dissertation präsentiert wurde, durch die folgenden vier neuen Aspekte zum Forschungsgebiet der SAR-basierten Objektrekonstruktion bei:

Rekonstruktion von Gebäudefassaden: Komplexwertiges, multiples Streuen und unterschiedliche Mikrowellenstreueigenschaften von Objekten in einer Szene, welche unterschiedliche geometrische und Materialmerkmale haben, führen dazu, dass TomoSAR-Punktwolken spezielle Charakteristiken aufweisen, wie zum Beispiel eine niedrige Ortsgenauigkeit, eine große Anzahl von Ausreißern, lückenhafte Daten und eine Fülle an Fassadeninformation, aufgrund der seitlichen Beobachtungsgeometrie. Im Anbetracht dieser zugehörigen Eigenschaften, insbesondere der Fülle an Fassadeninformation, wurde ein robustes Verfahren in dieser Dissertation entwickelt und präsentiert, das fortschrittliche robuste Schätzverfahren ausnutzt um 2-D/3-D Gebäudeumrisse (oder Grundrisse mit extrudierten Höhen) mittels Fassadenpunktanalyse zu rekonstruieren. Die grundlegende Idee ist Fassadenpunkte via einer Punktdichteanalyse zu extrahieren, diese in Clustern von Punkten zu segmentieren, die zu einzelnen Fassaden gehören, und letztendlich 3-D Gebäudemodelle mittels unabhängig Modellierung der einzelnen Fassaden zu rekonstruieren um den gesamten 2-D Gebäudegrundriss zu rekonstruieren, gefolgt von deren 3-D-Representation durch eine Höhentransformation. Der entwickelte Ansatz wird anhand von Beispielen mit TomoSAR-Punktwolken veranschaulicht und validiert, welche von einem Stapel von TerraSAR-X high-resolution spotlight Bildern generiert wurden, die aus einem aufsteigenden Orbit aufgenommen wurden und eine ungefähr 2 km² große Fläche mit

mehreren Hochhäuser in Las Vegas zeigen.

Rekonstruktion von Gebäudegrundrissen mit Dachpunkten: Der zuvor erwähnte Ansatz, der nur auf Fassaden fußt, ist eine gute Wahl um den Umriss von Hochhäusern zu detektieren, da viele Punkte ihrer Fassaden verfügbar sind. Allerdings gibt es auch Fälle, die vor allem bei niedrigeren Gebäuden auftreten, wo nur sehr wenige Streuer in den Daten vorhanden sind. Außerdem, da Satellitenorbits nahe den Polen der Erde verlaufen, könnte ein solcher Ansatz aufgrund der fehlenden Messungen bei der Rekonstruktion von Gebäuden, deren Fassaden nach Norden oder Süden gewandt sind, fehlschlagen. Darüber hinaus sind aufgrund der seitlichen Beobachtungsgeometrie mehrere Aufnahmen aus mehreren Richtungen erforderlich um komplette Gebäudeumrisse zu rekonstruieren. Um diese Probleme anzugehen wurde ein neues Framework entwickelt, das sowohl die Fassadenrekonstruktion als auch zusätzlich die verfügbaren Dachpunkte berücksichtigt um den kompletten Gebäudegrundriss zu bestimmen. Das entwickelte Framework ist modular und komplett datengesteuert und stellt keine Voraussetzungen oder Einschränkungen an den Gebäudeumriss (das heißt jeder beliebige Gebäudeumriss kann rekonstruiert werden). Dies führt dazu, dass eine robuste Rekonstruktion von sowohl hohen als auch niedrigen Gebäuden im Interessenbereich möglich ist, was sich daher gut für die Überwachung großer urbaner Gebiete aus dem Weltraum eignet. Der entwickelte Ansatz wurde anhand verschiedenen großflächigen Gebieten in Las Vegas, USA und Berlin, Deutschland, die hohe und mittelgroße Gebäude beinhalten, aufgezeigt und validiert.

Innovative Anwendungen: Um mögliche Anwendungen der rekonstruierten 2-D/3-D Gebäudeumrissmodelle zu untersuchen wurden drei innovative Anwendungen in dieser Dissertation demonstriert: 1) Ein erstes 4-D Gebäudemodell wurde präsentiert; 2) Eine modellbasierte Elevationsverbesserung wurde durchgeführt um die Elevationsposition von jedem unbearbeiteten TomoSAR-Punkt zu verbessern indem die genaueren Azimuth- und Rangekoordinaten zusammen mit dem korrespondierenden, rekonstruierten Modell benutzt werden; und 3) ein Konzept, das das rekonstruierte Modell benutzt um die TomoSAR-Inversion durch eine iterative Feedback-Methode verbessert (durch Benutzung des Grundrisses als a-priori Information für die Iterationen) wurde präsentiert.

Rekonstruktion einzelner Bäume: Neben Strukturen menschlichen Ursprungs, insbesondere Gebäuden, sind auch Bäume (oder Vegetation) für Stadtmodellierungen von großem Interesse. Allerdings können satellitengestützte repeat-pass TomoSAR-Daten nicht dazu benutzt werden zeitlich inkohärente Objekte (es sind keine Punkte für Vegetation und Bäume verfügbar) zu rekonstruieren. Mit Bezug darauf wurde eine modellbasierte Methode zur Rekonstruktion einzelner Bäume aus einer 3-D TomoSAR-Punktwolke entwickelt und validiert. Die benutzten multi-baseline Millimeterwellen-InSAR-Daten wurden mit dem luftgestützten single-pass MEMPHIS SAR-System aufgenommen. Dies führte zur ersten Rekonstruktion einzelner Bäume mit Millimeterwellen-SAR-Bilderstapeln.

Da diese Art von Punktwolken sehr neu sind und bis jetzt nicht genutzt wurden hoffe ich, dass die vorgestellten Methoden nicht nur die SAR-basierte Objektdetektion einen Schritt voranbringen, sondern auch wesentlich zur Vision dynamische Stadtmodelle zu entwickeln beitragen. Solche Modelle können helfen die dynamischen Prozesse urbaner Objekte sehr detailgenau zu überwachen und zu visualisieren.

Contents

Abstract	i
Zusammenfassung	iii
1 Introduction	1
1.1 Motivation and scientific relevance	1
1.2 Objectives and focus	2
1.2.1 <i>General goal</i>	2
1.2.2 <i>Methodological goals</i>	2
1.3 Reader's guide	3
2 SAR object reconstruction – Basics and state-of-the-art	5
2.1 SAR Imaging Fundamentals	5
2.2 Radiometric and geometrical effects of SAR imaging	6
2.3 Advanced SAR imaging	7
2.3.1 <i>SAR polarimetry</i>	7
2.3.2 <i>SAR interferometry (InSAR)</i>	8
2.3.2.1 <i>Differential InSAR</i>	8
2.3.2.2 <i>Phase-to-height conversion</i>	9
2.3.3 <i>SAR tomography (TomoSAR)</i>	9
2.4 Object reconstruction using SAR	10
2.4.1 <i>Mapping of buildings in VHR SAR images</i>	11
2.4.1.1 <i>Appearance and characteristics of buildings in SAR/InSAR</i>	11
2.4.1.2 <i>Buildings detection using single VHR SAR image</i>	12
2.4.1.3 <i>Height estimation and reconstruction</i>	13
2.4.1.4 <i>Building reconstruction using InSAR data</i>	15
2.4.1.5 <i>Combining SAR/InSAR with auxiliary information</i>	16
2.4.1.6 <i>Advanced InSAR (TomoSAR)</i>	17
2.4.2 <i>Reconstruction of Individual trees in VHR SAR</i>	17
2.4.2.1 <i>Appearance and challenges</i>	17
2.4.2.2 <i>Few recognition/reconstruction approaches</i>	18
3 Related work – Object reconstruction using 3-D (ALS) point clouds	19
3.1 3-D building reconstruction	19
3.1.1 <i>Buildings detection (or extraction)</i>	19
3.1.1.1 <i>Ground filtering techniques</i>	19
3.1.1.2 <i>Extraction of building points/pixels</i>	21
3.1.2 <i>Building boundary/footprint reconstruction</i>	26
3.1.2.1 <i>Rough footprint generation</i>	26
3.1.2.2 <i>Refinement or polygon simplification</i>	27
3.1.2.3 <i>Boundary regularization</i>	28
3.1.3 <i>Roof modeling/reconstruction</i>	29
3.1.3.1 <i>Model-driven approaches</i>	31
3.1.3.2 <i>Data-driven approaches</i>	31
3.1.3.3 <i>Hybrid approaches</i>	32
3.2 Trees reconstruction	32
3.2.1 <i>DEM based approaches</i>	33
3.2.2 <i>Point cloud based approaches</i>	33
3.3 TomoSAR vs LiDAR point clouds	35
4 Façades reconstruction	36
4.1 Façades detection/extraction	36

4.1.1	<i>Robust M-estimator based directional SD estimation</i>	37
4.1.2	<i>Computation of surface normals</i>	38
4.2	Segmentation	39
4.2.1	<i>Number of clusters for k-means</i>	39
4.2.2	<i>Proposed segmentation scheme</i>	39
4.3	Reconstruction	40
4.3.1	<i>Model identification (flat or curved)</i>	40
4.3.2	<i>Modeling of façades</i>	41
4.3.3	<i>Refinements</i>	42
4.4	Experimental results and validation	44
4.4.1	<i>Dataset</i>	44
4.4.2	<i>Results</i>	44
4.4.3	<i>Validation and discussion</i>	47
5	Reconstructing 2-D/3-D building shapes (or footprints)	50
5.1	Motivation (Problem formulation)	50
5.2	Building detection	52
5.2.1	<i>Region growing procedure</i>	52
5.2.2	<i>Energy minimization</i>	52
5.3	Buildings shape/footprint reconstruction	53
5.3.1	<i>Refinement (or generalization) of alpha shapes</i>	54
5.3.2	<i>Incorporating reconstructed façades</i>	54
5.3.3	<i>Addition of rectilinear constraints</i>	57
5.4	Experimental results and validation	58
5.4.1	<i>Datasets</i>	58
5.4.2	<i>Reference dataset</i>	59
5.4.3	<i>Validation of results and discussion</i>	61
6	Application examples	65
6.1	4-D (dynamic) building model	65
6.2	Elevation estimates refinement	66
6.3	Joint sparsity concept	69
6.3.1	<i>Dataset</i>	69
6.3.2	<i>Proposed segmentation strategy</i>	69
6.3.3	<i>Application of the extracted building parameters</i>	74
7	Segmentation and reconstruction of individual trees	75
7.1	Dataset	75
7.2	Methodological workflow	75
7.2.1	<i>Point cloud clustering</i>	75
7.2.2	<i>Ellipsoidal modeling</i>	78
7.3	Experimental results and evaluation	79
7.3.1	<i>Reference data</i>	79
7.3.2	<i>Optimal bandwidth parameter selection</i>	79
7.3.3	<i>Evaluation of reconstruction results</i>	81
8	Conclusions and outlook	84
8.1	Summary and conclusion	84
8.2	Future work	85
8.2.1	<i>Refinement of 2-D/3-D building shape reconstruction</i>	85
8.2.2	<i>Roof modeling</i>	85
8.2.3	<i>Object based fusion</i>	86
8.2.4	<i>New visualization methods – SAR street view concept</i>	86
8.2.5	<i>Improvement in individual tree reconstruction</i>	86

Appendices	87
A Zhu X. X., Shahzad, M., 2014. Façade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 52(6): 3541-3552	88
B Shahzad, M., Zhu, X. X., 2015. Robust Reconstruction of Building Façades for Large Areas Using Spaceborne TomoSAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 53(2): 752-769	101
C Shahzad, M., Zhu, X. X., 2015. Automatic detection and reconstruction of 2-D/3-D building shapes from spaceborne TomoSAR point clouds. IEEE Transactions on Geoscience and Remote Sensing PP(99): 1-19	120
D Zhu, X. X., Ge, N., Shahzad, M., 2015. Joint Sparsity in SAR Tomography for Urban Mapping. IEEE Journal of Selected Topics in Signal Processing 9(8): 1498 - 1509	140
E Schmitt, M., Shahzad, M., Zhu, X. X., 2015. Reconstruction of individual trees from multi-aspect TomoSAR data. Remote Sensing of Environment 165: 175-185	153
List of Abbreviations	165
List of Figures	167
References	175
Acknowledgments	194

1 Introduction

1.1 Motivation and scientific relevance

Recent advances in very high resolution (VHR) synthetic aperture radar (SAR) imaging have attracted attention of many remote sensing analysts in characterization and monitoring of Earth's surface from airborne and spaceborne platforms. As an imaging radar, SAR produces two-dimensional (2-D) reflectivity map of an illuminated scene by transmitting and receiving electromagnetic radiations particularly in microwave band with wavelengths ranging approximately from 1mm to 1m. These longer wavelengths, in comparison to optical and infrared remote sensing systems, enable SAR sensors to penetrate through clouds and precipitation with almost no distortions allowing them to perform well in virtually all-weather conditions. Moreover, SAR carries its own illumination and therefore can work independently of the sunlight. These encouraging attributes offer certain advantages of SAR over conventional optical remote sensing systems. However, microwave scattering characteristics render SAR images to contain different information. In optical and infrared domain, the observed backscattering is primarily caused due to molecular resonance occurring on the object surfaces while in microwave domain, the dielectric and physical properties (e.g., surface roughness, geometry, permittivity etc.) of the objects are mainly accountable for backscattering (Elachi, 1988). This on one hand makes interpretation of SAR images difficult while on the other hand sometimes provide better discriminative surface features than optical systems (Cumming and Wong, 2005). For instance, longer microwave wavelengths (e.g., L- or P- bands) allows penetration through volume structures (e.g., vegetation) enabling interior information extraction while shorter microwave wavelengths (e.g., Ka-band) depict low penetration through volume structures allowing exterior extraction of information.

SAR, similar to all other imaging techniques, also projects a three-dimensional (3-D) scene onto a 2-D image, making it "non injective" especially in urban scenarios due to the presence of vertical structures (e.g., building façades or other man-made objects) (Fornaro et al., 2012). SAR interferometry (InSAR) is an advance technique able to retrieve the *lost* third dimension (elevation perpendicular to the azimuth and range plane) by exploiting the coherent nature of complex SAR signals. It has been an established technique for measuring several geophysical quantities e.g., estimation of terrain topography, monitoring surface deformations, measuring ocean currents etc. (Bamler and Hartl, 1998) (Adam et al., 2009). More recently, InSAR has been extensively applied in detailed analysis of urban objects such as buildings, roads or even individual trees. However, information extraction and recognition of objects from SAR (or InSAR) data is highly challenging especially in urban environments due to special characteristics introduced by the side-looking geometry. The unwelcoming effects such as layover and shadowing seriously handicap the interpretation of SAR images. Layover leads to map multiple scatterers, having same distance to the sensor, into one azimuth-range pixel while shadowing corresponds to image regions not visible to the sensor.

Advanced InSAR techniques, such as persistent scatterer interferometry (PSI) and SAR tomography (TomoSAR), aim at SAR imaging in 3-D or even higher dimensions. Among them, PSI, developed by Ferretti et al. (Ferretti et al., 2000) (Ferretti et al., 2001) aims to perform long term monitoring of subsidence and uplift in urban infrastructures by exploiting highly coherent pixels, i.e., the bright long-term stable objects (called persistent or permanent scatterers) (Hooper et al., 2004) (Kampes, 2006) (Perissin and Ferretti, 2007) (Rocca, 2007) (Gernhardt and Bamler, 2012). The technique, however, assumes only a single scatterer in one azimuth-range pixel and therefore does not resolve the layover problem. To-

moSAR, on the other hand, aims at real and unambiguous 3-D SAR imaging (Pasquali et al., 1995) (Reigber and Moreira, 2000) (Lombardini et al., 2001) (Fornaro et al., 2005) (Zhu and Bamler, 2010b) (Fornaro et al., 2012). By exploiting stack(s) of SAR images taken from slightly different positions, like PSI, it builds up a synthetic aperture in the elevation direction that enables retrieval of precise 3-D position of dominant scatterers within one azimuth range SAR image pixel. Multiple layovered objects in any pixel are therefore separated from the reconstructed reflectivity profile in the elevation direction (Zhu and Bamler, 2010b) (Fornaro et al., 2005). Moreover, exploiting the fact that acquisitions are taken at different times, the synthetic aperture can also be extended to the temporal domain to enable four-dimensional (4-D) (space-time) focusing of SAR images. The technique is referred to as differential SAR tomography (D-TomoSAR), which combines the strengths of both TomoSAR and PSI (Zhu and Bamler, 2010b) (Zhu and Bamler, 2011) (Zhu and Bamler, 2014) (Lombardini, 2005) (Fornaro et al., 2005) (Fornaro et al., 2009) (Fornaro et al., 2010) (Fornaro et al., 2014) (Lombardini and Pardini, 2012) (Ma et al., 2015). The retrieved scatterer information when geo-coded into world coordinates produces TomoSAR point clouds, capable of containing not only the 3-D positions of the scatterer location but also the estimates of seasonal/temporal deformation, making them very attractive for generating 4-D city models. Motivated by these chances, the thesis aims to explore the object reconstruction from these point clouds. In comparison to conventionally used airborne Light Detection and Ranging (LiDAR) data, these point clouds are associated with special characteristics such as lower positioning accuracy, anisotropy and much varying point density stimulating the need for developing new dedicated algorithms related to object reconstruction.

1.2 Objectives and focus

1.2.1 General goal

As the first-ever attempt in the community, the overall general goal stimulating research in this thesis aims to

“Develop novel robust algorithms for object reconstruction using TomoSAR point clouds that could lead to open up new possibilities of developing 4-D (spatio-temporal) city models.”

Such models would be valuable in ensuring safety/security of growing urban population by monitoring of urban infrastructures against potential threats of damage and structural degradation caused by various factors, e.g., ground subsidence or uplift, bad construction, natural disaster, etc.

1.2.2 Methodological goals

Since these *dynamic* point clouds only recently became available, the development of new robust algorithms for object reconstruction is the essential focus in this thesis. Two target object categories including buildings and trees have been considered in this research. In this perspective, although several 3-D reconstruction techniques using airborne LiDAR point clouds have been developed, adaptation of these methods to TomoSAR point clouds is not straight forward and requires special in SAR imaging characteristics. In addition to sensor related issues e.g., varying point densities and gaps in the data, particular challenges in reconstruction are to cope with complex building structures, non-descriptive tree shapes and occlusion related issues. Aimed at developing 3-D reconstruction methods coping with these challenges, the following three research objectives – one major and two minor – have been addressed to reach the aforementioned general goal in this thesis:

Objective 1: *Development of 2-D/3-D building reconstruction methods from spaceborne TomoSAR point clouds*

Modern spaceborne SAR sensors such as TerraSAR-X/ TanDEM-X (Buckreuss et al., 2008) and COSMO-SkyMed (Mezzasoma et al., 2008) can deliver VHR resolution data that fit well to the inherent spatial scales of buildings. This very high resolution (VHR) data is particularly suited for detailed urban mapping Adam et al. (2009) (Zhu and Bamler, 2010b) (Zhu and Bamler, 2011) (Zhu and Bamler, 2012b) (Zhu and Bamler, 2014) (Gernhardt and Bamler, 2012) (Reale et al., 2011) (Reale et al., 2013) (Fornaro et al., 2012) (Fornaro et al., 2014). TomoSAR processing over such VHR SAR data offers tremendous improvement in detailed reconstruction and monitoring of urban areas, particularly man-made infrastructures e.g., buildings (Zhu and Bamler, 2010b) (Fornaro et al., 2014). E.g., experiments using TerraSAR-X high-resolution spotlight data stacks show that the scatterer density obtained from TomoSAR is on the order of 600,000-1,000,000 points/km² (Zhu and Bamler, 2010b) compared to a persistent scatterer (PS) density on the order of 40,000-100,000 PS/km² (Gernhardt et al., 2010). The rich scatterer information retrieved from TomoSAR allows generation of 4-D point clouds. The major goal of the thesis is to systematically exploit these unstructured *spaceborne* TomoSAR point clouds to develop a strategy that systematically allow reconstruction of 2-D/3-D building models fulfilling following requirements:

- ◇ The developed framework should exploit special characteristics of spaceborne TomoSAR data;
- ◇ Due to lower point density and gaps in the data, the method should be able to work directly over 3-D point clouds i.e., without rasterization to avoid resolution loss;
- ◇ No imposition of any constraint on the shape of the building to allow reconstruction of wider variety of architecture;
- ◇ To cope with poor geometric accuracy of the spaceborne TomoSAR data, presence of ghost-scatterers (i.e., outliers) (Auer et al., 2011) and other scene related complexities, the developed methods should be robust to allow automatic processing of larger areas.

Objective 2: *Explore the potential applications of reconstructed models*

As the reconstructed models only became available in this work, the goal is to explore possible applications. E.g., the reconstructed models can serve as a major component in the realization and generation of 4-D (space-time) or even higher dimensional dynamic city models. In addition, the reconstructed models can be potentially explored to enhance the knowledge about the nature of coherent scattering and/or improve the tomographic inversion procedure. Demonstration of such application examples may be not only scientifically interesting but would also further stimulate the relevant research in the field.

Objective 3: *Development of 3-D individual trees (i.e., non-coherent objects) reconstruction from airborne TomoSAR point clouds*

Besides man-made structures in particular buildings, trees (or vegetation) are also of key interest when it comes to city modeling. So far, the reconstruction of individual trees from SAR data has not yet met the interest of research community. Recently, preliminary investigations of wavelengths in millimeter domain have shown encouraging potential of individual tree reconstruction (Stilla et al., 2014). To this end, the goal is to develop a model based individual tree reconstruction procedure by exploiting 3-D point cloud obtained after tomographic processing of airborne millimeter wave single-pass multi-baseline InSAR data acquired from multiple aspect angles. It would be the first effort in the community to analyze the achievable accuracy of individual tree reconstruction using millimeterwave SAR.

1.3 Reader's guide

This is a *cumulative* dissertation where the above mentioned four objectives (one general and three methodological) are addressed in five peer-reviewed journal articles (see the Appendix) by the author – two as the first author and three as co-author. The remainder of this thesis is structured as follows:

Chapter 2 introduces the basics SAR imaging including an explanation of geometrical and radiometrical effects that occur due to side-looking geometry and different microwave scattering contributions. Subsequently, the concept of InSAR and TomoSAR is discussed and finally 2-D/3-D object reconstruction methodologies including appearances and challenges of both target categories are presented.

Chapter 3 then provides procedural overview of the existing techniques related to 3-D building and individual tree reconstruction that employ conventionally used LiDAR point clouds. Subsequently, special characteristics of TomoSAR point clouds in comparison to LiDAR are presented.

Chapter 4 and 5 provides a novel and complete data driven framework that systematically reconstruct 2-D/3-D building shapes from unstructured spaceborne TomoSAR point clouds generated from TerraSAR-X high resolution spotlight data stacks. The approach is modular as well as automatic (parametric) and has been developed keeping in view the requirements proposed in the previous section (Objective 1).

Chapter 6 then explore three potential applications of the reconstructed 2-D/3-D buildings models generated by the methodology described in Chapter 4 and 5. Three presented applications include: 1) A 4-D building model example; 2) Reconstructed model based elevation estimates refinement procedure; and 3) An idea aimed towards improving TomoSAR inversion by incorporating joint sparsity as prior is described.

Chapter 7 details a prototypical workflow to reconstruct individual trees from 3-D TomoSAR point cloud generated from multi-baseline InSAR millimeterwave data acquired from multi-aspects with an airborne single-pass MEMPHIS SAR system created by the Fraunhofer Institute for High Frequency Physics and Radar Techniques (Schimpf et al., 2002).

Finally, chapter 8 concludes the thesis and provides perspectives and outlook for future research directions.

2 SAR object reconstruction – Basics and state-of-the-art

2.1 SAR Imaging Fundamentals

SAR is a coherent side-looking imaging radar mounted on a moving platform such as an aircraft or satellite. The term “synthetic aperture” refers to an artificially extended antenna. SAR electronically simulates such an extended (or larger) antenna by utilizing flight path and exploiting the mounted antenna in a time-multiplex fashion to produce high resolution 2-D radar image (reflectivity or backscattering map) of an illuminated scene with an increased azimuth resolution. Theoretically, electromagnetic waves with any wavelength can be employed for transmission and measuring backscattered signals, however, practically wavelength in microwave domain (Bamler and Hartl, 1998) (Moreira et al., 2013) e.g., 2.4~3.8 cm (X-band), 3.8~7.5 cm (C-band), 8~15 cm (S-band) and 15~30 cm (L-band) are more commonly utilized.

The two orthogonal coordinates of a 2-D SAR image are (slant) *range* (or cross track) and *azimuth* (or along track). *range* corresponds to the line-of-sight (LOS) distance to the object under a certain incidence (or elevation) angle θ while *azimuth* refers to the flight direction. SAR imaging in the range direction follows the basic radar principle which consequently renders that the spatial range resolution ρ_r depends on the duration of transmitted pulse and is determined from chirp bandwidth f_W as (Cumming and Wong, 2005)

$$\rho_r = \frac{c}{2f_W} \quad (1)$$

where c is the speed of light. The spatial azimuth resolution ρ_a of a real aperture radar on the other hand is inversely proportional to the actual physical length L of the antenna which leads ρ_a to be on the order of hundreds of meters for airborne and several kilometers for spaceborne platforms. Thanks to the synthetic aperture principle which exploits the moving platform and coherent processing of radar echoes (backscattered signals), the spatial azimuth resolution ρ_a has been tremendously improved to be around half of the physical antenna length L i.e.,

$$\rho_a \approx \frac{L}{2} \quad (2)$$

with no dependence on slant range, sensor velocity and wavelength of the transmitted pulse (Bamler and Hartl, 1998). It is worth to mention here that achieved azimuth resolution in (2) is based on monostatic SAR sensors (i.e., same transmitting and receiving antenna) operating in standard (continuous) strip-map mode where the sensor viewing direction is kept fixed as the radar platform progresses forward. For a different mode, e.g., spotlight, even more fine azimuth resolution is achieved by steering the radar beam to increase the target illumination period resulting in a simulated effect of wider antenna beam or reduced antenna length (Cumming and Wong, 2005).

The azimuth-range resolution element of a conventional SAR is illustrated in Figure 1. The contributions from all subscatterers (e.g., ground, buildings, trees, rocks, etc.) present inside such a resolution cell are mapped onto a single complex-valued SAR pixel. The complex-valued measurement of each pixel of a focused SAR image thus contains the integral of the backscattered radar signal from all scatterers inside the corresponding resolution cell. Its amplitude represents the scene reflectivity while the phase is related to the distance of

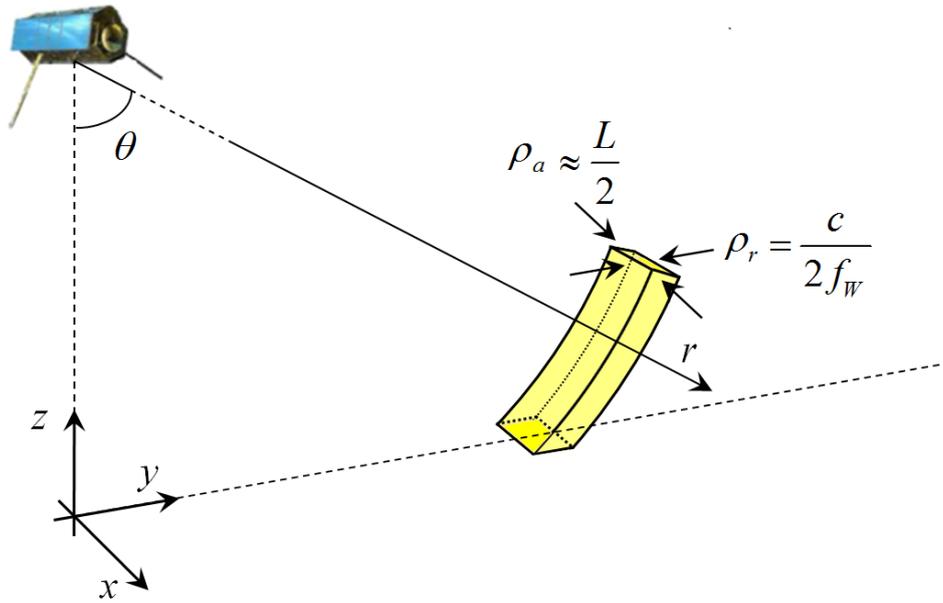


Fig. 1. Two dimensional resolution element of a conventional SAR image. ρ_r depends on the duration of transmitted pulse and is determined from chirp bandwidth f_W while ρ_a is approximately half of the physical antenna length L . x represents flight direction and is normally called azimuth or along track, r refers to the LOS direction of the sensor normally called range or slant range, θ denotes the elevation (or incidence) angle which actually defines the LOS direction, y refers to the ground range and z denotes the height direction that is perpendicular to the x - y plane. All scatterers lying within the yellow curve are mapped into the same azimuth-range pixel.

averaged scattering center to the sensor. Both amplitude and phase of SAR image depends on the physical (i.e., surface roughness, slope and geometry) and dielectric (i.e., permittivity) properties of the objects appearing in the illuminated scene (Moreira et al., 2013).

2.2 Radiometric and geometrical effects of SAR imaging

Radiometric and geometric effects are produced as a consequent of special SAR imaging technique. Most radiometric distortions are related to factors affecting radar backscattering coefficient (e.g., non-uniform antenna pattern, spreading loss effect etc.) and are systematically removed using sensor and orbit parameters (Frulla et al., 1998). However, a particular phenomenon affecting the radiometric appearance of SAR images that cannot be completely removed even from state-of-the-art approaches used for radiometric correction is the so-called speckle effect. It is more commonly visible in areas where surface roughness is comparable to the used radar wavelength (Goodman, 1975). The reason of speckle is the existence of more than one scatterer (i.e., multiple subscatterers) inside one resolution cell. Signal reflections from all subscatterers are coherently summed to represent one backscattered signal reflected from the corresponding resolution cell. The resulting amplitude (or brightness) of one SAR pixel thus not only depend on the physical characteristics of subscatterers but is also affected by constructive and destructive phase interaction from contributing subscatterers. This consequently renders amplitude and phase of backscattered signals to fluctuate from one pixel (or resolution cell) to another following Rayleigh and uniform distributions respectively.

Speckle is commonly mitigated by simple technique, known as multi-looking, which is essentially based on non-coherent local averaging of SAR intensities (or amplitudes) (Franceschetti and Lanari, 1999). Other more sophisticated adaptive speckle filters include Lee filter (Lee, 1980), Frost filter (Frost et al., 1982), maximum a posteriori (MAP) based filter (Lopes et al., 1990), or more recent wavelet- (Argenti and Alparone, 2002) (Chen et al., 2012b), nonlocal- (Deledalle et al., 2009) (Hu et al., 2015) and total variation- (Palsson et al.,

2012) based filtering methods. For an interested reader, a nice introduction and overview to speckle reduction techniques can be found in (Argenti et al., 2013). Although speckle is taken as noise in the image, it essentially carries information related to the subresolution structures (Bamler and Hartl, 1998). Also with improved sensor resolutions, the speckle effect tends to reduce due to decrease in number of subscatterers with in one resolution cell (Moreira et al., 2013).

The side-looking SAR imaging geometry introduces special geometric characteristics which appear in SAR images when the illuminated scene contains elevated objects. These geometric distortions renders scene interpretation from SAR images highly challenging especially in context of object recognition and 3-D reconstruction. In the following, the main geometric distortions commonly appearing in SAR images are briefly revisited (see Figure 2) (Lillesand et al., 2004).

- ◇ Forshortening occurs when the incidence (or look) angle θ of the SAR sensor is less steep than the object/terrain slope β (i.e., when $-\theta < \beta < \theta$ satisfies). This implies that the relative distance of the two points is shortened when projected onto the slant range direction. This causes backscattering energy to be concentrated in smaller region rendering affected intensity image parts to appear bright (Lillesand et al., 2004);
- ◇ For $\beta > \theta$ contributions from two or more scatterers located at the same distance to the sensor but at different spatial location are mapped to the same azimuth-range pixel of the SAR intensity image. This phenomena is called layover which causes elevated objects to appear bright and as being projected towards the sensor in the SAR image;
- ◇ Shadowing is seen for object parts that are not visible to the sensor (i.e., when $\beta < \theta - 90^\circ$ is satisfied). Image regions affected by shadowing appear dark in the SAR intensity image;
- ◇ Appearance of objects in urban areas are very much influenced by double- or (triple-) bounce scattering that is typically caused due to *di-(tri)-*hedral reflectors formed by vertical building façades (Thiele et al., 2007). The effect is clearly visible and appears in SAR images in the form of bright lines near the façade-ground transition regions.

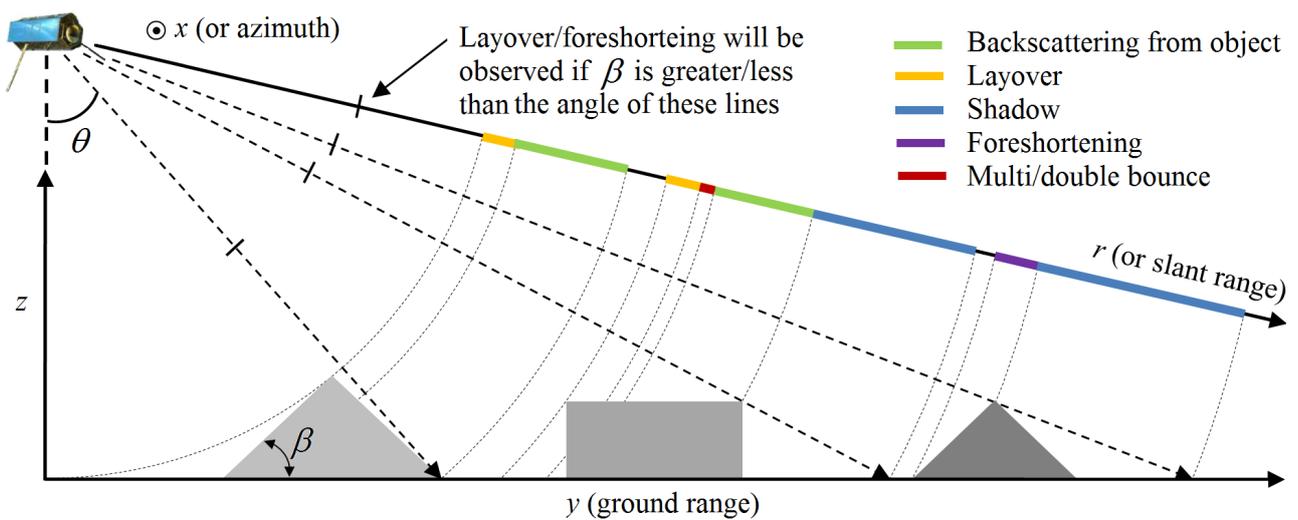


Fig. 2. Geometrical distortions appearing in SAR images. Three objects are shown in gray shades. θ refers to the incidence (or look) angle of the sensor while β represents the angle of the object surface with respect to the horizontal ground terrain (i.e., slope of the surface). The flight direction or azimuth axis x is perpendicular to the y - z plane i.e., coming out of the page in the depicted case.

2.3 Advanced SAR imaging

2.3.1 SAR polarimetry

SAR sensors have the ability to transmit and receive electromagnetic signals with particular polarization states. This enables them to provide, in contrast to single channel, dual- or fully- polarimetric SAR (PolSAR) images. Although, three different kind of wave polarizations e.g., circular, elliptic, and linear can be distinguished (Cloude and Pottier, 1996), linear wave polarization in two orthogonal H-V (horizontal-vertical) basis is most commonly used in remote sensing (Lee and Pottier, 2009). For a linearly polarized wave, the information content pertaining to the backscattering from one resolution cell can be characterized by a 2×2 complex scattering matrix represented as

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \quad (3)$$

where (in most cases) the second lower subscript indicates the polarization of the transmitted wave and the first one corresponds to the polarization of the received backscattered signal. For monostatic SAR systems, \mathbf{S} is symmetric due reciprocity theorem (i.e., $S_{VH} = S_{HV}$), resulting in three independent complex elements. Several methods to decompose \mathbf{S} , including the most common lexicographic (k_L) and the Pauli (k_P) decompositions, have been proposed in the literature for information extraction/retrieval purposes (Cloude and Pottier, 1996) (Freeman and Durden, 1998) (Yamaguchi et al., 2005) (Lee and Ainsworth, 2011) (Hong and Wdowinski, 2014) (Hong et al., 2015).

The scattering characteristics (geometric and material properties) of the object in the illuminated scene affects the polarization of the incident wave i.e., change of polarization may occur. This effect is consequently captured by the (measured) elements of the scattering matrix \mathbf{S} and forms the basis of enhanced information content in polarimetric data. Careful investigations of the complex elements of \mathbf{S} thus provide more information related to the scattering procedure which could be exploited to improve the results of features extraction for subsequently relevant object recognition and classification tasks (Cloude and Pottier, 1997) (Lee et al., 1999) (Chen et al., 2003) (Schou et al., 2003) (Guillaso et al., 2005) (Reigber et al., 2007) (He et al., 2008) (Haensch and Hellwich, 2010) (Sauer et al., 2011) (Guillaso et al., 2012) (Jin and Xu, 2013) (Haensch, 2014).

2.3.2 SAR interferometry (InSAR)

InSAR techniques utilize phase differences of (at least) two complex-valued SAR images acquired at different orbit positions and/or different times to extract topographic height or displacement information (Zebker and Goldstein, 1986) (Massonnet et al., 1993) (Bamler and Hartl, 1998). In a typical across track InSAR configuration, an interferogram is computed by pixel-wise multiplication of two precisely co-registered (up to sub-pixel level) single-look-complex (SLC) SAR images, acquired from two slightly different positions (separated by the distance referred to as spatial baseline B), in a complex conjugate manner i.e., multiplying one SAR image with the complex conjugate of the other SAR image. The resulting phase difference (or InSAR phase) is related to terrain topography and can be used to generate digital elevation model (DEM) (Li and Goldstein, 1990).

2.3.2.1 Differential InSAR

Apart from topographic mapping, the InSAR can also be extended to differential InSAR (D-InSAR) to precisely measure the LOS displacement (i.e., along slant range) (Zebker and Rosen, 1994) (Goldstein and Werner, 1998) (Bamler and Eineder, 2005). The technique principally works in a repeat-pass configuration where the illuminated scene is imaged twice with the same sensor from slightly different (but parallel) orbital positions at different times. Consequently, the InSAR phase difference not only consists of a topographic phase component but also contains another term that is independent of B and is caused by object displacements that occur during the time interval between the two acquisitions. This time interval is referred to as temporal baseline and is typically employed in the order of weeks, months or even years to achieve long-term deformation monitoring (Gabriel et al., 1989) (Zebker and Rosen, 1994). Note that in addition to these two phase terms, two additional (noisy) phase components, particularly existing in repeat-pass acquisitions, are also induced due to atmospheric propagation and changes in scattering behavior (Bamler and Hartl, 1998) where the former is typically mitigated by averaging multiple interferograms (Ferretti et al., 1997). Ignoring these phase terms, if an accurate DEM is available, the topography induced phase term can be computed and removed by subtracting it from the InSAR phase to measure subtle changes or displacements with even up-to millimeter accuracy (i.e., fraction of used wavelength) independently from sensor to scene distance which makes D-InSAR highly suitable technique for both airborne and spaceborne platforms (Moreira et al., 2013).

2.3.2.2 Phase-to-height conversion

In the context of topographic mapping, an important aspect in deriving correct terrain heights is to *unwrap* the InSAR phases that are ambiguous (as restricted within the interval $[-\pi \pi]$ i.e., modulo 2π). The wrapped phase actually corresponds to the so-called one fringe of the interferogram and its relationship to unwrapped phase can be expressed as follows:

$$\phi_{abs} = \phi_{InSAR} + 2\pi k \quad (4)$$

where k is an integer value. Finding value of k , a step referred to as phase-unwrapping (Goldstein et al., 1988), is a prerequisite for generating accurate InSAR DEMs (Li and Goldstein, 1990), and is solvable for continuous and smooth terrains but still remains highly challenging task in terrains that are discontinuous and contain frequent or abrupt height changes e.g., in urban areas (Soergel, 2010).

The unwrapped phase θ_{abs} depends approximately linearly with the height differences δh as

$$\delta h \approx \frac{1}{2\pi} \cdot h_{2\pi} \cdot \phi_{abs} = \frac{1}{2\pi} \cdot h_{2\pi} [\phi_{InSAR} + 2\pi k] = \frac{1}{2\pi} \cdot h_{2\pi} \cdot \phi_{InSAR} + h_{2\pi} \cdot k \quad (5)$$

where $h_{2\pi}$ represents the so-called height-of-ambiguity, i.e., height corresponding to one fringe (or 2π phase change of ϕ_{InSAR}) and is related to incidence (or look) angle θ and the perpendicular baseline B_{\perp} (component of B perpendicular to the look direction) as

$$h_{2\pi} = \frac{\lambda \cdot R \cdot \sin(\theta)}{p \cdot B_{\perp}} \quad (6)$$

where p is 1 for single-pass and 2 for repeat-pass measurements.

Magnitude of coherence (complex cross-correlation coefficient) gives an indication towards the measurement accuracy of the interferometric phase. Higher coherence refer to more accurate phase measurements which in turn leads to more accurate terrain height estimation. Similarly low coherence caused by various factors e.g., co-registration errors, thermal or receiver noise, temporal decorrelation etc. may lead to poor interferometric phase measurements resulting in producing low quality DEMs (Bamler and Hartl, 1998) (Moreira et al., 2013). A simple way to improve the measurement accuracy of phase estimates is multilooking (i.e., averaging interferometric samples) (Lee et al., 1994) which reduce the phase noise at the expense of increased loss in spatial resolution.

2.3.3 SAR tomography (TomoSAR)

TomoSAR is the most advanced InSAR method. In its differential form, i.e., D-TomoSAR, it is capable of 4-D, 5-D or even higher dimensional SAR imaging (Zhu and Bamler, 2010b) (Zhu and Bamler, 2011) (Lombardini, 2005) (Fornaro et al., 2009) (Fornaro et al., 2010) (Reale et al., 2011) (Lombardini and Pardini, 2012) (Fornaro et al., 2012) (Reale et al., 2013) (Fornaro et al., 2014) (Ma et al., 2015). In the following the TomoSAR imaging model is briefly described:

Let N represent the number of acquisitions (i.e., an image among stack of images), the complex-valued SAR azimuth-range pixel value g_n of n th ($n = 1, \dots, N$) acquisition with the corresponding perpendicular baseline b_n (see Figure 3) can be approximated as the weighted integral of reflectivity function $\gamma(s)$ (Fornaro et al., 2003) (Zhu, 2011)

$$g_n = \int_{\Delta s} \gamma(s) \exp(-j2\pi\xi_n s) ds \text{ with } \xi_n = -2b_n/\lambda r \quad (7)$$

where Δs denotes the span in elevation. Since it is well known that the far-field diffraction acts like a Fourier transform, the presented model is actually nothing but Fourier transform of $\gamma(s)$ sampled with spatial frequency (in elevation) ξ_n .

The continuous model in (7) can be discretized along elevation dimension into Q positions (i.e., $s_q \forall q = 1, \dots, Q$) by replacing the integral with the sum as follows:

$$g_n = \sum_{q=1}^Q \exp(-j2\pi\xi_n s_q) \gamma(s_q) + \varepsilon_n \quad (8)$$

or alternatively in matrix form as (Fornaro et al., 2003) (Zhu and Bamler, 2010a) (Zhu, 2011)

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (9)$$

where $\mathbf{g} \in \mathbb{C}^{N \times 1}$ is the measurement vector with $g_n \forall n \in \{1, \dots, N\}$, $\mathbf{R} \in \mathbb{C}^{N \times Q}$ is an irregularly sampled Fourier transform matrix with $R_{nq} = \exp(-j2\pi\xi_n s_q)$, $\boldsymbol{\gamma} \in \mathbb{C}^{Q \times 1}$ is the unknown discretized reflectivity vector with $\gamma(s_q)$, and $\boldsymbol{\varepsilon} \in \mathbb{C}^{N \times 1}$ is additive noise usually modeled as complex circular Gaussian random variable.

TomoSAR aim to invert the imaging model presented in (9) to retrieve the unknown discrete reflectivity vector $\boldsymbol{\gamma}$. Several TomoSAR inversion methods have been proposed in the literature based on conventional beam forming (Pasquali et al., 1995) (Lombardini and Reigber, 2003) (Fornaro et al., 2009), adaptive beam forming (e.g., Capon) (Lombardini et al., 2001), singular value decomposition (Fornaro et al., 2003), multiple signal classification (Gini and

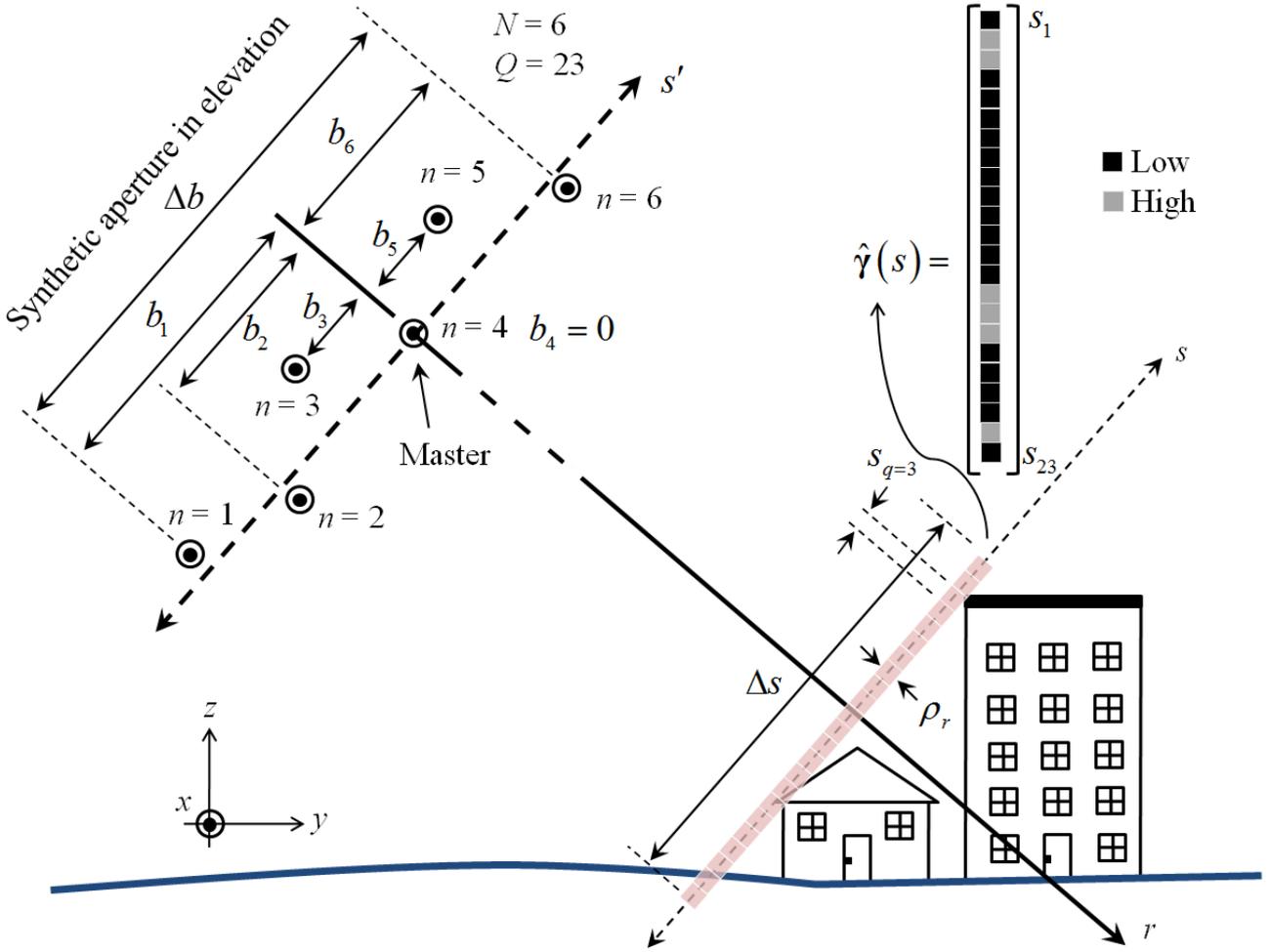


Fig. 3. Schematic illustration of the TomoSAR imaging geometry. The elevation aperture is built by exploiting multi-pass/multi-baselines (six in the depicted case) from slightly different viewing angles. It is shown that the backscattering contribution from the edge of two buildings and small portion of ground are mapped into a single azimuth-range SAR image pixel. TomoSAR aims to estimate the depicted reflectivity profile $\hat{\gamma}(s)$ for discretized (pink region) elevation extent Δs . Typically, the discretization factor is much higher i.e., $N \ll Q$ which renders (9) to be undetermined (i.e., more equations than unknowns). s denote the elevation axis which is actually a curve but is usually approximated as a straight line due to large range distances.

Lombardini, 2002) (Guillaso and Reigber, 2005) (Nannini et al., 2008) (Sauer et al., 2009), non linear least squares (Ferretti et al., 2005)) and compressive sensing based algorithms, e.g., the Scale-down by L_1 norm Minimization, Model selection, and Estimation Reconstruction (SL1MMER) algorithm (Zhu and Bamler, 2010a) (Zhu, 2011) (Zhu and Bamler, 2012b), that exploit sparsity of the signal. All of these mentioned algorithms treat TomoSAR inversion as spectral estimation problem by fulfilling following condition

$$\Delta s \ll \frac{\rho_r \cdot r}{\Delta b} \quad (10)$$

where Δb denotes the length of the elevation aperture.

The above condition (10) is particularly true for spaceborne TomoSAR due to large range distances and smaller angular diversity (Zhu, 2011). However, treating TomoSAR inversion as spectral estimation problem might not be reasonable for airborne SAR platforms and therefore other methods based on back projection (Reigber and Moreira, 2000) (Frey et al., 2008) or maximum likelihood estimation (Schmitt and Stilla, 2014b) needs to be employed.

2.4 Object reconstruction using SAR

Within the scope of this thesis, the following two subsections provides an overview related to SAR based 3-D building reconstruction methodologies of two target object categories namely *buildings* and *individual trees*.

2.4.1 Mapping of buildings in VHR SAR images

In the context of object recognition/modeling, SAR images have been exploited for the detection and reconstruction of three common man-made object categories, namely buildings (Bolter and Leberl, 2000) (Gamba et al., 2000) (Zhao et al., 2013) (Thiele, 2013), bridges (Soergel et al., 2007), and roads (Tupin et al., 1998). Among them, as resolution of SAR sensors improves, particular focus has been put in the domain of building boundary (foot-print) extraction and estimation of height in urban environments for 3-D reconstruction of buildings. Various approaches have been developed to tackle this problem in monoscopic (Tupin, 2003) (Quartulli and Datcu, 2004) (Barthelet et al., 2012) (Ferro et al., 2013) (Zhao et al., 2013) (Deng and Wang, 2014), radargrammetric (Simonetto et al., 2005) (Xu and Jin, 2007) (Dai et al., 2008) (Hill et al., 2008) (Dubois et al., 2013) and interferometric (Bolter and Leberl, 2000) (Soergel et al., 2003b) (Cellier et al., 2006) (Tison et al., 2007) (Thiele et al., 2007) (Thiele, 2013) (Dubois et al., 2013) (Dubois et al., 2014) frameworks. In the following, first a short discussion related to appearance of buildings is provided followed by an overview of state-of-the-art related to building detection and reconstruction based on SAR/InSAR datasets.

2.4.1.1 Appearance and characteristics of buildings in SAR/InSAR

Although geometrical effects introduced by the side-looking configuration cause difficulties in the interpretation of SAR images, they sometimes provide valuable hints which could potentially serve as meaningful features towards detection and reconstruction of a specific target category. E.g., layover and shadow information have been exploited to detect the building pixels (Tupin, 2003) (Bennett and Blacknell, 2003) (Soergel et al., 2003b) (Tison et al., 2004) (Quartulli and Datcu, 2004) (Hill et al., 2008) (Barthelet et al., 2012) (Brunner et al., 2010). Moreover, extracting the extent of layover and shadow areas, the height of individual buildings could be retrieved (e.g., using incidence angle, azimuth/range resolutions etc.). In InSAR domain, the appearance of buildings in phase images are also affected by layover and shadow effects (Thiele et al., 2007) (Thiele et al., 2013) (Dubois et al., 2014). Coherence plays a critical role when it comes to the analysis of phase images. Prior removal of noise and artifacts introduced by factors affecting coherence estimates is therefore prerequisite for efficient exploitation of InSAR phases.

Multi-bounce phenomenon is frequently observed in urban regions particularly between vertical structures (e.g., building façades) and ground (Quartulli and Datcu, 2004) (Thiele et al., 2007) (Soergel, 2010) (Barthelet et al., 2012) (Ferro et al., 2013) (Thiele, 2013). This effect appears in SAR images in the form of a bright line near the façade region (see Figure 4). Also with buildings having non-flat roofs e.g., gabled-roofs, the effect causes two bright parallel lines to appear at the two edges (or borders) of the layover regions (Thiele, 2013). Various line detection methods have been developed, e.g., (Touzi et al., 1988) (Schou et al., 2003) (Dai et al., 2004) (Wang et al., 2008a) (Ferraioli, 2010) (Baselice et al., 2014), to extract these bright lines in context of building detection/reconstruction from SAR imagery.

Surface roughness has direct effect on the backscattering energy. For instance, part of the building roof having smooth surface acts like a mirror and reflects away most of the energy (i.e., no signal is returned to the sensor due to the side-looking geometry). As a consequent,

the corresponding part appears dark. For relatively rough surfaces, Lambertian scattering - model of diffuse reflection - is quite often visible which renders their appearances to vary from less to full bright depending upon the (cosine of) local incidence angle (Thiele, 2013). Several researchers have simulated these backscattering characteristics together with material and roughness properties to provide semi-automatic solutions towards building detection and reconstruction. E.g., simulation based investigations of the double- and triple-bounce scattering effect occurring at building façades are depicted in (Brunner et al., 2010) and (Auer and Gernhardt, 2014) respectively.

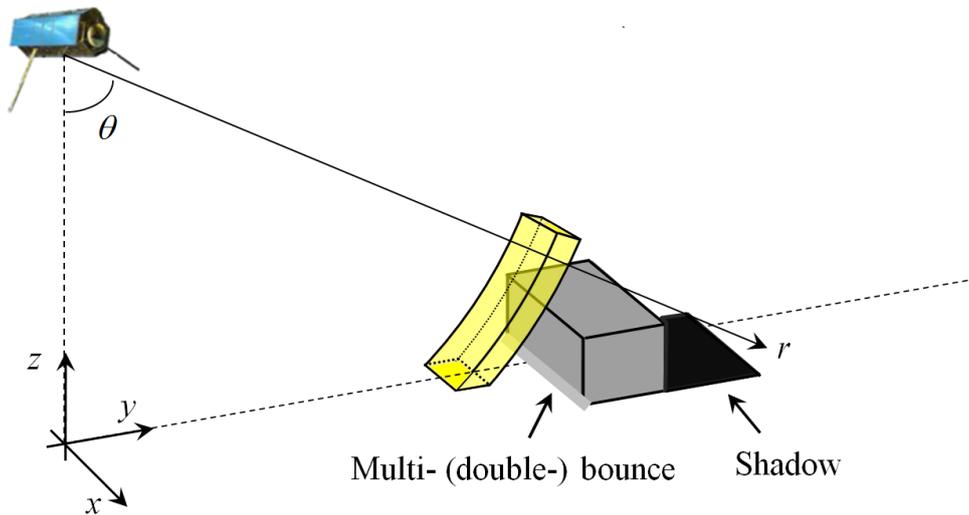


Fig. 4. Geometric illustration of multi- (double-) bounce and shadow effect. The light gray line depicts the double-bounce phenomenon that is frequently observed in urban regions particularly between vertical structures and ground. These lines give strong hint towards building detection. Shadow region (depicted as black) indicates the part not visible to the sensor. By knowing key parameters of the imaging radar e.g., θ , ρ_r , ρ_a etc., individual building height based on dimensions of shadow and/or layover could be retrieved.

2.4.1.2 Buildings detection using single VHR SAR image

Automatic detection and 3-D reconstruction of man-made objects in particular buildings from single VHR SAR image is of great practical interest especially when it comes to applications having stringent temporal restrictions e.g., emergency responses. However, due to inherent complexity of SAR images, literature on building detection and reconstruction from single SAR image is rather sparse. Among existing techniques, most approaches rely on extraction of bright lines due to already mentioned double bounce effect near the intersection of building façade and ground. For instance, one such approach was presented by (Quartulli and Datcu, 2004) which employed an automatic stochastic algorithm to reconstruct buildings from single SAR intensity image by modeling strong signals originated from dihedral scattering at the bottom of building and layover returns from the roof edges. Among recent approaches, (Zhao et al., 2013) proposed a building detection method based on marker controlled water shed algorithm to segment buildings from ground in a single SAR image. Both edge strength image and extracted markers were utilized that incorporated characteristics (bright pixel clusters) and contextual knowledge (shadows and the surrounding roads) pertaining to buildings. A similar approach that exploited layover and double bounce echoes to detect and determine the number of buildings from a single high resolution image was provided in (Cao et al., 2014). The approach detected bright patches via thresholding and morphological operations. Certain shape features were then employed to remove false alarms and later large patches were split using controlled water shed algorithm to determine the number of buildings. (Ferro et al., 2013) also developed a method that was primarily based on extracting set of low-level bright (lines) and dark (shadows) primitives. By exploiting probability of a primitive as belonging to a particular class (e.g.,

façade, double bounces etc.) using fuzzy membership functions, these primitives were then combined to produce more structured primitives consisting of reliable building footprint hypothesis. The approach was demonstrated to work on TerraSAR-X spotlight image containing both flat and gabled roof buildings. (Chen et al., 2015) introduced a more recent 1-D range detector to determine the 2-D building footprints. The approach sequentially scanned the range direction to detect buildings by their distinct profile. Each detected building wall (i.e., extracted line segment) visible to the sensor was then processed along the orthogonal azimuth direction to complete the 2-D shape of the building. The method could potentially well reconstruct simple symmetrical building footprints but might fail for scenes containing more complex non-symmetrical building shapes.

An interesting approach for extracting edges with application to building detection was presented by (Baselice et al., 2014). In contrast to many other edge detectors using SAR (Touzi et al., 1988) (Schou et al., 2003) (Dai et al., 2004) and InSAR (Ferraioli, 2010) images, their method was developed on Markovian stochastic framework and simultaneously exploited both real and imaginary parts of a SLC SAR image. They assumed that an edge (or discontinuity) present in the amplitude image would also exist at the same position in both real and imaginary parts of the complex SAR data.

So far, the mentioned approaches are based on single channel SAR image. Few publications exploiting multichannel (i.e., dual or fully polarimetric) SAR images have also been proposed in the literature. For instance, (He et al., 2008) proposed an approach to extract buildings from fully polarimetric SAR data. In it, mean shift was first applied to divide the image into homogeneous patches and later, by exploiting spatial connectivity, conditional random field (CRF) - probabilistic graphical model - was applied to segment each patch into shadow, layover and other areas. Comparison with basic logistic regression classifier revealed promising segmentation results. Other methods that exploited fully polarimetric SAR image include approaches based on supervised classification (Yan et al., 2011) via support vector machine (SVM) (Cortes and Vapnik, 1995) and integrated time-frequency decomposition technique with entropy/alpha- Wishart classifier (Deng and Wang, 2014).

2.4.1.3 Height estimation and reconstruction

SAR intensity (or amplitude) based building height estimation methods are typically proposed either in monoscopic or stereoscopic (i.e., radargrammetric) frameworks. In the following two subsections, an overview of methods lying in these two categories is presented.

2.4.1.3.1 Via radiometric effects - shadows & layover

One of the first investigations in determining object height from radar imagery was carried out in (Laprade and Leonardo, 1969). They provided initial basis to derive height estimates in both slant and ground range geometry from a single radar image. More recent approaches, relying on a single SAR image, estimated building heights either by exploiting information pertaining to building shadows (Bennett and Blacknell, 2003) (Tison et al., 2004) (Hill et al., 2008) (Barthelet et al., 2012) or layover (Tupin, 2003) (Quartulli and Datcu, 2004) (Brunner et al., 2010). Apart from approaches based on these two features, (Jiang et al., 2011) proposed a model based building height retrieval method based on hypothesis test procedure. In it, several building masks with various height estimates were first generated using predefined projection model and later an evaluation scheme based on simulated annealing was used to find the best height hypothesis. (Guida et al., 2010) also proposed a deterministic approach for extracting building height by exploiting both geometric and electromagnetic features from a single SAR image. The idea, initially presented in (Franceschetti et al., 2002), was based on modeling sound electromagnetic radar returns from isolated buildings. Initial building dimensions were assumed to be known in advance and the approach primarily

concentrated on the parameter inversion step. The method provided nice theoretical basis but the validation of height retrieval algorithm was only demonstrated using simulated images.

Another approach based on simulation and matching technique was presented in (Brunner et al., 2010). The approach simultaneously exploited major scattering characteristics (i.e., double bounce effects, shadows, and layover) for building height extraction. A hypothesis of building height was initially defined that was later matched with the simulated SAR image. The procedure was iterated until the optimization of the matching function with varying height propositions. Real data experiments on 40 flat and gabled roof buildings demonstrated the potential of the approach in extracting height of isolated buildings. Recently, (Wang et al., 2015) also developed a similar iterative model-based geometrical structure prediction and matching technique to estimate building height in a single VHR SAR image. In contrast to (Brunner et al., 2010), the method was able to perform under partial occlusion and required no assumptions related to the SAR backscattering distribution. The approach was validated over flat roof buildings using X-band airborne SAR and HH-polarized TerraSAR-X images.

2.4.1.3.2 Multi-aspect SAR based approaches (Radargrammetry)

The approaches mentioned so far that use a single SAR image only perform well in semi urban scenarios with isolated buildings, but may fail to work in dense urban areas where buildings are closely located together and smaller buildings are hidden (or occluded) by higher ones. This problem can be partially mitigated by applying stereoscopic (or stereo) principle to SAR imagery (Simonetto et al., 2005) (Soergel et al., 2006) (Xu and Jin, 2007) (Dai et al., 2008) (Hill et al., 2008) (Dubois et al., 2013) where two or more images of the same scene taken from different viewing angles are utilized to estimate building heights – a technique also referred to as radargrammetry.

Earlier SAR stereo based approaches were limited to only two configurations namely, same-side and opposite-side (Leberl, 1990). Same-side implies scene imaging from parallel flight orbits i.e., from the same aspect with different viewing angles (Soergel, 2010) whereas opposite-side refers to scene imaging from antiparallel orbits. An approach not limited to these two configurations was first proposed by (Michaelsen et al., 2005) and later extended in (Soergel et al., 2009) where the building heights were estimated using two orthogonal airborne SAR images. Images were separately analyzed to exploit certain building characteristics (e.g., rectangular structures, symmetries etc.) and later height was estimated by means of stereoscopic investigation. Instead of processing data individually from every aspect (Maksymiuk and Stilla, 2012) proposed a new concept to perform joint optimization using data from all aspects simultaneously. Discriminative CRF was used to model the data and the inference was carried out using loopy belief propagation algorithm. Two approaches employing airborne multi-aspect X-band PolSAR data from four orthogonal views to reconstruct 3-D buildings - modeled as cuboids or sets of cuboids - were presented in (Dai et al., 2008) and (Xu and Jin, 2007) (Jin and Xu, 2013). Recently, (Dubois et al., 2013) also proposed a pixel-based approach to estimate building height by analyzing building layover regions in the computed disparity map using data from TerraSAR-X.

2.4.1.4 Building reconstruction using InSAR data

SAR interferometry (InSAR) has been extensively used for detecting and reconstructing 3-D building model parameters (Bolter and Leberl, 2000) (Gamba et al., 2000) (Soergel et al., 2003b) (Thiele et al., 2007) (Brenner and Roessing, 2008) (Thiele, 2013) (Wegner et al., 2014). Similar to stereoscopy, InSAR also makes use of more than one SAR image in order to obtain height estimates. However, unlike approaches presented so far (i.e., approaches

based on SAR amplitude/intensity), they exploit the coherent phase information to determine building heights. Various InSAR based approaches have been developed to reconstruct buildings in 3-D. E.g., some approaches utilized InSAR DEMs (Gamba et al., 2000) while others e.g., (Bolter and Leberl, 2000) (Stilla et al., 2003) (Soergel et al., 2003a) (Thiele, 2013) exploited the full complex InSAR characteristics (i.e., phase, coherence and intensity) to detect and reconstruct buildings. In the following two sections, reconstruction approaches employing single and multiple InSAR pairs are presented.

2.4.1.4.1 Single-aspect approaches

In terms of methodology, the task of building detection and height estimation from a single (or multiple) InSAR image pair is either carried out jointly (Tison et al., 2007) or more commonly, the building is first detected in 2-D using amplitude (or intensity) values and later the third dimension (i.e., the building height) is estimated using InSAR phase measurements. Similar to building detection from single intensity images (mentioned earlier), the latter methods also rely on the occurrence of double bounce effect causing bright lines to appear between the building façade and the ground. E.g., (Soergel et al., 2003b) built quadrangular buildings objects by first extracting bright lines and shadow edges in intensity image and then assembled these primitive objects through an iterative production system. (Michaelsen et al., 2006) also extracted linear features (lines and edges) using Hough transform and later detected buildings by imposing rectangularity constraints which consequently allowed only orthogonal extracted bright lines to be selected as building features. This method offered advantage in reducing large number of bright lines usually appearing in an urban SAR image though, it might miss buildings having weaker response on one of its two visible sides.

Another interesting approach exploiting single-aspect X-band InSAR dataset was presented in (Cellier et al., 2006) to reconstruct large flat roof buildings. The approach first extracted the building characteristics (e.g., single/double bounce, shadows etc.) and later used an interferometric mixture model in the layover regions to detect front and back side of the illuminated building. A regularization procedure was then carried out to ensure extraction of geometrically correct (i.e., footprints with right angles only) buildings. The same approach was adapted and applied on fully polarimetric InSAR data in (Cellier and Colin, 2006). (Guillaso et al., 2005) also presented an approach to characterize building parameters using L-band fully polarimetric InSAR data. (Liu et al., 2013a) also developed a workflow to detect potential building layover regions using InSAR phase measurements. To incorporate contextual knowledge, over-segmentation was then carried out in the identified layover regions to retrieve superpixels which were subsequently merged via normalized cut algorithm (Shi and Malik, 2000) and a stochastic criterion to complete the detection procedure.

The recent launch of TanDEM-X has opened up new possibilities of obtaining single-pass InSAR measurements from space. By mitigating the decorrelation effects, it greatly enhanced the interferometric quality. (Thiele et al., 2013) presented a single-pass TanDEM-X interferograms to automatically extract building shapes. They developed a novel detector to investigate phase ramps in an interferogram computed from a single TanDEM pair. The basis of their shape detection approach was based on the fact that the layover of rectangular buildings could be characterized by two parallelograms whose two sides were parallel to the sensor's slant range direction. (Dubois et al., 2014) also exploited this idea and extended their approach such that the post processing to characterize building shapes was also performed on InSAR phases. With initial assumption of one visible façade, a bounding box was computed around the analyzed building which was iteratively sheared, scaled and translated to build a parallelogram. The parallelogram which fitted best in least square sense was used to represent the layover region. Building parameters were then computed by following the trend in the phase image i.e., looking up the phase values in the range direction.

2.4.1.4.2 Multi-aspect approaches

Problems pertaining to occlusion can be partially resolved using multi-view SAR data. In this context, few approaches employing multi-aspect InSAR have also been proposed. For instance, earlier approaches making use of the amplitude and InSAR height (and coherence) to detect and reconstruct buildings from multi-aspect InSAR datasets were proposed in (Bolter, 2001) and (Bolter and Leberl, 2000) respectively. (Soergel et al., 2003a) also presented an iterative and sequential method for building detection and reconstruction. In their workflow, first building hypotheses were generated and projected to ground range geometry. Then those hypothesis that fit a certain prior model criteria e.g., rectangularity, parallel lines etc. were further processed to reconstruct extruded building models. Subsequently these models were used to simulate the layover and shadow regions for iterative matching with the real image. Further development and extension to this approach was proposed in (Thiele et al., 2007) (Thiele, 2013) where multi-aspect fusion was realized on a lower feature level. As a consequent, it enhanced the capability to handle buildings even with rather weak feature support.

Multi-aspect approaches overcome the major limitation of single-aspect InSAR i.e., the occlusion problem. However, so far they still lack to deal with arbitrary building shapes and are only confined to reconstruct structured (e.g., rectilinear or rectangular) building footprints. Moreover the maximum height of the building that can be reconstructed is also limited by the derived height of ambiguity. To overcome these limitations, next, approaches that exploit auxiliary information together with SAR/InSAR datasets are presented.

2.4.1.5 Combing SAR/InSAR with auxiliary information

A number of approaches for building extraction and height retrieval have been proposed that combine SAR/InSAR data with auxiliary information such as optical imagery (Tupin and Roux, 2005) (Sportouche et al., 2011) (Denis et al., 2009) (Wegner et al., 2014), geographic information system (GIS) information (Tupin, 2003) (Thiele et al., 2010) (Liu et al., 2014) or other sources e.g., LiDAR digital surface model (DSM) (Tao et al., 2011). These approaches improved the feature extraction process by providing complimentary information. For instance, (Tupin, 2003) retrieved building heights of flat-roof industrial buildings by analyzing overlay regions in a single SAR image in combination with the manually extracted map data. Similarly, (Liu et al., 2014) performed layover analysis in single TerraSAR-X image together with available 2-D building outlines (or footprints) to estimate building heights. (Thiele et al., 2010) also combined available 2-D footprints and acquired height via InSAR phase to determine the after (or post) damage 3-D building shapes.

Apart from GIS data, high resolution optical imagery providing additional information about building appearance and shape potentially improve SAR based 3-D building reconstruction. In this context, an interesting and promising approach utilizing single nadir looking optical image and a couple of SAR images was presented in Tupin and Roux (2005). The SAR pair was used to extract 3-D information (e.g., lines, point-like scatterers etc.) in radar-grammetric framework which was projected to optical image. Structural information was then exploited by forming a region adjacency graph based on markov random field (MRF) and an elevation constraint was introduced in the prior term in MRF to ensure homogenous regions in the optical image to have similar heights. Similarly, (Sportouche et al., 2011) also demonstrated the use of building reconstruction using optical and SAR imagery. However in contrast to (Tupin and Roux, 2005), their approach exploited single optical and high resolution SAR image of the illuminated scene. Potential building footprints were detected in the optical image which were later projected and registered to the SAR image. Subsequently, these footprints were validated and their heights were retrieved through a joint optimization procedure based on log-likelihood function. In InSAR context, a recent and worth mention-

ing approach that combined an aerial orthophoto with a single-pass high resolution InSAR image pair to retrieve the height of flat roof buildings was presented in (Wegner et al., 2014). The key idea of their work was the combination of different height hypothesis estimated per building in a weighted least squares fashion. Although manual intervention was required in many steps but nevertheless the method provided nice theoretical basis towards joint estimation of building heights in data fusion perspective.

2.4.1.6 Advanced InSAR (TomoSAR)

Due to the complex urban scenes and inherent characteristics of SAR images due to geometrical projection caused by side-looking geometry, the previously presented approaches give solutions to 3-D building reconstruction but only to some extent. As already mentioned, advanced InSAR methods e.g., PSI and TomoSAR aim at SAR imaging in 3-D or even higher dimensions by resolving the distinct scatterer contributions within one azimuth-range pixel of a conventional 2-D SAR image. Preliminary investigations of TomoSAR were carried out to resolve volumetric scatterers (Reigber and Moreira, 2000) (Lombardini and Reigber, 2003) with applications typically in the field of forest mapping e.g., estimating biomass (Frey et al., 2008) (Tebaldini and Rocca, 2012) (Minh et al., 2014) and tree heights (Tebaldini, 2009) (Neumann et al., 2010) (d'Alessandro and Tebaldini, 2012) etc. In urban environments, using TomoSAR to resolve discrete scatterers is of more interest as it allows layover separation and has been demonstrated over C-band ERS data (Fornaro et al., 2005) and L-band airborne SAR data (Guillaso and Reigber, 2005) (Sauer et al., 2009). First demonstration of TomoSAR over spaceborne VHR TerraSAR-X data was presented in (Zhu and Bamler, 2010b) (Reale et al., 2011). SAR tomography of urban areas using CosmoSkyMed data in a multi-look framework has been proposed in (Fornaro et al., 2014). Recently, by exploiting the concept of joint sparsity, the state-of-the-art SLIMMER algorithm (Zhu and Bamler, 2010a) has been extended to multi-snapshot SLIMMER (M-SLIMMER) in (Zhu et al., 2015b).

Besides few approaches, e.g., (Guillaso et al., 2012) (Guillaso et al., 2013) (D'Hondt et al., 2012) (D'Hondt et al., 2013), aimed towards information extraction, 3-D object modeling/reconstruction from TomoSAR data is still a new field and has not been explored much. Preliminary investigations towards object modeling/reconstruction using spaceborne TomoSAR point clouds were demonstrated in (Shahzad et al., 2012) (Zhu et al., 2012) (Zhu and Shahzad, 2014) (Shahzad and Zhu, 2015b) (Shahzad and Zhu, 2015a) while TomoSAR point clouds generated over urban and vegetation areas using airborne SAR datasets were explored in (D'Hondt et al., 2012) (D'Hondt et al., 2013) and (Shahzad et al., 2015) (Schmitt et al., 2015) respectively.

2.4.2 Reconstruction of Individual trees in VHR SAR

2.4.2.1 Appearance and challenges

Single tree reconstruction from remote sensing data has been an active but highly challenging research topic due to the lack of geometrical features defining the shape of an individual tree. Especially in SAR imagery, apart from radiometric effects such as speckle and layover, the following factors considerably increase the complexity of individual tree detection/recognition:

- ◇ Trees show a manifold and possess highly non-descriptive appearance (Stilla et al., 2014);
- ◇ Exhibit weak backscattering signal;
- ◇ Often present in a group (or cluster) exhibiting significant overlap among each other;
- ◇ Rich variety of tree species rendering unarticulated diversity in tree shapes.

These factors make automatic recognition of individual tree from SAR imagery highly non-trivial. To this end, defining appropriate tree models corresponding to available SAR data

might be helpful in developing algorithms towards single tree recognition. For instance, 3-D ellipsoids (Lindberg et al., 2014) (Shahzad et al., 2015) (Schmitt et al., 2015) or paraboloids (Morsdorf et al., 2004) could be utilized to represent tree crowns as the first coarse approximation.

2.4.2.2 Few recognition/reconstruction approaches

Although recognition of individual trees (or vegetation) from remote sensing data has long been investigated from optical sensors (Hirschmugl et al., 2007) (Ardila et al., 2012) (Jiang and Lin, 2013) and LiDAR (Morsdorf et al., 2004) (Chen et al., 2006) (Höfle et al., 2008) (Jing et al., 2012) (Höfle et al., 2012) (Chang et al., 2013), there exist very few publications dealing the problem using SAR imagery. Among them, most approaches targeting towards forest mapping adopted SAR sensors to estimate the biomass (or volume) using L- and X-band SAR sensors (Englhart et al., 2011), L-band Polarimetric InSAR (Neumann et al., 2010), or via processing P-band SAR data using advanced InSAR methods e.g., TomoSAR (Frey et al., 2008) (Minh et al., 2014). Another approach to estimate stem volume using multi-view VHF-band SAR images was presented in (Hallberg et al., 2005). Use of X-band sensors were also explored in reconstructing canopy height model (Izzawati et al., 2006) and large-scale forest classification (Perko et al., 2011).

In the context of individual tree reconstruction, experiments using airborne millimeterwave SAR imagery have shown promising prospects (Schmitt et al., 2013). First approaches exploiting such millimeterwave SAR intensity image via morphological attribute filters were presented in (Maksymiuk et al., 2014) (Stilla et al., 2014). Recently, data from the same sensor has been utilized to produce a layover and shadow free 3-D point cloud via TomoSAR. After segmentation of the resulting point cloud, 3-D ellipsoid modeling has been applied to each segmented cluster to reconstruct individual trees. Relevant geometrical tree parameters is then extracted from these 3-D trees ellipsoids. Detailed processing scheme is illustrated in (Shahzad et al., 2015), (Schmitt et al., 2015) and Chapter 7 of this thesis.

3 Related work – Object reconstruction using 3-D (ALS) point clouds

3-D object reconstruction techniques from point clouds are widely employed using airborne laser scanning (ALS) data. This chapter first gives an overview related to 3-D building reconstruction methodologies. Subsequently, a brief overview about detection and reconstruction of individual trees is provided. Finally, special characteristics of TomoSAR point clouds in comparison to ALS point clouds are mentioned to emphasize the need of robust and dedicated approaches that have been developed in the scope of this thesis.

3.1 3-D building reconstruction

Automated building detection and reconstruction has been an active research area for at least two decades. Despite extensive research efforts, the topic is still of great interest due to ever increasing growth of urban population which gives rise to a wide range of potential applications in numerous fields. For instance, 2-D building footprints (or outlines) can be used in urban energy modeling (Sehrawat and Kensek, 2014), disaster management (Tomaszewski, 2014), devising emergency responses (Kolbe et al., 2008), noise pollution/control analysis (de Kluijver and Stoter, 2003), web-based mapping (Over et al., 2010), and flood risk modeling purposes (Shamaoma et al., 2006) etc. Additionally, extruding these footprints using height information leads to 3-D building models. Such models are essential for virtual city modeling (Döllner et al., 2006) and 3-D GIS applications (e.g., commercial software such as Google Earth, Apple Maps, etc.). Other possible usages may include urban planning (Koeninger and Bartel, 1998), augmented reality (Portales et al., 2010), exploration of solar potential over building roofs (Jochem et al., 2009), visibility analysis in urban environments (Engel and Döllner, 2009), analyzing electromagnetic reflections for placement of telecommunication antennas (Kirtner, 2000), gaming/entertainment industry (Pouliot et al., 2012) and many others.

A large amount of 2-D/3-D building detection and reconstruction techniques exist that employ ALS data. Interested reader may consult recent surveys over building extraction (Tomljenovic et al., 2015) and 3-D reconstruction (Haala and Kada, 2010) from ALS point clouds. Due to the large number of publications in the field, the presented overview is exhaustive, but certainly not complete, and mainly focuses on recent approaches and algorithms.

Methodologically, the problem of 3-D reconstruction is tackled by subdividing the task into two sequential steps, i.e. detection/classification of building points followed by their 2-D/3-D modeling and reconstruction. In the following two subsections, approaches lying in these two domains are sequentially presented.

3.1.1 Buildings detection (or extraction)

Building detection is generally carried out by first computing a digital terrain model (DTM) by filtering techniques e.g. via morphological filtering, slope or gradient analysis based filtering, linear prediction or interpolation based approaches, via iterative densification of triangular irregular network (TIN) structures or using segment based filtering. Interested readers may refer to (Meng et al., 2010) and (Vosselman and Maas, 2010) for an insight. In the following, these ground filtering techniques are briefly revisited:

3.1.1.1 Ground filtering techniques

3.1.1.1.1 Morphological based filtering

One of the early investigations on introducing morphological operations for ground points detection was carried out in (Lindenberger, 1993). Initially seed points having lowest height value within each horizontal structuring element were chosen by applying a morphological opening operation. Later, among neighbors of the chosen seeds, those points having height differences less than a certain defined threshold were extracted out in a recursive manner and characterized as belonging to the ground. A drawback of this approach was that its performance depends on the characteristics (e.g., size and shape) of the structuring element. Several researchers have focused on this issue and have proposed various morphological ground filtering techniques via progressively changing the size and the shape of the structuring elements. For instance, (Zhang et al., 2003) proposed such a filter to extract non-building points on almost flat slope terrains by gradually increasing the size of the structuring element. An extension to this approach able to work for more inclined terrains was presented in (Chen, 2007). The approach employed adaptive morphological opening operations using circular structuring element to detect larger buildings by prior removal of vegetation and other smaller objects. An even improved version minimizing the number of tuning parameters required in (Chen, 2007) from seven to two was proposed in (Chen, 2009). A progressive morphological approach that employs image inpainting methods together with simple slope thresholding and steadily (linearly) increasing the size of the structuring element was presented in (Pingel et al., 2013). (Li et al., 2014) also developed a brim filter that was embedded into (morphological) top-hat transformation to robustly extract ground points.

3.1.1.1.2 Slope or gradient analysis based filtering

Slope or gradient based filters rely on the assumption that ground and non-ground points have distinct slopes especially at the transition regions. As shown in (Vosselman, 2000), these filters are closely related to mathematical morphological (erosion) operations with the difference that they compute slope (or gradient) at each point and use it to detect non ground points. Similar to morphological based filters, each detected non-ground point is then further evaluated based on a defined elevation criterion. Use of fixed slope threshold worsens the filtering procedure especially in non-flat terrain surface. In this context, various approaches e.g., (Sithole, 2001) (Shan and Sampath, 2005) (Wang and Tseng, 2010) (Susaki, 2012) were proposed to adaptively set the slope thresholds to separate ground and non-ground points in complex urban scenes.

3.1.1.1.3 Linear prediction (or interpolation) based methods

These methods model the underlying terrain as a continuous (or at least piecewise) surface. An initial surface is first approximated by using all points in the dataset. Then, weights are computed for every point on the basis of its distance (or residual) from the surface. The weighting function is typically designed such that high weight is given to low objects (or points with negative residuals i.e., points lower than the initial approximation of the terrain surface) and low weight is assigned to the elevated objects (or points having higher positive residuals). In other words, points belonging to elevated objects are treated as outliers and the surface is iteratively refined using weighted linear regression techniques. A preliminary example of such a weighted interpolation was presented in (Kraus and Pfeifer, 1998). More recently, advanced methods based on repetitive interpolation (Kobler et al., 2007) and iterative thin plate spline technique (Mongus and Zalik, 2012) (Chen et al., 2013) have also been developed.

3.1.1.1.4 Progressive (or iterative) TIN densification filtering

In contrast to interpolation based approaches, these approaches model the ground terrain as a polyhedron. Instead of working on raw Lidar point clouds or rasterized DSMs, these methods first create a TIN or mesh by using sparse seed points with lower height values. Later the TIN is iteratively densified by adding more points that satisfy certain defined criterion. (Axelsson, 2000) first introduced TIN based ground filtering procedure. The algorithm worked by spatially tiling (i.e., 2-D gridding) of the whole point cloud. The ground terrain was then approximated by an initial TIN formed from a set of seed points. Each seed point was chosen as the lowest point in each grid. Densification of the sparse TIN was then made via addition of points by comparing the distances and angles of the points to the facet planes and nodes of the current polyhedron. The method worked well to handle discontinuous surfaces but lacked the ability to work in steep terrains. (Zhang and Lin, 2013) also proposed an algorithm to improve the performance of TIN densification based filters. After selecting initial seed points, their method performed segmentation based on smoothness constraint, proposed by (Rabbani et al., 2006), to grow the set of ground seed points prior to the subsequent densification. Experiments on benchmark data provided by ISPRS Commission III/WG 3 revealed that the idea of performing segmentation before the densification step helped in decreasing the omission errors during the filtering process. Recently, (Kang et al., 2014a) proposed a streaming framework for TIN densification. The proposed framework incorporated a parallelism scheme using multi-core architecture to improve the computational complexity of TIN related filters.

3.1.1.1.5 Segment based filtering

Except for (Zhang and Lin, 2013), so far the methods presented in this section are point-based. Although these methods have the strength of using an explicit model of the terrain surface, they lack the contextual knowledge (Sithole and Vosselman, 2005). Segment-based filters on the other hand first perform the segmentation of points to identify homogenous regions/patches. These patches (or segments) are then used as basic components for the subsequent filtering process. Initial attempt of applying segmentation based filtering directly over point clouds was presented in (Sithole and Vosselman, 2005). Use of contextual knowledge via segmentation allowed preserving discontinuities occurring especially in urban areas. Although the segment-based methods do have an advantage over point-based methods, the absence of an explicit surface can cause difficulties in correctly differentiating vegetation in the sloped terrain. To partially overcome these shortcomings, (Tóvári and Pfeifer, 2005) and (Lin and Zhang, 2014) proposed hybrid approaches that essentially adapted a (point-based) surface interpolation method to work on groups (or segments) of points. This combination improved the filtering accuracy by exploiting the strengths of both the approaches.

3.1.1.2 Extraction of building points/pixels

Nadir looking LiDAR points essentially give a DSM. Subtracting the DSM from the computed DTM (via ground filtering techniques from previous section) provides a normalized DSM (nDSM) which represents the height variation of filtered non-ground points (mostly comprising of man-made structures and vegetation). Subsequently, building points are extracted out from the nDSM (or filtered point cloud) by exploiting geometrical features such as deviations from the surface model, local height measures, roughness, and slope variations.

Methodologically, the building extraction approaches can be categorized into three groups where the first two groups of approaches extract buildings either *with* or *without* planarity assumption. The third group of approaches lies in supervised/unsupervised classification

framework. Although some of the classification methods also assume planarity of building roofs, these approaches have been separately categorized for better methodological clarity.

3.1.1.2.1 Building extraction *with* planarity assumption

Majority of approaches assume polyhedral roof structure and therefore make use of the planarity assumption to extract man-made structures in particular buildings from other objects in the scene. These approaches may be categorized into further two classes namely, segmentation-based (region growing / clustering) or fitting-based methods, and are presented as follows:

Segmentation based approaches

These approaches typically employ 3-D surface normals either as a similarity measure for region growing or as features during the clustering process. Commonly, surface normals are estimated via fitting “best” plane in least squares (LS) sense within some predefined local neighborhood v_c ; which is equivalent as performing principal component analysis (PCA) of the points in v_c (Hoppe et al., 1992). Several researchers have utilized the estimated normals to extract planar building segments. For instance, (Rottensteiner, 2003) extracted buildings from DSM via surface normals based region growing approach. (Forlani et al., 2006) also presented a similar region growing procedure to extract buildings by first extracting planar roof segments via slope analysis and later employing a rule based approach on the basis of computed topological and geometrical relationships of the extracted segments. (Zhang et al., 2006) and (Chen et al., 2012a) utilized a plane-fitting based region growing procedure to separate buildings from vegetation. (Sampath and Shan, 2007) presented a building boundary tracing algorithm from raw LiDAR point clouds. Their approach first performed ground filtering operation using 1-D slope filter. Then, a region growing based segmentation was adopted to extract building points. (Miliaresis and Kokkas, 2007) also proposed an object based segmentation method to extract buildings using LiDAR digital elevation models (DEMs). (Sohn et al., 2008) performed a model based filtering to first remove ground points. Connected component analysis is then carried out and buildings are extracted out based on characteristics features such as height, outline regularity, and smoothness of the surface. (Dorninger and Pfeifer, 2008) extracted building points by applying mean shift based clustering algorithm to segment planar roof points. (Höfle et al., 2009) also developed an approach to automatically extract buildings by exploiting strengths of both raster and point cloud methods. The method exploits nDSM and slope-adaptive echo ratio raster feature map to extract potential building candidates. Subsequently, an object based classification scheme based on computed surface roughness was adopted to remove the non-building regions. (Poullis and You, 2009) also proposed a probabilistic clustering algorithm based on computed surface normals. Their method was essentially a region growing procedure based on statistical analysis of geometrical properties of objects appearing in the scene. A segmentation approach to extract building directly from 3-D points was proposed in (Wang and Shan, 2009). Their method first determined the step (or jump) edges using nearest neighbors approach and later performed grouping of these edge points based on minimum spanning trees to determine building boundaries. Subsequently, separation of trees from buildings was achieved by employing a dimensionality learning method.

Typically, segments retrieved via region growing or clustering based on surface normals require a further post processing step to spatially separate coplanar segments. An approach that employed a density based concept (see Figure 5) to separate coplanar segments retrieved via fuzzy k -means algorithm to individually model distinct roof parts was presented in (Sampath and Shan, 2010). The approach worked well but needs the number of clusters (i.e., k) in advance. An elbow method was utilized to guess the initial number of clusters which however in most cases cannot be easily determined. Among recent approaches, (Kim

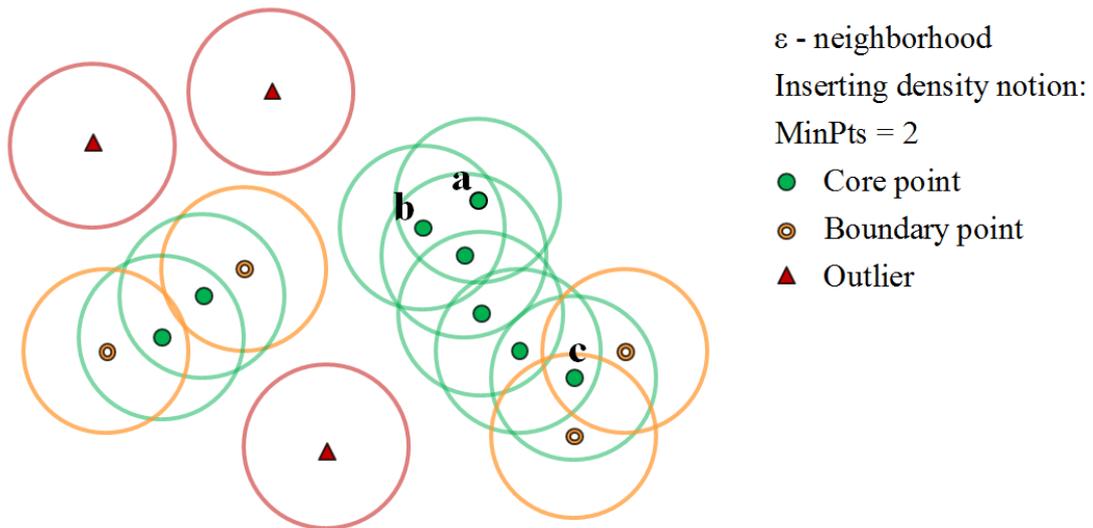


Fig. 5. Density based clustering algorithm (Ester et al., 1996). Points a and b are directly density connected to each other whereas points a and c are density connected to each other since there is a chain of points between them such that they all are directly density connected to each other. Two parameters that control the clustering process include the neighborhood parameter ϵ , i.e., the radius in case of sphere or cylindrical neighborhood, and the minimum number of points $MinPts$ in the ϵ -neighborhood for any particular point. The resulting clusters thus contains points such that all the points in any particular cluster are density connected to each other but are not density connected to any other point belonging to another cluster. Moreover, each point inside any particular cluster belongs to one of the three categories: 1) A point is labeled core point if it contains, within its ϵ -neighborhood, $MinPts$ number of points; 2) A point is considered border point if it is within ϵ -neighborhood of any core point but itself is not a core point and does not have $MinPts$ neighbors; 3) A point neither core point nor border point is termed as an outliers i.e., any point which do not have density (number of points) greater than $MinPts$ within its ϵ -neighborhood and also is not the neighbor of any other point.

and Shan, 2011) also performed planar roofs segmentation based on estimated normal vectors for 3-D building modeling by adopting an energy minimization approach formulated as multiphase and multichannel level set. (Liu et al., 2013b) also proposed a modified Locally Excitatory Globally Inhibitory Oscillator Networks (Wang and Terman, 1997) based segmentation to extract buildings from LiDAR DSM. In context to urban scene classification, (Vosselman, 2013) presented a multi-step segmentation approach based on planarity assumption. Later, merging and majority filtering techniques were employed for post processing. Similarly, (Jochem et al., 2012) also developed a roof plane segmentation approach for subsequent 3-D reconstruction by processing larger areas via spatially gridding of the whole dataset into smaller tiles. Within each tile, building regions were extracted using (Höfle et al., 2009) method. Surface normals based region growing procedure was then adopted within extracted regions to segment distinct roof surfaces. (Awrangjeb and Lu, 2014) presented another approach to extract building pixels by applying a segmentation method based on the connectivity based clustering technique. To allow real time extraction of buildings, an approach able to directly process scan lines of raw LiDAR data was presented in (Xu et al., 2014). The approach was based on discrete stationary wavelet transform (Demirhan and Gueler, 2011) and was able to extract polyhedral buildings in near real time (i.e., on the fly) during the airborne laser scanning process. An object based framework to extract buildings for subsequent classification into different types was proposed in (Belgiu et al., 2014). (Chen et al., 2014) also developed a multi-grid approach to detect and reconstruct buildings. A global optimization approach to segment individual roof segments for subsequent 3-D reconstruction was presented in (Yan et al., 2014). The key aspect of the approach was its ability to simultaneously extract multiple roof planes by formulating segmentation as a multi-label optimization problem.

Fitting based methods

Apart from the region growing/clustering methods, building pixels/points can also be extracted out by making use of plane fitting approaches followed by topological (contextual or geometric constraint based) analysis. Hough transform (Hough, 1962) (Duda and Hart, 1972) and random sample consensus (RANSAC) (Fischler and Bolles, 1981) are the two most commonly employed state-of-the-art algorithms that have been utilized/adapted to extract points belonging to planar roof surfaces. For instance, methods employing hough transform (or its variants) to determine planar roof facets have been proposed in (Vosselman and Dijkman, 2001) (Tarsha-Kurdi et al., 2007) (Sohn et al., 2008) (Borrmann et al., 2011) (Rau and Lin, 2011) (Hulik et al., 2014). Although hough transform is extensively employed to detect 2-D lines in images, their adaptation to 3-D plane detection is primarily restricted due to computational inefficiency and sensitivity to fitting parameters (Borrmann et al., 2011). RANSAC on the other hand is more generally applied for planar roof extraction. E.g., several variants of standard RANSAC algorithm for detecting planar roof structures have been presented e.g., in (Verma et al., 2006) (Schnabel et al., 2007) (Tarsha-Kurdi et al., 2007) (Kim and Shan, 2011) (Chen et al., 2012a). Unfortunately, similar to hough transform, RANSAC also suffers from the computational inefficiency. Although the algorithm is quite robust, it randomly selects the initial subset of points and presumes uniform prior probabilities which as a consequent in presence of many outliers may lead to an increasing number of iterations required for convergence. To overcome this, a bayesian version of RANSAC relying on conditional sampling methods has been proposed in (Botterill et al., 2009) (Kang et al., 2014b). Other versions of original RANSAC that improve the robustness and computation efficiency include MSAC (M-estimator SAMPLE and Consensus) (Henn et al., 2013), MLESAC (Maximum Likelihood Estimation Sample Consensus) (Torr and Zisserman, 2000), Guided MLESAC (Tordoff and Murray, 2005), PROSAC (PROgressive SAMPLE Consensus) (Chum and Matas, 2005) and many others.

3.1.1.2.2 Building extraction *without* planarity assumption

Although a lot of approaches employ planar roof assumption, there also exist some approaches relying only on geometrical properties (e.g., building shape, size, area compactness, height etc.). For instance, an earlier approach in this context was presented in (Weidner and Förstner, 1995) in which a pre-computed DEM was utilized to extract buildings by applying height difference (between DEM and the surface topography) thresholding based on prior knowledge pertaining to buildings geometries in the scene. More recently, (Meng et al., 2009) employed morphological operators to detect building pixels from ground filtered LiDAR data via gradually removing non-building points by incorporating prior knowledge related to minimum building height and area dimensions. (Mongus et al., 2014) also extracted buildings using contextual knowledge to distinguish objects with similar surface and geometrical properties by exploiting differential morphological profiles together with surface and regional attributes.

3.1.1.2.3 Classification based approaches

Building points can also be extracted out by explicitly labeling every point in the data set. For labeling purpose, local 3-D structure is typically inferred by employing eigenvalue analysis of the 3-D covariance matrix $\sum v_c$ derived from locally extracted neighbors v_c of a particular point p . The eigenvectors of $\sum v_c$ essentially give the orthonormal basis for the local neighborhood with their corresponding eigenvalues representing the magnitude (or variance) of expansion. Analyzing this magnitude gives an indication of the beneath surface, e.g., eigenvalue analysis of $\sum v_c$ with all points lying on the plane would ideally return only two nonzero eigenvalues. Similarly, all points lying on a 3-D line would give only one

nonzero eigenvalue. In addition to this, these eigenvalues are used to define a set of geometric features (e.g., anisotropy, eigenentropy, omnivariance, local surface variation or scatter etc.) (Mallet et al., 2011). These eigenvalue based features in conjunction with other geometric features such as spectral and directional features (Munoz et al., 2009), height/elevation variances and planarity (Mallet et al., 2011), point distances (Waldhauser et al., 2014), point density or density ratio (Guo et al., 2015), and distance ratio (Xu et al., 2012) etc. have been used to perform detailed 3-D analysis/classification. A more detailed description of these geometric (including eigenvalue) features can be found in (Mallet et al., 2011) and (Weinmann et al., 2015b).

Local neighborhood definition for feature computation

The definition of local neighborhood is crucial and plays an important role in estimating these features. Typically, the local neighbors are extracted from 3-D point clouds using spherical neighborhood (Munoz et al., 2009) (Weinmann et al., 2015b) where all points lying within a certain fixed radius from a particular point are taken as its neighbors. An alternative strategy is to compute neighbors via cylindrical neighborhood (Mallet et al., 2011) (Waldhauser et al., 2014) (Zhu and Shahzad, 2014) (Shahzad and Zhu, 2015b) where instead of sphere all points lying within a vertical cylinder (assumed to have infinite height) of defined radius centered at the point of interest are considered as the corresponding neighbors. Apart from these two, another common way to extract local neighbors is to consider fixed number of closest distance points as neighbors (i.e., k nearest neighbors) (Niemeyer et al., 2014) (Weinmann et al., 2015a). All of these strategies depend on one free scale parameter i.e., the radius of a sphere/cylinder or the constant parameter value of k . Selection of this scale parameter is not trivial due to its implicit dependency on the underlying 3-D structure and point density of the dataset and is therefore chosen either empirically or heuristically rendering it to be very much data specific/dependent (Weinmann et al., 2015b). Moreover, sensitivity and behavior of each feature is usually different with respect to the scale parameter and therefore one particular value may or may not be appropriate for efficient estimation of all considered features. To overcome these limitations, several researchers have focused on adaptive estimation of the scale parameter for each 3-D point (Weinmann et al., 2015a) e.g., via eigenentropy (i.e., Shannon's entropy) (Weinmann et al., 2014), normalized derivative (Pauly et al., 2003), local density and curvature (Mittra and Nguyen, 2004), TIN/voronoi (Sampath and Shan, 2010) (Yan et al., 2014) or dimensionality (Demantke et al., 2011) based scale selection techniques.

Supervised classification approaches

Object classification from extracted features is then attained mostly by using supervised learning methods e.g., SVM (Mallet et al., 2011) (Zhang et al., 2013a), random forests (Chehata et al., 2009), genetic algorithms (Waldhauser et al., 2014), artificial neural networks (Priestnall et al., 2000), AdaBoost (Lodha et al., 2007a), expected maximization (Lodha et al., 2007b) and bayesian discriminant classifiers (Khoshelham and Elberink, 2012). Furthermore, by bringing contextual knowledge into account, an approach based on combination of CRF and random forests to provide improved classification results was proposed in (Niemeyer et al., 2014). (Guo et al., 2015) also proposed a contextual constraint approach refining initial classification via joint boost classifier with graph-cut segmentation. An approach based on (non-) associative markov networks was also presented in (Shapovalov et al., 2010). (Weinmann et al., 2015a) also proposed a fully automated and comprehensive framework exploiting contextual knowledge.

Unsupervised classification approaches

Although both supervised paradigms (i.e., approaches with or without incorporating contextual knowledge) provides improved classification results, they require an already labeled dataset for training purposes which is not always available. In addition to it, the quality of

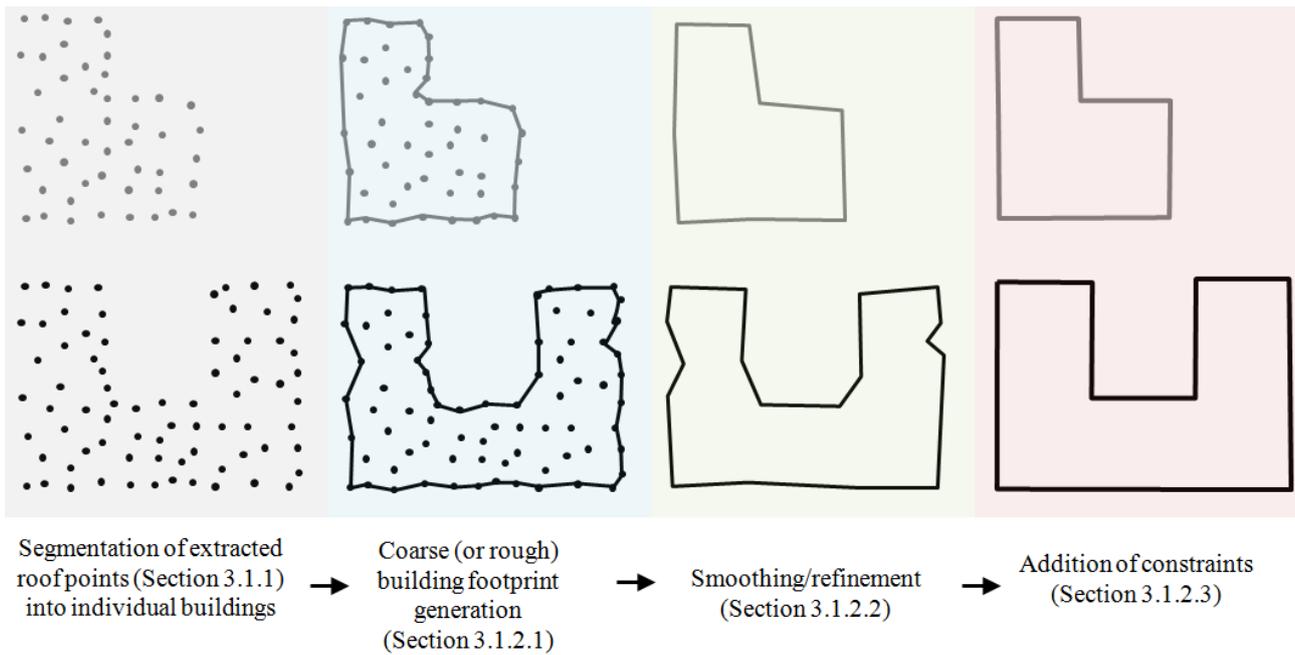


Fig. 6. Typical workflow related to building footprint reconstruction. The first column represents the segmentation process that separates the extracted building points into respective individual buildings (black and gray points belonging to the two buildings in the depicted case). Coarse shape of the building footprints are reconstructed by connecting the boundary points (second column). Different methods to determine this coarse shape are presented in Section 3.1.2.1. The initial shapes are smoothed using methods presented in Section 3.1.2.2 (third column). Finally constraints are added to yield geometrically correct and visually appealing building shapes via methods provided in Section 3.1.2.3 (fourth column). Each module is elaborated in the corresponding subsections.

training dataset has significant effect on the classification results. To cope with these issues, few researchers have provided solutions in fully unsupervised manner. For instance, (Lafarge and Mallet, 2012) proposed a graph cut based energy minimization approach to classify the input point cloud into four classes: building, vegetation, ground and clutter. A somewhat similar and seemingly improved approach with slightly different energy formulation to classify the whole point cloud into buildings, trees and ground was proposed in (Zhou and Neumann, 2013). Although both approaches were based on MRF, the main difference was in the definition of trees classification in the energy equation. Another unsupervised classification approach also based on graph cut optimization was proposed in (Sun and Salvaggio, 2013). Apart from graph cut based approaches, (Xu et al., 2012) also proposed a sequential three stage rule based strategy for classification of input point cloud LiDAR dataset into seven distinct categories. The features employed in the proposed classification approach were derived from three entities namely single points, planar segments, and segments resulting from mean shift algorithm.

3.1.2 Building boundary/footprint reconstruction

To reconstruct building footprints, a common strategy is to adopt a standard three-step workflow, depicted in Figure 6, in which the initial rough contours around extracted building points/pixels (Section 3.1.1) are first determined which are subsequently regularized using line simplification methods and additional geometric constraints (e.g., parallelism or orthogonality) to yield accurate and visually appealing building footprints.

3.1.2.1 Rough footprint generation

Several algorithms have been employed to reconstruct the rough shape of the building footprint. For instance, a computational geometry based alpha shapes (or α -shape) algorithm (Edelsbrunner et al., 1983) is commonly employed to extract building boundary/outline

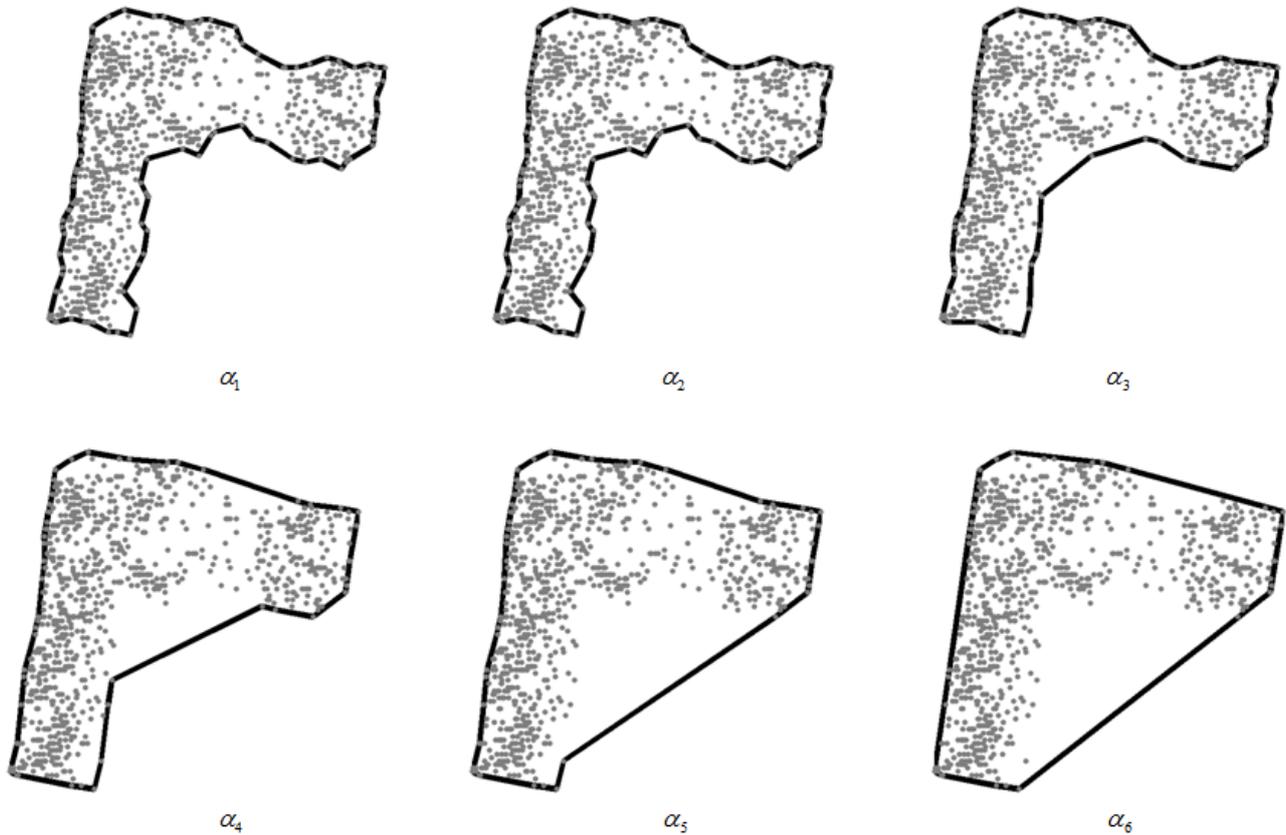


Fig. 7. Coarse building footprint reconstruction using alpha shapes method with varying α such that $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < \alpha_5 < \alpha_6$. It is demonstrated that with increasing α , the algorithm loses the ability to reconstruct concave building footprints and eventually with very large value of α or when $\alpha \approx \text{inf}$ returns a convex hull.

(Dorninger and Pfeifer, 2008) (Lach and Kerekes, 2008) (Kim and Shan, 2011) (He et al., 2014) (Shahzad and Zhu, 2015b). The method is essentially a generalization of convex hull and results in vertices that describe the coarse 2-D polygonal boundary of the building footprint. Unlike convex hulls, the method is also able to deliver both convex and concave boundaries with fine details. Moreover, both inner and outer polygonal boundaries can be extracted out. However, the reconstructed shape depends on a particular value of α (i.e., it controls the model complexity) and therefore has to be carefully chosen. For instance, an overlarge α could make it difficult to follow concave polygonal shapes, e.g., an L-shaped building (see Figure 7). Therefore, an estimate of α as proposed in (Dorninger and Pfeifer, 2008) that produces reliable building shapes, including smaller structures, may be chosen as twice of the mean of Euclidean distances computed for every point from its nearest building points. (Verma et al., 2006) employed a ball pivoting algorithm (Medeiros et al., 2004) to extract boundaries of the roof segments. The method was closely related to alpha shapes but result in triangular mesh of points. Another approach to trace boundary from building points extracted from raw LiDAR point clouds was proposed in (Sampath and Shan, 2007). Their method adopted a modified convex hull approach to trace the boundary points which were then connected to obtain coarse building outlines. (Poullis and You, 2009) also presented a modified boundary extraction method, originally proposed in border following context in (Suzuki and Abe, 1985), to extract building outlines via connected component analysis. Other approaches presented in the context of reconstructing building boundaries (i.e., via step edges, height jumps or tracing segmented roof faces) from building pixels/points extracted either from rasterized DSM (or nDSM) or directly from point clouds include (Maas and Vosselman, 1999) (Rottensteiner and Briese, 2002) (Rottensteiner, 2003) (Wang and Shan, 2009) (Kim and Shan, 2011) (Jarzabek-Rychard, 2012) (Sohn et al., 2012) (Liu et al., 2013b).

3.1.2.2 Refinement or polygon simplification

Generally due to measurement noise, low sampling rate and varying point density, the initial coarse footprints are usually irregular (or rough) and exhibit zigzag patterns. Typically, line simplification methods are employed to achieve smoothed/refined footprints. One of the most commonly employed method is the Douglas-Peucker (DP) algorithm proposed in (Douglas and Peucker, 1973). The algorithm begins by connecting the first and last point of the ordered set containing the vertices of the initial footprint. Distance of all in-between points are computed from the connected line segment. If the distance of the farthest point is less than a certain predefined threshold, it is discarded. Otherwise this point becomes a new end point. The algorithm then runs recursively with the first point and the new end point and then with new end point and the last point. Once the recursion stops, the algorithm outputs a new ordered set of vertices containing the same initial vertices except the discarded ones. Approaches employing DP method to refine initial rough building boundaries were presented in (Zhang et al., 2006) (Jwa et al., 2008) (Poullis and You, 2009) (Kim and Shan, 2011) (Galvanin and Poz, 2012) (Sohn et al., 2012) (Liu et al., 2013b) (Yan et al., 2015). A slightly modified vertex-driven DP version to remove erroneous vertices was presented in (He et al., 2014). In contrast to DP, the vertex-driven DP concentrated on polygonal complexity and generated a distinct polygonal hypothesis based on certain n number of vertices. An energy minimization based optimal polygon selection procedure was adopted to determine the optimal value of n . Another polyline simplification algorithm based on sleeve fitting technique was proposed in (Zhao and Saalfeld, 1997). (Dorninger and Pfeifer, 2008) also proposed refinement procedure to smooth the rough building footprints. Their method was based on mean angular deviation computed at each edge (i.e., line connecting a vertex with its next consecutive vertex) of the initial polygon.

3.1.2.3 Boundary regularization

Many approaches apply perpendicularity or parallelism constraints to yield better looking and visually correct building shapes (Weidner and Förstner, 1995) (Maas and Vosselman, 1999) (Alharthy and Bethel, 2002) (Gross et al., 2005) (Wang et al., 2006) (Zhang et al., 2006) (Sampath and Shan, 2007) (Dorninger and Pfeifer, 2008) (Jarzabek-Rychard, 2012) (Perera et al., 2012) (Arefi and Reinartz, 2013) (Liu et al., 2013b) (He et al., 2014) (Awrangjeb and Fraser, 2014) (Seo et al., 2014) (Zhu et al., 2015a) (Yan et al., 2015). Boundary regularization methods may be categorized into two categories i.e., approaches that do or do not need estimates of dominant building directions. The first set of approaches initially determine the dominant building directions and later utilize them to regularize refined footprints whereas the second set of approaches typically formulate the regularization process into a constraint energy minimization framework. Solution of such a formulation is then usually obtained via optimization algorithms e.g., simulated annealing (gradient descent) or graph cuts. In the following, both types of regularization methods are briefly discussed:

3.1.2.3.1 Approaches based on dominant direction estimation

These approaches typically assume that buildings are composed of only two dominant directions. Once the main orientation is estimated, the refined edges are then adjusted (or projected) either onto the estimated principal or its orthogonal direction depending upon the smaller angular difference. Retrieval of the main orientation is thus the crucial step to enforce such a regularization. To this end, minimum bounding rectangle (MBR) can be employed to determine the dominant/principal orientations of the building footprint (Arefi and Reinartz, 2013). I.e., the two orthogonal axes of the MBR provide the direct estimate of the desired dominant orientations. (Gross et al., 2005) also proposed a MBR based method to generate rectangular building footprints via recursive rectangular approximation. (Seo

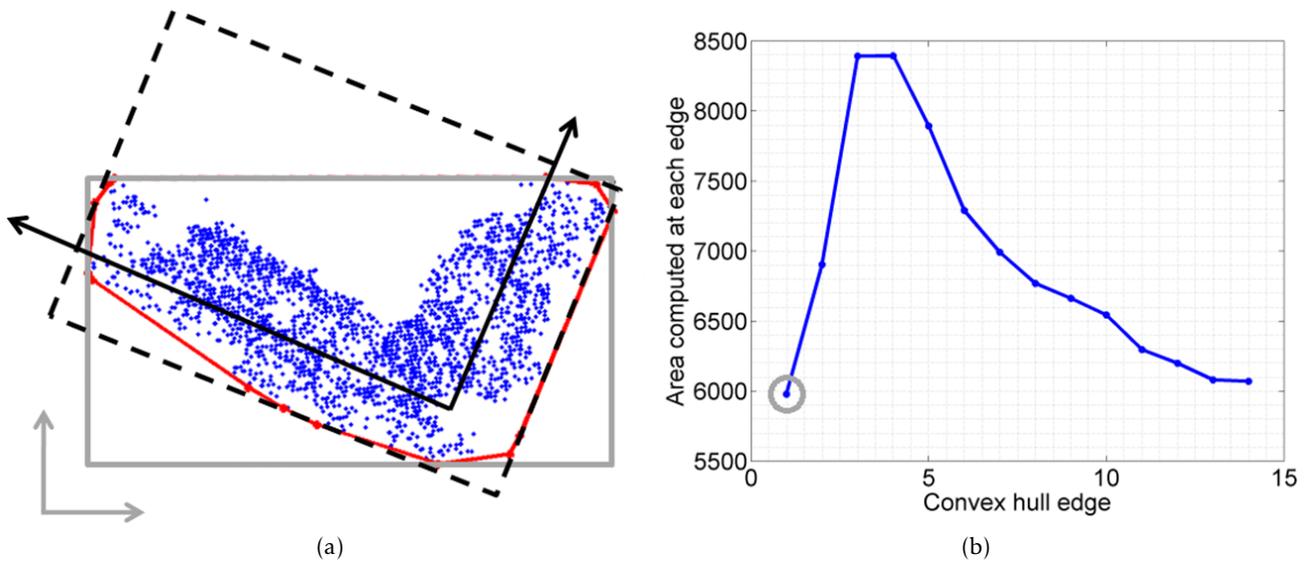


Fig. 8. Example depicting failure of MBR to correctly determine the dominant direction of the building footprint: (a) 2-D points of buildings (blue color) are depicted. Convex hull with 14 edges around these points are drawn as red polygon. Grey rectangle is the minimum area bounded rectangle computed via rotating callipers method. It can be seen that the dominant directions estimated via grey MBR do not correctly represent the true dominant orientation of the building points. The desired MBR that correctly represent the dominant orientation is shown as black dotted rectangle; (b) the areas of the MBR by rotating MBR around its center of gravity at angles equal to every edge of the convex hull. The grey circle shows the minimum area bounded rectangle corresponding to grey MBR in (a) (Shahzad and Zhu, 2014).

et al., 2014) and (Zhu et al., 2015a) also employed MBR based constraints to individually model every extracted planar patch. Commonly, MBR is computed by a method known as “rotating calipers” (Toussaint, 1983) which is based on the theorem, proved in (Freeman and Shapira, 1975), that any minimum area bounded rectangle is collinear with at least one of the sides of the convex hull (surrounding the same 2-D points). The convex hull is therefore first computed and later the bounding rectangle is sequentially computed by rotating the convex hull polygons in a way such that each side of the convex hull becomes parallel to a fixed axis e.g., x - (horizontal) axis. In each rotation, the area of the minimum bounding box around 2-D points is computed and the rotation angle that provides a minimum bounding area is used to determine the vertices of the desired MBR. Although the use of MBR provides reasonable estimates of the principal orientation of the buildings, there exist cases as demonstrated in Figure 8 where MBR fails in accurate determination of building dominant directions. An alternative to MBR, (Alharthy and Bethel, 2002) proposed an image cross-correlation matching with rotating template and histogram analysis based approach to estimate two orthogonal dominant directions. (Zhou and Neumann, 2008) also performed a similar histogram analysis to estimate principal directions. Their method in contrast to (Alharthy and Bethel, 2002) was able to detect more than two directions. (Zhang et al., 2006) also presented an approach to estimate the dominant orientations of the building based on weighted line segment lengths. The authors proved that their approach was able to provide robust estimates of the dominant directions as long as the total length of orthogonal segments was greater than the total length of non-orthogonal segments. Few approaches e.g., (Dorninger and Pfeifer, 2008) and (Jarzabek-Rychard, 2012) considered the longest polygonal edge of the refined footprint while (Maas and Vosselman, 1999) regarded line representing ridgeline edges as an estimate for the principal orientation of the building footprint.

The estimated principal direction is subsequently utilized to regularize refined footprints. Figure 9 illustrates the mechanism of practically adding rectilinear constraints to a building polygon. Apart from the depicted procedure, few other methods also exist in the literature. E.g., (Sampath and Shan, 2007) presented an approach that regularized the rough building outlines by following a hierarchical least squares solution. (He et al., 2014) proposed an

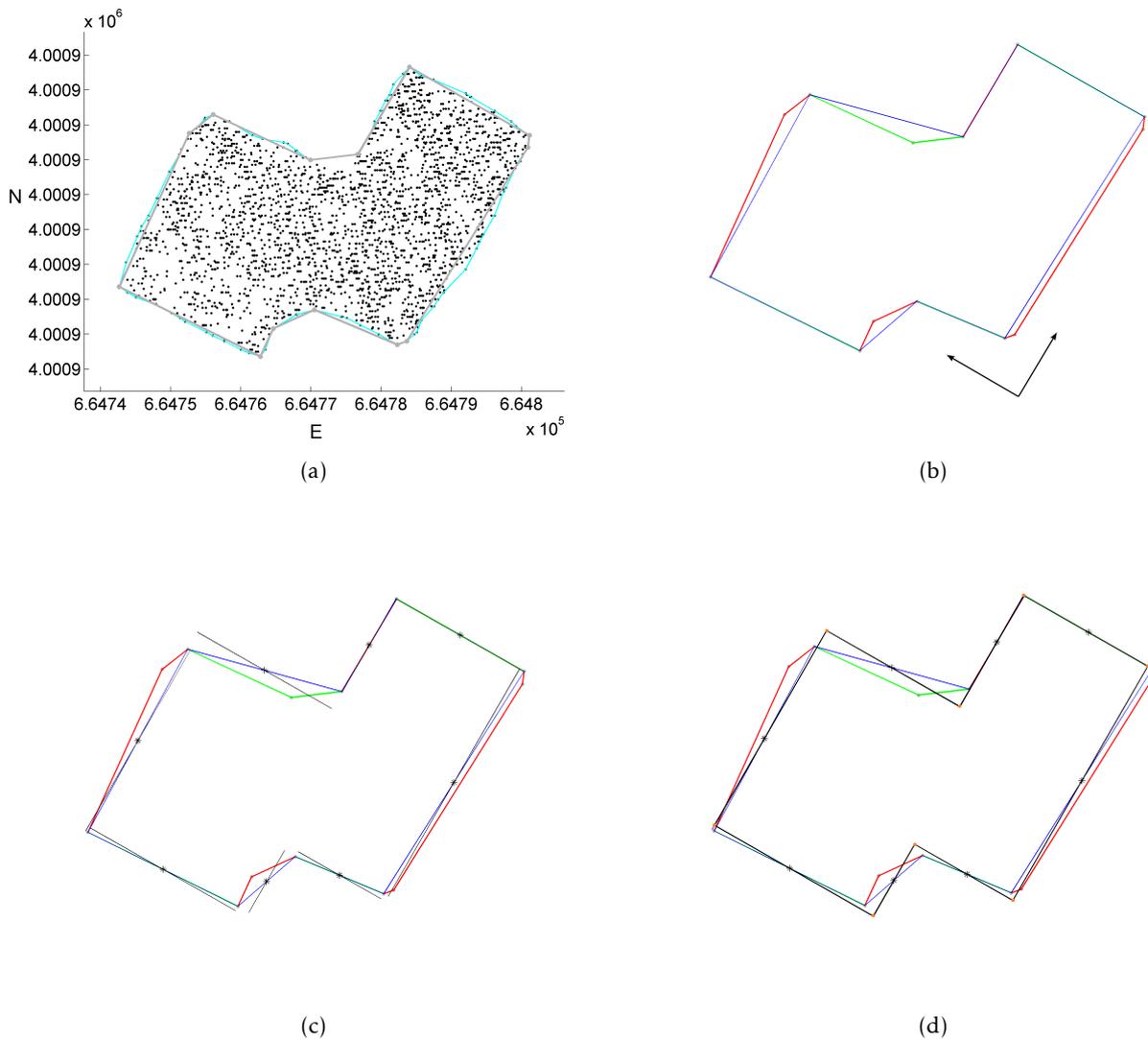


Fig. 9. Procedure to add rectilinear constraints to any one particular building polygon: (a) Black dots indicate the extracted building points, cyan polygon represents the initial coarse outline obtained using alpha shapes algorithm and gray polygon depicts the refined/smoothed polygon to which rectilinear constraints are added; (b) Each edge of the refined gray polygon of (a) is segmented to belong to one of the two estimated dominant directions indicated by black arrows. The segmentation is based on their angular deviation (i.e., an edge is associated to that principal direction with whom the angular difference is less). Segmented edges are shown in red and green colors. The blue polygon then depicts the reduced refined polygon of (a) by merging all adjacent edges that belong to the same principal axis. (c) Each edge of the blue polygon is then rotated/projected around its midpoint onto its corresponding principal axis; (d) Finally the intersection (vertex) points among the adjacent edges of the projected blue polygon edges of (c) are computed for rectilinearization. Note that the depicted polygon contains no associated façade and therefore belongs to case 3. For case 2 polygons, the associated façades are first fused to the refined polygon and, if identified as rectilinear, constraints are added to complete the reconstruction procedure.

approach that employed a global adjustment procedure based on data fitting errors and made use of the estimated dominant direction to reconstruct regular building footprints.. Recently, (Yan et al., 2015) also presented a 2-D snake algorithm that performed an energy minimization of a refined (via DP) 2-D vertex topology represented as a graph in order to regularize building footprints for subsequent 3-D building modeling.

3.1.2.3.2 Approaches without dominant direction estimation (Markovian/energy-minimization based methods)

Some approaches enforce rectilinear (or other) regularization constraints without explicitly estimating the main building orientation. For instance, (Wang et al., 2006) presented a

Bayesian approach to reconstruct footprints from pre-segmented building points. The cubic spline function was used as prior for MAP estimation to give preference to straight, orthogonal or oblique edges. (Lafarge et al., 2008) also proposed an automated object based method that employed marked point process and utilized an input DEM to extract/reconstruct structured building footprints. An extension to this approach was presented in (Lafarge et al., 2010) which also allowed possibility of an interactive operation to obtain more accurate reconstruction results. A similar workflow of obtaining structural footprints was provided in (Brédif et al., 2013). (Galvanin and Poz, 2012) also presented a building footprint identification procedure by adopting MRF based approach. Their method adopted a contour following technique described in (Ballard and Brown, 1982) together with DP method to reconstruct refined contours of above ground objects. Subsequently, building contours were identified by optimizing the MRF based energy function; comprising of four energy terms: rectangularity, area, spatiality and entropy; via simulated annealing algorithm.

3.1.3 Roof modeling/reconstruction

Extruding 2-D footprints using the height information results in prismatic (i.e., having flat roofs) 3-D building models (Zhou and Neumann, 2008) (Poullis and You, 2009) (Ledoux and Meijers, 2011) (Chen et al., 2014) (Yan et al., 2015). Further appropriate roof modeling enables extension from prismatic to more specific building models. In this context, most roof reconstruction methods make use of the fact that man-made structures such as buildings usually have either parametric shapes (model-driven/top-down strategy) or are composed of polyhedral structures only (data-driven/bottom-up strategy) (Maas and Vosselman, 1999). The latter is however more common in the literature where distinct roof faces are first determined using segmentation-based (region growing or clustering) or fitting methods (3-D hough transform or RANSAC) algorithms (as mentioned in Section 3.1.2 and later reconstructed by employing topological or contextual constraints. Moreover, with high point density, data-driven approaches are able to reconstruct variety of complex building shapes. However, with low point density and noisy datasets, existing data-driven algorithms result in either erroneous or incomplete building models. In such scenarios, model-driven strategies provide more elegant and accurate solution. To improve reconstruction results, few approaches have also attempted to combine both model-driven and data-driven approaches giving rise to hybrid approaches. In the following, model-driven, data-driven and hybrid approaches for 3-D roof modeling/reconstruction are sequentially addressed.

3.1.3.1 Model-driven approaches

Model-driven approaches employ a predefined library of building primitives to reconstruct 3-D building models. An advantage of this strategy is that the pre-defined constraints of adjacent roof surfaces allow generation of regularized and watertight models. In addition, it is relatively easier to define a set of rules with few primitives (or roof shapes) in contrast to organizing individual roof facets derived from data-driven approaches (Huang et al., 2013). Several model-driven approaches have been proposed in the literature in the past two decades. For instance, among early investigations (Weidner and Förstner, 1995) modeled simple rectangular building footprints with symmetric sloped roofs being parametrically reconstructed. (Maas and Vosselman, 1999) also derived parameters for buildings with gabled roofs by analyzing static (or invariant) moments of point clouds. (Brenner and Haala, 1999) also employed 2-D GIS data as an additional data source to reconstruct common roof categories: flat, gable, hip and desk. For more complicated buildings, the concept of predefined building primitives with 2-D ground maps is typically employed (Brenner, 2000). The basic idea is to decompose the whole building/roof into smaller structures which could be easily modeled using a catalog of defined primitives. In this context, (Kada and McKinley,

2009) proposed a cell decomposition approach based on existing 2-D building footprints to reconstruct the complete roof structure by grouping (or assembling) standard individual building blocks, i.e., simple parameterized shapes/primitives (flat, shed, gabled, hipped, berliner), stored in library. The cell decomposition referred to division of a footprint into quadrangular and nonintersecting regions (or cells) depending on estimated plane equations. Each cell was then modeled by respective shape via similarity analysis of estimated normal vectors. (Lafarge et al., 2010) also utilized a library of 3-D blocks to “place” an individual best fitting block onto each 2-D footprint reconstructed either interactively or automatically. A stochastic Gibbs model based on constructive solid geometry representation was employed to control the fitting and assembling process and subsequently a Bayesian framework was adopted to determine the optimal configuration. (Huang et al., 2013) also proposed a similar strategy of reconstructing 3-D roof structure with the difference that they allowed primitive overlapping and in this context proposed new merging and assembling rules. (Arefi and Reinartz, 2013) presented a ridge line based simple parametric model fitting method for tilted (i.e., gable and hipped) roofs. (Henn et al., 2013) exploited 2-D available footprints and presented a fully automatic model-driven approach able to work over sparse point clouds. In their work, the cell decomposition approach presented in (Kada and McKinley, 2009) together with a library of standard roof shapes were employed and model selection was performed in a supervised classification framework using SVM. In similar context, (Zheng and Weng, 2015) also utilized a catalog of seven roof primitives and employed a decision tree classifier over building points extracted from nDSM to choose optimum roof type.

3.1.3.2 Data-driven approaches

With increasing point densities, data-driven approaches provide more generic solution towards 3-D building roof reconstruction. These approaches typically employ polyhedral building structure and therefore follows a common strategy i.e., search for planar segments by segmentation or fitting methods (presented in section 3.1.2.1, estimate model parameters (i.e., to extract linear feature) by determining height jumps (or step edges) and computation of ridge lines via adjacent planes intersections, followed by subsequent regularization to incorporate geometric and topological constraints (Maas and Vosselman, 1999) (Vosselman and Dijkman, 2001) (Rottensteiner, 2003) (Gross et al., 2005) (Forlani et al., 2006) (Dorninger and Pfeifer, 2008) (Sohn et al., 2008) (Sampath and Shan, 2010) (Kim and Shan, 2011) (Rau and Lin, 2011) (Sohn et al., 2012). For the latter two steps, adjacency of extracted roof segments is typically determined by building up a roof topological graph (Verma et al., 2006) (Elberink and Vosselman, 2009) (Perera et al., 2012) (Hron and Halounová, 2015) (Xiong et al., 2015) (also termed as roof adjacency graph (Forlani et al., 2006) or adjacency matrix (AM) (Sampath and Shan, 2010)). Topological relationships of individual (adjacent) roof surfaces are then adjusted according to the set of rules imposing certain constraints in the reconstruction procedure. Instead of explicitly (pre-)specifying rules based on a priori knowledge, (Sohn et al., 2008) proposed a binary space partitioning tree serving as a mid-stage grouping technique to globally adjust (adjacent) linear features for (geometric) topological reconstruction of building roofs. (Sohn et al., 2012) further improved the topological reconstruction by employing a minimum description length based global optimization procedure to select the optimal model among different generated hypothesis. (Lafarge and Mallet, 2012) presented a MRF based approach to retrieve adjacency of primitive labels. (Rau and Lin, 2011) also proposed a line based TIN merging and reshaping method for 3-D roof modeling based on side projection of ALS data. Initial roof topology was reconstructed via geometric analysis of side (ground-elevation axis) projected data. The method then utilized 3-D structural lines to remodel the initial topology while simultaneously reshaping individual roofs. Instead of making certain assumptions related to angles between neighboring roof surfaces (Zhou and Neumann, 2008) (Zhou and Neumann, 2010) utilized

principal directions to adjust the individual roof planes to reconstruct building footprints. Their method essentially supported flat roof tops however interactive extension to non-flat roof surfaces (e.g., cones, cylinder etc.) were also anticipated. (Poullis and You, 2009) and (Ledoux and Meijers, 2011) (Zhu et al., 2015a) also proposed extruded 3-D building models based on 2-D footprints estimated via boundary tracing/regularization and by imposing geometric topological constraints respectively.

3.1.3.3 Hybrid approaches

Some researchers also employed model-driven strategies over initial models obtained from data-driven methods. For instance, (Scholze et al., 2002) employed (model-driven) geometric modeling of polyhedral roofs using probabilistic relaxation labeling technique. (Verma et al., 2006) performed segmentation to retrieve planar roof patches and defined three simple parametric roof topology models which were exploited to reconstruct 3-D models for more complex buildings. (Elberink and Vosselman, 2009) also proposed a target based graph matching algorithm that topologically matched common parametric roof models with the extracted features (both represented as topology graphs). (Lafarge and Mallet, 2012) presented a detailed method that combined 3-D primitives (cones, cylinders and planes) to model buildings using adjacency of primitive labels established via MRF based approach. Based on the assumption that most non-flat roofs could be modeled or decomposed into a set of gabled roof parts, (Fan et al., 2014a) employed RANSAC to detect ridge lines which were later utilized to perform plane fitting based segmentation to obtain individual roof facets. In presence of noise or low point densities, segmentation of individual roof surfaces fails consequently leading to incorrect roof topology graphs. In this context, (Brédif, 2010) and (Lafarge and Alliez, 2013) proposed solutions to cope with incorrect roof topologies prior to 3-D modeling. A graph edit dictionary based technique to automatically correct erroneous roof topologies for 3-D reconstruction of complex roof structures was proposed in (Xiong et al., 2014) (Xiong et al., 2015). In their work, the dictionary containing simple basic elements was exploited to restrict (or enforce) certain configurations ensuring retrieval of correct roof topologies.

3.2 Trees reconstruction

Analysis of single trees has been an active and exciting field of research within remote sensing community since mid 1990's. This is mainly triggered due to the existence of variety of application areas especially in the domain of sustainable forest management, generation of city tree cadastres or to reconstruct complete 3-D city models and/or GIS (Straub and Heipke, 2001). In addition, manual extraction of important forest inventory parameters particularly related to single trees e.g., mean tree height, trunk location, crown diameter, timber volume, identification of tree species etc. is quite cumbersome (Shiver and Borders, 1996) urging the need to develop methods that extract such information automatically. In this regard, numerous approaches have been proposed that particularly employ ALS datasets for individual tree reconstruction.

A common workflow for such existing techniques is to apply segmentation/clustering based approaches e.g., *k*-means (Morsdorf et al., 2004) (Gupta, 2010), watershed-segmentation (Koch et al., 2006) (Chen et al., 2006) (Reitberger et al., 2009) (Zhang et al., 2013b) (Zawawi et al., 2015), region growing (Persson et al., 2002), mean shift (Shahzad et al., 2015) (Schmitt et al., 2015), object-based raster analysis (Höfle et al., 2008) (Höfle et al., 2012), normalized cut (Reitberger et al., 2009) (Yao et al., 2012) etc. either on to computed canopy height model (CHM) (Persson et al., 2002) – analogous to nDSM in urban regions – (also termed as digital tree height model (DTHM) (Hyypä et al., 2001) or tree canopy model

(TCM) (Chang et al., 2013)) or directly over raw LiDAR point clouds to obtain individual tree clusters. Subsequently each segmented cluster is appropriately modeled to derive aforementioned tree parameters. For comparison of different ALS based single tree extraction methods, interested readers are referred to recent benchmarking studies in (Vauhkonen et al., 2012) (Kaartinen et al., 2012) and (Eysn et al., 2015).

3.2.1 DEM based approaches

One of the early investigations of individual tree detection from airborne LiDAR data was presented more than one and a half decade ago (Brandtberg, 1999) (Hyypä and Inkinen, 1999). Since then several researchers have attempted to automatically detect and reconstruct individual trees. For instance, (Persson et al., 2002) identified individual trees by first computing a CHM via active contour based surface interpolation method. Subsequently, tree tops were detected via local height variations using smoothed (via 2-D Gaussian filter) CHM. Tree modeling was then performed via fitting in each segment a second order parabolic surface from which core parameters such as height, crown diameter etc. were estimated. (Koch et al., 2006) also proposed a similar workflow to determine the footprint of individual tree crowns. From detected tree tops, a pouring algorithm similar to classical watershed algorithm (Soille, 1999) was applied to estimate coarse outlines of multiple trees which were later regularized by imposing certain prior knowledge constraints based on tree shapes. (Chen et al., 2006) also employed a marker controlled watershed segmentation algorithm to detect individual trees from smoothed nDSM. Instead of using CHM, identification of tree tops was carried out by searching for local maximum with varying neighborhood (i.e., local window sizes) in a canopy maxima model which represented a modified CHM where each pixel was assigned maximum height value within its local neighborhood. In (Jing et al., 2012), a multi-scale analysis based approach was proposed to delineate tree crowns of different sizes. In this perspective, their approach applied a series of morphological opening operations to the computed CHM to produce multiple marker-controlled watershed based segmentation maps which were subsequently fused to generate a complete map of tree crowns. An approach based on random forests to predict distinct tree attributes including stem volume, tree height and diameter at breast height was proposed in (Yu et al., 2011). (Chang et al., 2013) also identified individual tree crowns by exploiting geometric characteristics such as distributions of local maxima and minima. These characteristics helped in determining treetops and respective crown boundaries which were subsequently utilized to model tree crowns as appropriately fitting circles. (Zhang et al., 2013b) also utilized marked point process model to represent tree crown projections as circles in a CHM. An energy term, consisting of a data term to measure the goodness of fit and a prior term to constrain unrealistic tree shapes, was defined and minimized using a steepest gradient descent algorithm.

3.2.2 Point cloud based approaches

Approaches presented so far work over rastered interpolated image (i.e., CHM) which in addition to possible loss of information may possess inherent errors/uncertainties that essentially degrades the accuracy of the individual tree segmentation (Guo et al., 2010). To overcome this problem, several researchers have opted to segment and reconstruct individual trees directly from 3-D LiDAR point clouds. For instance, (Morsdorf et al., 2004) utilized k -means clustering algorithm to perform segmentation of 3-D points. To overcome the certain limitations of k -means e.g., the dependency on the initial location of cluster centers and the value of k (i.e., number of clusters/trees), initial seed points providing this information were obtained through positions of local maxima in the DSM determined using conventional approach (Hyypä et al., 2001). Since the tree spans are usually more in the height (or z -

) dimension, a scaling factor of 3 in the z-direction was also introduced to allow retrieval of cluster shapes other than favored spherical ones. Geometric reconstruction of individual trees via rotational paraboloid was then employed using derived trunk location, height and crown diameter information from distinctly segmented clusters. (Lindberg et al., 2014) also employed k -means algorithm with the provision to allow more number of clusters than the number of detected treetops. Instead of using raw LiDAR points, (Wang et al., 2008b) and (Li et al., 2012) utilized normalized LiDAR points, obtained from subtraction of heights from raw points, and the rastered DTM, also generated using raw points via active contour algorithm (Weinacker et al., 2004) and clustering approach (Lee et al., 2010) respectively. (Reitberger et al., 2009) also proposed a segmentation approach to detect single trees. The stem positions were first identified in the segmented CHM image via hierarchical clustering procedure and later reconstructed using RANSAC algorithm. A bipartite graph was set up in the voxel space created in the tree regions and normalized cut segmentation (Shi and Malik, 2000) was utilized to retrieve segments representing individual tree points. The voxel representation allowed recovering smaller trees which would possibly be missed if searched in CHM via conventional watershed segmentation. An improved version of the approach which also took into account the prior knowledge about tree location and spatial distribution of LiDAR points was presented in (Yao et al., 2012). In similar domain, (Yao et al., 2014) also carried out a detailed sensitivity analysis of the modules/parameters used for normalized cut segmentation in relation to single tree detection. A graph partitioning based segmentation algorithm that exploited the topological relationships of tree crowns was proposed in (Strîmbu and Strîmbu, 2015). The algorithm captured the tree crown topology in a weighted graph by exploiting adjacent cell patches generated from a rastered height map. Weak edges were then cut to partition the graph and subsequently used to separate distinct tree crowns.

3.3 TomoSAR vs LiDAR point clouds

The aforementioned methods and the majority of other techniques present in the literature that make use of 3-D LiDAR data cannot be directly applied to TomoSAR point clouds due to different microwave scattering properties of the objects appearing in the scene reflecting different geometrical and material features. Consequently, TomoSAR point clouds possess following special peculiarities (see Figure 10) in comparison to ALS point clouds (Zhu and Shahzad, 2014) (Shahzad and Zhu, 2015a):

Accuracy and Errors:

- ◇ TomoSAR point clouds deliver moderate 3-D positioning accuracy on the order of 1 m (Zhu and Bamler, 2012b) as compared to (airborne) LiDAR systems having an accuracy on the order of 5-10 cm (May and Toth, 2007).
- ◇ Ghost scatterers (Auer et al., 2011) may be generated due to multiple scattering that appears as outliers far away from a realistic 3-D position.
- ◇ A small number of images and limited orbit spread render the location error of TomoSAR points highly anisotropic, with an elevation error typically one or two orders of magnitude higher than in range and azimuth (Zhu and Shahzad, 2014).

Coherent Imaging:

- ◇ Due to coherent imaging nature, temporally incoherent objects such as trees cannot be reconstructed from multipass *spaceborne* SAR image stacks. However, for configurations with more than one transmitting and receiving antennas capable to acquire InSAR data in a single-pass may lead to highly coherent data even for vegetation regions by minimizing the effect of temporal decorrelation. The use of such single-pass data to reconstruct individual trees is demonstrated in Chapter 7 of this thesis.

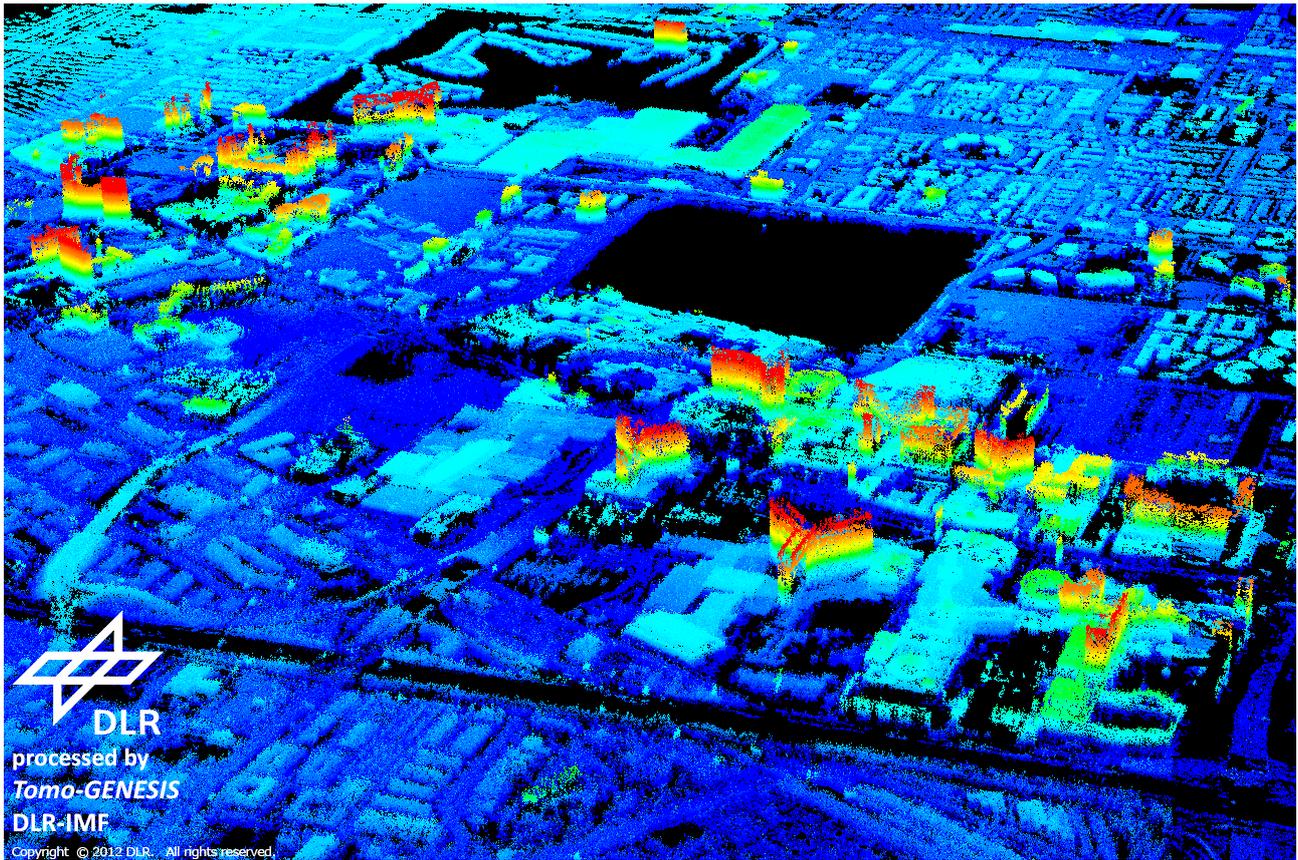


Fig. 10. Spaceborne TomoSAR point clouds of high rise area in Las Vegas processed by DLR's Tomo-GENESIS system (Zhu, 2011) (Zhu et al., 2013). Height is color-coded. It can be seen that the side-looking geometry render rich amount of points/scatterers on building façades. Also due to the coherent imaging nature of SAR and the repeat-pass data-takes, temporally incoherent objects such as trees or vegetation cannot be reconstructed, depicted in black.

Side-Looking SAR Geometry:

- ◇ Separation of layover on vertical structures renders geocoded TomoSAR point clouds to possess higher density of points on building façades;
- ◇ In order to obtain the full structure of individual buildings from space, multiple views are required.

Higher Dimensional Imaging:

- ◇ In addition to 3-D spatial information, spaceborne TomoSAR point clouds also possess the 4-D information, i.e., temporal or seasonal deformation estimates, making them very attractive for dynamic city modeling.

Keeping in view the above special characteristics, the new object reconstruction methodologies have been developed as part of this dissertation. The following three chapters are dedicated to these novel techniques exploiting for the first time this class of data. Chapter 7 then details a prototypical workflow to reconstruct individual trees from 3-D TomoSAR point cloud generated from multi-baseline InSAR millimeterwave data acquired from multi-aspects with an airborne single-pass MEMPHIS SAR system created by the Fraunhofer Institute for High Frequency Physics and Radar Techniques (Schimpf et al., 2002). Subsequently, chapter 8 summarizes and presents future perspectives.

4 Façades reconstruction

This chapter essentially summarizes two journal papers A and B (see the Appendices) which demonstrates for the first time the potential of automatic façade reconstruction from space-borne TomoSAR point clouds generated from data acquired from either single- (ascending only) and/or multi- (ascending and descending) viewing perspectives. In addition to this first demonstration, following are the key highlights/contributions of the developed methodology:

- ◇ Two façade extraction procedures have been proposed where the first method employs a simple grid based scatterer (or point) density (SD) estimation and later apply morphological operation to extract façade pixels/points. The second method on the other hand provides a more robust solution towards façade extraction. It is based on directional SD estimation which incorporates the local façade geometry and estimates the SD for each point in a directional window. Later, robustly estimated 3-D surface normal information is utilized to extract façade points;
- ◇ Automatic segmentation procedures based on the well-known k -means and mean shift algorithms have been proposed to group the extracted façade points into clusters representing individual façades;
- ◇ Façades are modeled in 2-D using general (first and second order) polynomial equations to cater for a wide variety of building footprint shapes. The modeling procedure is able to cater arbitrarily (rotated) orientated building façades while the coefficients of the model are estimated using weighted total least squares (WTLS) method to cope with the localization errors of TomoSAR points in both xy directions;
- ◇ During the reconstruction procedure, the presence of smaller clustered segments occurring at façade transition regions handicaps the accurate determination of vertex points from the adjacent façade pair and may cause the reconstruction procedure to fail. To deal with this problem, smaller “conflicting segments” are automatically identified and removed for an accurate and robust reconstruction of the adjacent façades;
- ◇ A partial solution to the occlusion problem is presented which refines the reconstructed façade footprints via insertion (of additional segments) and extension operations.

4.1 Façades detection/extraction

Due to the side-looking SAR geometry, TomoSAR point clouds when projected onto ground plane exhibit higher SD at vertical façade regions. It is mostly true due to the presence of strong corner reflectors (e.g., window frames) on the building façades. Taking this fact into account, the first step towards extraction of façade points is to obtain point density estimates. To this end, an approach that exploits the idea of orthogonally projecting the points onto the 2-D ground plane as presented in (Dorninger and Pfeifer, 2008) has been developed. However, instead of estimating local planes to refine the building outline, the 2-D scatterer density in the horizontal xy (ground) plane is used to extract the façade points. SD is locally estimated for each grid point defined on the ground plane by first accumulating the number of points within a local window and then dividing by the window size. A specific threshold value TH is then used to remove low density grid cells. As a consequent, a binary mask is produced which is later exploited via morphological dilation operation to extract building façade cells. Connected component analysis then leads to the identification of individual building masks.

This simple method of SD estimation works well for high-rise buildings having a much higher point density on façades (compared to non-façade areas) but limits the extraction of

façade points from relatively lower buildings. Moreover automatic selection of a particular threshold becomes crucial. To resolve this issue, a robust M-estimator based SD estimation is employed in this section which provides much better estimates of façade regions compared to previously developed grid- (or sliding window-) based SD estimation method by incorporating the façade geometry (Wang and Zhu, 2015). Moreover, instead of rejecting non-façade points via 2-D morphological operations, robust 3-D surface normal information is utilized. The use of additional-dimensional along with the vertical façade assumption helps in better rejecting non-façade points (please refer to B for comparison).

4.1.1 Robust M-estimator based directional SD estimation

For each 3-D TomoSAR point p , points within its local neighborhood v_c are used for the SD estimation. v_c includes all those points that lie inside a vertical cylinder (of infinite height) horizontally centered at p . To emphasize the building façades, the façade geometry is incorporated in the SD estimation by determining the façade position and orientation direction of the local neighborhood via line fitting using robust M-estimator. The method computes M-estimates by iteratively applying weighted least squares to the objective function $\arg \min_{\hat{\beta}} \sum_{p_i \in v_c} w_{p_i}(\hat{\beta}) |y_{p_i} - f_{p_i}(\hat{\beta})|^2$ where weights w_{p_i} are computed using bisquare function of the form (Street et al., 1988):

$$w_{p_i} = \begin{cases} (1 - u^2)^2 & \text{for } \text{abs}(u) < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{with } u = \frac{|y_{p_i} - f_{p_i}(\hat{\beta})|}{4.685 \hat{\sigma} \sqrt{1 - t}} \quad (11)$$

where β are estimated line parameters that iteratively updated, t is the leverage computed from least squares fit, and $\hat{\sigma}$ is the scale of the error term computed by $\hat{\sigma} = 1.483 * \text{MAD}$, where MAD is the median absolute deviation of the residuals from their median. The term 1.483 is used to make the estimator consistent for the standard deviation estimation of Gaussian distribution and has been practically used for robust initializations in M-estimates (Street et al., 1988) (Huber, 1981).

The estimated line describes the main principal axis of the cylindrical footprint of the local neighborhood. Orthogonal distance of every point in v_c with respect to the principal axis (shifted in parallel to the point p) is then calculated. The points having distances less than d are taken as “inliers” and used in SD estimation. SD for each point is thus defined as the number of points within a directional (cylindrical) neighborhood in 3-D divided by the area in 2-D:

$$\text{SD} = \frac{\text{number of points in } v_d}{\text{Area of } v_d} \quad (12)$$

where $v_d \subseteq v_c$ but includes only those points that lie close to the principal axis of points in v_c .

Figure 11 shows the graphical procedure of the proposed robust SD estimator. Applying a hard threshold to the estimated SD, similar to grid-based approach (i.e., using higher TH value for thresholding) would lead to miss-detection of façade points of lower buildings whose SD estimates would be relatively low with respect to high rise buildings. To avoid such miss-detection, the façade points are extracted in a sequential way: I.e., firstly, a lower TH value is used to detect not only higher buildings but also the lower ones. However, a softer threshold would also introduce false positives, i.e. roof points or ground points with

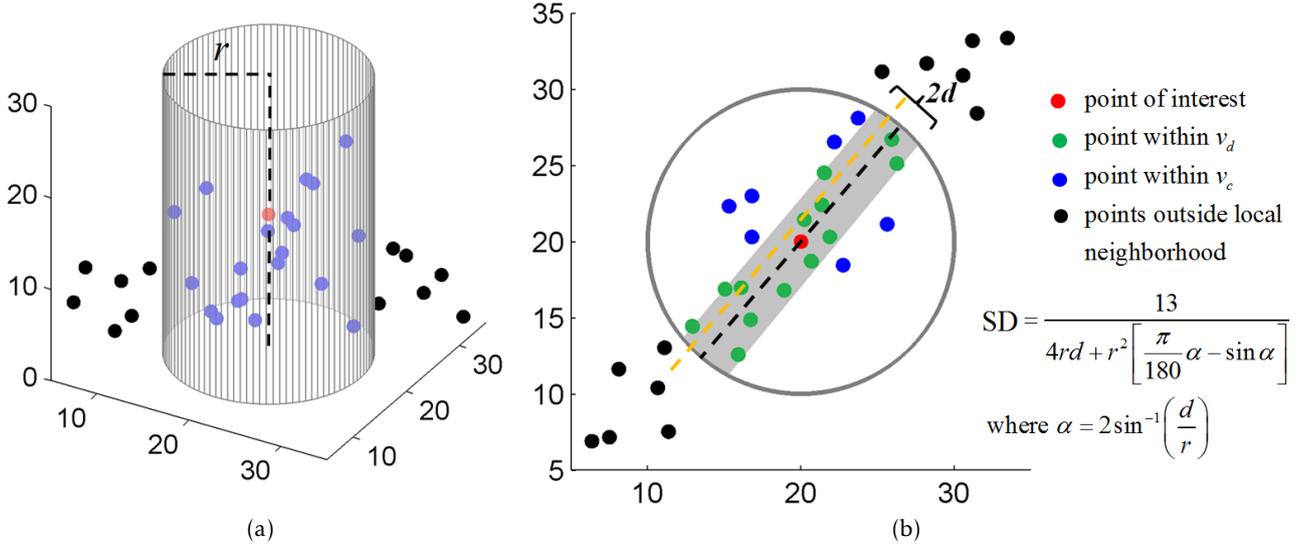


Fig. 11. Illustration of SD estimation: (a) 3-D view of the local cylindrical neighborhood around the point of interest; (b) Top view of (a). The coefficients of the dotted yellow regression line are estimated via M estimation. Black dotted line shows the shift of yellow line to the point of interest. Shaded area shows the region of v_d within v_c .

a local point density comparable to those of lower buildings. Therefore, a second step is introduced which utilizes 3-D surface normals information as prior knowledge (i.e., façades are assumed to be vertical surfaces which separate them from non-vertical ground plane and roofs) to eliminate those false positives.

4.1.2 Computation of surface normals

As mentioned in (Section 3.1.2.1), surface normals can be directly estimated for each 3-D point via eigenvalue/eigenvector analysis of the local 3-D (i.e., 3×3) covariance matrix Σ_{v_c} . Use of eigenvalue/eigenvector analysis via classical PCA may however fail to give precise 3-D surface normal estimates using TomoSAR point cloud due to considerable amount of outliers and localization errors in the data (see Section 3.3). Robust covariance matrix estimation is therefore needed. In this work, robust minimum covariance determinant (MCD) method (Hubert et al., 2005) is utilized.

The covariance matrix $\hat{\Sigma}_{v_c}$ estimated using the local neighboring points $p_i \in v_c$ around (in cylinder) the point of interest $p_o(x_o, y_o, z_o)$ is then used to determine the local 3-D surface normal at p_o . If a plane which fits the neighboring points p_i is denoted as $n_x x + n_y y + n_z z + \rho = 0$ with $\rho = -n_x x_o - n_y y_o - n_z z_o$, then $N_o(n_x, n_y, n_z)$ depicts the local 3-D surface normal at p_o . N_o is thus directly estimated from Σ_{v_c} by computing the eigenvector associated to the smallest eigenvalue of Σ_{v_c} (here v_c includes points in the vicinity of p_o) i.e.,

$$\begin{aligned} & \text{if } \Sigma_{v_c} \cdot \mathbf{v}_{\lambda_j} = \lambda_j \cdot \mathbf{v}_{\lambda_j}, j = 1, 2, 3 \text{ (descending order)} \\ & \text{then surface normal of the underlying surface at point } p_o: N_o(n_x, n_y, n_z) = \mathbf{v}_{\lambda_3} \end{aligned} \quad (13)$$

From (13), 3-D surface normals are robustly computed for each point that is obtained after SD thresholding. Ideally, the direction of surface normal should be parallel to the ground for points on the vertical façades which separate them from non-vertical ground plane and roofs. Taking this fact into account, façade points are extracted out by retaining only those points having normals are close to the horizontal axis (i.e., parallel to ground for points belonging to a vertical surface). The proposed two step approach allows to robustly extract façade points over a large area where both high and low buildings are present.

4.2 Segmentation

For reconstruction, the extracted points need to be segmented such that each cluster represents points belonging to an individual façade. For this purpose, an approach utilizing k -means clustering with a criterion for prior guessing the number of clusters has been initially developed.

4.2.1 Number of clusters for k -means

A major shortcoming of k -means is that it is required to provide in advance the value of k (i.e., the number of clusters) which is not practical. To deal with this problem, a method to determine k is devised which works by estimating the within-cluster dispersion in successive clustering runs for varying numbers of clusters. I.e., if D_r is defined as the mean deviation of points in r th cluster from its respective center:

$$D_r = \sum_{i=1}^{n_r} \frac{d_i}{n_r} \quad (14)$$

where n_r is the number of points in r th cluster and d_i is the Euclidean distance of the i th point in r th cluster from its center. The dispersion index I_k for k clusters can then be determined as (Sampath and Shan, 2010):

$$I_k = \sum_{r=1}^k \frac{D_r}{k} \quad (15)$$

A plot of such dispersion index against the number of clusters gives an indication how to choose an appropriate number of clusters (Sampath and Shan, 2010). The dispersion index I_k usually decreases significantly with increasing number of clusters and becomes steady afterwards. The location of the elbow point can be considered as a good estimate of the number of clusters (Tibshirani et al., 2001).

4.2.2 Proposed segmentation scheme

Although the elbow point can be used to estimate k and has been applied in (Sampath and Shan, 2010) (Zhu and Shahzad, 2014), it is not always trivial to determine it automatically especially when applied to data of larger areas. Moreover with k -means certain shape of clusters is not very well recognized. For these reasons, an alternative three-step automatic (unsupervised) clustering approach has been developed that is able to work directly on bigger areas without requiring any prior knowledge about the number of clusters. Following is the brief description of the approach:

- (1) First a density-based clustering method (Ester et al., 1996) is applied to extracted façade points. The approach involves the notion of density connectivity between the points., i.e., two points are considered to be directly density connected to each other if one is in the neighborhood vicinity of the other point. If the two points are not directly connected to each other, still they can be density connected to each other if there is a chain of points between them. Thus starting from a point, all points that are density connected to each other are clustered into single group representing an individual building. These clustered points are then removed and the procedure is repeated for remaining points until all the points are assigned to a particular cluster;

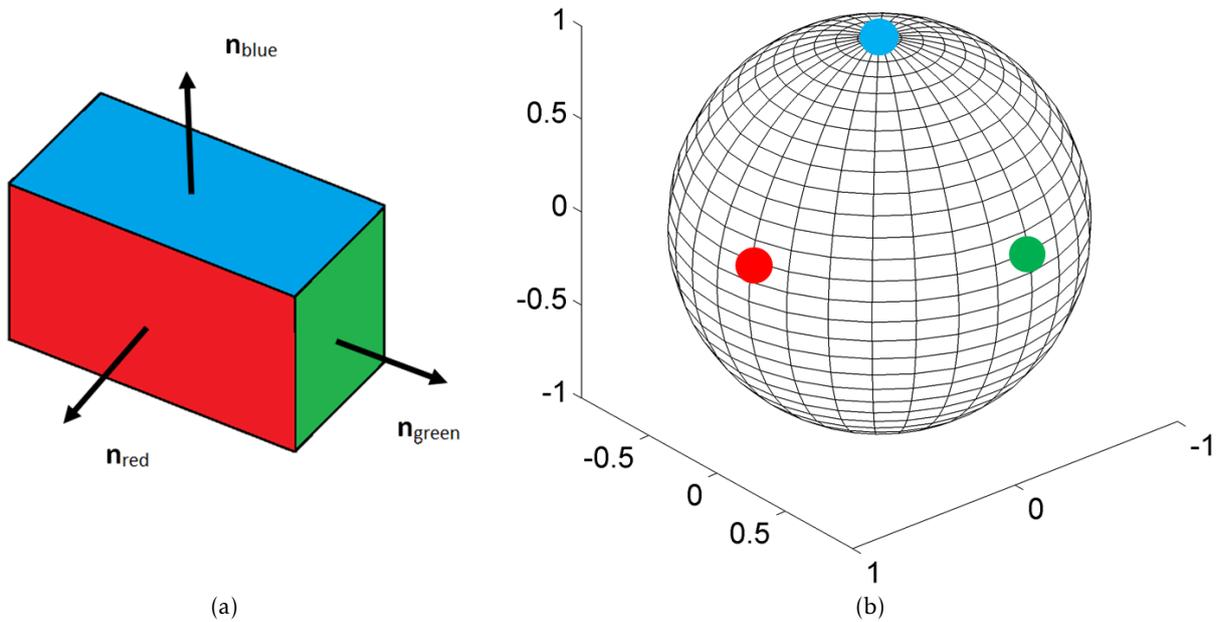


Fig. 12. Gaussian image of three connected planar surfaces: (a) Arrows indicate surface normal vectors ($n_{\text{red}}, n_{\text{green}}, n_{\text{blue}}$) to the respective surfaces; (b) All points belonging to one particular surface are mapped to same identical point in GI (ideal scenario).

- (2) Gaussian image (GI) based mean shift clustering algorithm (Cheng, 1995) (Liu and Xiong, 2008) has been developed which is applied to each density clustered segment representing points belonging to an individual building. Let's assume that a coarsely clustered segment K_i consist of one or more vertical adjacent façades $F_j, j = 1, \dots, J$. An image of a map $M: F \rightarrow F^2$ that assigns each point in F to its respective unit surface normal is known as Gaussian image (or map) GI of F (Carmo, 1976). Flat F (i.e., planar surface) should ideally be represented by a point in GI. Figure 12 conceptually illustrates this in an ideal scenario. In practice, surface normals are estimated locally and may fluctuate from one point to another as practical data often contains errors in 3-D positions. But, if the estimation of normals is robust enough, a surface mapped to GI should be represented as a dense cluster of points in GI. The shape of clusters in GI corresponds to the geometry of connected surfaces (Liu and Xiong, 2008). Mean shift algorithm is employed to perform clustering in the GI.
- (3) Mean shift when applied in GI produce clusters whose corresponding points in spatial domain represent different façades. However, it is also possible that spatial points corresponding to any one particular normal cluster in GI may belong to two or more different façades. This can happen if points of two or more façades that are "nearly" parallel to each other (i.e., having close normal directions) are present in K_i . To resolve this, density based clustering is performed again in the resulting clusters for spatial separation of parallel façades points clustered into one group. Finally, clusters with very few points are removed from further processing for robust reconstruction.

4.3 Reconstruction

4.3.1 Model identification (flat or curved)

Each cluster is further classified into flat or curved surface by analyzing derivatives θ (see Figure 13). θ for each 3-D point is equal to the azimuthal angle of the corresponding com-

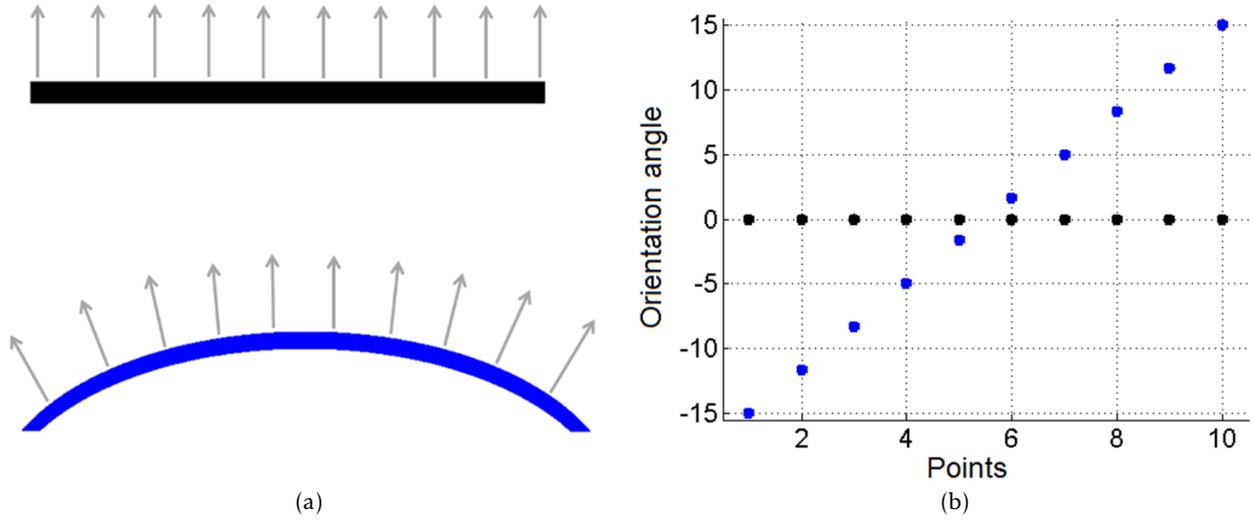


Fig. 13. Illustration of orientation angle for flat and curved vertical footprints (top view); (a) Arrows indicate pattern of change in orientation (azimuthal) angles of ten points on each vertical surface; (b) plots their respective orientation angles.

puted surface normal:

$$\theta = \arctan\left(\frac{\lambda_{3y}}{\lambda_{3x}}\right) \quad (16)$$

where λ_{3x} and λ_{3y} represents the x and y components of the surface normal \mathbf{v}_{λ_3} of any 3-D point. Ideally, the flat surfaces should have constant orientations, i.e., zero derivatives compared to the curved surfaces that have gradually changing orientations (see Figure 13). This fact is exploited and the first derivative θ' of the orientation angle for each façade footprint is computed. Since the original orientation derivatives θ' are usually noisy, all the points are first projected to the first principal axis and polynomial fitting is later applied for denoising. Based on the behavior of θ' , façade footprints are classified as flat or curved.

4.3.2 Modeling of façades

Identified façade clusters in xy plane are then modeled using the following general polynomial equation:

$$f_p(x, y) = \sum_{q=1}^p a_q x^i y^j \quad i + j \leq q \quad (17)$$

where i and j are permuted accordingly, p is the order of polynomial, the number of terms in the above polynomial is equal to $\frac{(p+1)(p+2)}{2}$. Cross terms are introduced in the model in case of rotated local coordinate system. Eq. (17) is solved by restricting to only 1st and 2nd order (i.e., flat with $\max(i, j) = 1$ & curved with $\max(i, j) = 2$). The coefficients a_q are estimated using WTLS method where total least squares is utilized to cope for localization errors of TomoSAR points in both xy directions and the weight of each point is assigned equal to its corresponding SD. The weighted polynomial fitting (residual) error f_{err} is minimum when the local coordinate system in which the façade points exist lies in axis or parallel (i.e., unrotated) with the global coordinate system. In this case, the right hand side of (17) reduced (for $p = 2$) to $\sum_{i=0}^p a_i x^i$ (i.e. with no cross terms). For a rotated local coordinate system (which is

often the case), the following steps are performed to obtain consistent parameter estimates of all façades in a global coordinate system:

- ◇ Rotate the points by rotation angle ω and compute polynomial fitting error f_{err} by applying WTLS method;
- ◇ Consider coefficients computed with ω_{min} that gives the minimum polynomial fitting error f_{err} as polynomial terms depicting unrotated points in the global coordinate system. ω_{min} is computed by using an unconstrained nonlinear optimization procedure to find the minimum of the error function f_{err} by varying ω over 0~360 degree range via Nelder-Mead simplex algorithm (Nelder and Mead, 1965);
- ◇ Rotate the computed polynomial by replacing the unrotated $(x-, y-)$ axis terms by their rotation counterparts $(x \cos \omega + y \sin \omega, -x \sin \omega + y \cos \omega)$ to yield polynomial terms a_q in global coordinates.

After estimation of model parameters, the next step is to describe the overall shape of the building footprint by further identifying adjacent façade pairs and determining the intersection of the façade surfaces. The adjacency of façades is usually described by AM that is built up via connectivity analysis (Sampath and Shan, 2010). Identified adjacent façade segments are used to determine the vertex points (i.e., façade intersection lines in 3-D) by computing the intersection points between any adjacent façade pair. Determination of these intersection points can sometimes become difficult if the transition points are segmented as isolated small clusters (also referred to as conflicting segments) rather than part of the corresponding adjacent façade segments. As a consequence, it gets complicated to find a legitimate adjacent façade pair from which intersection points should be computed. To resolve this issue, the vertex points are computed from the intersection of *valid* adjacent segments that are identified as the two largest segments present at the façade transition regions (see Figure 14).

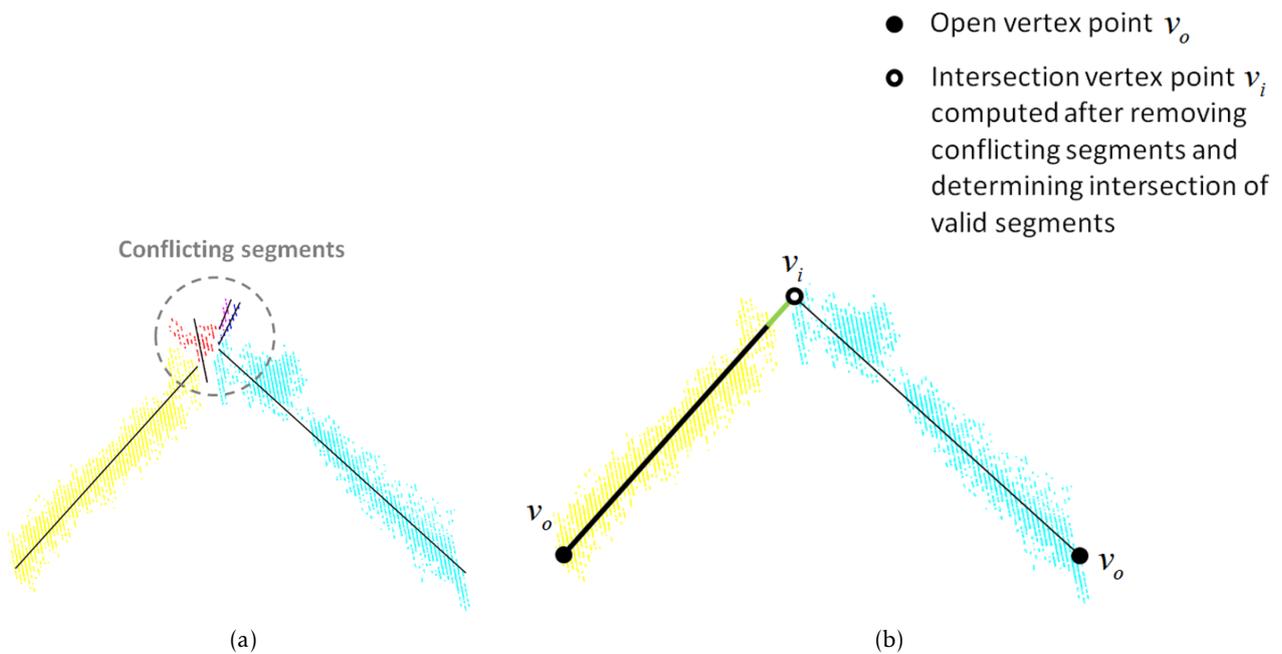


Fig. 14. Removal of conflicting segments and vertex point computation: (a) shows the case of five clustered point segments belonging to one particular building. At the transitional regions, there are conflicting (smaller) clusters, shown in grey dotted circle, that behave as noisy segments and should be removed before computing the vertex points; (b) shows the retained segments and their corresponding vertices after removing conflicting segments.

4.3.3 Refinements

Sometimes the reconstructed façades remain either incomplete or are broken into more than one segment due to the following reasons: 1) Higher building structures present nearby can partly (or fully) occlude the façades of lower buildings, and 2) due to the geometrical shape, only very few points are available at some parts of building façades. In order to refine the reconstruction, a heuristic procedure has been proposed that tries to refine the reconstructed façades by inserting additional segments between the broken regions and extend those façades that remain incomplete. Computed vertex points are therefore first categorized into two types: First type consists of vertices that are computed from the intersection of two adjacent façades, while the second type consists of the other vertices representing “open” endpoints. Now, $Seg_i = \{s_j | j = 1, \dots, n\}$ represents a series of n segments (i.e., $s_1 \sim s_n$) that are adjacently connected. E.g., in Figure 15(a) two such series exist where the first comprises of segments 1 and 2 while the second consist of segment 3 only. In addition to this, for each series Seg_i , there exists two open vertices and $(n - 1)$ intersection vertices. An example where no open vertex would exist is when all façades are visible to the sensor (i.e., a multi-view configuration) in which the complete footprint of the building (i.e., a closed loop polygon) would be reconstructed.

For broken façades, two segments are considered as part of the same (broken) building façade if both segments are not far enough from each other and at the same time possess data points in between that have close maximum height values. For this task, following three conditions comprising of local orientations and height values of 3-D points within neighborhood vicinity of v_o , \hat{v}_o and mid_{v_o} are checked:

$$\begin{aligned}
 C1 : & \|\hat{v}_o - v_o\|_2 < 2\varepsilon \\
 C2 : & abs(h_{\max \text{ at } v_o} - h_{\max \text{ at } mid_{v_o}}) < T_h \ \& \ abs(h_{\max \text{ at } \hat{v}_o} - h_{\max \text{ at } mid_{v_o}}) < T_h \\
 C3 : & abs(\theta_{v_o} - \theta_{\hat{v}_o}) > 45^\circ
 \end{aligned} \tag{18}$$

v_o represents one of the two open vertex points of the considered series Seg_i , \hat{v}_o represents the nearest open end vertex point from v_o that belongs to another adjacently connected series Seg_j ($j \neq i$), mid_{v_o} is the midpoint of v_o and \hat{v}_o , and h_{\max} is taken as the mean of at least ten maximum height values. The two parameters ε and T_h represent thresholds to the distance and the height difference respectively. A single façade segment with vertices (v_o, \hat{v}_o) is inserted to fill the empty (i.e., broken) regions of the same façade if both C1 and C2 are true but C3 is false. On contrary, if all three conditions are met indicating that the open vertex pair $v_o \hat{v}_o$ is not part of the same façade but rather belong to two different façade segments (determined via difference in the local orientation angle $> 45^\circ$), then two segments with vertices $(v_o, p_{v_o \hat{v}_o})$ and $(\hat{v}_o, p_{v_o \hat{v}_o})$ are inserted where $p_{v_o \hat{v}_o}$ denotes the point of intersection. Figure 15(b) graphically depicts such a situation where grey open vertices of segment 4 and 5 are (assumed to be) within 2ε distance but have difference in the local orientation angle of 90° . The grey dotted line shows the addition of new segment without checking condition C3. When C3 is taken into account, two segments shown in black dotted line are inserted.

Subsequent to insertion operation or when any of C1 or C2 fails, the extension operation at the open vertex point v_o as depicted in Figure 15(c) is carried out by imposing following two

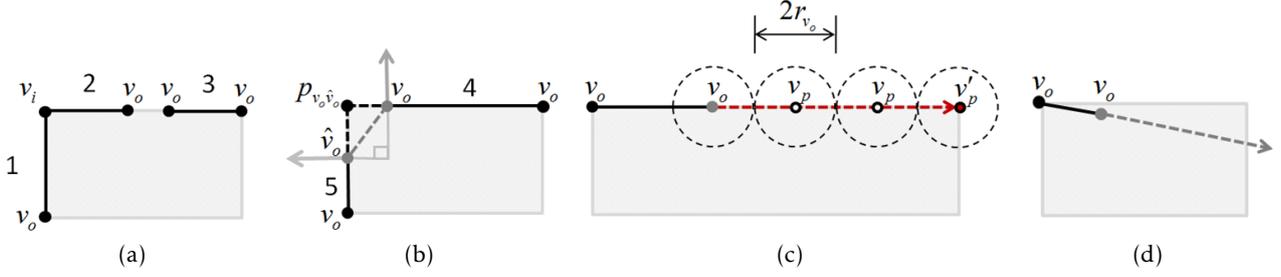


Fig. 15. Vertices for refinement. Grey rectangles depicts the 2-D building footprint from the top: (a) shows the total of 5 vertices out of which 4 are open endpoint vertices and one is intersection vertex computed from the intersection of segments 1 and 2; (b) depicts the situation where ignoring condition C3 would yield false segment addition shown as grey dotted line. The grey arrows indicate the local orientation angle θ at open vertices v_o and \hat{v}_o . Two black dotted lines represent the two correct inserted segments between v_o and \hat{v}_o ; (c) depicts the recursive extension procedure of the open vertex v_o . v_p represent the intermediate extension points where as the v'_p denotes the final point; (d) illustrates the direction of extension of grey open vertex over the roof region. This can happen if we only consider the h_{\max} and ignore the local standard deviation of height h_σ .

constraints:

$$\begin{aligned}
 C4 &: \text{abs}(h_{\max \text{ at } v_p} - h_{\max \text{ at } v_o}) < T_h \\
 C5 &: \text{abs}(h_{\sigma \text{ at } v_p} - h_{\sigma \text{ at } v_o}) < T_\sigma
 \end{aligned} \tag{19}$$

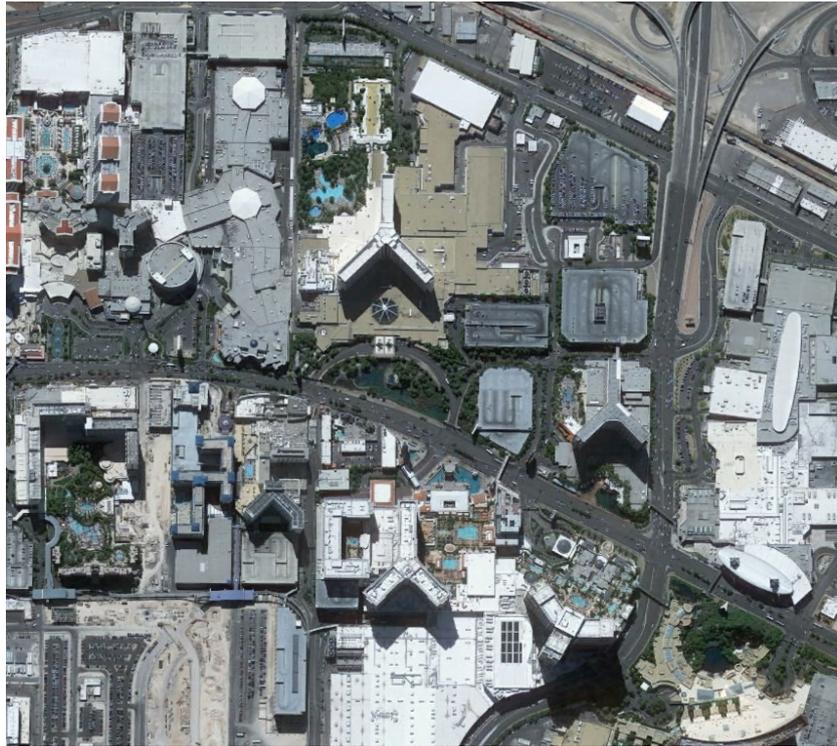
Similar to C2, the condition C4 ensures that the extended point have the closer maximum height value. The condition C5 ensures that the local 3-D points around v_p have certain standard deviation in the third (i.e., height) dimension denoted as h_σ at v_p . It is necessary to make sure that the extension is not carried out in the direction that deviates from the façade footprint. I.e., it avoids the extension if the local 3-D points around v_p belongs to other non-façade objects, e.g., roofs etc., rendering h_σ at v_p to exhibit low value consequently failing condition C5. The problem is illustrated in Figure 15(d) where the grey open vertex can potentially extend over the roof region if the condition C5 is ignored. Thus imposing this constraint helps in limiting this false extension.

Finally, the computed vertex points (i.e., the intersection vertices and the open vertices before and after refinement) along with their estimated model parameters are used to reconstruct the 3-D model of the building façades.

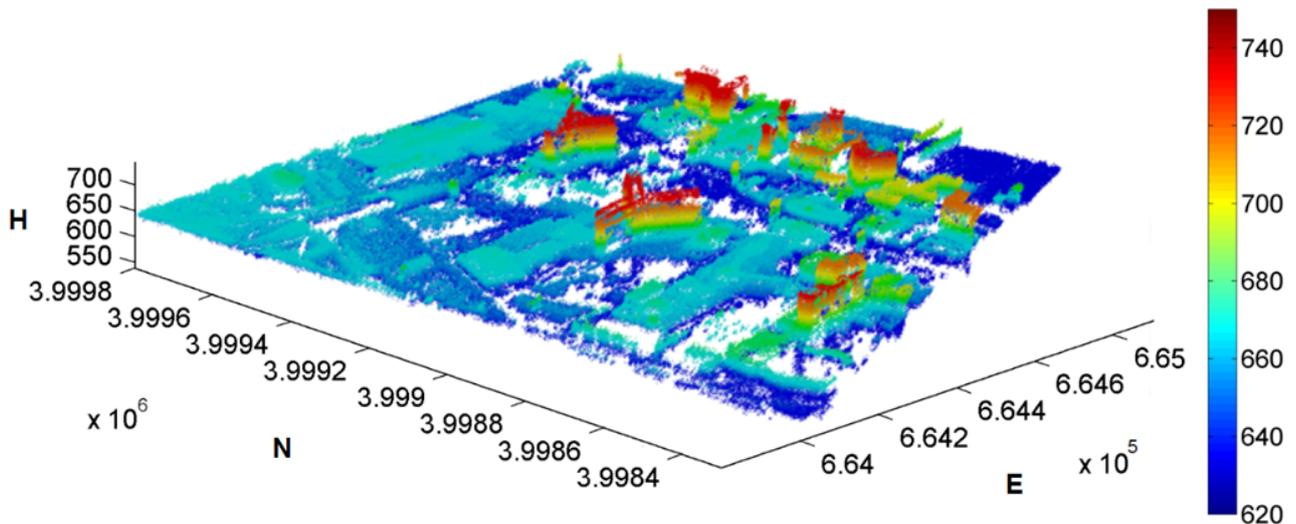
4.4 Experimental results and validation

4.4.1 Dataset

To validate the approach, the algorithm is tested on TomoSAR point clouds generated from a stack of 25 TerraSAR-X high spotlight images from ascending orbit only using the TomoSAR module of the GENERIC System for Interferometric SAR (Tomo-GENESIS) software developed at the German Aerospace Center (DLR) (Zhu, 2011) (Zhu et al., 2013). Figure 16(a) depicts the test area covering approx. 2 km² high rise part of the city of Las Vegas. The number of TomoSAR points in the area of interest is about 1.2 million. Figure 16(b) shows the corresponding TomoSAR point cloud in universal transverse mercator (UTM) coordinates.



(a)



(b)

Fig. 16. Dataset: (a) Optical image of the test area in Las Vegas. Copyright Google. (b) TomoSAR points in UTM coordinates of the corresponding area in Las Vegas generated from a stack acquired from ascending orbit only. The height is color-coded.

4.4.2 Results

The result of applying SD estimation procedure is illustrated in Figure 17. The two parameters r (radius of the neighborhood cylinder) and d are empirically set to 5m and 0.9m respectively according to the point density of the data set. The TH value influences the number of extracted façade points. Lower TH value results in higher completeness but lower correctness. The empirical evaluation of estimating SD with varying TH values result in best trade-off of $TH = 2$ pts/m² in terms of completeness and correctness with this class of data (Please see A for details of empirical evaluation). 2 pts/m² works well for high rise buildings but might ignore relatively smaller façades. Therefore to extract lower façades (and also to automate the procedure), the TH value is set to the maximum of SD histogram value (see

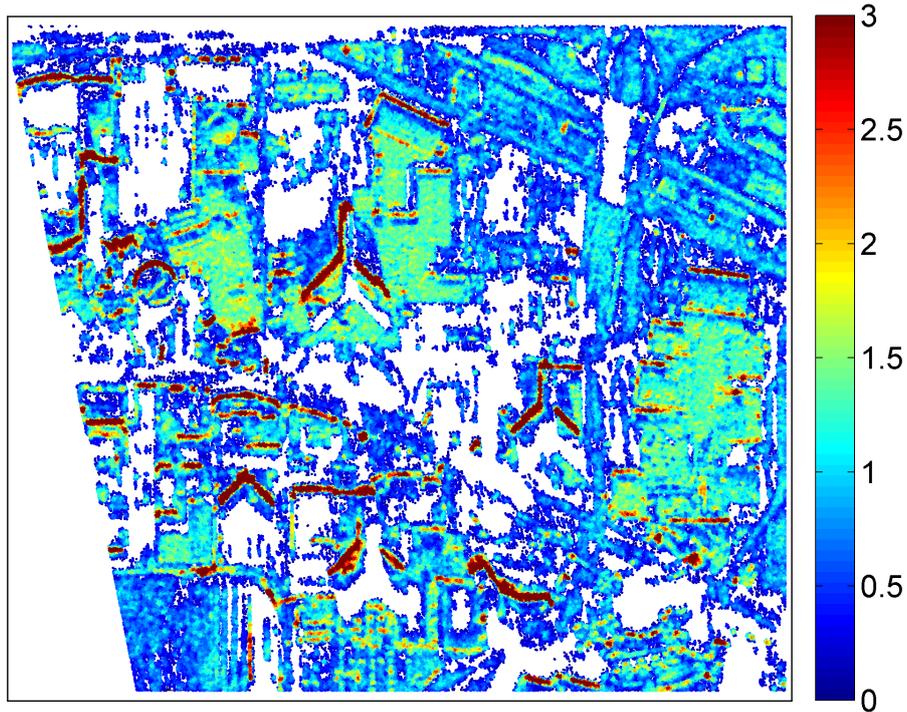


Fig. 17. SD estimated with radius $r = 5\text{m}$ and inliers $d = 0.9\text{m}$. Colorbar indicates computed SD.

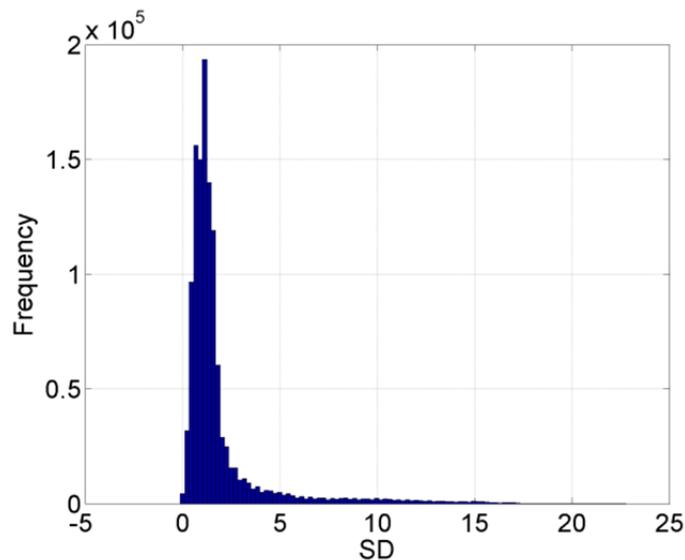


Fig. 18. Automatic selection of TH value. SD histogram for the TomoSAR points of the corresponding test area depicted in Figure 16. The maximum SD histogram value used for thresholding is 1.132. Number of bins is set to $\text{bins} = \text{min}(\text{SD}) : (\text{max}(\text{SD}) - \text{min}(\text{SD})) / 100 : \text{max}(\text{SD})$. It is worth mentioning here that for a rural or sub-urban area, it is possible that the maximum of histogram occurs at $\text{SD} = 0$ which would consequently render all points to be retained. However for urban regions, this usually never happens due to existence of higher point densities over man-made structures.

Figure 18). This, as described in Section 4.1.1, includes not only the façade points but additionally also some non-façade points with relative high SD, e.g., roof points. To reject these points from the set of extracted points after SD thresholding, surface normals information is utilized.

Figure 19 shows the extracted façade points by retaining only those points having normals between ± 15 degrees from the horizontal axis (or equivalently ± 90 degrees from the vertical axis). Extracted façade points are then coarsely clustered based on connectivity of points. Subsequently, in order to reconstruct individual façades, mean shift clustering is applied



Fig. 19. Extracted façade points from TomoSAR point cloud depicted in Figure 16 overlaid onto the corresponding optical image (©Google).

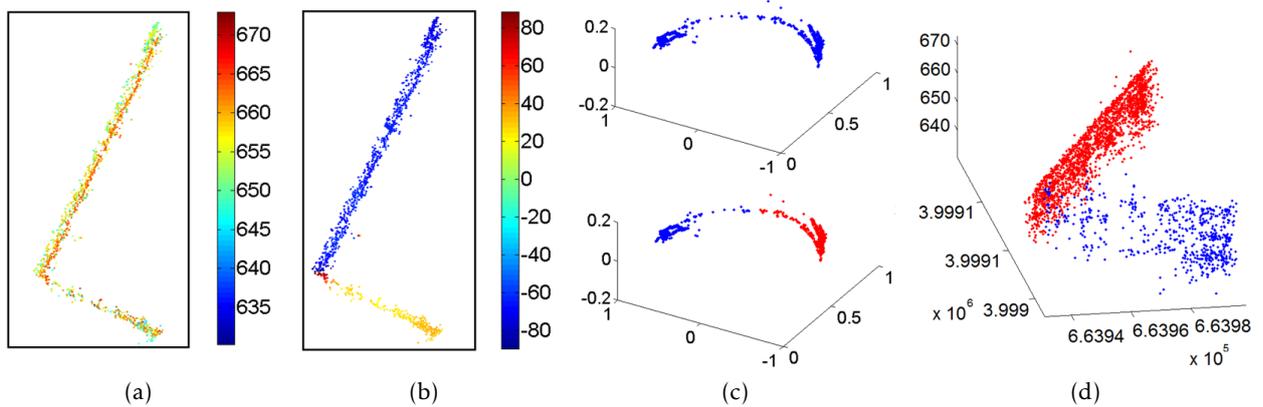
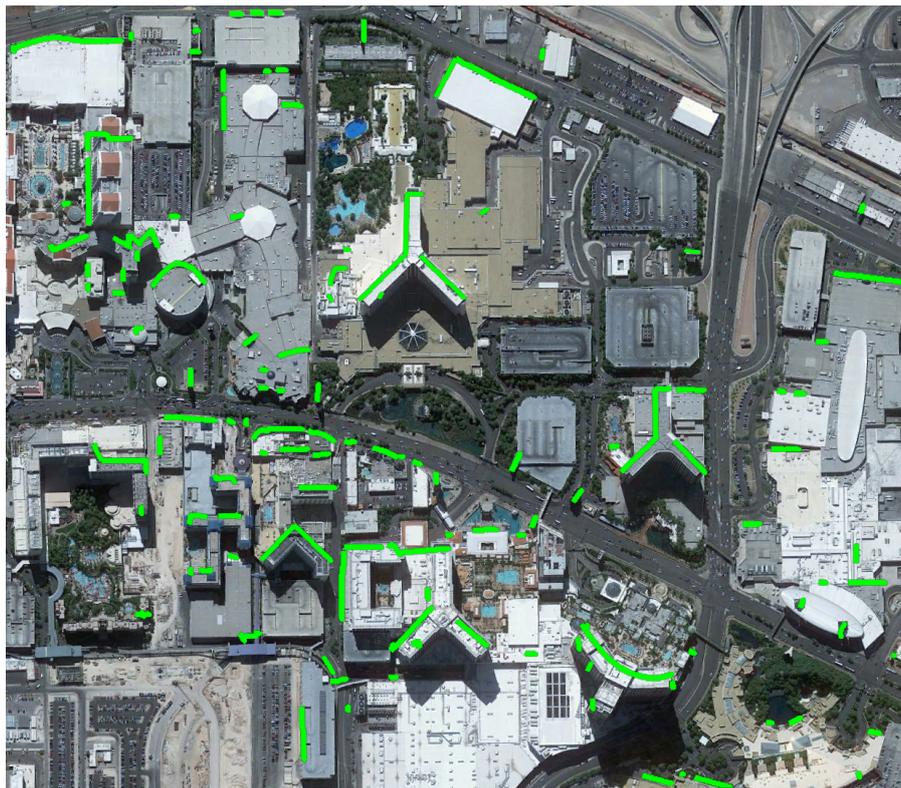


Fig. 20. Fine clustering results after applying mean shift clustering using Gaussian kernel with bandwidth $b = 0.4$ to the coarsely clustered segments in their normal feature space (in GI domain): (a) TomoSAR points of one particular density connected cluster (top view). Colorbar indicates height in meters; (b) Corresponding orientation angle in degrees; (c) Non clustered (top) and clustered (bottom) points in the Gaussian image of points in (a); (d) Resulting clustered points in 3-D.

in normals feature space (in GI domain) to the obtained density based coarse clusters. Figure 20(b) shows the estimated orientation angle θ for extracted façade points from single building shown in Figure 20(a). The variation in orientation angle is quite evident and allows mean shift to cluster points having similar orientations together. Further separation of points in the spatial domain is also required in some cases where the spatially separated points are clustered into one segment. This happens when these points belonging to different façades have similar normals and are spatially closer. Density based clustering is therefore again applied for spatial separation of the clusters within clusters.

For reconstruction, the segmented façades are first classified to flat and curved surfaces by analyzing derivatives of the local orientation angle θ . A slope value θ' of 0.3 (≈ 17 degrees) is set by empirically testing the computed orientation angles of all the buildings in the area of interest to distinguish flat and curved surfaces. After identification, appropriate model parameters are estimated from the core points of the individual clusters. Figure 21 (a) and (b)



(a)



— Reconstructed façades without refinement
 — Additional façades added between the broken facade regions
 — Extension of façades after adding additional segments

(b)

Fig. 21. Reconstructed façades: (a) shows the 2-D view of the façade footprints overlaid onto the optical image prior to refinement; (b) shows the 2-D view of the façade footprints overlaid onto the optical image after refining with parameters settings $T_h = 5\text{m}$ and $T_\sigma = 2.5\text{m}$.

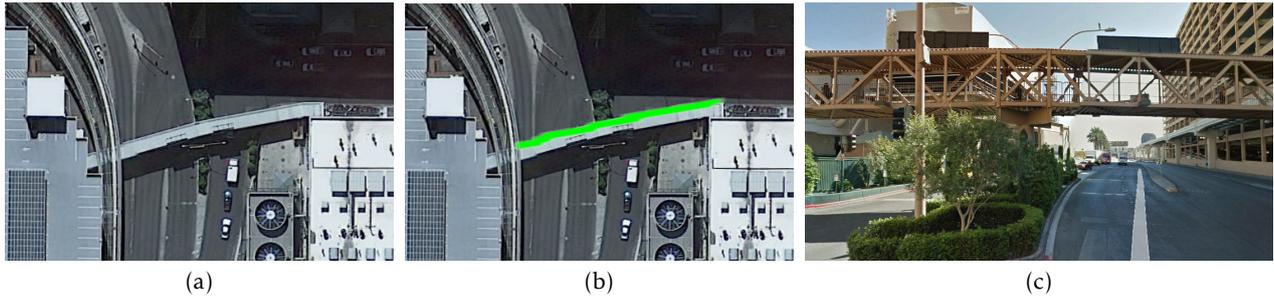


Fig. 22. A case depicting wrong reconstruction of a pedestrian bridge: (a) shows the optical image of the bridge; (b) overlays the reconstructed segment in green onto the optical image of (a); (c) side view of the bridge (© Google Street View).

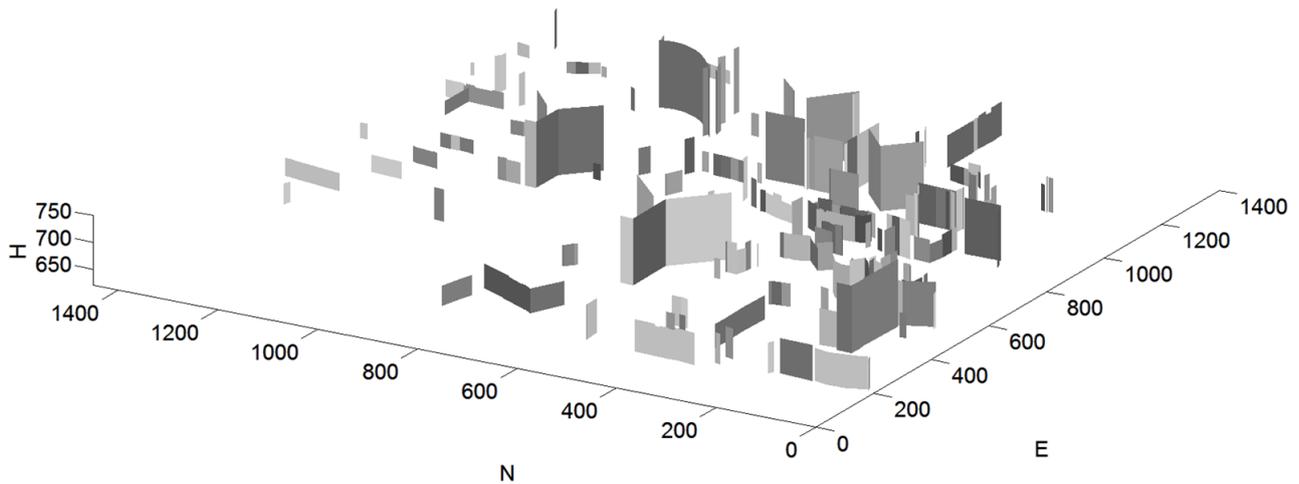


Fig. 23. 3-D view of the final façade reconstruction. The axis is in meters range and has been translated to the origin for better metric clarity by subtracting UTM easting and northing values by their respective minimum values present in the reconstructed vertices.

depicts the reconstructed façades models in the area of interest before and after refinement, respectively.

4.4.3 Validation and discussion

The actual ground truth data is missing for exact qualitative evaluation of the approach. In order to provide some quantitative measures, the actual number of façades that had to be reconstructed were manually counted. Total of 141 façades are present in the dataset out of which 7 are curved façades and remaining 134 are flat. Prior to refinement, the algorithm reconstructed in total of 176 façades, i.e., higher than the actual façades present in the dataset. As already stated in Section 4.3.3, this is because some individual façades have been broken down into two or more segments due to discontinuity in the number of points available in the dataset. After refinement, 29 insertion segments (27 single and 2 double based on the condition $C3$ in eq. (18)) are added between the broken façade regions where as 43 façades have been extended. In the final reconstruction, 147 reconstructed façades are obtained i.e., all 141 façades are successfully reconstructed; among them 5 façades remain broken (counted as additional 5 façades) and there is one case of false alarm which is actually covered metallic pedestrian bridge (see Figure 22(c)). Besides these 5 cases, there exist 7 façades that are not extended and therefore remain incomplete. This is however due to the inadequate number of points available in the data. Finally, Figure 23 provides the 3-D view of the reconstruction results. Such a reconstructed façade model can be used to refine the elevation estimates of the raw TomoSAR points as will be demonstrated in Chapter 6 of this thesis.

5 Reconstructing 2-D/3-D building shapes (or footprints)

This chapter summarizes the journal paper C (see the Appendix) which propose a novel data driven approach that systematically allows automatic reconstruction of 2-D/3-D building shapes (or footprints) using unstructured TomoSAR points clouds generated from one incidence angle only. The approach proposes new methods and aims at finding a more general and systematic solution towards automatic reconstruction of the whole city area. Following are the innovative contributions specific to the proposed approach:

- ◇ A hybrid technique based on region growing and energy minimization framework is proposed to automatically extract building roof/façade points directly from the 3-D TomoSAR points;
- ◇ A recursive angular deviation based approach is presented to smooth/refine the initial coarse building polygons obtained using alpha shapes (generalization of convex hulls);
- ◇ A novel façade-roof fusion procedure is proposed which is robust and fuses the legitimate *Façade-Polygon* pair together by interpreting the refined/smoothed polygon of each building as a graph. Series of operations taking part in reduction of graph (or smoothed polygon) include identification of points on the graph that are nearest to all the associated/paired reconstructed façades, computation of longest and shortest paths on the graph using these identified points, simplification of shortest paths by representation as line segments and rotation/projection of certain line segments (simplified shortest paths that are identified as positive paths) to match all the reconstructed façades belonging to the same graph/smoothed building polygon;
- ◇ An effective and robust procedure is developed for rectilinear identification of building polygons. To this end, dominant direction of each building is first determined based on weighted line segments approach. Angular deviations at each adjacent or consecutive vertex of the polygon are computed from the dominant direction and histogram analysis is then carried out for rectilinear building footprints identification. Later rectilinear constraints are added to the identified building polygons;
- ◇ Finally, due to the high inclination angle of the TerraSAR-X orbit i.e., near-polar orbit, the approach presented in the previous chapter (or A and B) may fail to reconstruct building façades facing North or South due to the missing of measurements. The solution to this problem is inherently provided by exploiting roof points in determining the complete shape/footprint of the building.

5.1 Motivation (Problem formulation)

As depicted in the previous chapter, façade reconstruction turns out to be an appropriate first step to detect and reconstruct building shape from these point clouds when dense points on the façade are available. Especially, when data from multiple views e.g., from both ascending and descending orbits, are available, the full shape of buildings can be reconstructed using extracted façade points. However, there are cases when no or only few façade points are available. This happens usually for lower height buildings and renders detection of façade points/regions very challenging. Moreover, problems related to the visibility of façades mainly pointing towards the azimuth direction can also cause difficulties in deriving the complete structure of an individual building. These problems motivate us to reconstruct full 2-D/3-D shape of building footprints via façade-roof analysis. In this regard, based on different object contents illuminated by side looking SAR, following three cases could be derived using data acquired from one incidence angle, e.g., in this case, image

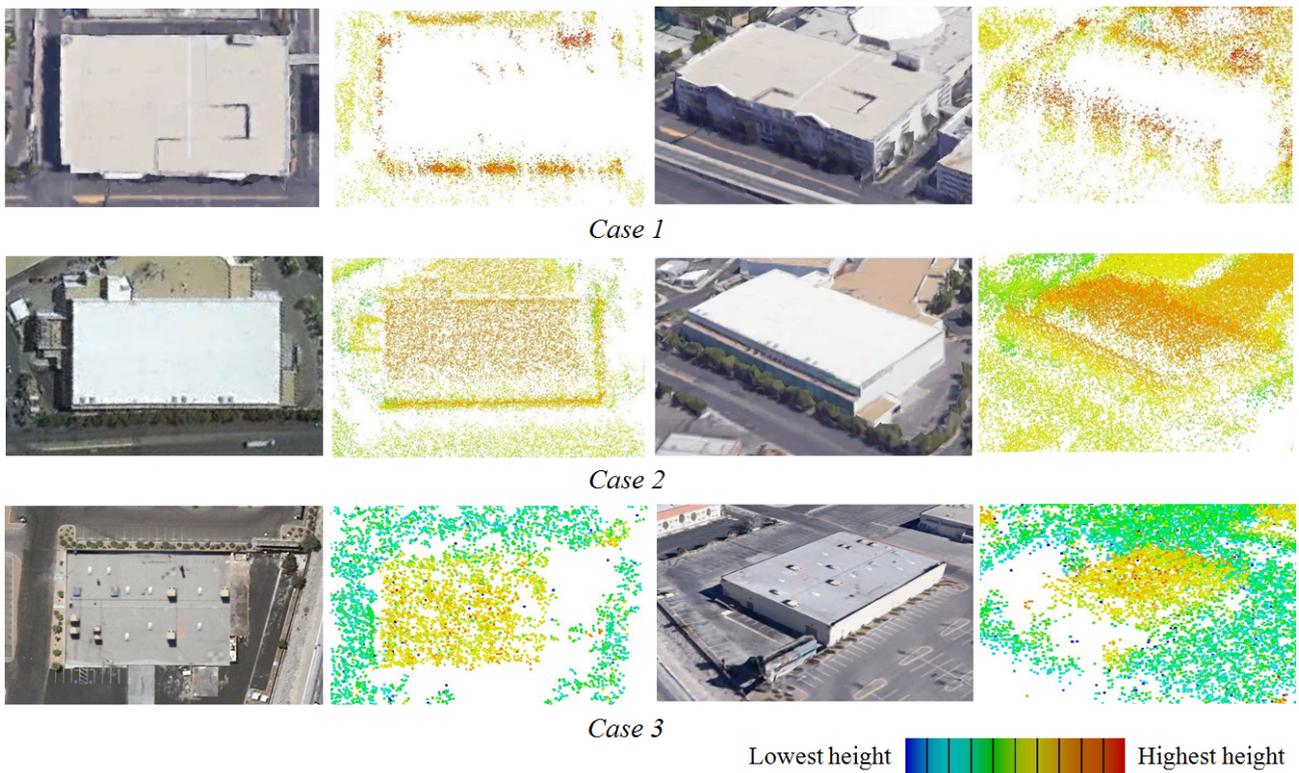


Fig. 24. Depicting three different cases over simple buildings commonly occurring in TomoSAR point clouds. Heights of the points are color-coded. Optical images © Google.

stacks from ascending orbit only (Figure 24):

Case 1: Higher density of façade points present with no or very few roof points – In this case, the complete 2-D/3-D building shapes could be fully reconstructed by adding points from multiple incidence angles. The solution to this case is demonstrated in (Shahzad and Zhu, 2015b) where 3-D façades model have been reconstructed for high rise buildings using one incidence angle only and in (Zhu and Shahzad, 2014) where full shape of the building was derived by prior fusion of two point clouds (Wang and Zhu, 2015) from ascending and descending stacks.

Case 2: Higher density of façade points present together with existence of relatively higher density of roof points – This case allows to reconstruct full shape of the building footprints from a single data stack by making use of both façade and roof points. Thus, the side of the building visible to the sensor could be reconstructed as the first step using façade points and later the other side of the building could be completed by exploiting the available roof points.

Case 3: No or very few façade points available but enough roof points exist – This case particularly appears for lower height buildings rendering detection of façade points/regions very challenging. This motivates us to obtain the full 2-D/3-D footprint of these buildings via roof point analysis only using conventional techniques as applied by LiDAR community. Even though these techniques are very much matured, still their adaptation to TomoSAR point clouds is not straight forward due to different object contents illuminated by side looking SAR together with problems related to less and varying point density and much less positioning accuracies of TomoSAR point clouds in comparison to airborne LiDAR.

In this chapter, solutions for the latter two cases (i.e., case 2 and case 3) is provided by extending (or utilizing) the solution provided for case 1 in the previous chapter.

5.2 Building detection

5.2.1 Region growing procedure

The reconstructed façades provide direct hints for the desired building regions. The idea is then to search in the nearby vicinity of the reconstructed façade to select seed points and later employ a surface normals based region growing algorithm to extract existing roof points. In this regard, the seed points are selected by first computing the midpoint of the reconstructed façade segment and then analyzing the local height statistics in the two opposite directions perpendicular to the reconstructed segment. Figure 25 provides the graphical illustration of the seed selection procedure.

Extracted seed points from the above procedure are then used in the region growing procedure to extract existing roof points based on the similarity of their surface normals (robustly estimated via (Hubert et al., 2005)) i.e., points having angular difference between their surface normals less than $\theta_{normals}(= 15^\circ)$ are added to the grown cluster. Using surface normals only however could also add ground points during growing procedure. An example of such a case will be where both roof and the neighboring ground are flat. A minimum height constraint is therefore needed and is incorporated in the growing procedure to restrict addition of points below a certain height h_{min} which is adaptively computed for each seed point as $m_p + (\text{height of } \mathbf{s}_p - m_p) \cdot fac$ where $m_p = m_b$ if $m_g > m_b$ or $m_p = m_g$ if $m_b > m_g$ (see Figure 25). The factor fac (empirically set to 0.55 in this work) can be adjusted between 0 ~ 1 to adjust h_{min} .

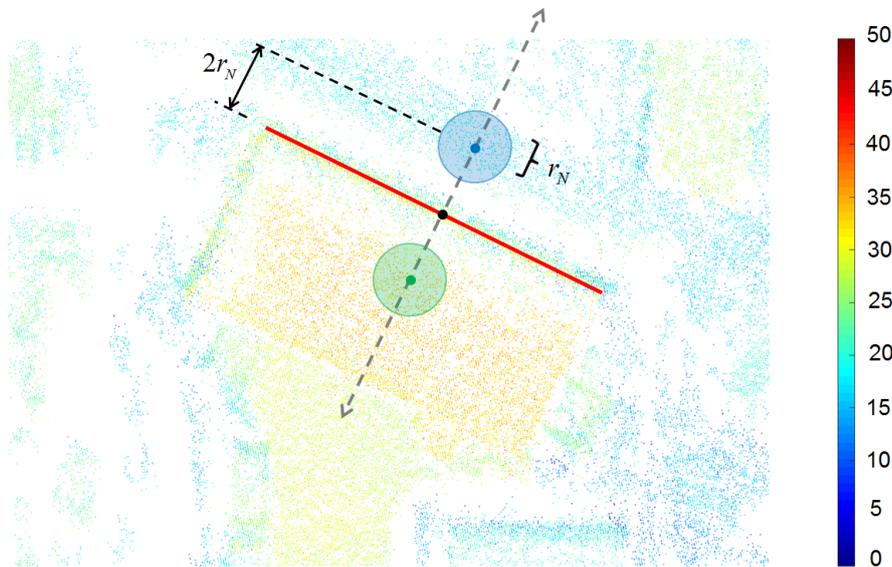


Fig. 25. Graphical illustration of the seed point selection procedure. Top view of 3-D TomoSAR points of a small region containing one building is depicted. The height in meters of each TomoSAR point is color-coded. The red line shows the reconstructed façade segment with black point as its midpoint. The green and blue points show the selected points in two opposite orthogonal directions with respect to the reconstructed façade (depicted as gray dotted lines). r_N is cylindrical radius used to determine local neighbors (i.e., points within shaded circles in respective colors). Local mean height of the blue and green points are computed, denoted as m_b and m_g respectively. Seed point s_p is chosen as the maximum height point among the local neighbors of blue points if $m_b > m_g$ or among the neighbors of green points if $m_g > m_b$ where latter is true in the depicted case.

5.2.2 Energy minimization

The above procedure extract building roof points for which information pertaining to façades is available i.e., *case 2*. To extract those building/roof points where no façade infor-

mation is available (i.e., resolve *case 3*), the problem of building detection is sequentially formulated into following simple energy minimization framework (Boykov et al., 2001) (Boykov and Kolmogorov, 2004) (Lafarge and Mallet, 2012) (Zhou and Neumann, 2013) (Yan et al., 2014):

$$E(l) = \sum_{p_i \in \mathbf{P}} D_{p_i}(l_{p_i}) + \sum_{(p_i, p_j) \in \mathbf{N}} w_{p_{ij}} \delta(l_{p_i} \neq l_{p_j}) \quad (20)$$

where $\mathbf{P} = \{p_i | i = 1, \dots, n\}$ denotes the set of remaining n 3-D points, \mathbf{N} denotes the set containing pairs of neighboring points (p_i, p_j) , l_{p_i} represents the label assigned to the point p_i , i.e. roof or non-roof point, $\delta(\cdot)$ is an indicator function, and $w_{p_{ij}}$ is the weight of each pair of neighboring points (p_i, p_j) and is defined as $w_{p_{ij}} = \exp(-\|\mathbf{p}_i - \mathbf{p}_j\|)$. The second summation term in (20) ensures the spatial smoothness, i.e., it favors consistent labeling between the neighboring point pairs whereas the first term in (20) denotes the data discrepancy term that measures the suitability/unsuitability of a particular labeling l assigned to the set of points \mathbf{P} . It is defined as (Lafarge and Mallet, 2012):

$$D_{p_i}(l_{p_i}) = \begin{cases} (1 - h_{p_i}) + \eta \cdot r_{p_i} & \text{if } l_{p_i} = \text{building roof point} \\ h_{p_i} + \eta \cdot (1 - r_{p_i}) & \text{if } l_{p_i} \neq \text{building roof point} \end{cases} \quad (21)$$

where r_{p_i} is the plane residual distance feature computed for each point p_i by locally fitting robust RANSAC plane among its local neighbors and h_{p_i} is the differential height of the point p_i obtained after subtracting the height of the underlying terrain, which is approximated by fitting a cubic polynomial surface to the non-building/ground points via robust least absolute (LAR) residuals method. Non-building/ground points are extracted by following the similar height constraint region growing procedure as described in (Shahzad and Zhu, 2014).

Both features h_{p_i} and r_{p_i} are normalized to the scale of $0 \sim 1$ by adopting the forms (Lafarge and Mallet, 2012): $h_{p_i} = \min(1, h_{p_i}/\varepsilon)$ and $r_{p_i} = \min(1, r_{p_i}/r_N)$ where ε is the tuning factor adjusting the sensitivity of the height feature (i.e., it ensures that all points having relative heights greater than ε provides minimum data discrepancy term in (20), and r_N is the radius size used to extract local neighbors v_c . η is the relative importance factor for the feature r_{p_i} , i.e., η defines the relative importance of r_{p_i} with respect to h_{p_i} . Value of η less than 1 (e.g., $\eta = 0.5$ used in this work) means more importance is given to the height feature. This is reasonable since r_{p_i} essentially depicts the degree of planarity and is lower for both building roofs and flat ground surfaces. Thus to differentiate between points on a flat planar ground segment, e.g., parking area and on the flat roof, their heights become the only discriminative factor in the labeling process and therefore should be given relatively higher importance. Moreover, planar objects with higher height are more probable to be part of the building structure than to the ground surface.

The above energy formulation in (20) is solved (minimized) via graph cuts based optimization library using $\alpha\beta$ -swap move algorithm (Boykov et al., 2001) (Boykov and Kolmogorov, 2004) (Kolmogorov and Zabih, 2004) (Fulkerson et al., 2009). The minimum energy corresponds to the labeling l such that higher planar points are detected as building roof points. Combing them with the set of roof points extracted in the previous section via façade information completes the extraction procedure.

5.3 Buildings shape/footprint reconstruction

Extracted building (roof) points are further segmented into individual building points via density based clustering as mentioned in Chapter 4 (Section 4.2.2). Reconstruction of building shape is initially obtained by employing alpha shapes (or α -shape, i.e., generalization of convex hull) around each segmented building (Edelsbrunner et al., 1983). This results in vertices that describe the 2-D polygonal boundary of the building footprint. Due to lower point density of TomoSAR points, alpha shapes only define the coarse outline of an individual building which is usually rough and therefore needs to be refined/smoothed (or generalized).

5.3.1 Refinement (or generalization) of alpha shapes

The refinement algorithm begins by first computing the angular deviations β_k at each vertex point of the alpha polygon \mathbf{V}_j as:

$$\beta_k = \begin{cases} \theta_k & \text{if } \theta_k \leq 90^\circ \\ 180 - \theta_k & \text{if } \theta_k > 90^\circ \end{cases} \quad \text{with } \theta_k = \cos^{-1} \left(\frac{\mathbf{d}\mathbf{v}_k \cdot \mathbf{d}\mathbf{v}_{k+1}}{\|\mathbf{d}\mathbf{v}_k\| \|\mathbf{d}\mathbf{v}_{k+1}\|} \right) \quad (22)$$

where ‘ \cdot ’ denotes the dot product and $\mathbf{d}\mathbf{v}_k$ is the direction vector computed at each edge formed by connecting two consecutive vertices \mathbf{v}_k and \mathbf{v}_{k+1} of the building polygon. Subsequently all those vertices (or edges) having angular deviations less than the threshold θ_{ang} are removed. The polygons before and after current iteration are compared and the process repeats itself if any vertex is removed (i.e., polygons before and after current iteration do not contain same number of vertices). Finally, the process terminates when there is no further removal of vertices.

5.3.2 Incorporating reconstructed façades

To improve the geometrical accuracy of the footprints, the reconstructed façades are fused with the refined building polygons. For this purpose, the façade associated to each refined building polygon is categorized into following two types:

- ◇ Type I façade: Façade fully or partly inside the refined polygon;
- ◇ Type II façade: Façade lying completely outside but associated to the refined polygon.

Identification of type I façades is easily achieved by checking if the endpoints of the reconstructed façades lie inside the polygon. Thus if both or at least one of the façade endpoints lie inside the building polygon, it is categorized to be type I façade. To identify façades of type II, following procedure is adopted:

- (1) First the midpoint of the reconstructed façade is computed and two points are chosen in opposite directions orthogonal to the reconstructed façade at a distance d from the midpoint;
- (2) Compute intersections of line₁ and line₂ with all the building polygons. Here line₁ denotes the line segment formed from by connecting midpoint to one of the chosen points and similarly line₂ is the line segment formed by connecting midpoint to the other opposite point;
- (3) If there exists an intersection of line₁ or line₂ with any of the building polygons, façade is assigned to the polygon with which the intersection occurs. In case there are more than one line-polygon intersections or both line₁ or line₂ intersects with different polygons, the façade is assigned to the polygon having the intersection point nearest to it.

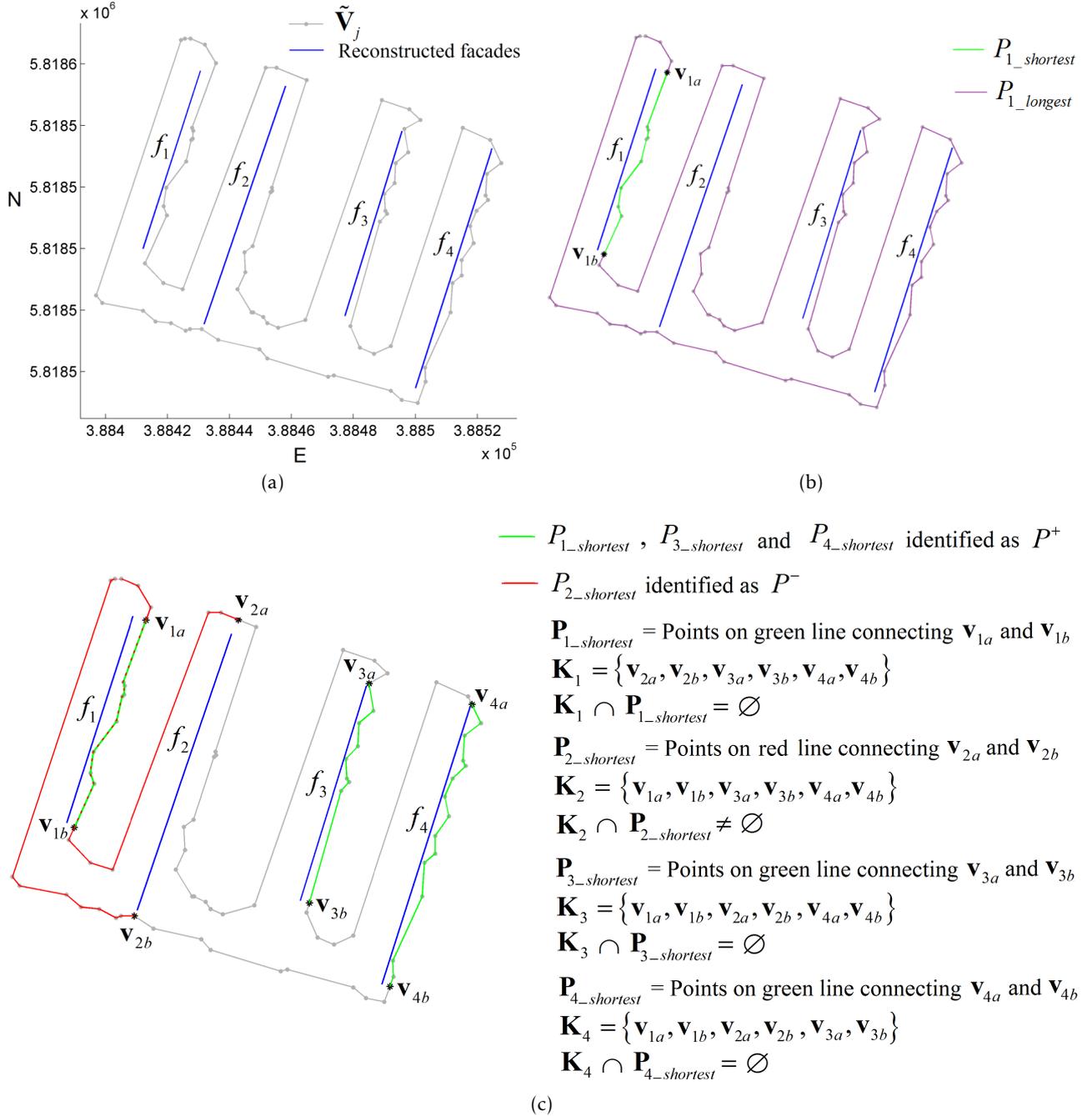


Fig. 26. Procedure describing the concept of $P_{t_shortest}$ and P^+ : (a) Refined polygon of one particular j th building \tilde{V}_j having vertices $\mathbf{v}_{k=1,\dots,m}$ with four reconstructed façades f_1, f_2, f_3 and f_4 that are to be incorporated; (b) illustrates the concept of shortest and longest paths associated for a particular façade f_1 . \mathbf{v}_{1a} and \mathbf{v}_{1b} denotes the closest points on the polygon/graph \tilde{V}_j to the two endpoints of façade f_1 respectively; (c) Example illustration of the concept of positive path P^+ . $P_{2_shortest}$ is identified as P^- as there exist points in \mathbf{K}_2 that are also present in $P_{2_shortest}$.

Implementation-wise, steps 1-3 are performed in a recursive manner. I.e., d is initialized to 1m and steps 1-3 are carried out. In case, there exists no line-polygon intersection (i.e., façade is not assigned to any polygon), the procedure repeats itself but this time d is incremented by 1m. The recursion stops if either the façade is assigned to any polygon or the distance d exceeds a certain threshold which is set to fixed 20m in this work. Thus a façade is only associated/paired to any building polygon if it lies at a distance less than 20m, otherwise it is regarded to have no polygon associated to it (i.e., categorized to *case 1*).

Once the façade types are identified, they are fused with the refined building polygon in

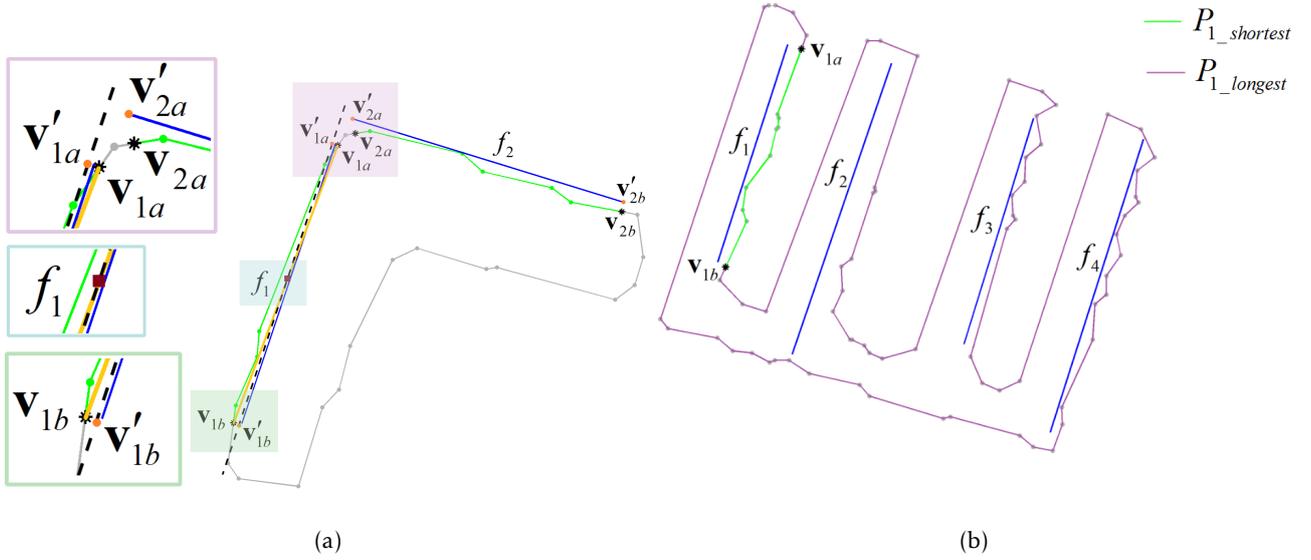


Fig. 27. Fusion of both type I and II façades. The refined polygon of one particular j th building \tilde{V}_j with two reconstructed façades f_1 and f_2 with types I and II respectively is shown; (a) depicts the procedure of computing points \mathbf{v}'_{1a} and \mathbf{v}'_{1b} . Note that for type I façade, \mathbf{v}'_{1a} and \mathbf{v}'_{1b} are computed by projecting \mathbf{v}_{1a} and \mathbf{v}_{1b} onto black dotted line which is parallel to the reconstructed façade f_1 and passes through the midpoint of line segment L formed by connecting points \mathbf{v}_{1a} and \mathbf{v}_{1b} while for the other façade having type II, \mathbf{v}'_{2a} and \mathbf{v}'_{2b} are simply the endpoints of the façade f_2 . All vertices of \tilde{V}_j on $P_{1_shortest}$ and $P_{2_shortest}$ (i.e., points in $\mathbf{P}_{1_shortest}$ and $\mathbf{P}_{2_shortest}$) are replaced by vertices \mathbf{v}'_{1a} , \mathbf{v}'_{1b} and \mathbf{v}'_{2a} , \mathbf{v}'_{2b} respectively; (b) Resulting polygon in black after fusing façades f_1 and f_2 with types I and II respectively.

slightly different manners as explained in the following: Let $V_{refined} = \{\tilde{V}_{i=1,\dots,N}\}$ denote the set containing N matrices of refined building polygons with $\tilde{V}_j = \{\mathbf{v}_{k=1,\dots,m}\}$ ($j \in i$) being the matrix containing 2-D vertices of the j th refined polygon having m vertices and $f_{r=1,\dots,s}$ as the corresponding s number of (paired) reconstructed façades. If \tilde{V}_j is interpreted as a graph, then a path P_t for any particular façade f_t is defined as a way consisting of polygonal chain of vertices that connect two points \mathbf{v}_{ta} and \mathbf{v}_{tb} lying on the graph/polygon i.e., the polygonal segment comprising of all the points of the polygon \tilde{V}_j within the interval $[\mathbf{v}_{ta}, \mathbf{v}_{tb}]$ defines path P_t . If \mathbf{v}_{ta} and \mathbf{v}_{tb} denote the closest points from the two end points of any particular façade f_t , then P_t can be further categorized into either $P_{t_shortest}$ or $P_{t_longest}$ (see Figure 26(b)). If the path length of P_t is denoted as $PL_{\mathbf{v}_{ta}\mathbf{v}_{tb}}$, then P_t is $P_{t_shortest}$ only if $PL_{\mathbf{v}_{ta}\mathbf{v}_{tb}} < \frac{TL}{2}$ where TL is the total path length (i.e., perimeter) of the polygon.

$P_{t_shortest}$ is further classified into two types: Positive path P^+ and negative path P^- . If the set of points on the polygons that are nearest to the end points of all façades of the same building other than f_t are denoted as \mathbf{K}_t (i.e., \mathbf{K}_t contains points nearest to the endpoints of façades $f_{r'=1,\dots,t-1,t+1,\dots,s}$ such that $r' \cup t = s$), then $P_{t_shortest}$ of the reconstructed façade f_t is defined to be P^+ if set of points $\mathbf{P}_{t_shortest}$ belonging to the path $P_{t_shortest}$ does not contain any element of \mathbf{K}_t i.e., $\mathbf{P}_{t_shortest} \cap \mathbf{K}_t = \emptyset$. Thus all façades whose paths are identified as positives are incorporated in the fusion process while façades having negative paths are not considered any further. An example illustration of the concept of positive path is presented in Figure 26(c).

Fusion of type I façade is carried out by

- ◇ First computing the midpoint of the line segment L_t formed by connecting the two points \mathbf{v}_{ta} and \mathbf{v}_{tb} and then determine two new points \mathbf{v}'_{ta} and \mathbf{v}'_{tb} by projecting \mathbf{v}_{ta} and \mathbf{v}_{tb} onto another line parallel to the respective façade f_t but passing through the midpoint of L_t (see Figure 27(a));
- ◇ Subsequently replace all vertices within $P_{t_shortest}$ (i.e., points in $\mathbf{P}_{t_shortest}$) by \mathbf{v}'_{ta} and \mathbf{v}'_{tb} ;

- ◇ An area constraint is then employed to ensure that the change in the polygonal area after incorporating particular façade is not greater than the certain fraction (set to 15%) of the previous polygonal area;
- ◇ The last step is necessary to make sure that the inner structures of the building do not interfere during the fusion procedure or in other words only façades that are exterior and define the building outlines are utilized.

Fusion of type II façade is straight forward. The points \mathbf{v}'_{ta} and \mathbf{v}'_{tb} are directly taken as the endpoints of the reconstructed façade.

The reason for this difference in computation of \mathbf{v}'_{ta} and \mathbf{v}'_{tb} while fusing the type I and II façades is due to the fact that point density on building roofs is quite varying and can contain gaps in between. This could lead to *under* reconstruct the building footprint i.e., part of the building roof region could not be reconstructed due to unavailability of points. Presence of type II façades implicitly validates this plausible phenomenon and therefore fusion of refined polygons by fully incorporating the reconstructed façades (of type II only) result in improved overall accuracy of reconstruction. Doing same for type I façades, on the other hand, may affect the footprint polygon in presence of façades belonging to inner building structures. Thus, only the orientation of type I façade is essentially incorporated by the proposed procedure. Figure 27 graphically illustrates the fusion procedure for both façade types.

5.3.3 Addition of rectilinear constraints

The façade-fused polygons are further identified whether they are rectilinear or not. The identification procedure is based on estimated dominant directions of the buildings and proceeds by determining angular difference β_i ($0 \leq \beta_i \leq 90^\circ$) of all the edges of the polygon with respect to the dominant/principal directions. Then the histogram of these angular differences is computed and all the edges whose angular differences fall within the bin intervals $[0^\circ, 20^\circ]$ and $[70^\circ, 90^\circ]$ are found. Subsequently, the polygon is identified to be rectilinear if the total sum of lengths of these edges are more than a certain fraction L_f of the total polygonal length TL . (L_f is fixed to 0.75 in our work i.e., 75% of the total polygonal length TL).

Following steps are performed to rectilinearize the identified polygons:

- ◇ Classify each edge of the building polygon such that it belongs to one of the two orthogonal principal axes based on its angular deviation (i.e., an edge is associated to that principal direction with whom the angular difference is less);
- ◇ Merge all adjacent edges that share same class i.e., associated to the same principal direction;
- ◇ Apply rectilinear transformation to every merged polygonal edge by projecting it onto its corresponding principal axis/vector;
- ◇ Computing intersection (or vertex) points between the consecutive vertices.

The identification and addition of rectilinear constraints depends on the estimated dominant direction of the façade-fused polygon. For building polygons belonging to *case 2* (i.e., having one or more reconstructed façades associated to each polygon), the principal direction is easily determined by assigning it to the direction vector computed by subtracting the endpoints of the longest reconstructed façade paired to it. For *case 3* building polygons, the principal direction is directly estimated from the polygon itself. Since no façade is associated to them, a weighted edge length based method is employed to estimate the two orthogonal principal directions of the building polygon.

Following objective function is minimized (Zhang et al., 2006):

$$PD = \sum_{i=1}^n g_1(l_i) g_2(\varphi_i(\beta_i, \phi_{\mathbf{d}\mathbf{v}_q})) \quad (23)$$

where n is the total number of vertices of the polygon and β_i ($0 \leq \beta_i \leq 90^\circ$) is the angular deviation of each edge l_i with respect to the direction vector $\mathbf{d}\mathbf{v}_q$. β_i is computed similar to (22) with the difference that the two direction vectors are $\mathbf{d}\mathbf{v}_i$ and $\mathbf{d}\mathbf{v}_q$ instead of direction vectors of consecutive edges $\mathbf{d}\mathbf{v}_j$ and $\mathbf{d}\mathbf{v}_{j+1}$. $\phi_{\mathbf{d}\mathbf{v}_q}$ is anticlockwise rotation angle which $\mathbf{d}\mathbf{v}_q$ makes with the unrotated coordinate system. $\varphi_i(\cdot)$ is a function that maps the angular deviations β_i to one of the two orthogonal directions (or axes) as defined by direction vector $\mathbf{d}\mathbf{v}_q$ and its corresponding normal vector. It is computed as

$$\varphi_i(\beta_i, \phi_{\mathbf{d}\mathbf{v}_q}) = \begin{cases} \beta_i & \text{if } \beta_i \leq 45^\circ \\ 90 - \beta_i & \text{if } \beta_i > 45^\circ \end{cases} \quad (24)$$

Both $g_1(\cdot)$ and $g_2(\cdot)$ are the weighting functions. $g_1(\cdot)$ assigns weight to each edge based on its relative length with respect to the overall length of the polygon edges. It is constructed in a way such that edges with longer lengths contribute less in (23) as compared to shorter edges lengths. Following linear function is used to describe $g_1(\cdot)$:

$$g_1(l_i) = 1 - \frac{l_i}{\sum_{i=1}^n l_i} \quad (25)$$

Similarly, $g_2(\cdot)$ assign weights to each edge based on its φ_i value. Assignment of weight is directly proportional to φ_i i.e., lower weight is given to an edge with lower φ_i inferring that edges close to one of the two orthogonal directions are given less weight as compared to the ones that are deviating. Since the span of φ_i for each edge is defined to be within the interval $[0, 45^\circ]$, therefore $g_2(\cdot)$ is computed by adopting the following linear function as:

$$g_2(\varphi_i(\beta_i, \phi_{\mathbf{d}\mathbf{v}_q})) = \frac{\varphi_i}{45} \quad (26)$$

Solution of (23) is obtained by rotating $\phi_{\mathbf{d}\mathbf{v}_q}$ within the interval $[0, 90^\circ]$ and then finding an optimum (or minimum) $\hat{\phi}_{\mathbf{d}\mathbf{v}_q}$ by comparing PD for each $\phi_{\mathbf{d}\mathbf{v}_q}$ value. The direction vector and its corresponding normal vector associated to the optimum (or minimum) $\hat{\phi}_{\mathbf{d}\mathbf{v}_q}$ $\left(= \arg \min_{\phi_{\mathbf{d}\mathbf{v}_q}} (PD) \right)$ thus describe the two orthogonal principal directions for *case 2* building polygons.

5.4 Experimental results and validation

5.4.1 Datasets

For validation, the proposed approach is tested on two different datasets. One is composed of TomoSAR point clouds generated from a stack of 25 TerraSAR-X high resolution spotlight

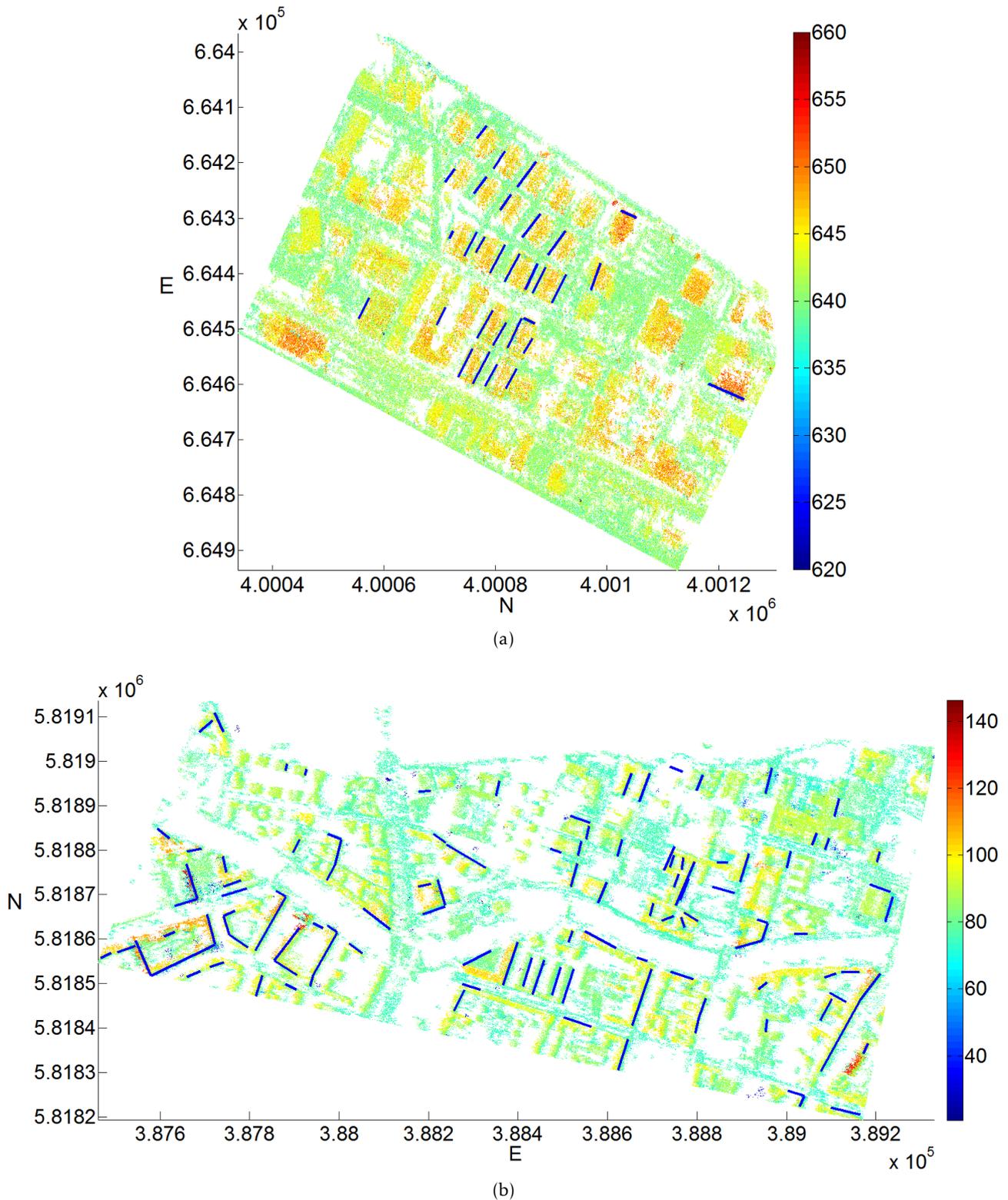


Fig. 28. Datasets: Top view of the three dimensional TomoSAR points in UTM coordinates of the area of interest in Las Vegas, USA (a) and Berlin (b). Blue lines depict the reconstructed façade segments (longer than 10 meters). The height of TomoSAR points is color-coded [unit: m].

images covering approximately $(900 \times 600 \approx) 0.54 \text{ km}^2$ area in the city of Las Vegas, USA. It contains 0.48 million points and consists of moderate sized buildings with relatively simple geometry. The second dataset is added to test the generality of the proposed algorithms and is composed of TomoSAR point clouds of complex building structures produced from a stack of 102 TerraSAR-X high resolution spotlight images covering around $(1750 \times 900 \approx) 1.5 \text{ km}^2$



Fig. 29. Building extraction results – Las Vegas: (a) Extracted building points in red are overlaid onto the optical image (© Google) of the area of interest. Yellow circles indicate extracted points originating from sources like advertisement boards, monuments etc. Large black circle encloses two buildings which remain undetected due to lower relative heights; (b) Finally, reference footprints (in green) overlaid onto the extracted building points. Red points are building points while black points are non-building points.

area in the city of Berlin, Germany. Number of points in Berlin dataset are approximately 0.52 million. Both TomoSAR point cloud datasets are generated from images taken from ascending orbit using the Tomo-GENESIS software developed at DLR (Zhu et al., 2013).

5.4.2 Reference dataset

For the area of interest in Las Vegas, the reference footprints for Las Vegas dataset were acquired from Cybercity3-D (CyberCity3D, 2015). These footprints are highly precise with positional accuracy up to ± 15 cm and are generated using automated and semi-automated photogrammetry-based techniques with data source derived from aerial, oblique, or satellite stereo imagery.



(a)



(b)

Fig. 30. Results of building extraction – Berlin: (a) Extracted roof points in red are overlaid onto the optical image (© Google) of the area of interest; (b) Red and black points depict building and non-building points, respectively. The overlaid green polygons are reference buildings downloaded from OSM (Geofabrik, 2015). Blue polygons are manually extracted buildings not present in OSM data. Gray polygons are newly constructed buildings that are not present in our dataset whereas magenta colored polygons are buildings that do not actually exist but present in OSM data. Both gray and magenta polygons are not included in the evaluation.

For Berlin dataset, the building extraction results are compared with the reference polygons downloaded from the OpenStreetMap (OSM) (Geofabrik, 2015). The OSM data is free to download and comes under the open license Open Data Commons Database License (ODbL). Since it is a volunteered geographic information project, the data quality may vary from region to region. Recently, the building footprints have been evaluated for their completeness (Hecht et al., 2013) and correctness (Fan et al., 2014b) for various cities of Germany. The analysis of OSM data with surveying datasets reveals fairly precise positioning accuracies varying within 4 meters (Haklay, 2010) (Fan et al., 2014b).

5.4.3 Validation of results and discussion

Figure 28 shows the result of applying façade reconstruction procedure over both the datasets. The final extracted building points overlaid onto the optical images are depicted in Figure 29(a) and Figure 30(a). To validate the extraction results, the reference polygons are plotted onto the extracted building points in Figure 29(b) and Figure 30(b). It can be visually seen that the extracted building points fit very well to these reference polygons. For Berlin dataset, few buildings were missing in the OSM dataset and therefore by analyzing the detected buildings from TomoSAR point clouds and validating using optical data, these few missing buildings were completed in OSM dataset i.e., polygons depicted in blue in Figure 30(b). The performance of the (detection) extraction procedure in both the datasets is then assessed by employing standard (completeness/correctness/quality) metrics in (%): (93.923/86.173/81.622) – Las Vegas, (94.779/85.316/81.487) – Berlin.

In both the datasets, all the buildings having relative heights (with respect to ground) more than 5m are detected by the extraction procedure. However, an example of two smaller buildings for which no points could be extracted is highlighted by a black circle in Figure 29(a). These buildings have footprint area of approximately $(28 \times 5 =) 140 \text{ m}^2$ but possess relative heights of only 2m. Since during the extraction procedure seed points are chosen based on local height jumps of 5m, therefore no seed point could be chosen for these buildings. As a consequence, they remain undetected. In terms of false alarms, it is worth to mention that it might happen that points belonging to some small vertical structures on ground (e.g., advertisement boards, monuments, etc.) also get detected during our extraction procedure. Few examples of such cases are also highlighted by yellow circles in Figure 29(a). The reason for this occurrence of false positives is our implicit definition (or assumption) pertaining to buildings i.e., higher points with higher degree of planarity are detected as buildings by the extraction procedure.

The extracted points are in turn utilized for 2-D/3-D reconstruction of building footprints using procedure explained in Section 5.3. In order to evaluate the reconstruction results, both the reconstructed and reference polygonal footprints are rastered onto an image with pixel resolution of 1m (i.e., 1 pixel corresponds to 1m^2 spatial area). Figures 31 and 32 shows the common and difference images created from intersection and subtraction of the reconstructed footprint image from the reference footprint image respectively. The evaluation results based on these images result in following (commission/omission) errors in (%): (10.79/17.20) for Las Vegas and (19.43/14.57) for Berlin datasets.

Figure 33 depicts the final building footprint reconstruction in 3-D on both Las Vegas and Berlin datasets, respectively. Hypothetically, the reconstruction results will be improved with higher density of TomoSAR points because more points would be available for param-



Fig. 31. Common images computed using reference footprint image and final reconstructed footprints.

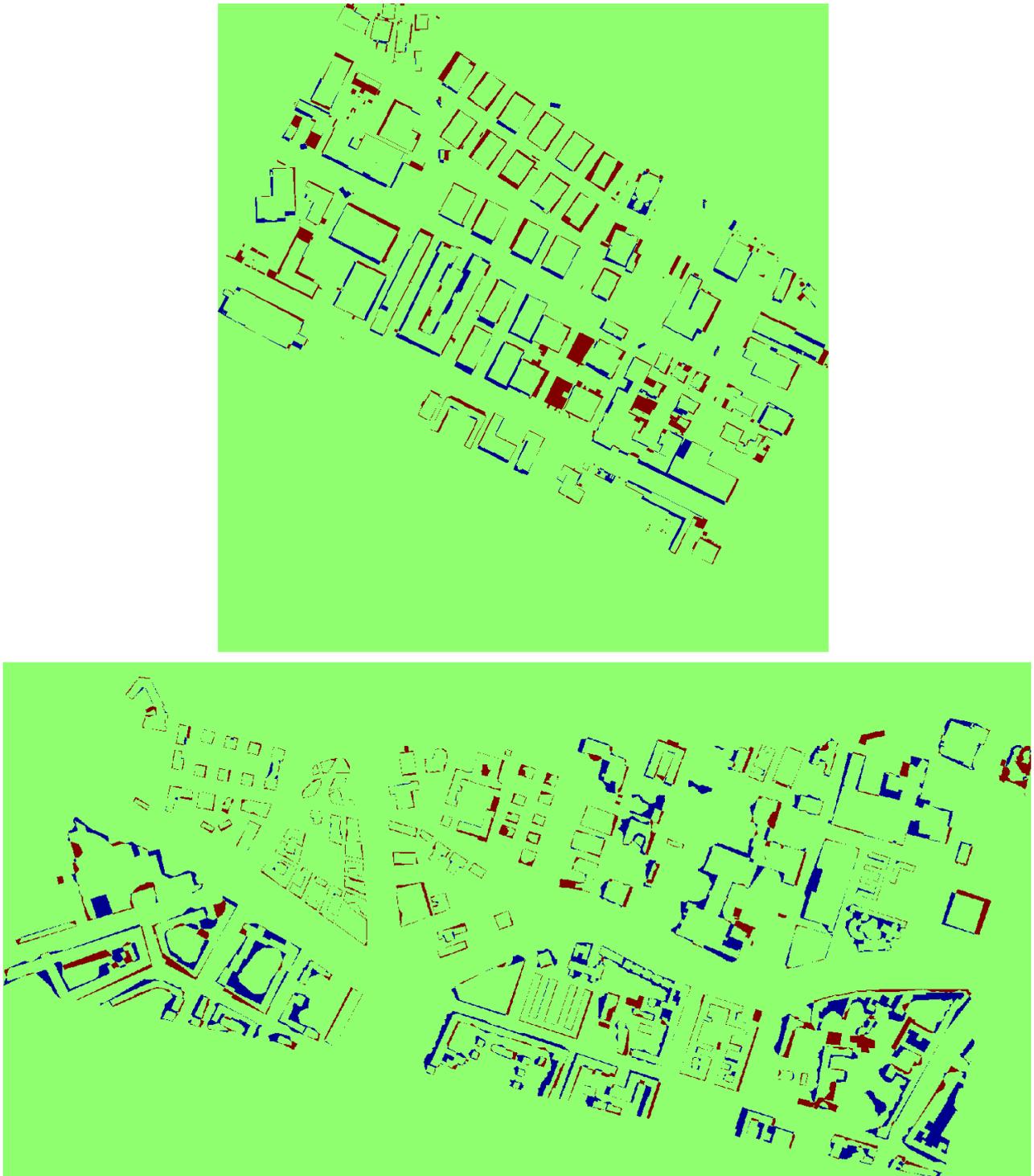


Fig. 32. Difference images computed by subtracting the final reconstructed footprint image from the reference image. Red pixels indicate building regions that are not reconstructed by the proposed algorithm while blue pixels are over reconstructed regions i.e., pixels not part of the reference footprint image but present in the reconstructed image. Note that green pixels indicate no difference between reconstructed and reference rastered images.

eter estimation. Numerical experiments also demonstrated that reconstruction accuracy is better for buildings with higher density of roof points. For low density roof regions, the reconstruction accuracy is however restricted by the number of available points that consequently reduces omissions errors. Further improved model based approach might be helpful in this regard.

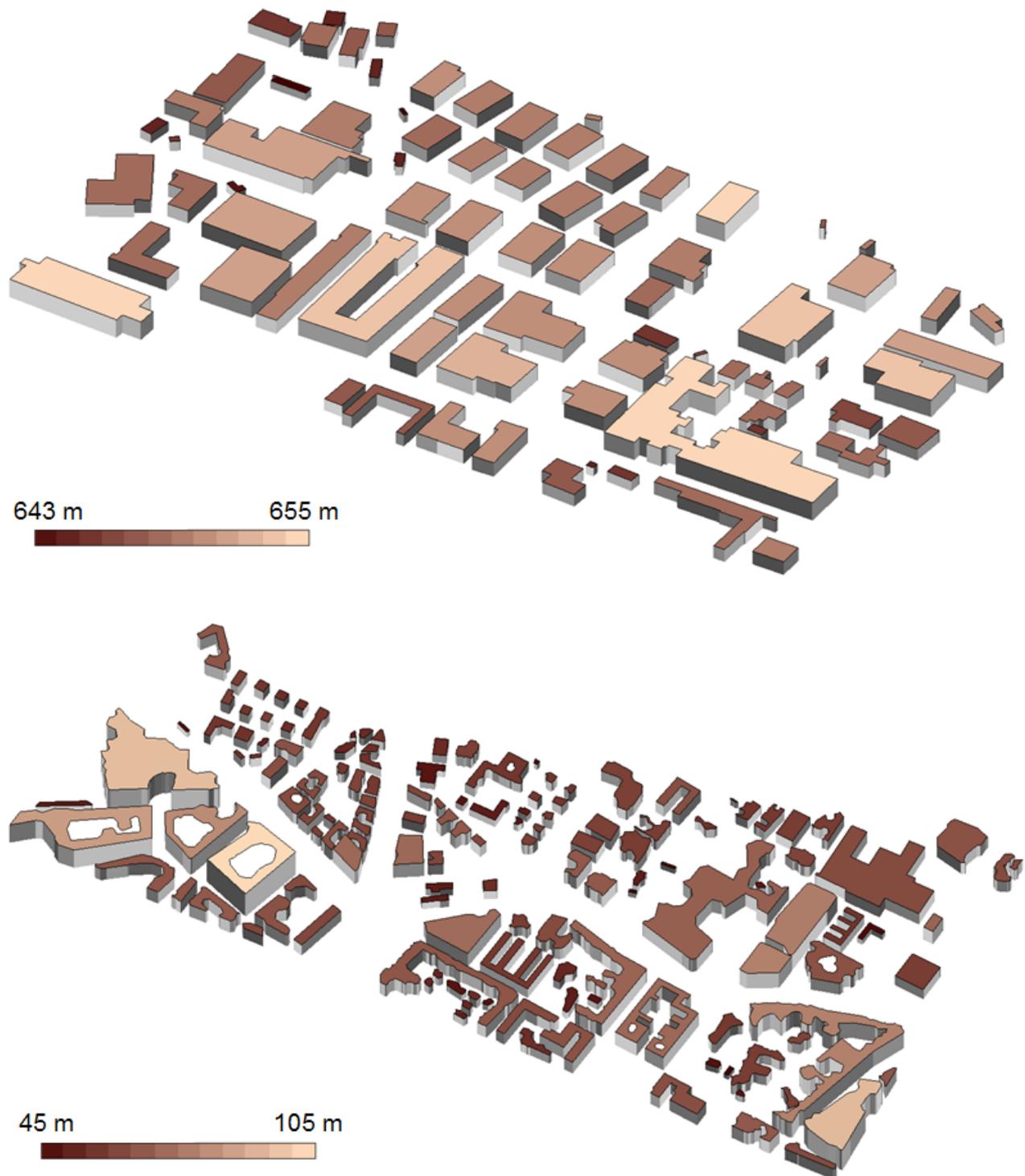


Fig. 33. Final reconstructed 3-D building footprints: Las Vegas (top) and Berlin (bottom) datasets.

6 Application examples

This chapter exemplifies potential applications of the building models reconstructed using the framework proposed in the previous two chapters. Three innovation applications are identified in A and D including

- ◇ **4-D (dynamic) building model:** For improved monitoring of detailed structures of individual buildings, an example of 4-D building model is presented;
- ◇ **Elevation estimates refinement:** Model based elevation refinement is carried out to refine elevation of each raw TomoSAR point by using its more accurate azimuth and range coordinates together with the corresponding reconstructed model;
- ◇ **Joint sparsity concept:** An algorithm that is able to segment overlaid building pixels in the SAR image with the help of openly available 2-D GIS data has been developed and presented in detail. The idea behind this is to utilize the extracted building characteristics (e.g., building mask, orientation etc.) to further improve the TomoSAR inversion process by exploiting the concept of joint sparsity as illustrated in D. For areas where 2-D GIS is not available, the motivation is to utilize the footprint of reconstructed models instead to improve the TomoSAR inversion in a feedback iterative manner i.e., to use reconstructed footprints as iterative priors.

6.1 4-D (dynamic) building model

The dataset used to show the first example consists of TomoSAR point clouds generated from two stacks (each comprising 25 images) of TerraSAR-X high-resolution spotlight images from ascending (36° incidence angle) and descending (31° incidence angle) orbits provided by (Zhu and Bamler, 2012a). Due to the different scattering properties from diverse geometries, there is little chance to identify a common reference point for both stacks. This problem results in a shift in the elevation directions of both point clouds reconstructed from these two stacks with different viewing angles. To obtain the full structure of individual buildings from space, the point clouds are first geodetically fused by determining this shift in elevation direction (Gernhardt et al., 2012) (Wang and Zhu, 2015). Figure 34(a) shows the optical image of our test buildings, the Bellagio hotel complex in Las Vegas. The corresponding fused input TomoSAR point cloud in UTM coordinates is depicted in Figure 34(b). The size of the test area is about $520 \times 570 \text{ m}^2$. The number of TomoSAR points is approximately 0.4 million.

Figure 35 depicts the SD estimates obtained for the test area using the procedure explained in Chapter 4 (see Section 4.1.1). As can be seen, much higher SD estimates are obtained at façade locations. Applying softer threshold to them and utilizing 3-D surface normals information (see Section 4.1.2) allow extraction of façade points as depicted in Figure 36(a). Figure 36(b) shows the result of segmenting façade points as belonging to individual façades which are later utilized to reconstruct 3-D façade model. Figure 36(c)(d) depicts the fused point clouds overplotted onto the reconstructed façade model. The height of the points is color-coded. The corresponding estimated motion parameter, in this case the amplitude of seasonal motion caused by thermal dilation, is illustrated in Figure 37. This information, if properly visualized/textured, can be used in the development of dynamic building models that may help to monitor individual buildings and even the whole city from space. In addition to 4-D building model, if the points are color-coded according to their reflectivity values, a *real* 3-D SAR can be generated. Figure 38 shows the corresponding reconstructed 3-D SAR image, i.e., the reflectivity map overlaid on the façade model. Such an image visualizes in detail how the Bellagio hotel would look like in X-band for our eyes, if they could

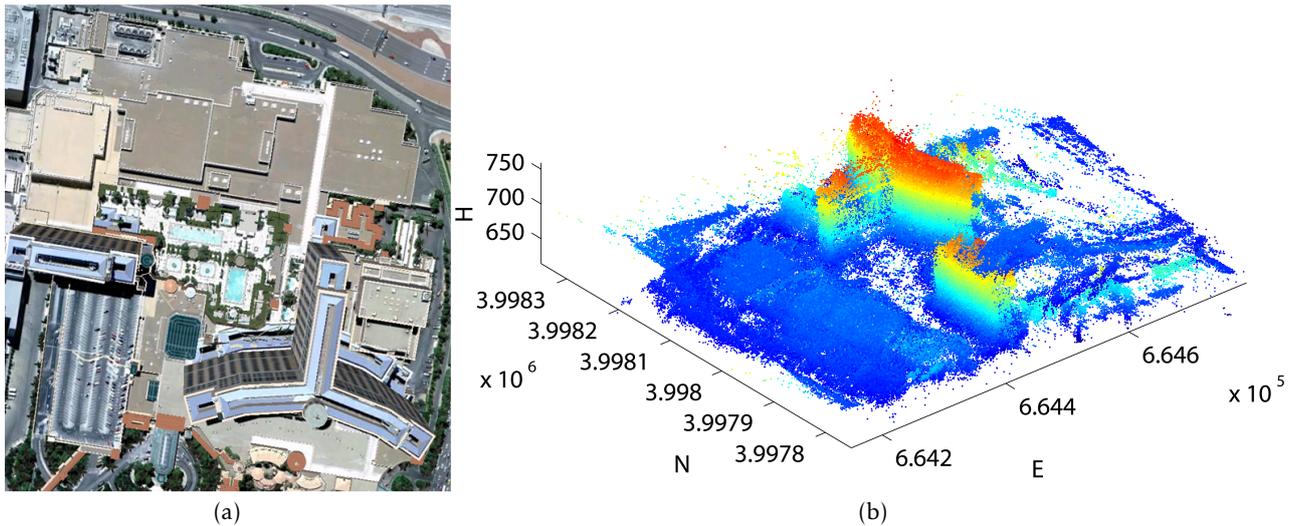


Fig. 34. Test buildings - Bellagio hotel, Las Vegas: (a) Optical image (© Google); (b) Fused TomoSAR point clouds from both ascending and descending orbits in UTM coordinates.

sense microwaves, from the position of the SAR satellite. Also, such visualizations may be helpful to develop an intuition about the multiple bounce effect. Moreover, the very bright individual scatterers that behave as corner reflectors can be precisely located which consequently helps in better understanding of the nature of scattering.

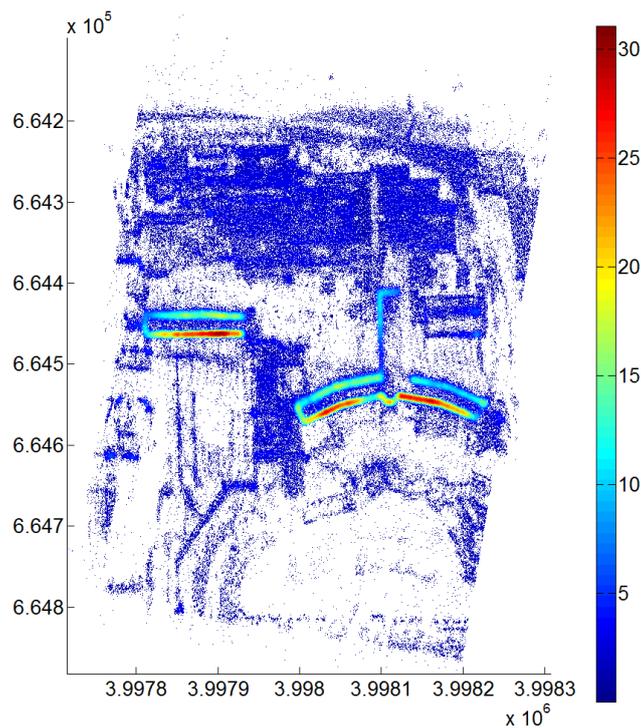


Fig. 35. SD estimates of TomoSAR point cloud depicted in Figure 34(b).

6.2 Elevation estimates refinement

As briefly mentioned in Chapter 3 (Section 3.3), due to limited orbit spread and the small number of images, the location error of TomoSAR points is highly anisotropic with an elevation error typically one to two orders of magnitude higher than in range and azimuth (Zhu, 2011). For example, for TerraSAR-X high resolution spotlight images with typical param-

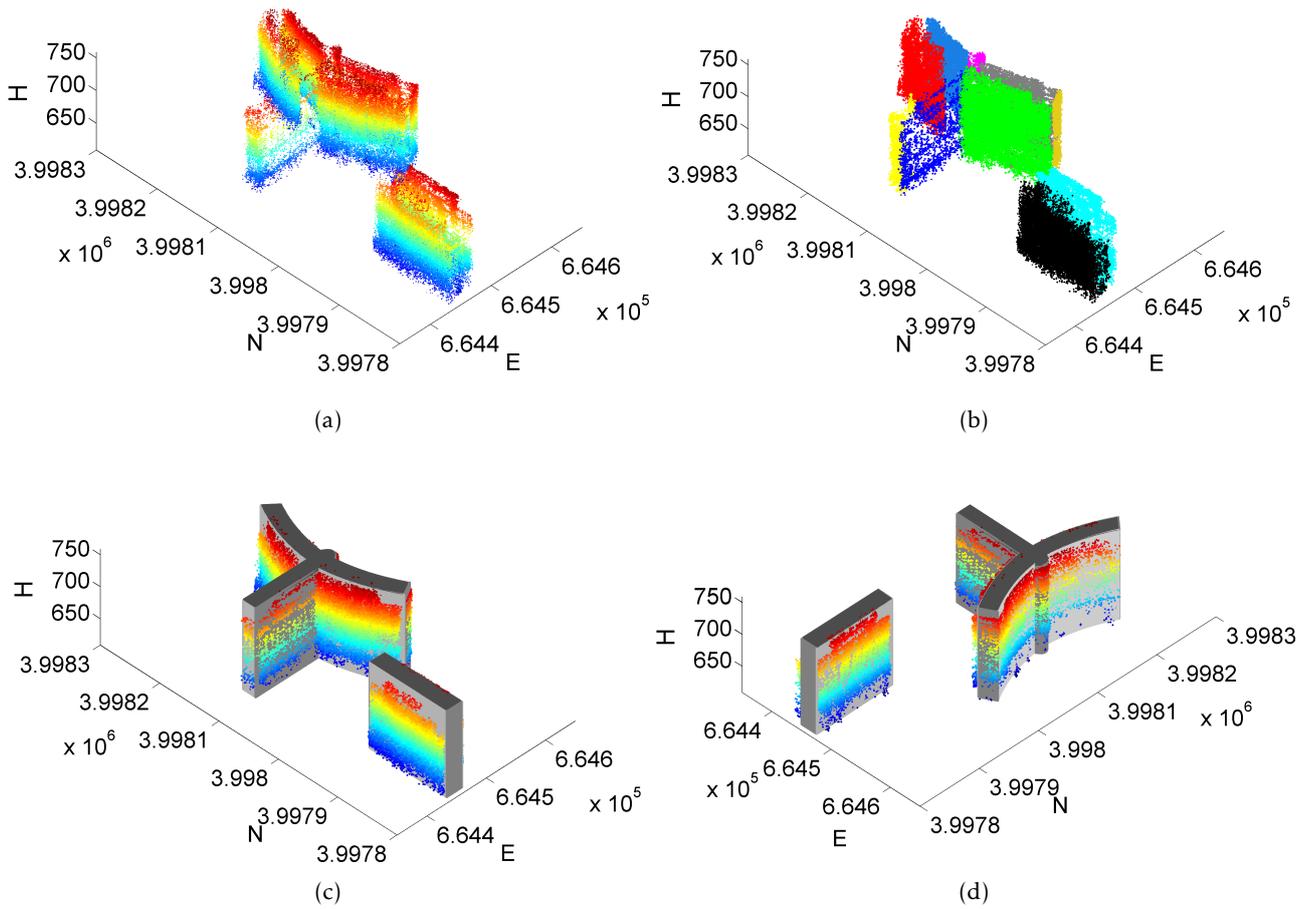


Fig. 36. 3-D Façade reconstruction results: (a) Extracted façade points; (b) Segmented façade points; (c) Reconstructed 3-D façade model with extracted façade points. For (a), (c) and (d), the height of the points is color-coded from blue (lowest) to red (highest) [unit: m].

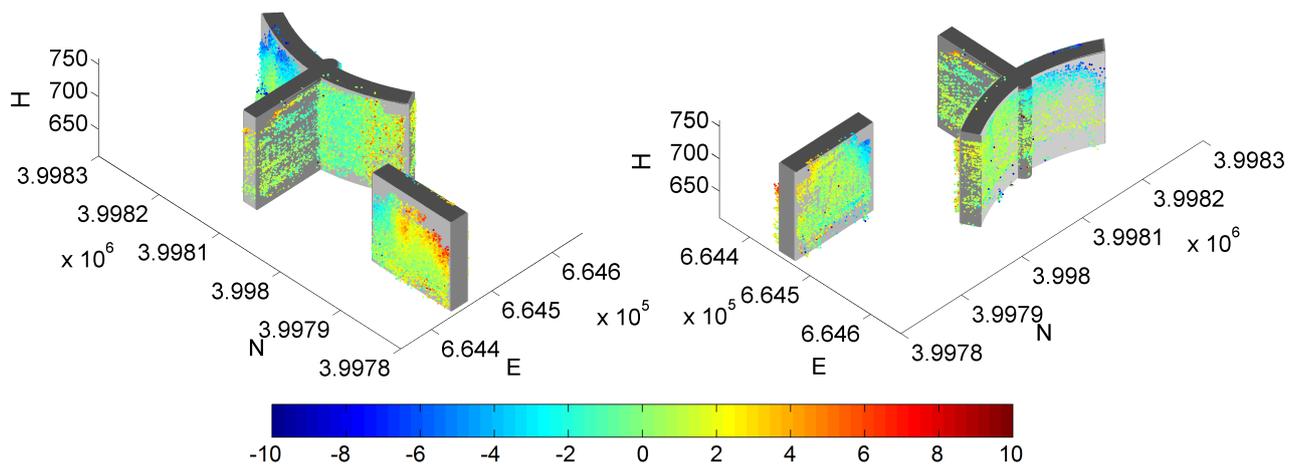
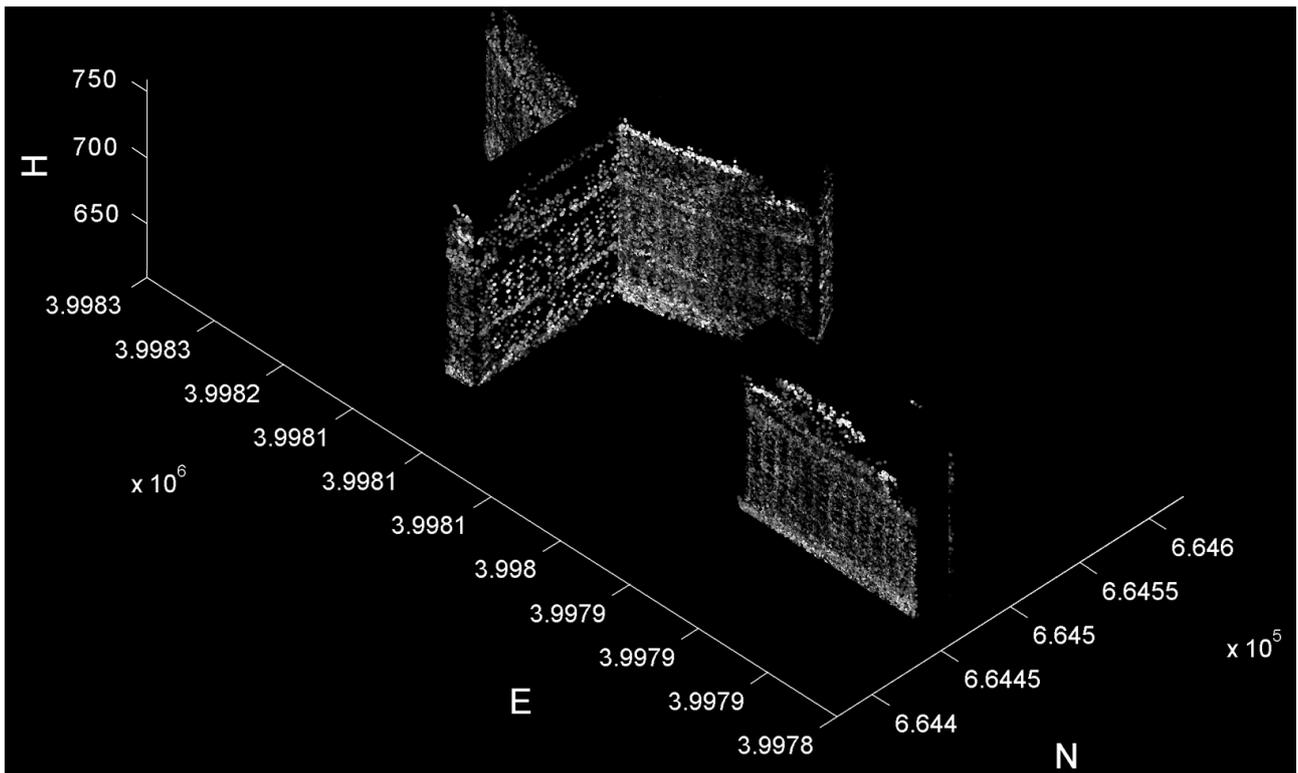


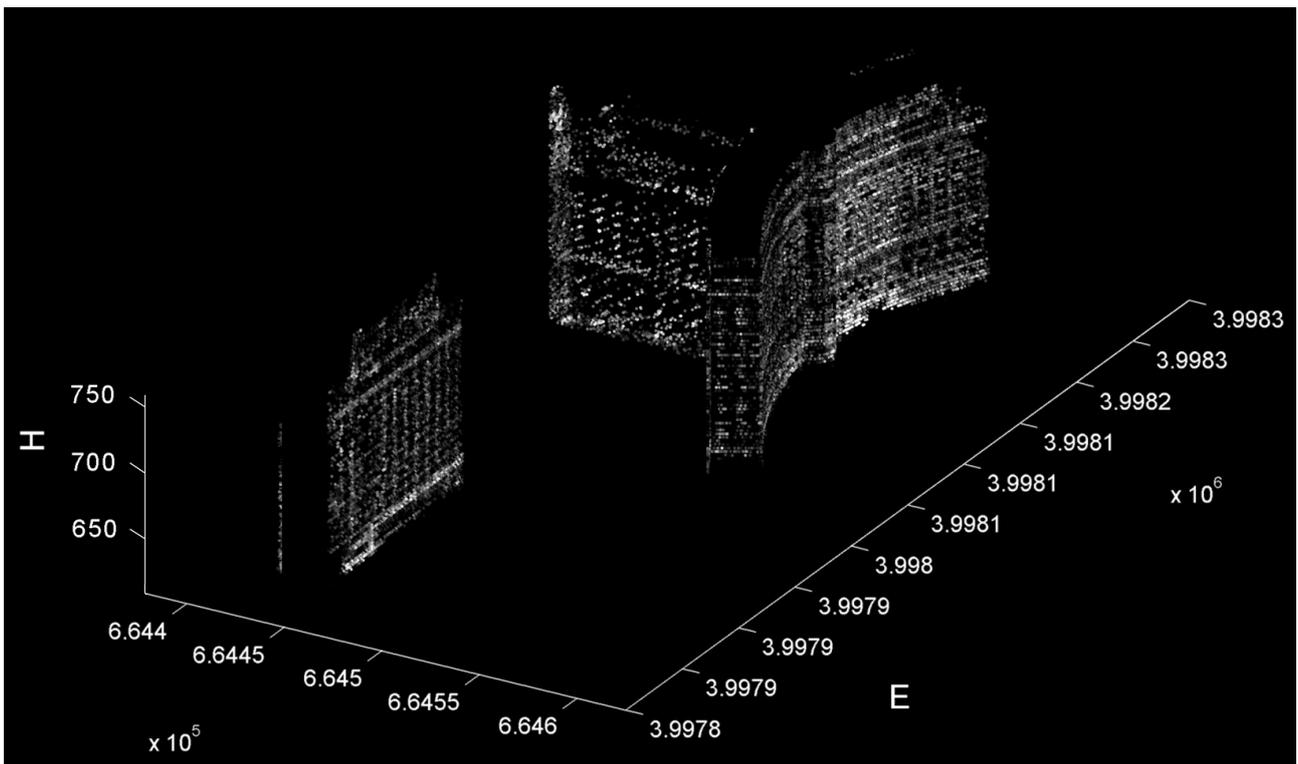
Fig. 37. Reconstructed 4-D building façade model, amplitude of seasonal motion is color-coded [unit: mm].

ters, the theoretical relative localization precision of a persistent scatterer is (Gernhardt and Bamler, 2012): 1.7 ~ 2.1cm in range; 3.2 ~ 3.8 cm in azimuth and 62 ~ 139 cm in elevation.

The elevation estimates of the TomoSAR points can be refined by using their more accurate azimuth and range coordinates and the reconstructed façade surfaces as depicted in Figure 39. This sketch illustrates the refinement principle in range-elevation plane. The red points represent the raw TomoSAR point locations at different heights along a façade. The ellipse indicates the error ellipse of the TomoSAR estimate in range and elevation direction,



(a)



(b)

Fig. 38. Reconstructed 3-D SAR image overplotted on the reconstructed façade model. Note that this is not only a projection of the SAR image onto the building models. Rather, the lay-overed brightness contributions from façade and ground have been separated in the tomographic reconstruction step.

i.e. much poorer accuracy in elevation compared to range. The black line indicates the reconstructed façade surface. The corresponding iso-azimuth-range lines of each point along elevation (directions are indicated by the arrows) are projected to the identified and modeled façade surface it belongs to. The final refined 3-D position is obtained by taking the

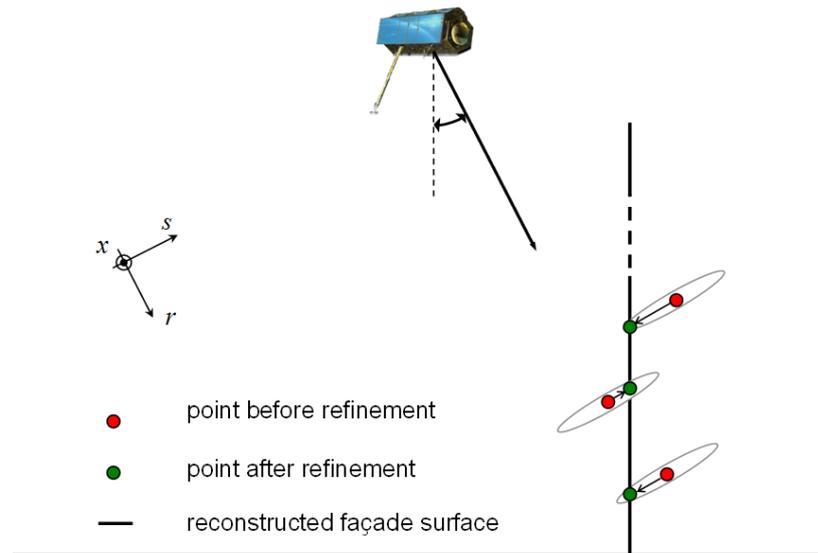


Fig. 39. The basic principle for the elevation estimates refinement of the TomoSAR points by using their more accurate azimuth and range coordinates and the reconstructed façade surfaces.

elevation coordinate of the intersection point. This is an approximation of an optimal linear estimate. The green points represent the positions after elevation refinement. In this way, it is expected to achieve much better elevation estimation accuracy that is in the order of several centimeters, though it is still slightly worse than the ones in azimuth and range due to error propagation.

To validate this improvement, a row of bright points from the intensity image belonging to a façade portion of constant height are manually selected as shown in Figure 40(a). Figure 40(b) and (c) compare the height estimates of the analyzed points before and after the refinement. It is obvious that their height estimates are improved significantly. The standard deviations before and after the refinement are 190 cm and 5.5 cm, respectively. An improvement by a factor of 35 corresponds quite nicely to the ratio of inherent resolutions in elevation (in the order of 30 ~ 50m) and range (1.1 m).

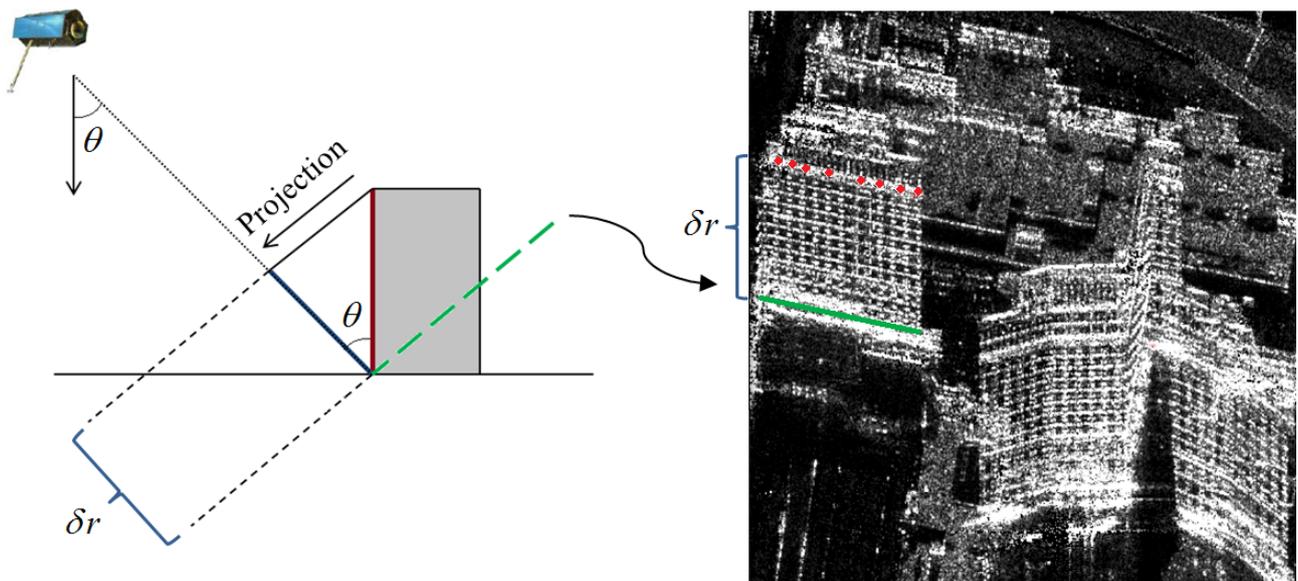
6.3 Joint sparsity concept

6.3.1 Dataset

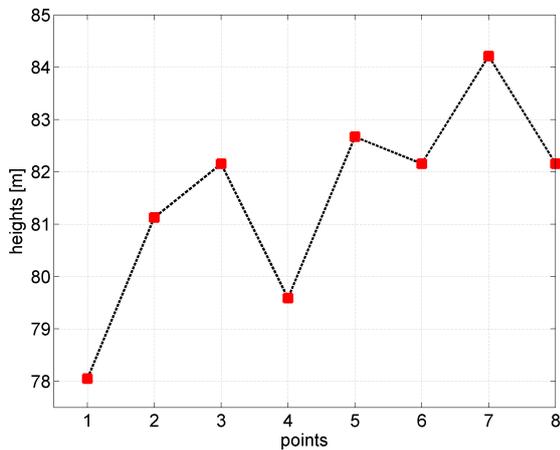
A single SAR intensity image acquired from German SAR satellites TanDEM-X and 2-D building outlines downloaded from openly available OSM website (Geofabrik, 2015) are utilized to demonstrate the building segmentation procedure. Figure 41(a) depicts the optical image of the test buildings while Figure 41(b) depicts the corresponding SAR image. Downloaded 2-D building outlines are shown in Figure 42.

6.3.2 Proposed segmentation strategy

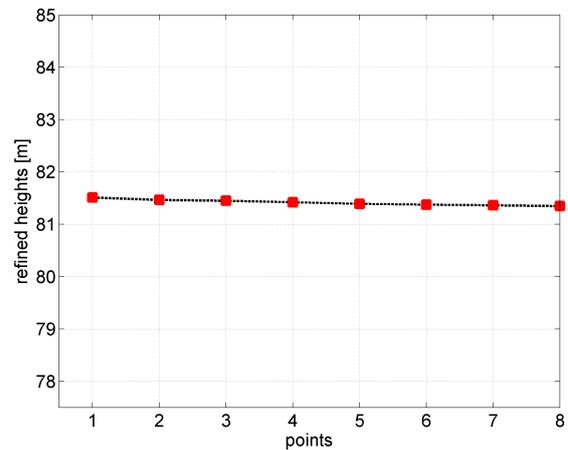
The key idea is to make use of the aforementioned online freely assessable 2-D building footprints to extract detailed high rise building features including building masks, orientations, and the iso-height lines in SAR image data stacks. The extracted information can be further incorporated as a prior knowledge into the estimation for a more accurate tomographic SAR inversion. For this purpose, in the following a sophisticated approach is proposed that is tolerable to moderate errors in the input GIS data for automatic extraction of high rise



(a)



(b)



(c)

Fig. 40. Elevation estimates refinement: (a) TerraSAR-X mean intensity map from ascending stacks (the red dots are the analyzed points) along with the projection geometry; Height estimates of the analyzed points before (b) and after (c) refinement.

building features (e.g., building mask, iso-height lines, orientation etc.) in the SAR image data stacks.

Following steps are performed to extract the interesting building features (i.e., the mask and iso-height lines) of the considered buildings in a single SAR image with the help of 2-D footprints:

- (1) Available building footprints from OSM in world (latitude/longitude) coordinates are transformed/geocoded into SAR (azimuth/range) imaging coordinate system. Figure 43(a) shows the projected reference polygons overlaid onto the buildings of interest in the corresponding SAR image shown in Figure 41(b);
- (2) Due to the side-looking geometry, SAR illuminates only one side of the building. Therefore, the complete building footprint of individual buildings is further segmented into two parts by means of a simple 2-D visibility test i.e., the part illuminated by the sensor which will be further used for iso-height pixel extraction (as depicted by red polylines in Figure 43(a)) and the part in the shadow area not visible to the sensor which will not be used in further processing (as depicted by green polylines in Figure 43(a));
- (3) Finally, errors in the identified red polylines, caused by inaccuracies of the input GIS

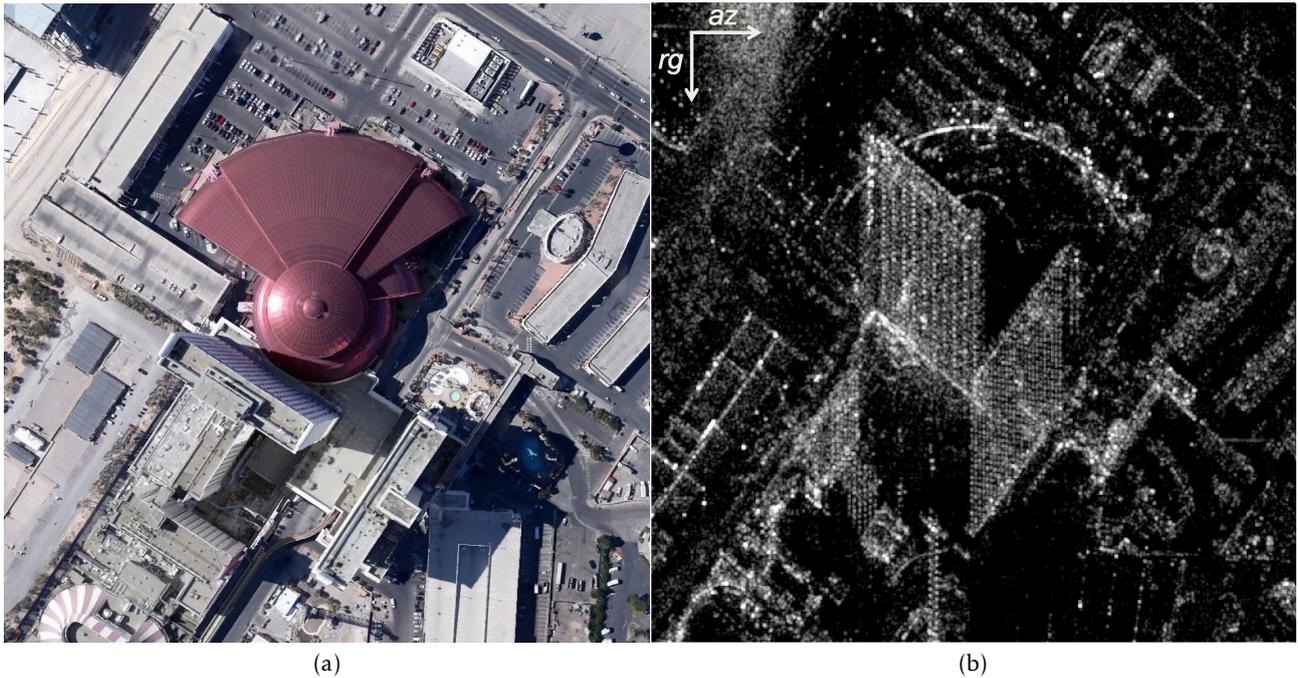


Fig. 41. Test Area: (a) Optical image of the test area © Google; (b) Corresponding SAR intensity map with spatial resolution of 1.1×0.588 m in azimuth (az) and range (rg) respectively.

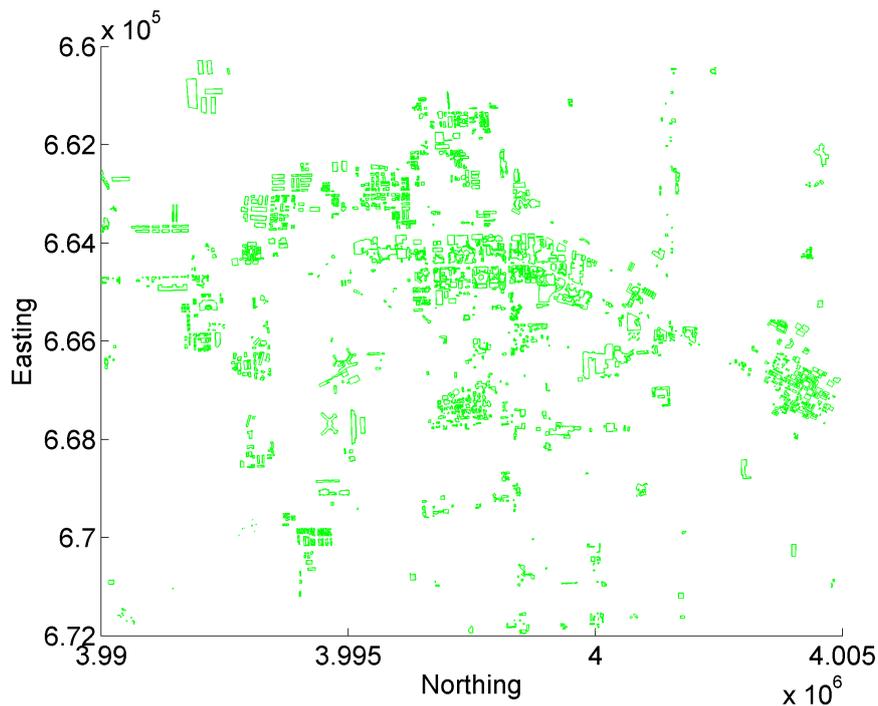


Fig. 42. GIS data (2-D building footprints) of Las Vegas from OSM.

data in both orientation and translation, are compensated by adopting following sequence of actions:

- (a) Shift/translate the identified polyline (red polyline in Figure 43(a)) in 2-D sliding window fashion within the intervals: range shift $[-10 \ 10]$ and azimuth shift $[-5 \ 5]$;
- (b) Within each shift, rotate the polyline between interval $[-7.5 \ 7.5]$ degrees and compute median of intensities along the rotated polyline. This is accomplished by selecting equally spaced points along the rotated polyline, and then using nearest neighbor interpolation to find the intensity value for each point. Finally, the me-

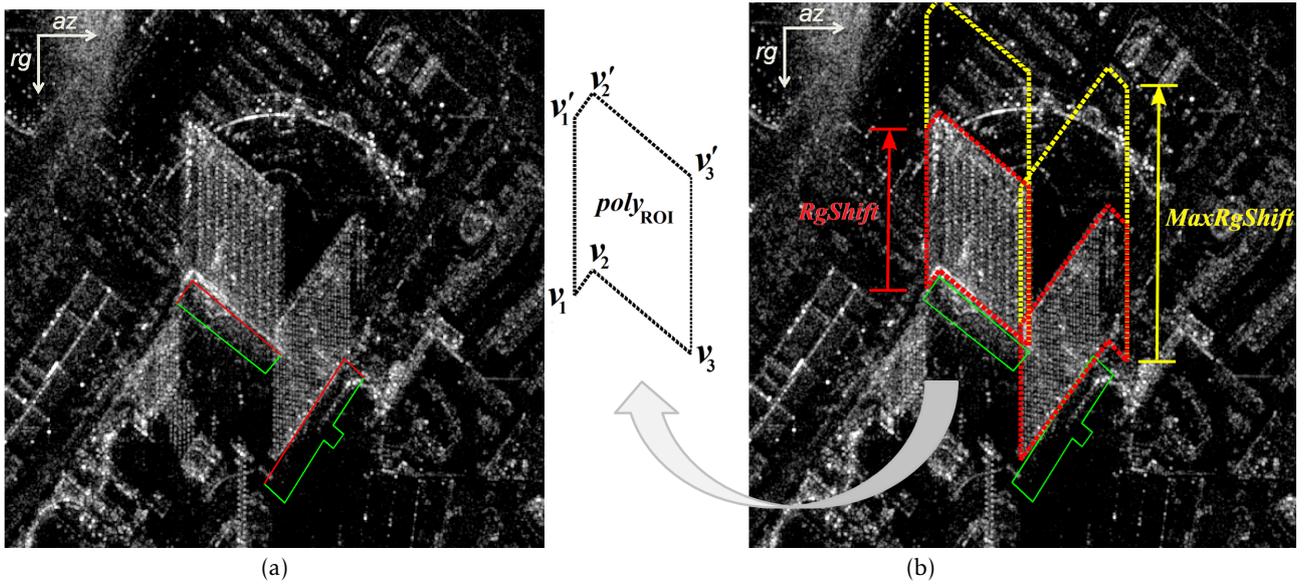


Fig. 43. Building mask extraction: (a) Reference polygons (shown in red and green polylines) of two buildings in the area of interest overlaid onto the SAR intensity map after geocoding. Side of the buildings facing the sensor are shown in red while the other side not visible to the sensor in green; (b) After rotation and range-azimuth shift compensation, the red polylines in (a) are shifted towards the sensor. The yellow dotted lines indicate the maximum range shift of $MaxRgShift$ ($= 318m$) whereas the red dotted lines indicate the $RgShift$ obtained by taking the maximum of the approximate derivative of C .

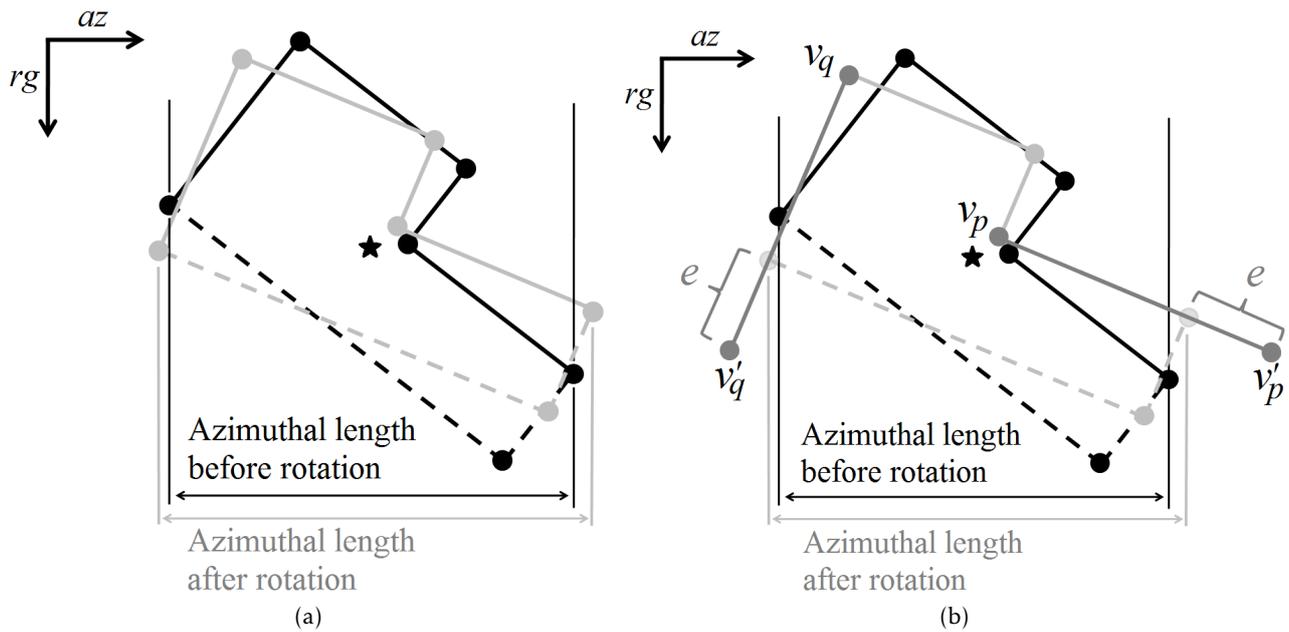


Fig. 44. Graphical illustration of adjusting polyline length: (a) Black and gray polygons indicate polygons before and after rotation (around centroid depicted as black star) respectively with circles representing corresponding vertices. The dotted polylines represent building side not visible to the sensor. It is shown that after orientation the azimuthal length is changed; (b) Length of the outer (first and last edge depicted in dark gray) edges of the polyline facing the sensor is slightly extended by distance e ($= 5m$ in this work). Intensities (interpolated) over these extended edges are analyzed and first and last extended points (i.e., v'_p and v'_q) are respectively replaced by the points on the edges $v_p v'_p$ and $v_q v'_q$ where the approximate derivative (or change in intensities) is maximum.

- dian value of all these intensities is computed;
- The polyline is rotated and shifted with the rotation angle and the azimuth-range shifts which give the maximum of computed median intensities (from the previous step);
 - Finally, the rotation causes the change in the azimuthal length of the polyline which needs to be adjusted (see Figure 44(a)). This is accomplished by first slightly extend-

ing the polyline and later adjusting the lengths of the outer (first and the last) edges of the polyline by analyzing their (interpolated) intensities (same as step 3b except that only the first and last edge of the polyline is used). Figure 44(b) graphically illustrates the adjustment procedure.

- (4) The mask of an individual building is further generated by adopting following steps:
- (a) Iteratively shift the corrected polylines towards the sensor (i.e., one pixel in the range direction). The iterations stop when number of iteration the number of iterations equals $MaxRgShift$ ($= 595$). This value of $MaxRgShift$ is computed on basis of height of tallest building in Las Vegas, the Stratosphere Tower having height of approx. 350 m, and range resolution of the sensor (i.e., maximum building size along elevation appearing in the SAR image of Las Vegas city, computed as $350/0.588(\approx 595$ pixels) where 0.588m is the approx. range resolution). Within each iteration, median of (interpolated) intensities along the shifted polylines are computed and stored in a column matrix C ;
 - (b) Take approximate derivative of C (i.e., calculate differences between adjacent elements of C) and store the result in matrix D ;

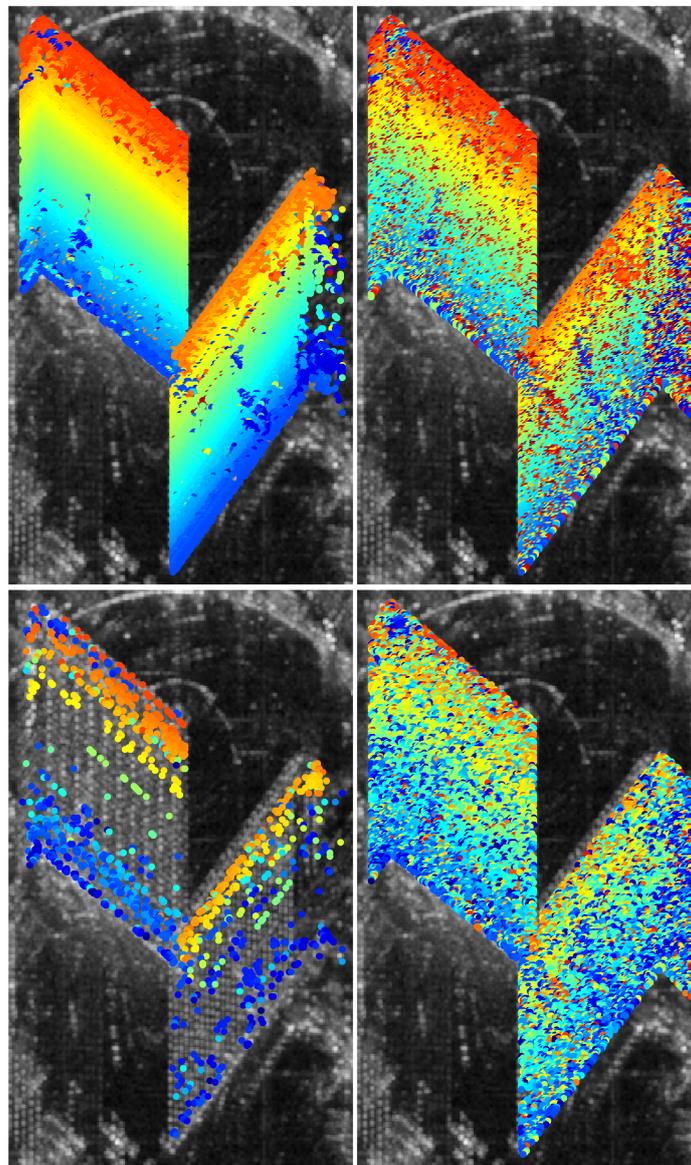


Fig. 45. Reconstructed and color-coded elevation of the two test buildings using 6 interferograms, visualized in two layers, overlaid with intensity. From top to down: first and second layer, respectively; from left to right: M-SL1MMER and SL1MMER, respectively (Zhu et al., 2015b).

- (c) Compute $\arg \max_{RgShift} (D_{RgShift})$ where $RgShift (= 1, \dots, d-1)$ denotes the maximum change point in \mathbf{D} ;
- (d) Determine $poly_{ROI}$ using $RgShift$ and vertices of the identified (compensated) polyline facing the sensor to extract the mask of the particular building.

To elaborate how polyROI is computed, consider a building having three adjacently connected vertices $v_1 - v_2 - v_3$ of the polyline facing the sensor where ‘-’ denotes the adjacency (i.e., v_2 is adjacently connected to v_1 and v_3 , and so on). Assuming that the polyline has been compensated for rotation and range-azimuth shifts, the $poly_{ROI}$ is then simply formed as $v_1 - v_2 - v_3 - v'_3 - v'_2 - v'_1 - v_1$ where $\mathbf{v}'_j(az, rg) = \mathbf{v}_j(az, rg - RgShift)$ with $j = 1, 2$ and 3 (see Figure 43(b)). Finally, $poly_{ROI}$, describing the polygon surrounding the overlaid pixels of the same building in the SAR image, is used to generate the building mask of an individual building.

6.3.3 Application of the extracted building parameters

The downside of advanced repeat-pass InSAR techniques, like PSI and TomoSAR, are their high demand on the data, i.e., typically a stack of 20~100 images over the illuminated area are required. Even using most efficient algorithms, like non-linear least squares and SL1MMER, a minimum number of 11 acquisitions are at least required achieving a reasonable reconstruction in the interesting parameter range of spaceborne SAR. The extracted building parameters (i.e., the mask and iso-height lines) from the previous section may be used in significantly reducing the required number of images while achieving the same quality of reconstruction.

Assuming the height of the building across extracted iso-height line stays constant, the concept of group/joint sparsity can be exploited as e.g., in D, where experiments on bistatic TanDEM-X data stacks depict highly accurate tomographic reconstruction using six interferograms only (see Figure 45). The developed algorithm extends the state of the art SL1MMER algorithm (Zhu and Bamler, 2010a) (Zhu and Bamler, 2012b) to M-SL1MMER, i.e., the multi-snapshot case, by exploiting the joint sparsity concept. It uses multiple snapshots of iso-height pixels identified in SAR images using 2-D OSM data to achieve better reconstruction capability (for details, please refer D. Although the availability of high quality 2-D GIS data is continuously increasing with time, there are places where it is not available. In that case, alternatively one can use the 2-D footprints reconstructed from a preliminarily retrieved TomoSAR point cloud, as demonstrated in previous two chapters.

7 Segmentation and reconstruction of individual trees

Unlike most of the current state of the art approaches which typically employ optical and/or LiDAR data, the work in this chapter is to demonstrate the potential of millimeterwave SAR remote sensing for the reconstruction of individual trees, as well as to preliminarily analyze the achievable accuracy. This chapter essentially summarizes the methodological and experimental sections from the journal paper E (see the Appendix).

The proposed approach delivers first 3-D reconstruction results on individual tree level using airborne millimeterwave TomoSAR data fused from multiple aspects. The essential components of the proposed approach are as follows:

- (1) An unsupervised mean shift clustering has been employed to segment the 3-D point cloud;
- (2) Modeling of trees is adopted via minimum volume enclosing 3-D ellipsoids;
- (3) Core parameters including position, height, and crown diameter of individual trees are extracted out automatically.

7.1 Dataset

The dataset used in this work is acquired from the German MEMPHIS system created by the Fraunhofer Institute for High Frequency Physics and Radar Techniques (Schimpf et al., 2002) during a campaign over Munich, Germany in June 2013 (Schmitt and Stilla, 2014a). The MEMPHIS system is equipped with four receiving antennas enables single-pass multi-baseline InSAR data, consequently leading to highly coherent data even for vegetation by minimizing the effect of temporal decorrelation. Furthermore, the system operates in millimeterwave domain which provides less canopy penetration (i.e., less underestimation of tree heights) in comparison to longer wavelengths typically used in radar remote sensing. This data is particularly suited for individual tree reconstruction (Stilla et al., 2014) (Schmitt et al., 2015):

By fusing TomoSAR reconstruction using data acquired from two opposing aspects (flight heading angles 20° and 200° respectively, a layover- and shadow-free fused 3-D point cloud over the area of interest is generated in (Stilla et al., 2014) (Schmitt et al., 2015). More details of the MEMPHIS sensor and 3-D point cloud generation procedure can be found in E. The selected area of interest is “Alter Nordfriedhof”, an abandoned cemetery, which is used as a public park today, with the target coordinates $48^\circ 09' 13''$ N, $11^\circ 34' 13''$ E. As can be seen in Figure 46(a), this approximately 5 ha large area is mainly characterized by a light planting of deciduous trees, resembling a grove or little wood. The corresponding SAR intensity image and the resulting 3-D TomoSAR point cloud is displayed in Figure 46(b) and Figure 46(c) respectively.

7.2 Methodological workflow

7.2.1 Point cloud clustering

The 3-D points retrieved from TomoSAR processing are clustered by mean shift clustering algorithm to extract individual trees. Since the tree crowns represent denser regions in the point cloud, mean shift algorithm is therefore applied to cluster points in the spatial domain

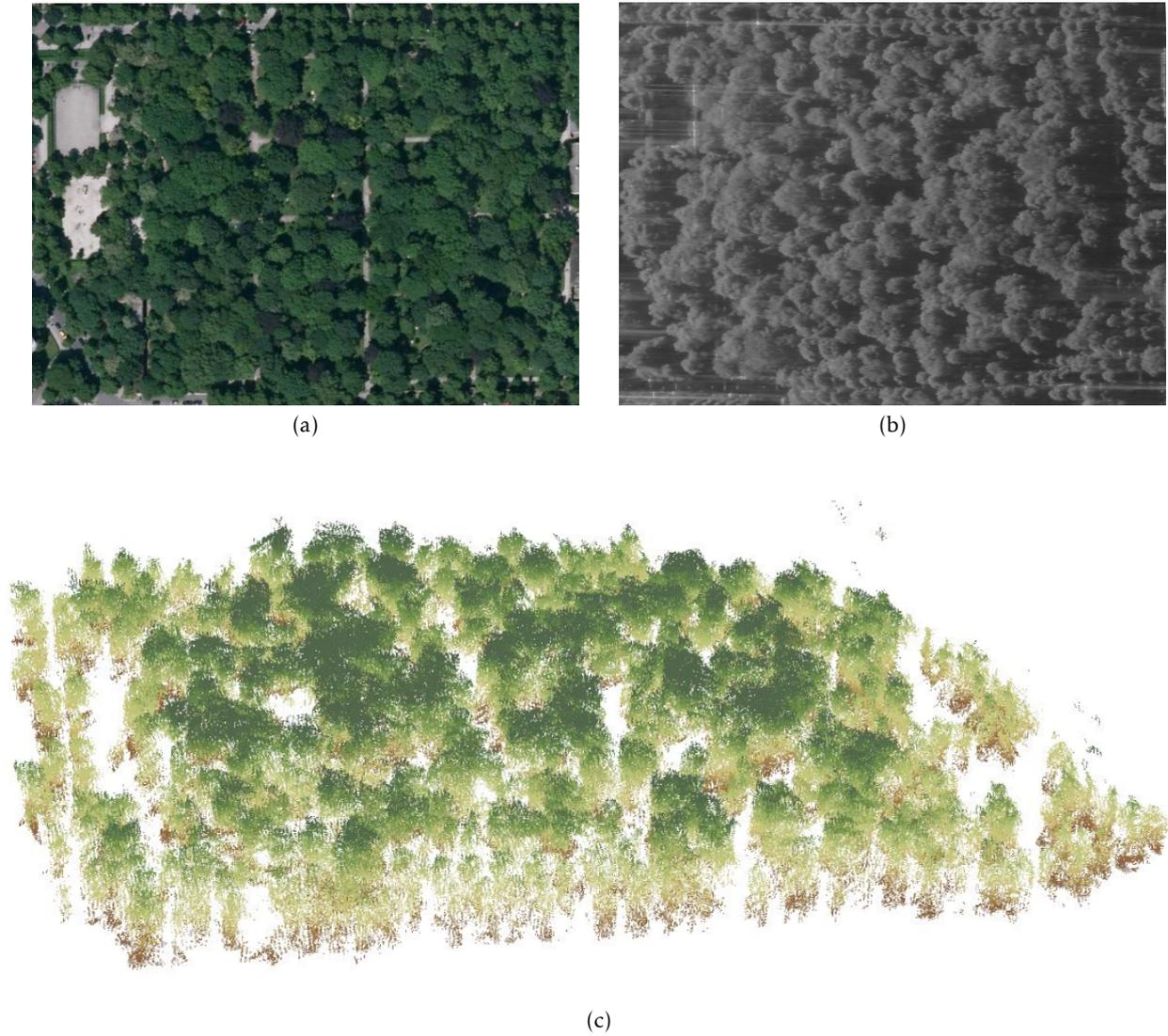


Fig. 46. Test scene “Nordfriedhof” in Munich, Germany: (a) Optical image © Google; (b) MEMPHIS intensity image, range direction from left to right; (c) Corresponding airborne 3-D TomoSAR point cloud.

(i.e., our feature space comprises of spatial coordinates in the Euclidean space). If $p_{j=1,\dots,n}$ denote the indices of 3-D points, then the kernel density estimate at any point \mathbf{p}_i ($i \in j$) is given by the well-known expression (Comaniciu and Meer, 2002):

$$D_{\mathbf{p}_i} = \frac{c}{nb^3} \sum_{j=1}^n g\left(\left\|\frac{\mathbf{p}_i - \mathbf{p}_j}{b}\right\|^2\right) \quad (27)$$

where b is the bandwidth parameter and $g(x)$ is a nonnegative, non-increasing, piecewise continuous function with definite integral i.e., $\int_0^{\infty} g(x) dx < \infty$. From the concept of kernels (Cheng, 1995) (Comaniciu and Meer, 2002), the function $g(x)$ is defined as the profile of the radially symmetric kernel $G(x)$ satisfying $G(x) = cg(\|x\|^2)$ where c is a normalization constant ensuring that $G(x)$ integrates to 1. Different kernels, such as the Epanechnikov kernel and the Gaussian kernel can be used to define the density $D_{\mathbf{p}_i}$. Mean shift essentially seeks modes of the kernel density estimates and works iteratively by shifting every data point toward the weighted mean of points within its neighborhood (defined to be cylindrical in our case). The shift vector $m(\mathbf{p}_i)$ always points toward the direction of the maximum

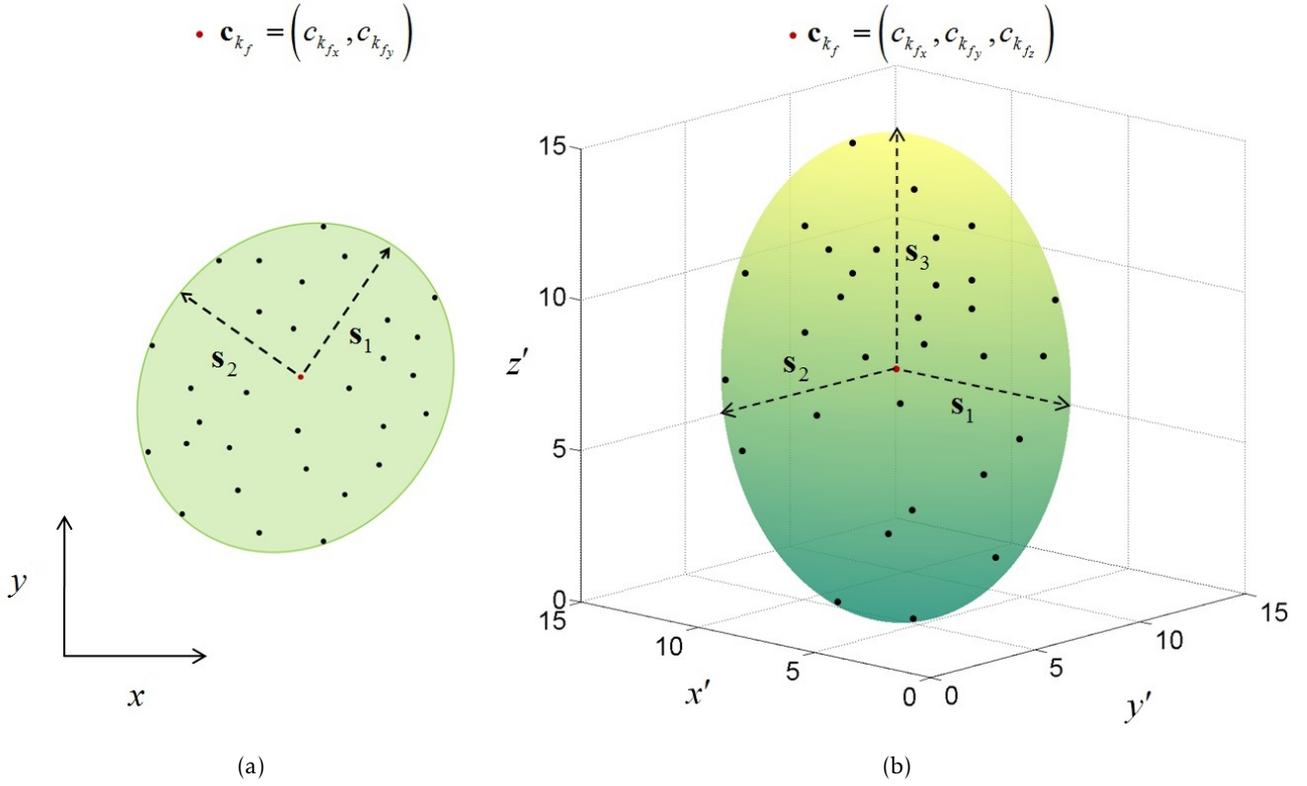


Fig. 47. Illustration of the ellipsoidal modeling: (a) MVEE computed using 3-D points denoted as black dots (projected onto the xy plane). s_1 and s_2 are the computed semi axes of the MVEE; (b) MVEE of (a) is extruded in z direction both upwards and downwards forming a 3-D ellipsoid with third semi axes denoted as s_3 . x' and y' in (a) represents axes of the global coordinate system while x and y denotes the coordinates in local coordinate system. The orientation and axes of the rotated ellipsoid are computed using singular value decomposition of matrix A . Red point in both (a) and (b) represents the ellipsoid center.

increase in the density $D_{\mathbf{p}_i}$ (Comaniciu and Meer, 2002) and is computed as

$$m(\mathbf{p}_i) = \frac{\sum_{j=1}^n \mathbf{p}_j \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{b^2}\right)}{\sum_{j=1}^n \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{b^2}\right)} - \mathbf{p}_i \quad (28)$$

The process of iteration continues until there is no or very little shift in $m(\mathbf{p}_i)$ (i.e., the length of shift vector $m(\mathbf{p}_i)$ is very small). Due to the gradient ascent nature, mean shift return clusters using the concept: attraction of basin. I.e., those points whose trajectories lead to the same mode forms the basin of attraction for that mode (Comaniciu and Meer, 2002) and are clustered into one group. The clustering procedure is repeated till all the points are assigned to their respective modes.

Clustering via mean shift is non parametric in a sense that it does not require the number of clusters in advance nor it needs any defined model for the shape of the resulting clusters. But still, it does require a bandwidth parameter (radius of the kernel) that affects the number of clusters (= number of modes) returned by the algorithm. However, unlike other clustering algorithms such as k -means, fuzzy c -means, Expectation Maximization etc., the bandwidth parameter has some physical meaning and could be set while having some prior knowledge e.g., in our case, average diameter of the tree crowns.

7.2.2 Ellipsoidal modeling

Once the clustering is done, the individual tree clusters are modeled in 3-D using generalized (tri-axial) ellipsoids that are aligned to z -axis. For this purpose, parameters of an arbitrarily oriented minimum volume enclosing ellipse (MVEE) are estimated by first projecting points belonging to individual tree clusters onto the xy plane followed by extruding the 2-D xy ellipse in z direction to form 3-D ellipsoid aligned to z -axis. The motivation for having an z -axis aligned ellipsoid is to obtain geometrically correct tree models such that resulting modeled ellipsoids may have arbitrary orientation in xy plane but remain upright or vertical with respect to the ground (see Figure 47). This is based on the assumption that the trunks of trees to be modeled are vertical to the ground surface.

Following procedure is adopted for the computation of MVEE: If $\mathbf{K} = \{k_l | l = 1, \dots, m\}$ denote m clusters returned by mean shift algorithm and $\mathbf{Q} = \{q_u | u = 1, \dots, r\}$ denote the set of r points belonging to any one particular cluster $k_f (f \in l)$, then any arbitrarily oriented ellipse ε can be candidate for MVEE(\mathbf{Q}) only if all points in \mathbf{Q} lie on or inside its boundary i.e., satisfying (Kumar and Yildirim, 2005):

$$(\mathbf{q}_u - \mathbf{c}_{k_f})^T \mathbf{A} (\mathbf{q}_u - \mathbf{c}_{k_f}) \leq 1 \text{ for } u = 1, \dots, m \quad (29)$$

In the above equation, \mathbf{A} is a $d \times d$ positive finite matrix (d refers to dimension which is 2 in our case) and \mathbf{c}_{k_f} is the center of the ellipse surrounding the clustered points \mathbf{Q} . Semi axes \mathbf{s}_v of such an ellipse are given as $\mathbf{s}_v = (\lambda_v)^{-1/2} \mathbf{v}_v$ where \mathbf{v}_v denote the eigenvectors of \mathbf{A} giving the direction of semi axis whereas λ_v are eigenvalues of \mathbf{A} that are related to the length of these axes (i.e., length of each axis is equal to the $1/\sqrt{\lambda_v}$). The volume of the above ellipsoid is given by $\frac{4\pi}{3} \det(1/\sqrt{\mathbf{A}})$. Thus in order to obtain an MVEE(\mathbf{Q}), $\det(1/\sqrt{\mathbf{A}})$ has to be minimized such that (29) is satisfied together with $\mathbf{A} > 0$. To solve this, Khachyan's first order algorithm (Khachyan, 1996) has been employed which solves this primal problem by formulating it as an optimization problem using Lagrangian duality.

The computed MVEE(\mathbf{Q}) is extended to the third dimension by extruding it in z axis to form a 3-D ellipsoid. The z coordinate of the ellipsoid center and its semi axis length s_3 in z direction are estimated as:

$$\begin{aligned} \mathbf{c}_{k_{fz}} &= \text{median}(h_{k_{f\min}}) + s_3 \text{ where} \\ s_3 &= \frac{1}{2} (\text{median}(h_{k_{f\max}}) - \text{median}(h_{k_{f\min}})) \end{aligned} \quad (30)$$

where $h_{k_{f\max}}$ and $h_{k_{f\min}}$ are the vectors containing maximum and minimum of N height points in any particular cluster k_f .

Once the modeling is complete, the tree parameters e.g., tree heights, crown diameter and trunk location can be directly estimated from the ellipsoid tree model. For instance, the tree height is assumed to be the maximum height of the ellipsoid in z direction, tree crown (or span) radii are given by the x and y semi axes of the ellipsoid while the xy coordinates of the center point of the ellipsoid gives the location of tree trunks. Although this simple parametric model suffices for deciduous trees having approximately an ellipsoidal shape, an extension to more general tree models e.g., as described in (Sheng et al., 2001) seems potentially viable.

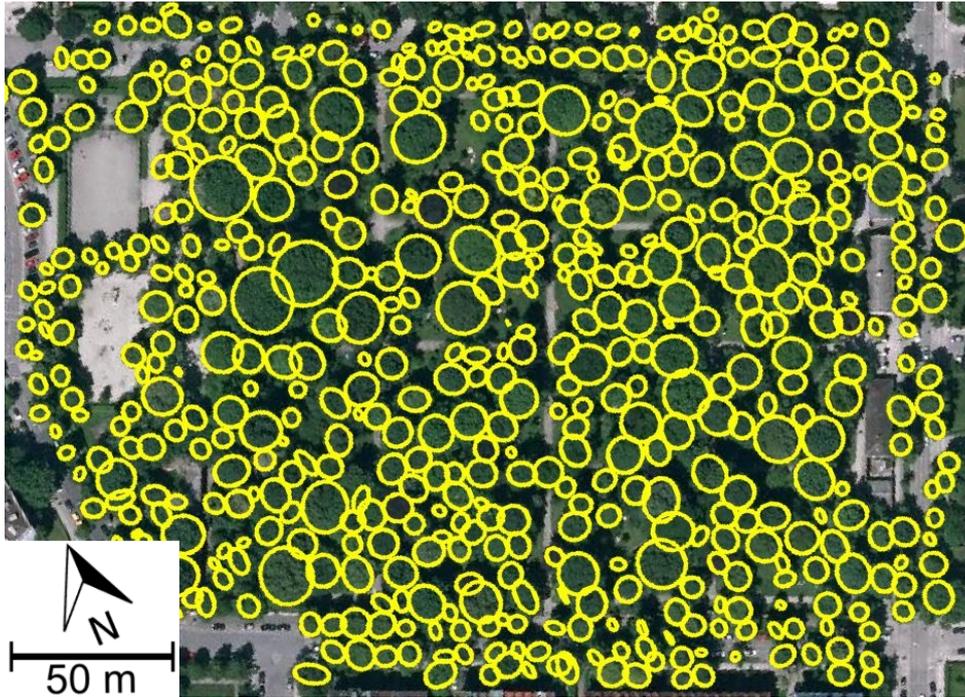


Fig. 48. Reference data of the test scene, created from a LiDAR point cloud and a co-registered orthophoto. Each yellow ellipse indicates one manually extracted reference tree.

7.3 Experimental results and evaluation

7.3.1 Reference data

In order to evaluate, the reconstruction results are compared with manually generated reference dataset. The dataset is acquired from helicopter-borne LiDAR point cloud, containing approximately 0.16 million points (i.e., 3 points/m²) with position accuracy of around 1~20 cm (Schmitt, 2014), co-registered with the orthophoto, provided by the Bavarian Surveying Administration (official name: Landesamt für Digitalisierung, Breitband und Vermessung) with pixel resolution of about 50 cm, co-registered with the orthophoto of Figure 46(a). Reference trees from the dataset were extracted out by human operator based on following sequence of steps:

- ◇ Visual identification of individual trees from the orthophoto and LiDAR data;
- ◇ Manually measuring approximate tree parameters;
- ◇ Assigning maximum LiDAR height point of each tree as its corresponding height;
- ◇ Circular plus LiDAR-derived tree height parameterization of each tree;
- ◇ Ellipsoidal modeling of each tree using the method described in Section 7.2, using only the LiDAR points located within these preliminary tree circles

In total, the reference dataset consists of 570 trees. Figure 48 shows the reference ellipses overlaid onto the corresponding optical (orthophoto) image.

7.3.2 Optimal bandwidth parameter selection

Figure 49 compiles the clustering results for varying kernel bandwidth parameter with four potential cases:

- ◇ One-to-one detection refers to the case such that each reference tree corresponds to one particular cluster.
- ◇ Together with all reference trees, which are oversegmented, i.e. more than one cluster is assigned to them, this adds to the overall rate of detected trees.

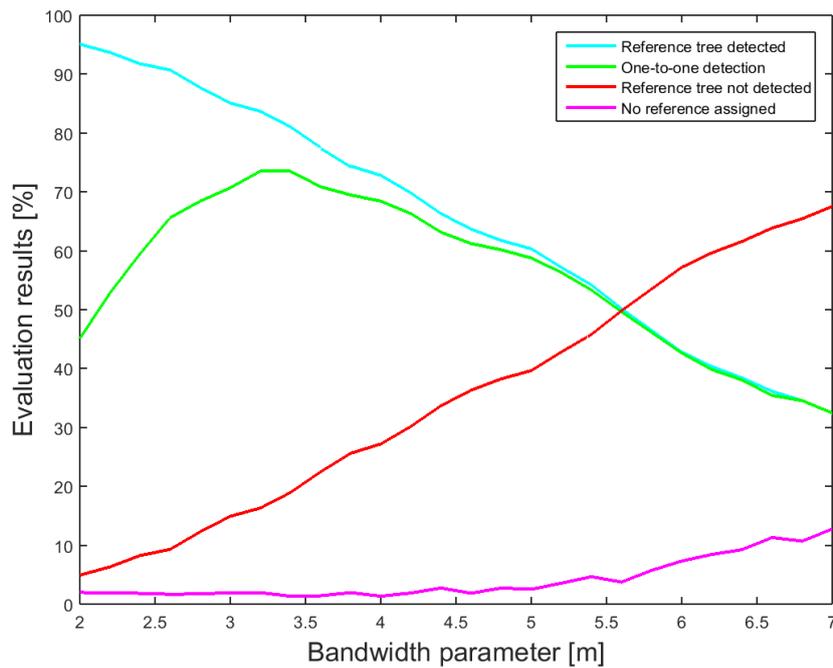


Fig. 49. Determination of the optimal bandwidth parameter by analysis of tree segmentation accuracy for all 570 reference trees. Green indicates perfect matches of one cluster to one reference tree (i.e., optimal detection), red missed reference trees, and purple tree clusters that cannot be assigned to any reference tree. In addition, cyan summarizes the overall detected trees, including oversegmented trees.

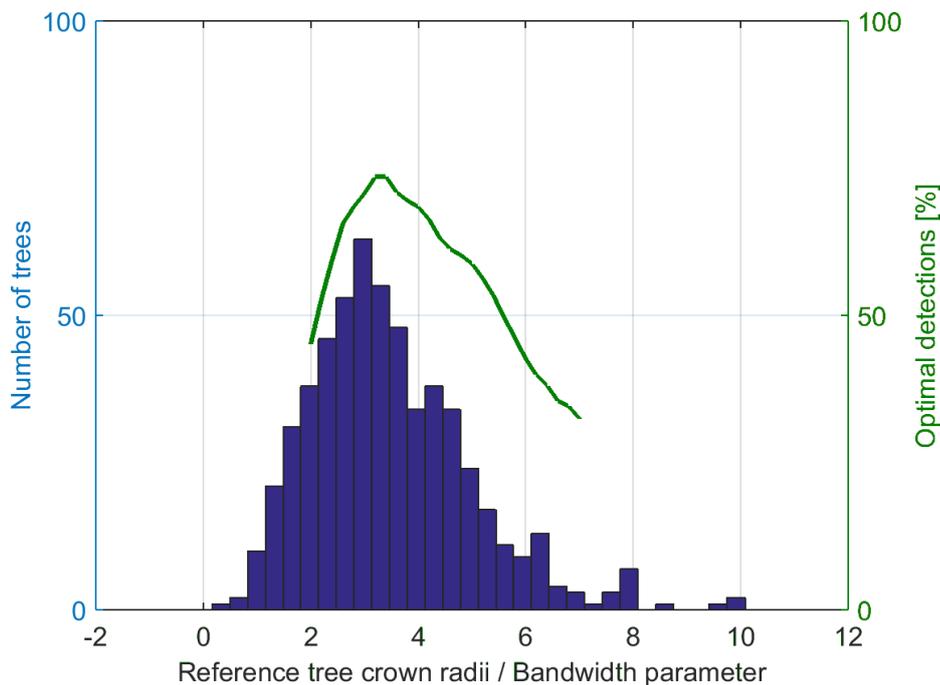


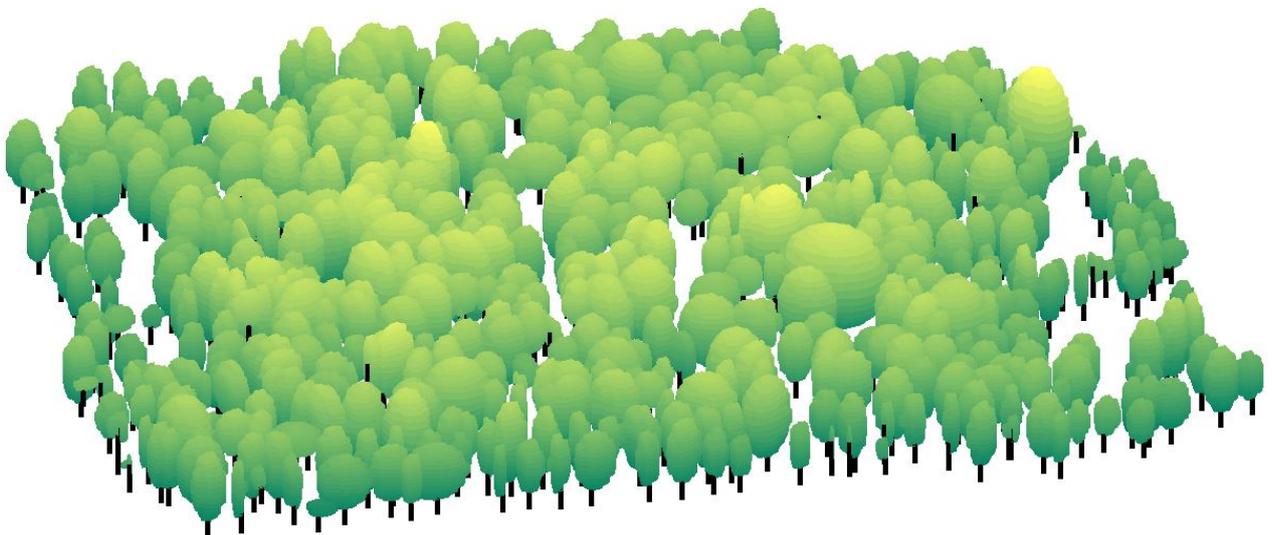
Fig. 50. Correlation of bandwidth parameter and distribution of reference tree radii.

- ◇ When no cluster center falls within 2-D crown outline of the reference tree, it is considered as the case of a missed hit.
- ◇ Finally, some clusters cannot be assigned to any reference tree at all, which is often referred to as false positives.

It is important to mention here that all remaining evaluations are based on one-to-one detections only, whereas oversegmented detections are counted as non-detected trees in order



(a)



(b)

Fig. 51. Tree reconstruction results: (a) Clustering results of applying mean shift algorithm with bandwidth 3.2 m over TomoSAR point clouds depicted in Figure 46(c); (c) Reconstructed 3-D ellipsoidal tree models.

to provide a fair assessment. From Figure 49, one can easily see that the optimal bandwidth parameter is about 3.2 m. It is interesting to mention that the median reference tree radius is 3.25 m. Figure 50 nicely correlates the peak of tree radius histogram with the optimal bandwidth parameter. Based on this, it can be easily inferred that prior knowledge of expected tree radii of the area of interest helps in automatic tuning of the clustering procedure.

7.3.3 Evaluation of reconstruction results

The result of the mean shift clustering of the remaining point cloud with optimal bandwidth parameter of 3.2 m is displayed in Figure 51(a). The points have been segmented into 566 clusters, which already resemble individual trees by visual impression. The final result of the ellipsoid modeling process is depicted in Figure 51(b), including tree crowns of different shape and hypothetical stem positions.

Figure 52 shows a projection of the reconstructed ellipsoids onto the 2-D reference data. The visual map in xy (or ground) plane of reference and estimated tree locations is depicted

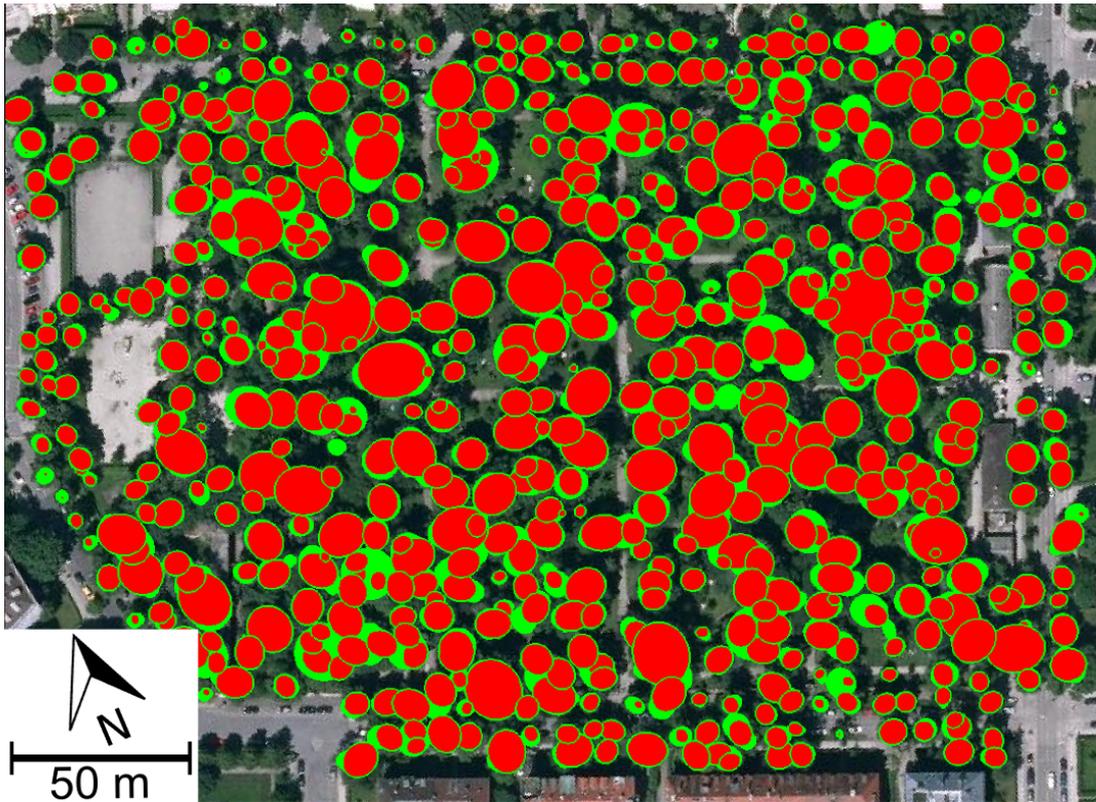


Fig. 52. Ellipsoid models projected onto the 2-D reference dataset for one-to-one comparison. The green circles indicate the reference trees, the circles filled in red correspond to the reconstructed trees.

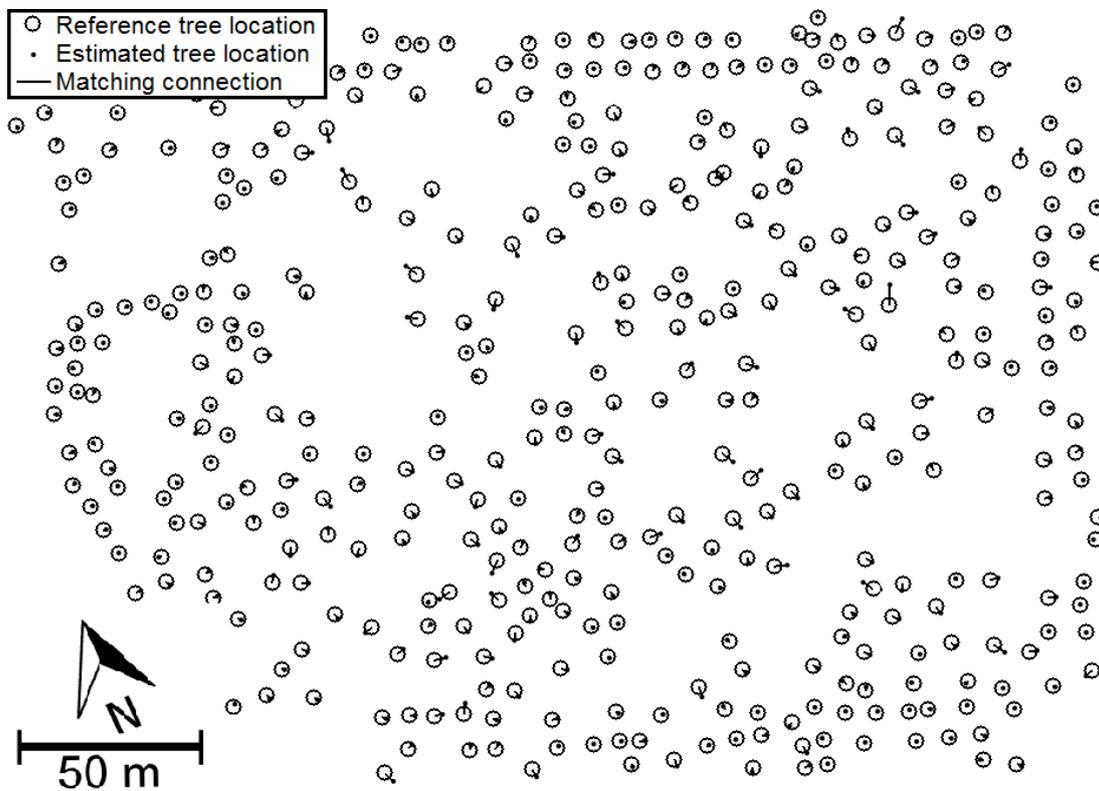


Fig. 53. Localization errors of reconstructed trees and reference trees.

in Figure 53. Additionally, Figure 54 provides the error distributions of the tree heights and the crown radii. The evaluation of the proposed reconstruction strategy with respect to manually acquired LiDAR data set reveals that around 74% of all trees are correctly segmented

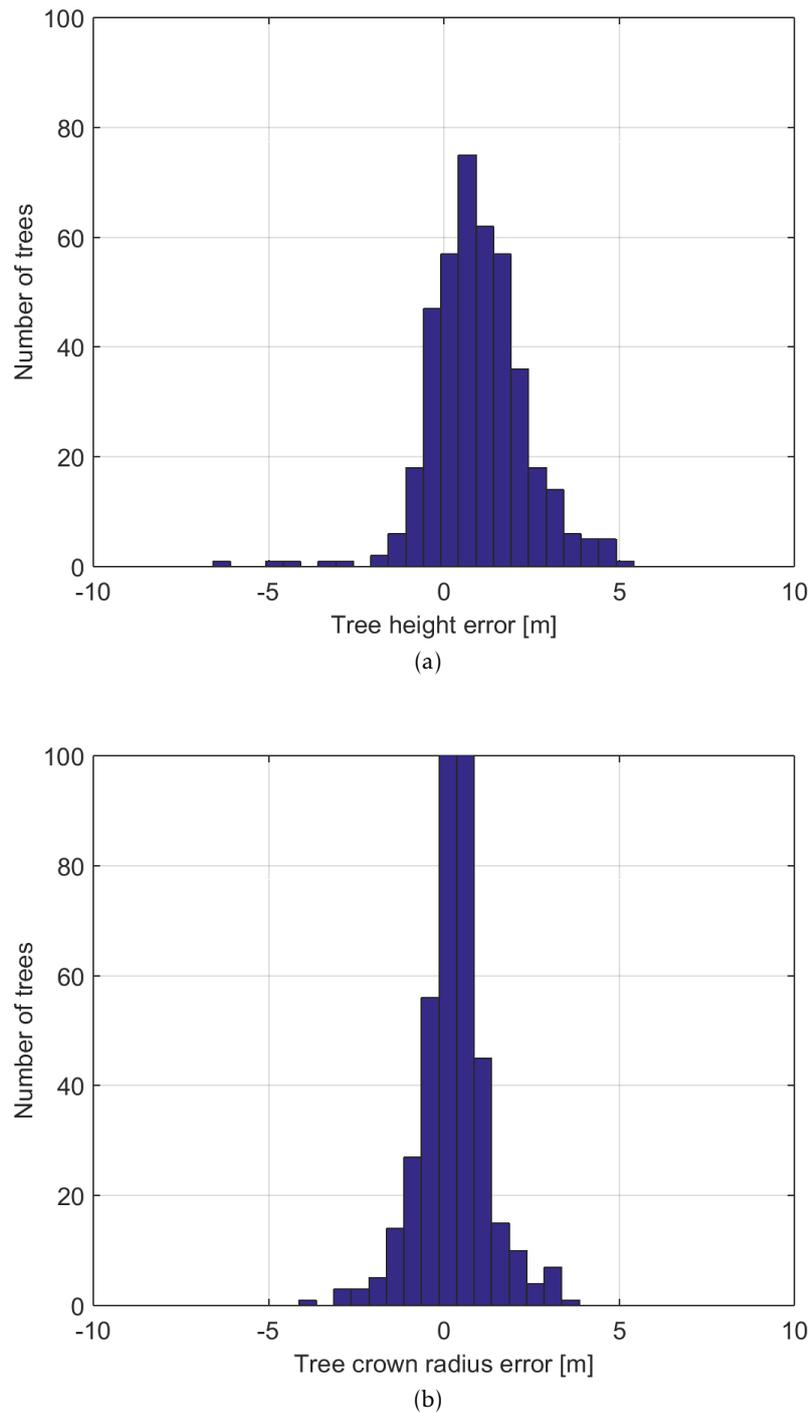


Fig. 54. Distributions of (a) the tree height errors and (b) the crown radii errors. It can be seen that both tree heights and crown radii tend to be slightly overestimated.

and subsequently reconstructed. This not only demonstrate the potential of millimeterwave SAR remote sensing but also provides a promising perspective to further ignite researchers towards individual tree recognition from SAR data.

8 Conclusions and outlook

8.1 Summary and conclusion

The dissertation aims to *develop object reconstruction methods using tomographic (TomoSAR) SAR point clouds* that can be used to monitor and visualize the dynamics (i.e., long-term slow/seasonal motion and deformation in the mm- and cm-range) of a city area with very high level of detail. In order to realize this overall goal/vision, four objectives – including one general and three methodological objectives – summarized in Chapter 1 (Section 1.2) have been pursued. The following conclusions can be sequentially drawn in relation to the work presented in this thesis:

- ◇ Modern SAR satellites (including TerraSAR-X/TanDEM-X and CosmoSkyMED) provide VHR data beyond the inherent spatial scales (in the order of 1m) of buildings constituting an invaluable data source for detailed urban mapping on large scale. Processing this VHR data with advanced interferometric techniques such as SAR tomography (TomoSAR) allows generation of 4-D point clouds; containing not only the 3-D positions of the scatterer location but also estimates of temporal deformation on the scale of centimeters or even millimeters; making them very attractive for generating dynamic city models from space;
- ◇ Due to the side-looking geometry and different microwave scattering characteristics of objects appearing in the scene, TomoSAR point clouds exhibit special characteristics that render conventionally employed 3-D object reconstruction techniques based on ALS point clouds not directly applicable;
- ◇ Façade extraction procedures are based on the assumption of having a high number of scatterers on the building façades and hence used the SD as the basis for various operations, including segmentation, orientation parameter estimation, and façade parameter estimation. In most cases, the assumption is valid because of the existence of strong corner reflectors, e.g., window frames, on the building façades. However, there are exceptional cases: 1) the façade structure is smooth, i.e., only very few scatterers can be detected on the façades; and 2) the building is low. In these cases, SD might not be the optimum choice. Alternatively, other scatterer characteristics such as intensity and SNR may be exploited for extraction and reconstruction purposes;
- ◇ In similar context, since the satellite orbits are bound to pass close to the poles of Earth, the approach based on SD (i.e., the presented *façade-only* approach) from one incidence angle only may fail to reconstruct building façades facing North or South due to the missing of measurements. The solution to this problem is provided in this thesis by also incorporating roof points in determining the complete shape/footprint of the building. The approach is modular and completely data driven that does not impose any restrictions on the shape of the building (i.e., any arbitrarily shaped footprint could be reconstructed) which as a consequence allows for a robust reconstruction of both tall and low buildings, and therefore is well suited for urban monitoring of larger areas from space;
- ◇ Availability of roof points is however dependent on the structure and geometry (area) of the roofs. For part of the roof that is smooth or completely invisible to the sensor, no points would be obtained which may result in *under-reconstruction* of the footprint (i.e., the footprint size is under estimated). Such situations can be rectified by using points from other viewing angles (e.g., use of TomoSAR points generated from both ascending and descending orbits) and/or by incorporating some prior model knowledge of the shape to obtain complete building footprints;
- ◇ The reconstructed 2-D/3-D building shape models could be utilized for further innovation applications as demonstrated in this thesis. For instance, 1) By enriching the recon-

structured 3-D model by its deformation history, new kind of 4-D buildings models could be generated; 2) The elevation estimates of raw TomoSAR points could be refined by using their more accurate azimuth and range coordinates together with the corresponding reconstructed model; 3) By extracting key building features in SAR image, e.g., iso-height lines, the concept of joint sparsity could be exploited to improve the TomoSAR inversion procedure;

- ◇ Spaceborne repeat-pass SAR data lacks to reconstruct temporarily incoherent objects (i.e., no points are available over vegetation or trees). To this end, an approach has been developed in the scope of this thesis to reconstruct individual trees from 3-D TomoSAR point cloud generated from multi-baseline InSAR millimeterwave data acquired from multi-aspects with the airborne single-pass MEMPHIS SAR system. The presented experiments highlight three important aspects: 1) The millimeterwave InSAR data is suitable for individual tree reconstruction; 2) Unsupervised clustering of such an airborne point cloud is possible by mean shift clustering if an expectation value of the typical tree crown radii in the investigated scene is available; 3) Ellipsoids can be used to model the shapes of deciduous trees, thus providing approximate estimates for core tree parameters such as location, height and diameter;
- ◇ As a final remark, although the work presented in this thesis primarily focuses on object reconstruction methods using TomoSAR point clouds generated from spaceborne- and airborne- SAR sensors, the developed methods are also applicable to work on unstructured 3-D point clouds generated from different sensors with similar configuration (i.e., oblique geometry) with both low and high point densities.

8.2 Future work

In relation to the current status of the work presented in this dissertation, a few topics for further study are outlined which mainly concern 1) Refinement of 2-D/3-D building shape reconstruction; 2) Roof modeling; 3) Object based fusion; 4) New visualization methods – SAR street view concept; and 5) Improvement in individual trees reconstruction.

8.2.1 Refinement of 2-D/3-D building shape reconstruction

The extraction of building points is based on few assumptions: I.e., façades are assumed vertical while and roof points are extracted based on two assumptions, namely planarity and higher relative heights. This assumption also enables detection/reconstruction of other man-made vertical structures e.g., advertisement boards, monuments etc. Imposing dimensional constraints on the reconstructed footprints or other geometric measures may help in further distinguishing them from buildings.

8.2.2 Roof modeling

In this research, both façades and roof points have been utilized to build the overall 2-D shape of the building footprint which after incorporating the height information produce extruded 3-D building models. Explicit modeling of roof structure has not been exploited yet. In this regard, a possible future research direction is to explore the potential of extending the algorithm towards generation of automatically reconstructed complete watertight prismatic (or polyhedral) 3-D/4-D building models from space.

8.2.3 Object based fusion

Unknown absolute height values of different reference points, which are chosen independently while processing individual VHR SAR data stacks, hinder in direct fusion of TomoSAR point clouds from multi-views. To this end, 2-D/3-D footprints independently produced from each single aspect TomoSAR point cloud could be matched/correlated to develop a more precise object based geometric TomoSAR fusion method compared to two existing ground points based (Gernhardt et al., 2012) and feature based (Wang and Zhu, 2015) approaches. An alternative approach is geodetic TomoSAR (Zhu et al., 2015c) where the absolute 3-D coordinates of the reference points are retrieved using imaging Geodesy (Eineder et al., 2011) and stereoSAR (Gisinger et al., 2015).

8.2.4 New visualization methods – SAR street view concept

◇ TomoSAR point clouds are very attractive for dynamic (i.e., 4-D) city modeling. In this context, the combined visualization of 3-D city models with 4-D TomoSAR point clouds is challenging as it requires combination of different visualization techniques such as point-based visualization for displaying 3-D points and polygon based visualization for plotting the reconstructed model of the object. Furthermore, in order to meet the needs of different users with different SAR and TomoSAR experience, volume and/or texture based visualization methods may also need to be incorporated to cope with users operating on various systems with different graphics processing capabilities.

◇ Proper visualization may also help in developing a new concept which may be called "SAR Street View", similar to Google Street View, where instead of displaying the photorealistic 3-D models only, following two layers or modes of operation could be incorporated:

- 1) *Deformation mode* – Reconstructed 3-D model with deformation estimates visualized over it;
- 2) *3-D SAR mode* – 2-D radar reflectivity mapped onto the reconstructed 3-D model of the object of interest.

The first mode may be better for monitoring and visualization of the dynamics of urban infrastructure. Changes and deformation of different parts of individual buildings would be accessible for different types of users, e.g. geologists, civil engineers, decision makers, etc.

The second mode could be helpful to develop an intuition about the multiple bounce effect. Moreover, the very bright individual scatterers that behave as corner reflectors can be precisely located consequently helping in better understanding of the nature of scattering.

8.2.5 Improvement in individual tree reconstruction

Despite of the nice reconstruction results provided in Chapter 7 (Section 7.3), there is still room for further improvement:

- ◇ Results of final tree reconstruction to derive core tree parameters are essentially dependent on the clustering algorithm. In this context, a more robust clustering procedure needs to be developed that is able to handle varying tree sizes and shapes (or types). An extension to already proposed mean shift approach would be to employ an adaptive bandwidth parameter selection based on local neighborhood characteristics. Moreover, additional information such as SAR intensity (or amplitude) values may also be utilized to achieve robust detection/segmentation of individual trees;
- ◇ Simple ellipsoidal model may not be appropriate for variety of different tree shapes. In this regard, a generalized ellipsoid model also accounting for varying crown curvature could be utilized to develop a more generic reconstruction approach;

- ◇ To resolve problems related to missed trees due to occlusion, volumetric tomography based methods might be employed;
- ◇ Moreover, a more robust tree parameter estimation strategy may be employed to significantly reduce the over-estimation bias;
- ◇ Finally, since no analysis of any kind of data can replace in-situ observations, a more accurate reference data may be generated for fair evaluation.

Appendices

- A Zhu X. X., Shahzad, M., 2014. Façade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 52(6): 3541-3552**

<http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6573417>

Facade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds

Xiao Xiang Zhu, *Member, IEEE*, and Muhammad Shahzad, *Student Member, IEEE*

Abstract—Recent advances in very high resolution tomographic synthetic aperture radar inversion (TomoSAR) using multiple data stacks from different viewing angles enables us to generate 4-D (space-time) point clouds of the illuminated area from space with a point density comparable to LiDAR. They can be potentially used for facade reconstruction and deformation monitoring in urban environment. In this paper, we present the first attempt to reconstruct facades from this class of data: First, the facade region is extracted using the density estimates of the points projected to the ground plane, the extracted facade points are then clustered into individual facades by means of orientation analysis, surface (flat or curved) model parameters of the segmented building facades are further estimated, and the geometric primitives such as intersection points of the adjacent facades are determined to complete the reconstruction process. The proposed approach is illustrated and validated by examples using TomoSAR point clouds generated from stacks of TerraSAR-X high-resolution spotlight images from two viewing angles, i.e., both ascending and descending orbits. The performance of the proposed approach is systematically analyzed. To explore the possible applications, we refine the elevation estimate of each raw TomoSAR point by using its more accurate azimuth and range coordinates and the corresponding reconstructed building facade model. Compared to the raw TomoSAR point clouds, significantly improved elevation positioning accuracy is achieved. Finally, a first example of the reconstructed 4-D city model is presented.

Index Terms—Facade reconstruction, point cloud, TerraSAR-X, tomographic synthetic aperture radar (SAR) inversion (TomoSAR), 4-D city model.

I. INTRODUCTION

THE automatic detection and reconstruction of buildings and other man-made structures from space is becoming increasingly important with the growing number of population in urban areas. Reconstructed models can serve as a major component in the realization and generation of 4-D (space-time) or even higher dimensional dynamic city models. Urban planning and management [1], tourism [2], architecture [3], damage assessment [4], and disaster management [5] are few among their various potential application areas.

Manuscript received December 7, 2012; revised May 4, 2013; accepted June 19, 2013. Date of publication August 1, 2013; date of current version March 3, 2014. This work is part of the project “6.08 4D City” funded by the International Graduate School of Science and Engineering, Technische Universität München, and the German Research Foundation (DFG, Förderkennzeichen BA2033/3-1).

X. Zhu is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany, and also with the Chair of Remote Sensing Technology, Technische Universität München, 80333 Munich, Germany (e-mail: xiao.zhu@dlr.de).

M. Shahzad is with the Chair of Remote Sensing Technology, Technische Universität München, 80333 Munich, Germany.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2013.2273619

Recent advances in very high resolution synthetic aperture radar (SAR) imagery and its key attributes—self-illumination and all-weather capability—have attracted the attention of many remote sensing analysts in the characterization of urban environments. Various techniques have been developed that make use of SAR imagery for building detection and reconstruction. Complex building shapes surrounded by roads and other structures make building detection a challenging problem. One possible solution is to discriminate buildings from other objects using the building height and width measurements extracted from SAR imagery [6]. The key issue is then the building height retrieval. For this purpose, various methods have been developed, including using sound electromagnetic models [7], layover [8] or shadow analysis [9] and simulation-based methods [10]. In [11], an approach particularly suited for the detection and extraction of large buildings based on information acquired from interferometric SAR (InSAR) data is proposed. Stochastic model-based and low level feature-based approaches for extracting and reconstructing buildings from a single SAR intensity image are presented in [12] and [13], respectively. Wang *et al.* [14] presented an approach for building extraction from high-resolution single-aspect polarimetric SAR data. Since, in urban areas, the structures are densely packed, the appearance of one particular building is dependent on the viewing angle of the sensor. Using a single-view SAR image, it is difficult to detect buildings that have no orientation component in the sensor’s azimuth direction [15]. To overcome this limit, multiview SAR acquisitions are required. In [16], an approach for estimating building dimensions using multiview SAR images is presented. Bolter and Leberl [17] and Thiele *et al.* [18] proposed methods for building reconstruction based on multiview InSAR data. Building reconstruction in context to stereoscopic SAR radargrammetric and multiview polarimetric SAR acquisitions has also been used in [19] and [20], respectively.

Due to the complex urban scenes and inherent problems of SAR images such as speckle effect and layover [21], the previously presented approaches give solutions to building reconstruction but only to some extent. Spaceborne meter resolution SAR data, together with multipass InSAR techniques, including persistent scatterer interferometry (PSI) and tomographic SAR inversion (TomoSAR), allow us to reconstruct the shape and the undergoing motion of individual buildings and urban infrastructures [22]–[25]. PSI exploits bright and long-term stable objects, i.e., the persistent scatterers (PSs). However, it is restricted to single scatterers in an azimuth–range pixel. TomoSAR, on the other hand, extends the synthetic aperture principle into the elevation and temporal domain

for 3-D and 4-D imaging [24]–[29]. It resolves the layover problem by separating multiple scatterers along the elevation direction [24]–[26], [28]. Without any preselection of pixels as PSI does, TomoSAR offers tremendous improvement in detailed reconstruction and monitoring of urban areas, particularly man-made infrastructures [24]. Experiments using TerraSAR-X high-resolution spotlight data stacks show that the scatterer density obtained from TomoSAR is on the order of 600 000–1 000 000/km² compared to a PS density on the order of 40 000–100 000 PS/km² [23], [24]. The rich scatterer information retrieved by TomoSAR from multiple viewing angles enables us for the first time to generate 3-D point clouds of the illuminated area with a point density comparable to LiDAR [23], [30]. These point clouds can be potentially used for building facade reconstruction in urban environment from space with the following considerations:

- 1) TomoSAR point clouds reconstructed from spaceborne data have a moderate 3-D positioning accuracy on the order of 1 m [31], while (airborne) LiDAR provides accuracy typically on the order of 0.1 m [32]. Due to limited orbit spread and the small number of images, the location error of TomoSAR points is highly anisotropic with an elevation error typically one or two orders of magnitude higher than in range and azimuth. Another peculiarity of TomoSAR and PSI point clouds is that, due to multiple scattering, ghost scatterers may be generated that appear as outliers far away from a realistic 3-D position [33].
- 2) Due to the coherent imaging nature and side-looking geometry, TomoSAR point clouds emphasize different objects than LiDAR: 1) The side-looking SAR geometry enables TomoSAR point clouds to possess rich facade information, and results using pixelwise TomoSAR for the high-resolution reconstruction of a building complex with very high level of detail from spaceborne SAR data are presented in [34]; 2) temporarily incoherent objects, e.g., trees, cannot be reconstructed from multipass spaceborne SAR image stacks; and 3) to obtain the full structure of individual buildings from space, facade reconstruction using TomoSAR point clouds from multiple viewing angles is required [35], [36].
- 3) Complementary to LiDAR and optical sensors, SAR is so far the only sensor capable of providing the fourth dimension information from space, i.e., temporal deformation of the building complex [37], and microwave scattering properties of the facade reflect geometrical and material features.

However, in order to provide a high-quality spatiotemporal 4-D city model, object reconstruction from these TomoSAR point clouds is emergent. Motivated by these chances and needs, in this paper, we attempt to detect and reconstruct the building facades from TomoSAR point clouds.

Three-dimensional object reconstruction techniques from point clouds are widely employed using LiDAR data. They mostly make use of the fact that man-made structures such as buildings usually have parametric shapes. After selecting local sets of coplanar points using 3-D Hough transform or random sample consensus algorithms, 3-D objects are recon-

structed by surface fitting in the segmented building regions [38]. Numerous methods are employed for building roof segmentation and reconstruction such as unsupervised clustering approaches [39], region growing algorithms [40], and graph-based matching techniques [41]. These techniques, however, cannot be directly applied to TomoSAR point clouds due to different object contents captured by the side-looking SAR as mentioned earlier.

In this paper, we present an approach for the detection and reconstruction of building facades from these unstructured TomoSAR point clouds. It consists of three main steps, including facade detection and extraction, segmentation, and reconstruction: First, the facade region is extracted by analyzing the density of the point projected to the ground plane, the extracted facade points are then clustered into segments corresponding to individual facades by means of orientation analysis, and surface (flat or curved) model parameters of the segmented building facades are further estimated. Furthermore, we refine the elevation estimate of each raw TomoSAR point by using its more accurate azimuth and range coordinates and the corresponding reconstructed surface model of the facade. The proposed approach is illustrated and validated by examples using TomoSAR point clouds generated from stacks of TerraSAR-X high-resolution spotlight images from two viewing angles, i.e., both ascending and descending orbits.

This paper is structured as follows. Section II introduces the facade surface model assumptions and describes the data set used in this paper. In Section III, the proposed approach is presented, and the different processing steps are described in detail. In Section IV, the results obtained on the test buildings using TomoSAR point clouds generated from multiple viewing angles are presented, and the performance of the proposed approach is analyzed. Two application examples of the reconstructed facade models are presented in Section IV. Finally, in Section V, some conclusions are drawn, and future perspectives are outlined.

II. MODEL ASSUMPTION AND THE DATA SET

A. Model Hypotheses/Assumptions

Many existing approaches assume polyhedral building structure, i.e., roofs as planar surfaces and facades as vertical flat planes. The building model is then described by vertex points determined from intersections of ridges and boundary line segments. In most cases, the building footprint is assumed to be a rectangle polygon. As a consequence, boundary tracing algorithms usually regularize the identified boundary points to straight line segments such that the building footprints represent a polygonal shape. In our work, we assume the facades to be vertical but model their footprints by polynomial lines to allow a wider variety of architecture.

B. Data Set

The data set used in this paper is TomoSAR point clouds generated from two stacks (each comprising 25 images) of TerraSAR-X high-resolution spotlight images from ascending (36° incidence angle) and descending (31° incidence angle) orbits as reported in [34]. Due to the different scattering properties

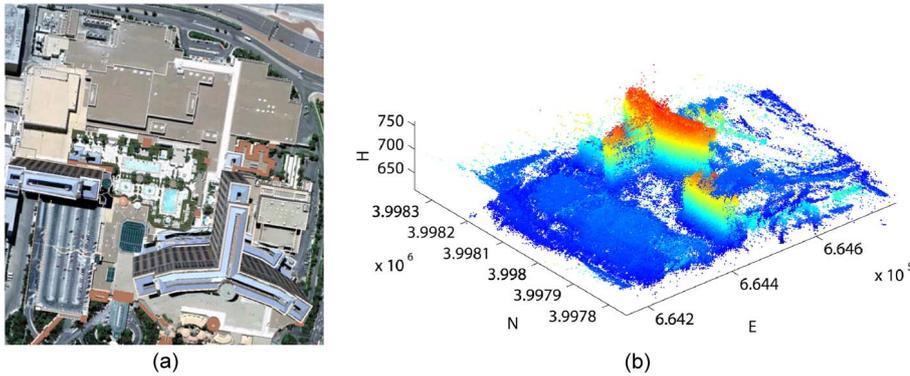


Fig. 1. Test buildings—Bellagio hotel, Las Vegas. (a) Optical image (Google). (b) Fused TomoSAR point clouds from both ascending and descending orbits in UTM coordinates.

from different geometries, there is little chance to identify a common reference point for both stacks. This problem results in a shift in the elevation directions of both point clouds reconstructed from these two stacks with different viewing angles. To obtain the full structure of individual buildings from space, the point clouds are first geodetically fused by determining this shift in elevation direction [30], [42]. The proposed facade reconstruction approach is then applied to the resulting fused point clouds. Fig. 1(a) shows the optical image of our test buildings, the Bellagio hotel complex in Las Vegas. Fig. 1(b) gives an overview of the fused input TomoSAR point cloud in universal transverse mercator (UTM) coordinates. The size of the test area is about $520 \times 570 \text{ m}^2$. The number of TomoSAR points is approximately 0.4 million.

III. METHODOLOGY

As illustrated in Fig. 2, the proposed approach consists of three main steps, including facades detection and extraction, segmentation, and reconstruction.

A. Facade Detection and Extraction

Building facade detection and extraction is generally the first and important step toward the reconstruction of 3-D building models from point clouds generated from aerial or spaceborne acquisitions. A common approach in LiDAR point cloud processing is to first compute a (or use an already existing) digital terrain model (DTM) by filtering techniques, e.g., morphological filtering [43]–[45], gradient analysis [46], or iterative densification of triangular irregular network structure [47], [48], and then use the DTM to extract nonground points [49], [50] from the rasterized point cloud data. The rasterized nadir-looking LiDAR point cloud gives a digital surface model (DSM). The difference of DSM and DTM provides us a normalized DSM that gives us the height variations among nonground points. By exploiting geometrical features such as deviations from the surface model [51], local height measures [32], [45], roughness [45], and slope variations [43], [46], building points can be extracted out. Some methods support the building detection problem by explicitly using 2-D footprints [52], [53]. They help in reducing the building detection problem by providing the building regions but can suffer from inaccurate positioning accuracy [54] and artifacts introduced during data acquisition [38].

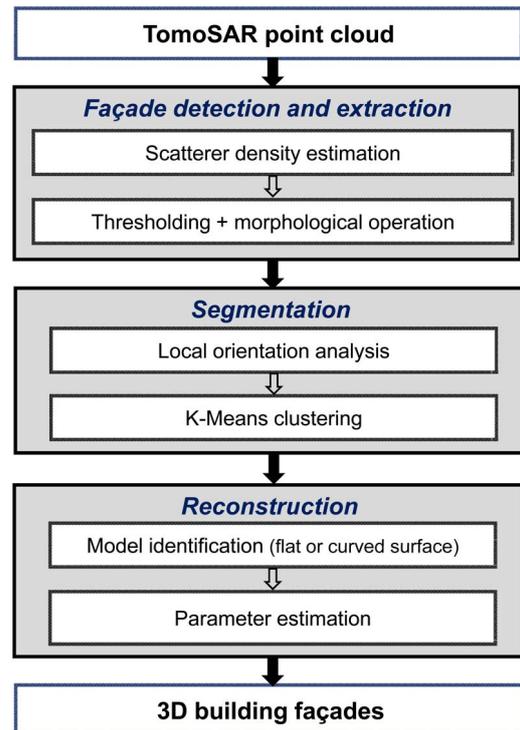


Fig. 2. Workflow of the proposed approach.

Our proposed approach for extracting building facades exploits the idea of orthogonally projecting the points onto the 2-D ground plane as presented in [38]. However, instead of estimating local planes to refine the building outline, the 2-D scatterer (point) density (SD) in the horizontal x - y (ground) plane is used to extract facade points. The proposed method works directly on the unstructured 3-D TomoSAR points. SD is locally estimated for each grid point defined on the ground plane by first accumulating the number of points within a local window and then dividing by the window size. By exploiting the fact that, for a side-looking instrument like SAR, the point density is much higher for vertical structures (value depending on the building height), the building facades are extracted.

Fig. 3(a) shows the SD map of the input TomoSAR point cloud shown in Fig. 1(b). The grid spacing is set to $1 \times 1 \text{ m}^2$. The grid cells having a point density less than a specified threshold TH are removed. A mask is then generated after a morphological dilation which, in turn, is used for building

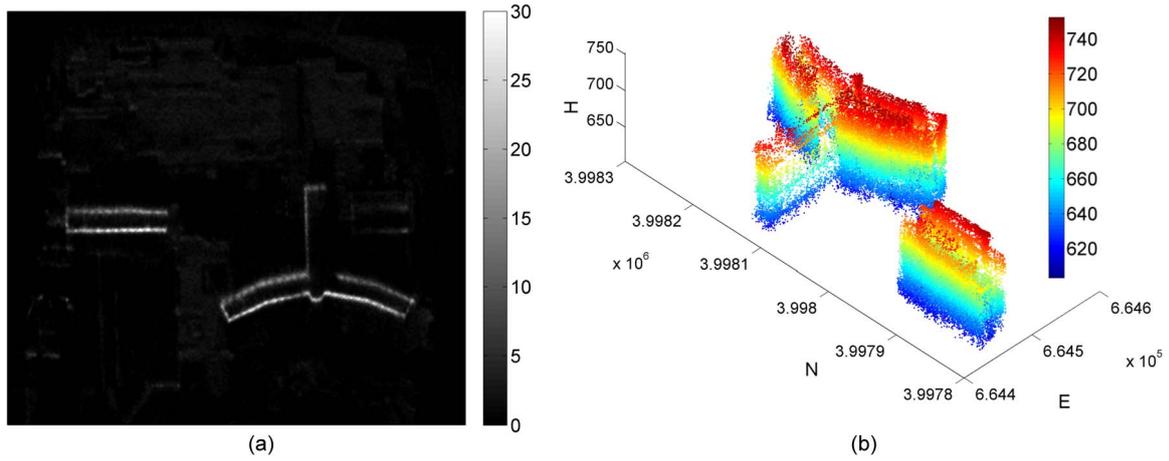


Fig. 3. Facade (detection) extraction. (a) Scatterer density map in the ground plane on a $1 \times 1 \text{ m}^2$ grid. (b) Extracted building facade points. The color bar indicates the (a) number of points/ m^2 . (b) Sea-level height in meters of each 3-D building facade point.

(facade) point extraction in each grid cell. Fig. 3(b) shows the extracted points belonging to the facades of two different buildings. The number of buildings in the scene is found by analyzing the footprint in the generated mask (which is two in our case).

B. Segmentation

To reconstruct individual facades, the segmentation of the points belonging to the same facade is required. Most segmentation approaches make use of unsupervised clustering techniques. They typically search for local plane features and then perform neighborhood analysis using the detected features [38], [55]. Only considering the planar segments can be too restrictive as in the appearance of the curved surfaces that can be better modeled using second-order or higher order polynomials. Therefore, we search for both planar and curved surfaces and further distinguish them by local footprint orientation analysis.

a) Local Orientation Estimation: Given the set of pixels representing the building regions in the x - y plane, the local orientation angle θ is estimated by weighted least squares (WLS) adjustment. The corresponding weight of the facade pixels within the estimation window is given by the corresponding estimated SD. If there is no point inside the window other than the considered point, that point is no longer considered as part of any facade footprint and hence removed. The estimated local orientation along the facade footprints for a $10 \times 10 \text{ m}^2$ window size is shown in Fig. 4. The orientation change between different facades is quite evident.

b) Feature Vector Selection: Extracted facade points from the preceding step are further clustered into segments corresponding to individual facades. As mentioned earlier, the orientation estimates of different facade footprints are used to cluster the points. To distinguish grid pixels that are spatially far apart but having similar orientation, spatial parameters are also incorporated as features for clustering, i.e., a 3-D feature space (x, y, θ) is adopted.

c) K-Means Clustering: The well-known K-means clustering algorithm is used here for segmentation with the aforementioned 3-D feature vector incorporating spatial features (x, y) and orientation angle θ depicted in Fig. 4.

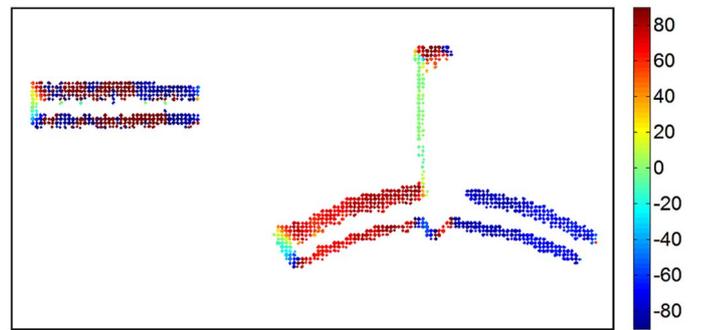


Fig. 4. Orientation estimates in degrees on each grid point of Fig. 3(a). The color bar indicates the degree range $[-90 \sim 90]$.

A common problem related to k -means is to know the number of clusters (facades) k in advance that is not pragmatic in our case. To overcome this limitation, an initial guess about the number of clusters is first computed such that it underestimates k . The points are then clustered with the initial guess using k -means. Points within each resulting cluster that are spatially disconnected are then further separated as smaller clusters. Based on this more detailed clustering result, the points near boundaries of adjacent facades that are normally far away from the corresponding cluster center are further finely clustered.

Initial guess about number of clusters: To determine the initial number of clusters, the within-cluster dispersion is determined in successive clustering runs for varying numbers of clusters.

Let us define D_r as the mean deviation of points in cluster r from its respective center

$$D_r = \sum_{i=1}^{n_r} \frac{d_i}{n_r} \quad (1)$$

where n_r is the number of points in cluster r and d_i is the Euclidean distance of the i th point in r from its center. The dispersion index I_k for k clusters can then be determined as [55]

$$I_k = \sum_{r=1}^k \frac{D_r}{k}. \quad (2)$$

A plot of such dispersion index against the number of clusters gives us an indication on how to choose an appropriate number of clusters [55]. The dispersion index I_k usually decreases significantly with increasing number of clusters and becomes steady afterward. The location of the elbow point can be considered as a good estimate of the number of clusters [55], [56].

Separation of clusters Within clusters: With initially guessed and underestimated k , it is common that facades having similar orientation estimates and relatively small spatial distances have been clustered into one group. It is therefore necessary to extract and treat clusters within clusters that are not spatially connected as separate clusters. For this reason, we perform a connectivity analysis to determine the number of contours and treat each contour as a separate cluster. However, if the contour is very small, i.e., the number of pixels is less than T_p , it is omitted.

Finer clustering: After separating clusters within clusters, the following procedure is adopted for refinement.

- 1) Appropriate polynomial models (first for flat and second for curved) are fitted to estimate the cluster footprints. Model parameters are estimated by L_1 norm minimization that is robust against outliers.
 - 2) Accept points in each corresponding cluster that are within the 2σ distance from the estimated facade, where σ is the standard deviation of the residual.
 - 3) Connectivity analysis is then carried out for the rejected points. Three possible cases can exist.
 - a) The point is isolated. In this case, it is removed.
 - b) The point is connected to other discarded points. In this case, the number of points is counted, and in case the number is less than T_p , they are removed. Otherwise, they are merged together to form a new cluster.
 - c) The point belongs to another cluster. In this case, the discarded point is assigned to another existing cluster.
- By following the aforementioned procedure, extracted facade points are clustered into segments.

C. Reconstruction

A facade is normally characterized by a flat or curved surface, edges (facade boundary), and the corresponding vertices. These features will be reconstructed in this section.

a) Model Identification (Flat or Curved Facade Surfaces): The facade surfaces to be modeled are first classified to flat and curved surfaces by analyzing derivatives of the local orientation angle θ . The curved surfaces have gradually changing orientations across their footprint compared to flat surfaces that have ideally constant orientations. We first compute the first derivatives θ' of the orientation angle θ for each facade footprint. Since the locally estimated θ' is usually noisy, second-order polynomial fitting is applied for denoising. The decision on whether an individual facade footprint is flat or curved is based on the behavior of θ' . Facade footprints with too small orientation variation are considered to be flat while facade footprints with gradually changing orientation are considered to be curved.

TABLE I
PROCEDURE TO FIND THE ADJACENT SEGMENTS
OF A PARTICULAR FACADE SEGMENT

Given: S_1, S_2, \dots, S_k , where $k =$ number of facade segments

Select a facade segment S_i whose adjacency is to be determined

for $m = 1, \dots, k$

Determine a fused segment S' by appending the segment

$$S_m \text{ to } S_i \text{ (i.e., } S' = S_m \cup S_i \text{)}$$

Compute a mask of the points belonging to S'

Find the number of contours in the mask

if number of contours = 1

S_m and S_i are adjacent to each other

end if

end for

b) Parameter Estimation: Finally, model parameters for each segmented facade are estimated. Each extracted point in Fig. 3(b) is assigned a weight corresponding to its SD depicted in Fig. 3(a). Two-dimensional facade footprints are then reconstructed by a WLS method. Polynomials are used to model the facade footprints in the x - y plane

$$f_p(x, y) = \sum_{q=1}^p a_q x^i y^j \quad i + j \leq q \quad (3)$$

where i and j are permuted accordingly, p is the order of the polynomial, and the number of terms in the above polynomial is equal to $(p+1)(p+2)/2$. Flat and curved surfaces are modeled using first-order ($p=1$) and second-order ($p=2$) polynomial coefficients, respectively. Higher order polynomials could be used to model more complex building structures.

c) Vertex Determination: Once the facade model parameters are estimated, the final step is to describe the overall shape of the building footprint by further identifying adjacent facade pairs and determining the intersection of their facade surfaces.

The adjacency of facades is described by an adjacency matrix that is built up via connectivity analysis. Table I shows the procedure that we propose to decide the adjacency of one particular facade segment with other facade segments of the building.

Identified adjacent facade segments are then used to determine the vertex points in 2-D (i.e., facade intersection lines in 3-D). They are found by computing the intersection points between any adjacent facade pair. Since polynomial models are used for facade parameter estimation, the problem of finding vertex points boils down to find the intersection point between the two polynomials corresponding to the two adjacent facades. Depending on flat or curved facades, two cases needed to be discussed.

- 1) Adjacent facades are flat: In this case, there is only one intersection point which is taken as the vertex point.
- 2) One of the two facades is curved, or both facades are curved: There is more than one solution. In these cases, mean Euclidean distances of the possible intersection

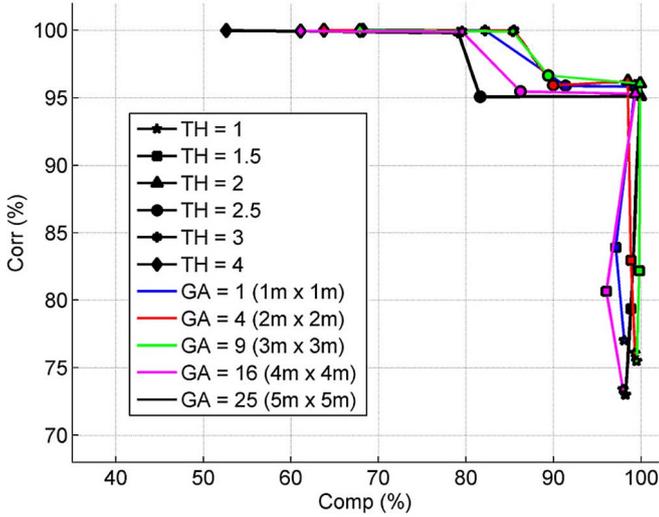


Fig. 5. Completeness versus correctness with the varying TH and GA parameters.

points to both adjacent facades are computed. The point having the smaller mean distance to both facades is considered as the vertex point.

The computed vertex points and the estimated model parameters are then used to finally reconstruct the 3-D model of the building facades.

IV. EXPERIMENTAL RESULTS AND VALIDATION

In this section, we discuss and assess the overall performance of the proposed method: The performance of the building facade extraction is evaluated, clustering issues are discussed, and the positioning accuracy of the facade models is further assessed.

A. Performance Assessment of Facade Extraction

To evaluate the quality of our facade extraction procedure, a pointwise comparison method is employed. Extracted facade points are compared to the reference data, and the results are analyzed for quantitative and qualitative evaluation.

Pseudoreference data: The absence of exact reference data representing the actual facade area restricts the accurate evaluation of the facade point extraction procedure, but since we are already incorporating prior knowledge (i.e., the fact that higher point density areas represent vertical structures) in estimating few facade parameters from a large number of points belonging to an individual building facade, it is highly probable that points closer to the reconstructed facade footprint are indeed points belonging to the corresponding facades. We therefore consider all points that are within 1.5σ of the estimated facade footprints reported in Section IV-C to be true facade points and use them to assess the performance of the proposed facade extraction procedure.

Evaluation Metrics: To evaluate the performance of the proposed building facade extraction procedure, all the points in the input point cloud data are assumed to belong to one of the two categories, i.e., facade or nonfacade points. Any point detected as a facade point by the algorithm that also

TABLE II
OVERALL SUCCESS OF THE EXTRACTION PROCEDURE IN TERMS OF QUALITY FOR SIX TH AND FIVE GA VALUES

$GA [m^2]$	1×1	2×2	3×3	4×4	5×5
$TH [points/m^2]$					
1	75.89	75.71	75.22	72.29	72.06
1.5	81.87	82.16	82.05	78.08	78.65
2	95.20	94.82	95.94	94.71	95.02
2.5	87.92	86.62	86.71	82.82	78.29
3	82.18	85.47	85.32	79.49	79.01
4	68.24	63.77	67.86	61.07	52.62

corresponds to a facade in the reference data set is taken as true positive (TP). Similarly, a point labeled as a facade point but is not actually a facade point in the reference data set is treated as false positive (FP). A false negative (FN) corresponds to a point which belongs to the facade in the reference data set but is wrongly labeled as a nonfacade point by the facade extraction procedure. The performance of the (detection) extraction procedure is then assessed by employing the following evaluation metrics [57], [58]:

$$\left. \begin{aligned} \text{Completeness (\%)} : comp &= 100 \times \left(\frac{TP}{TP+FN} \right) \\ \text{Correctness (\%)} : corr &= 100 \times \left(\frac{TP}{TP+FP} \right) \\ \text{Quality (\%)} : Q &= \frac{comp \times corr}{comp+corr-comp \times corr} = \frac{TP}{TP+FP+FN} \end{aligned} \right\} (4)$$

The metrics mentioned above assess the overall performance of the extraction algorithm. Completeness tells up to what percentage the algorithm has detected the facade points while correctness provides a measure of correct classification. Among them, completeness is particularly important in our application in order to preserve intact facade footprints. The quality Q is crucial when comparing the results obtained from different algorithms [57].

Dependency on window size GA and thresholding parameter TH : Two parameters that influence the number of facade points extracted by the algorithm are as follows: the threshold TH and window size GA . In order to assess the effect of these two parameters on the extraction procedure, the evaluation was carried out using the following sequence of GAs and THs , with $GA = \{1 \times 1, 2 \times 2, 3 \times 3, 4 \times 4, 5 \times 5\} m^2$ and $TH = \{1.0, 1.5, 2.0, 2.5, 3.0, 4.0\} points/m^2$.

To characterize the performance of the extraction procedure, the dependence of completeness and correctness metrics on these two parameters is analyzed. A tradeoff between completeness and correctness can then be chosen based on the adjustable setting parameters. Fig. 5 depicts the completeness and correctness achieved with different TH and GA parameters. It is obvious that a lower TH value results in higher completeness. The higher the TH value, the less the false positives observed, which results in higher correctness. This simply lies in the fact that it is more probable that the grid point with higher SD belongs to a facade. Table II gives us the overall success of the extraction procedure in terms of quality (Q) for six TH and five GA values defined earlier. The performance of the extraction procedure is best under the following parameter settings: $TH = 2 points/m^2$ and $GA = 3 \times 3 m^2$. The overall quality of which

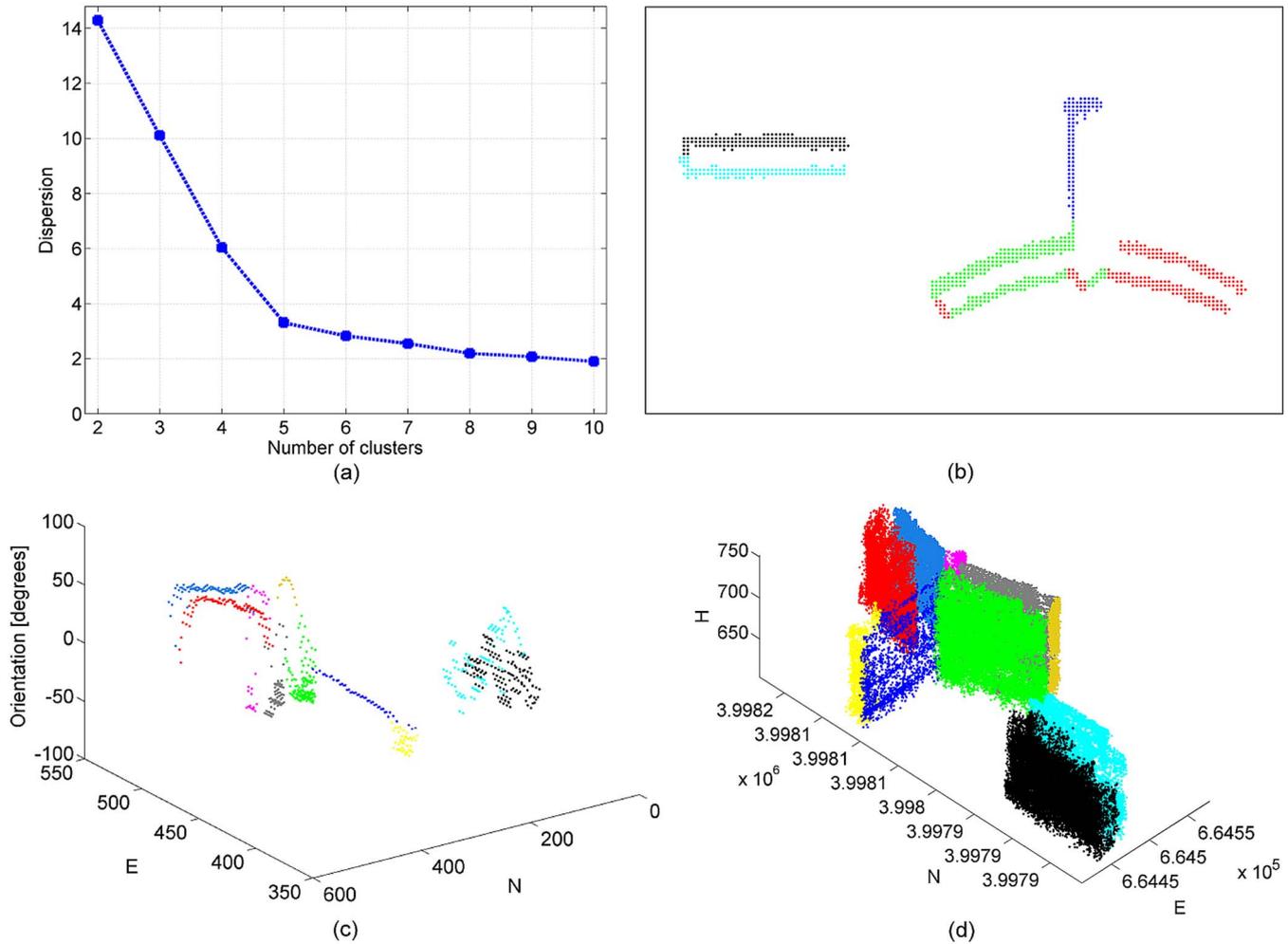


Fig. 6. Segmentation results. (a) Dispersion plot with elbow point at $k = 5$. (b) Corresponding clustering with $k = 5$. (c) Clustered grid points in feature space, i.e., orientation (in degrees) and spatial positions on the ground plane (in meters). (d) Corresponding clustered TomoSAR points.

is 95.94%. Generally better correctness and completeness can also be observed from Fig. 5 with this parameter setting, i.e., a completeness of 99.89% and a correctness of 96.04%.

Based on the aforementioned discussion, in our experiment, we estimate the SD on the ground plane with a grid distance of 1 m. The window size is set to be $3 \times 3 \text{ m}^2$. A facade mask on the ground plane is obtained based on the resulting SD map by setting a threshold of 2 points/m² that are finally used for facade extraction. It is worth to mention that aforementioned parameters are tuned to point clouds generated from TerraSAR-X high-resolution spotlight data, i.e., a resolution of $1.1 \text{ m} \times 0.6 \text{ m} \times 30 \text{ m}$, and the particular incidence angles. For other configurations, the optimal GA will be different.

B. Clustering

Extracted facade points from the previous step are further clustered into segments corresponding to individual facades. For a reasonable initial guess of the number of clusters, Fig. 6(a) shows the plot of I_k by assuming different numbers of clusters $k = 2, \dots, 10$. We can observe that the dispersion index I_k decreases significantly with decreasing number of clusters with k up to 5 and becomes steady afterward. The number of clusters at this elbow point has been chosen as the initial number

of clusters. Fig. 6(b) shows the preliminary clustering results with $k = 5$. It is evident that different facades having similar orientation estimates and relatively small spatial distance have been clustered in one group. Therefore, small clusters that are not spatially connected are separated. Very small contours, i.e., the number of pixels is less than T_p ($T_p = 10$ in our case), are omitted. By following the procedure of the final fine clustering, the extracted points are clustered into ten segments. Fig. 6(c) and (d) shows the color-coded clustering of grid points in feature space after refinement and their corresponding TomoSAR points in UTM coordinates, respectively.

C. Three-Dimensional Facade Reconstruction

By analyzing the orientation derivatives as described in Section III-C, the ten clustered facades in Fig. 6(d) are identified as five curved and five flat facades. Each extracted facade footprint point in 2-D is assigned a weight corresponding to its SD depicted in Fig. 3(a). Two-dimensional facade footprints are then reconstructed using the WLS method.

Once the facade model parameters are estimated, the next step is to find the intersection of these facade surfaces to describe the overall shape of the building footprint. Following the procedure depicted in Table I, the corresponding adjacency

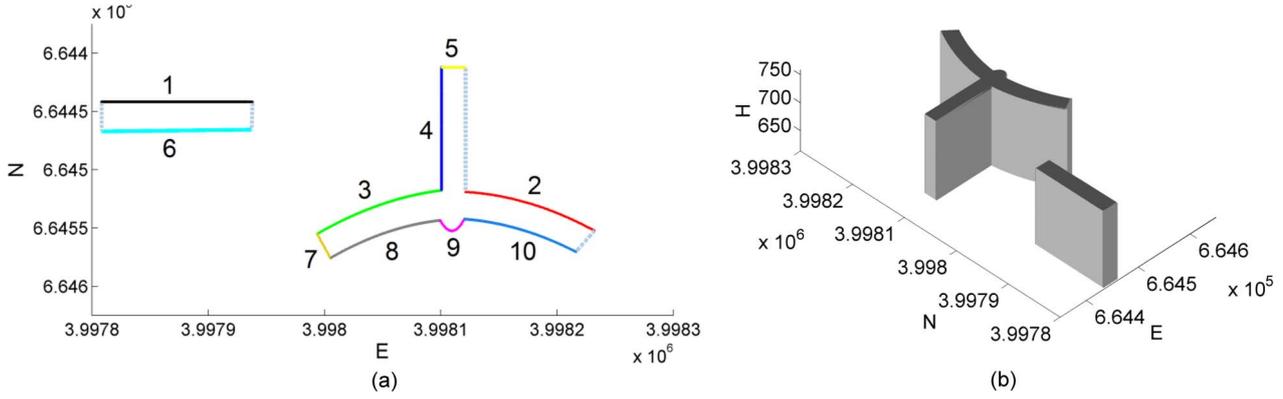


Fig. 7. Three-dimensional view of the reconstructed full building facade model in (a) vertical and (b) 3-D view.

TABLE III

ADJACENCY MATRIX OF THE FACADE SEGMENTS DEPICTED IN Fig. 7(a)

Façades	1	2	3	4	5	6	7	8	9	10
1	-									
2		-								
3			-	x			x			
4			x	-	x					
5				x	-					
6						-				
7			x				-	x		
8							x	-		
9									-	x
10									x	-

TABLE IV

STATISTICS AND ACCURACY OF THE RECONSTRUCTED FACADES

Façades ID	Number of points	Model	RMSE (weighted) [m]	RMSE (unweighted) [m]
1	19,102	Flat	3.81	4.21
2	11,498	Curve	9.51	10.67
3	20,516	Curve	3.89	4.37
4	2,912	Flat	1.27	1.29
5	2,060	Flat	5.05	5.09
6	18,964	Flat	1.56	2.11
7	1,773	Flat	1.15	1.16
8	13,334	Curve	1.64	2.07
9	3,148	Curve	1.91	1.92
10	16,863	Curve	6.88	7.03
Mean RMSE			3.67	3.99

matrix for all the facade segments labeled in Fig. 7(a) is depicted in Table III. The crosses indicate the mutual adjacent facade pairs. The vertex points are found by computing the intersection points between any adjacent facade pair.

Fig. 7 shows the final reconstructed building facade models using the estimated parameters and the determined vertex points. Table IV depicts the statistics of the reconstructed facades. It also assesses the accuracy of the estimated facade models by computing the root-mean-square error (RMSE) of all the points from their respective reconstructed facade. For comparison, facade parameters are also estimated by assigning

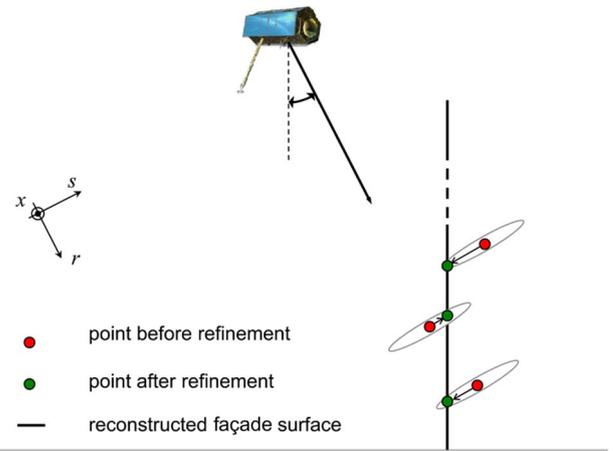


Fig. 8. Basic principle for the elevation estimate refinement of the TomoSAR points by using their more accurate azimuth and range coordinates and the reconstructed facade surfaces.

uniform weights. Compared to facade parameters estimated by assigning weights according to their SD, bigger mean RMSE is observed.

V. APPLICATION EXAMPLES

In this section, the reconstructed model presented in Section IV is used to refine the elevation estimates of the raw TomoSAR point clouds, and an example of the reconstructed 4-D building model is presented.

A. Elevation Estimate Refinement

As briefly mentioned in Section I, due to the limited orbit spread and the small number of images, the location error of TomoSAR points is highly anisotropic with an elevation error typically one to two orders of magnitude higher than in range and azimuth. For TerraSAR-X high-resolution spotlight images with typical parameters, the theoretical relative localization precision of a PS is as follows [59]: 1.7–2.1 cm in range; 3.2–3.8 cm in azimuth, and 62–139 cm in elevation.

The elevation estimates of the TomoSAR points can be refined by using their more accurate azimuth and range coordinates and the identified and modeled facade surfaces as depicted in Fig. 8. This sketch illustrates the refinement principle in

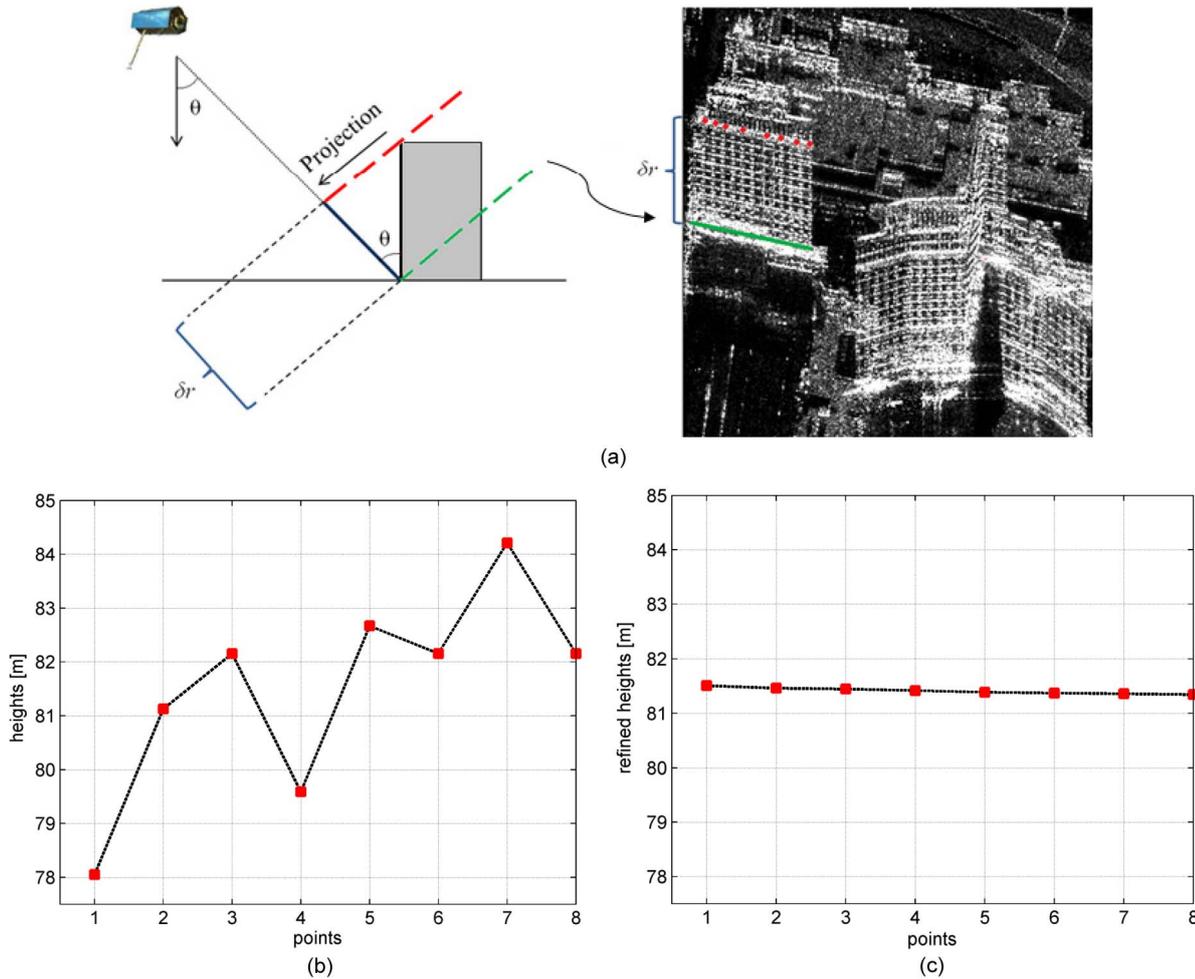


Fig. 9. Elevation estimate refinement. (a) TerraSAR-X mean intensity map from ascending stacks (the red dots are the analyzed points) along with the projection geometry. Height estimates of the analyzed points (b) before and (c) after refinement.

the range–elevation plane. The red points represent the raw TomoSAR point locations at different heights along a facade. The ellipse indicates the error ellipse of the TomoSAR estimate in the range and elevation direction, i.e., much poorer accuracy in elevation compared to range. The black line indicates the reconstructed facade surface. We project the corresponding iso-azimuth-range lines of each point along elevation (directions are indicated by the arrows) to the identified and modeled facade surface that it belongs to. The final refined 3-D position is obtained by taking the elevation coordinate of the intersection point. This is an approximation of an optimal linear estimate. The green points represent the positions after elevation refinement. In this way, we expect to achieve much better elevation estimation accuracy that is on the order of several centimeters, although it is still slightly worse than the ones in azimuth and range due to error propagation.

To validate this improvement, we selected a row of bright points from the intensity image belonging to a facade portion of constant height as shown in Fig. 9(a). Fig. 9(b) and (c) compares the height estimates of the analyzed points before and after the refinement. It is obvious that their height estimates are improved significantly. The standard deviations before and after the refinement are 190 and 5.5 cm, respectively, an im-

provement by a factor of 35 which corresponds quite nicely to the ratio of inherent resolutions in elevation (on the order of 30–50 m) and range (1.1 m).

B. 4-D Building Model

To better monitor the detailed structures of individual buildings, an example of the reconstructed 4-D building model is presented in Figs. 10–12. In Fig. 10, the fused point clouds with refined elevation are visualized by overplotting them onto the reconstructed facade model. The height of the points is color coded. The corresponding estimated motion parameter (in this case, the amplitude of seasonal motion caused by thermal dilation) is illustrated in Fig. 11. This information can be used for developing dynamic building models from spaceborne SAR data that can help to monitor individual buildings and even the whole city. Fig. 12 shows the reconstructed 3-D SAR image, i.e., the reflectivity map overlaid on the facade model. Such an image visualizes in detail how the Bellagio hotel would look like in X-band for our eyes, if they could sense microwaves, from the position of the SAR satellite. Such visualizations may lead to a better understanding of the nature of scattering.

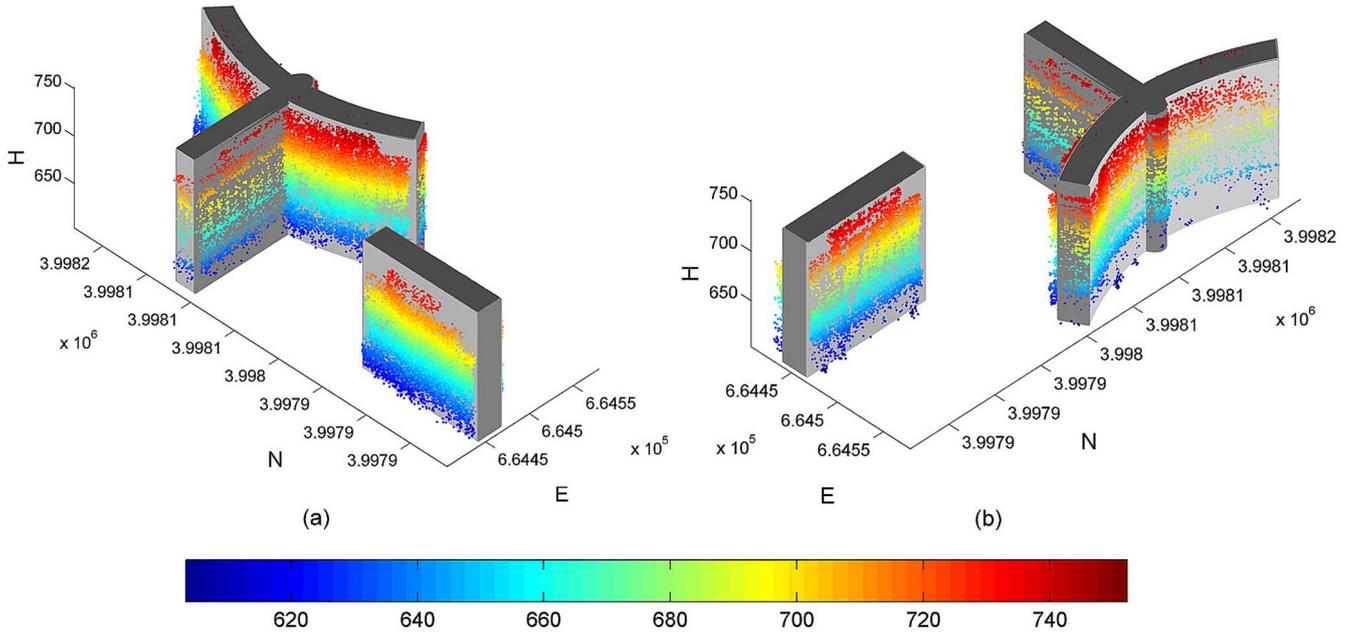


Fig. 10. Point clouds with refined elevation overlaid on the reconstructed facade model. The height of the points is color coded (unit: meters).

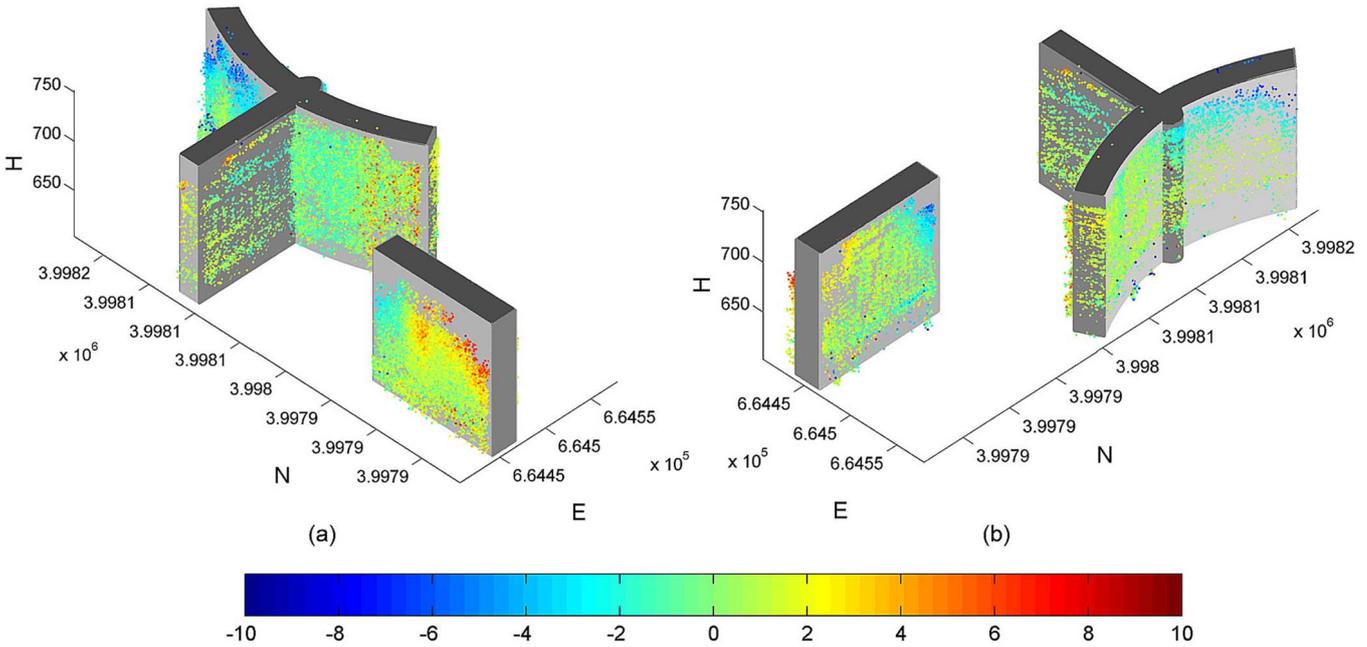


Fig. 11. Reconstructed 4-D building facade model. The amplitude of seasonal motion is color coded (unit: millimeters).

VI. OUTLOOK AND CONCLUSION

TomoSAR point clouds are very attractive for dynamic city model generation. As the first attempt, a facade reconstruction approach tailored to this class of data is proposed in this paper. It consists of three main steps: facade extraction, segmentation, and reconstruction. The proposed approach is illustrated by using fused TomoSAR point clouds from two stacks (ascending and descending) of TerraSAR-X high-resolution spotlight data. We use the reconstructed facade model to refine the

TomoSAR elevation estimates. Compared to the raw TomoSAR point clouds, significantly improved elevation positioning accuracy on the order of several centimeters is achieved. A first example of the reconstructed 4D building model is also presented.

There are several aspects of the proposed reconstruction procedure that can be improved in the future. Among them, the proposed approach is based on the assumption that facades are vertical and the footprint of each segment can be represented by

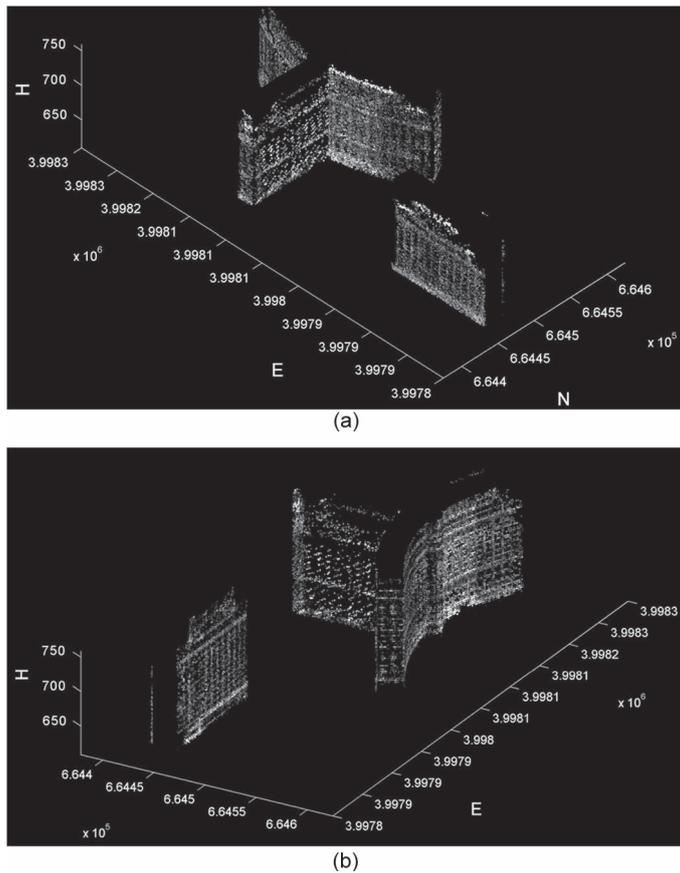


Fig. 12. Reconstructed 3-D SAR image overlapped on the reconstructed facade model. Note that this is not only a projection of the SAR image onto the building models. Rather, the lay-overed brightness contributions from facade and ground have been separated in the tomographic reconstruction step.

a set of polynomial coefficients. The fact that facade parameters are estimated from the segmented points makes the reconstruction performance strongly dependent on the quality of the segmentation. In our experiment, we rely on the assumption of having a high number of scatterers on the building facades and hence used the SD as the basis for various operations, including segmentation, orientation parameter estimation, and facade parameter estimation. In most cases, the assumption is valid because of the existence of strong corner reflectors, e.g., window frames, on the building facades. However, there are exceptional cases: 1) the facade structure is smooth, i.e., only very few scatterers can be detected on the facades, and 2) the building is low. In these cases, SD might not be the optimum choice. Alternatively, we can use other scatterer characteristics such as intensity and SNR for extraction and reconstruction purposes.

In the future, we will also concentrate on object-based TomoSAR point cloud fusion, building roof reconstruction, and automatic object reconstruction for large areas.

ACKNOWLEDGMENT

The authors would like to thank R. Bamler for proofreading the manuscript.

REFERENCES

- [1] J. Benner, A. Geiger, and K. Leinemann, "Flexible generation of semantic 3D building models," in *Proc. Int. ISPRS/EuroSDR/DGPF-Workshop Next Gen. 3D City Models*, K. Gröger, Ed., Bonn, Germany, 2005, pp. 18–22, EuroSDR Publication No. 49.
- [2] J. Döllner, T. H. Kolbe, F. Liecke, T. Sgouros, and K. Teichmann, "The virtual 3D city model of Berlin—Managing, integrating, and communicating complex urban information," in *Proc. 25th UDMS*, Aalborg, Denmark, May 15–17, 2006, pp. 1–12.
- [3] X. Yin, P. Wonka, and A. Razdan, "Generating 3D building models from architectural drawings: A survey," *IEEE Comput. Graph. Appl.*, vol. 29, no. 1, pp. 20–30, Jan./Feb. 2009.
- [4] J. Lee and S. Zlatanova, "A 3D data model and topological analyses for emergency response in urban areas," in *Geospatial Information Technology for Emergency Response*, L. Zlatanova, Ed. New York, NY, USA: Taylor & Francis, 2008.
- [5] S. Kemec, S. Duzgun, S. Zlatanova, D. I. Dilmen, and A. C. Yalciner, "Selecting 3D urban visualization models for disaster management: Fethiye tsunami inundation case," in *Proc. 3rd Int. Conf. Cartogr. GIS*, Nessebar, Bulgaria, Jun. 15–20, 2010, pp. 1–9.
- [6] A. Bennett and D. Blacknell, "Infrastructure analysis from high resolution SAR and InSAR imagery," in *Proc. GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion Over Urban Areas*, 2003, pp. 230–235.
- [7] R. Guida, A. Iodice, and D. Riccio, "Height retrieval of isolated buildings from single high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2967–2979, Jul. 2010.
- [8] F. Tupin, "Extraction of 3D information using overlay detection on SAR images," in *Proc. GRSS/ISPRS Joint Workshop Data Fusion Remote Sens. Over Urban Areas*, 2003, pp. 72–76.
- [9] A. J. Bennett and D. Blacknell, "The extraction of building dimensions from high resolution SAR imagery," in *Proc. Int. Radar Conf.*, 2003, pp. 182–187.
- [10] D. Brunner, G. Lemoine, L. Bruzzone, and H. Greidanus, "Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1487–1504, Mar. 2010.
- [11] P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 611–617, Jan. 2000.
- [12] M. Quartulli and M. Datcu, "Stochastic geometrical modeling for built-up area understanding from a single SAR intensity image with meter resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1996–2003, Sep. 2004.
- [13] A. Ferro, D. Brunner, and L. Bruzzone, "An advanced technique for building detection in VHR SAR images," in *Proc. SPIE Conf. Image Signal Process. Remote Sens.*, 2009, vol. 7477, pp. 74770V-1–74770V-12.
- [14] Y. Wang, F. Tupin, C. Han, and J. M. Nicolas, "Building detection from high resolution PolSAR data by combining region and edge information," in *Proc. IEEE IGARSS*, 2008, pp. IV-153–IV-156.
- [15] J. D. Wegner, U. Soergel, and A. Thiele, "Building extraction in urban scenes from high resolution InSAR data and optical imagery," in *Proc. Urban Remote Sens. Event*, 2009, pp. 1–6.
- [16] R. Hill, C. Moate, and D. Blacknell, "Estimating building dimensions from synthetic aperture radar image sequences," *IET Radar, Sonar Navig.*, vol. 2, no. 3, pp. 189–199, Jun. 2008.
- [17] R. Bolter and F. Leberl, "Detection and reconstruction of human scale features from high resolution interferometric SAR data," in *Proc. ICPR*, 2000, pp. 291–294.
- [18] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, "Building recognition from multispect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007.
- [19] E. Simonetto, H. Oriot, and R. Garelo, "Rectangular building extraction from stereoscopic airborne radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2386–2395, Oct. 2005.
- [20] F. Xu and Y. Q. Jin, "Automatic reconstruction of building objects from multispect meter-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2336–2353, Jul. 2007.
- [21] U. Stilla, U. Soergel, and U. Thoennessen, "Potential and limits of InSAR data for building reconstruction in built-up areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1/2, pp. 113–123, Jun. 2003.
- [22] R. Bamler, M. Eineder, N. Adam, S. Gernhardt, and X. Zhu, "Interferometric potential of high resolution spaceborne SAR," in *Proc. PGF*, 2009, pp. 407–419.
- [23] S. Gernhardt, N. Adam, M. Eineder, and R. Bamler, "Potential of very high resolution SAR for persistent scatterer interferometry in urban areas," *Ann. GIS*, vol. 16, no. 2, pp. 103–111, Jun. 2010.

- [24] X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
- [25] D. Reale, G. Fornaro, A. Paucillo, X. Zhu, and R. Bamler, "Tomographic imaging and monitoring of buildings with very high resolution SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 661–665, Jul. 2011.
- [26] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.
- [27] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," in *Proc. IEEE IGARSS*, Toulouse, France, 2003, pp. 1206–1208.
- [28] G. Fornaro and F. Serafino, "Imaging of single and double scatterers in urban areas via SAR tomography," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 12, pp. 3497–3505, Dec. 2006.
- [29] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 224–237, Jan. 2009.
- [30] S. Gernhardt, X. Cong, M. Eineder, S. Hinz, and R. Bamler, "Geometrical fusion of multitrack PS point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 38–42, Jan. 2012.
- [31] X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, Jan. 2012.
- [32] F. Rottensteiner and C. Briese, "A new method for building extraction in urban areas from high resolution LIDAR data," in *Proc. ISPRS Photogram. Comput. Vision*, Graz, Austria, Sep. 9–13, 2002, pp. 295–301.
- [33] S. Auer, S. Gernhardt, and R. Bamler, "Ghost persistent scatterers related to multiple signal reflections," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 919–923, Sep. 2011.
- [34] X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.
- [35] M. Shahzad, X. Zhu, and R. Bamler, "Facade structure reconstruction using spaceborne TomoSAR point clouds," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 467–470.
- [36] X. Zhu, M. Shahzad, and R. Bamler, "From TomoSAR point clouds to objects: Facade reconstruction," in *Proc. TyWRRS*, Naples, Italy, 2012, pp. 106–113.
- [37] X. Zhu and R. Bamler, "Let's do the time warp: Multi-component non-linear motion estimation in differential SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, Jul. 2011.
- [38] P. Dorninger and N. Pfeifer, "A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds," *Sensors*, vol. 8, no. 11, pp. 7323–7343, Nov. 2008.
- [39] S. Filin, "Surface clustering from airborne laser scanning data," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 3A, pp. 119–124, Sep. 2002.
- [40] F. Ahmed and S. James, "Reconstructing 3D buildings from Lidar data," in *Proc. ISPRS Commis. III Symp.*, Graz, Austria, Sep. 9–13, 2002, pp. 1–6.
- [41] G. Forlani, C. Nardinocchi, M. Scaioni, and P. Zingaretti, "Complete classification of raw LIDAR data and 3D reconstruction of buildings," *Pattern Anal. Appl.*, vol. 8, no. 4, pp. 357–374, Feb. 2006.
- [42] Y. Wang, X. Zhu, Y. Shi, and R. Bamler, "Operational TomoSAR processing using multitrack TerraSAR-X high resolution spotlight data stacks," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 7047–7050.
- [43] G. Sithole and G. Vosselman, "Experimental comparison of filtering algorithms for bare-earth extraction from airborne laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, no. 1/2, pp. 85–101, Aug. 2004.
- [44] K. Zhang, S.-C. Cheng, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing non-ground measurements from airborne LIDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 872–882, Apr. 2002.
- [45] F. Rottensteiner, J. Trinder, S. Clode, and K. Kubik, "Building detection by fusion of airborne laser scanner data and multi-spectral images: Performance evaluation and sensitivity analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 2, pp. 135–149, Jun. 2007.
- [46] G. Vosselman, "Slope based filtering of laser altimetry data," in *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XXXIII (Pt. B3)*, 2000, pp. 935–942.
- [47] G. Sohn and I. Dowman, "Terrain surface reconstruction by the use of tetrahedron model with the MDL criterion," in *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XXXIV (Pt. 3A)*, 2002, pp. 336–344.
- [48] P. Axelsson, "DEM generation from laser scanner data using adaptive TIN models," *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 33, no. B4/1, pp. 110–117, 2000.
- [49] F. Rottensteiner and C. Briese, "A new method for building extraction in urban areas from high-resolution LIDAR data," *Proc. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 34, no. 3A, pp. 295–301, 2002.
- [50] U. Weidner and W. Förstner, "Towards automatic building extraction from high-resolution digital elevation models," *ISPRS J. Photogramm. Remote Sens.*, vol. 50, no. 4, pp. 38–49, Aug. 1995.
- [51] N. Pfeifer, A. Kostli, and K. Kraus, "Interpolation and filtering of laser scanner data-implementation and first results," in *Proc. Int. Arch. Photogramm. Remote Sens. XXXII (Pt. 3/1)*, 1998, pp. 153–159.
- [52] J.-Y. Rau and B.-C. Lin, "Automatic roof model reconstruction from ALS data and 2D ground plans based on side projection and the TMR algorithm," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. S13–S27, Dec. 2011.
- [53] J. Overby, L. Bodum, E. Kjems, and P. M. Iisoe, "Automatic 3D building reconstruction from airborne laser scanning and cadastral data using Hough transform," *Int. Arch. Photogramm. Remote Sens.*, vol. 35, pt. B3, pp. 296–301, 2004.
- [54] N. Haala and M. Kada, "An update on automatic 3D building reconstruction," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 6, pp. 570–580, Nov. 2010.
- [55] A. Sampath and J. Shan, "Segmentation and reconstruction of polyhedral building roofs from aerial LiDAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1554–1567, Mar. 2010.
- [56] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 63, no. 2, pp. 411–423, 2001.
- [57] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, "A comparison of evaluation techniques for building extraction from airborne laser scanning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 11–20, Mar. 2009.
- [58] G. Sohn and I. Dowman, "Data fusion of high-resolution satellite imagery and LIDAR data for automatic building extraction," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 1, pp. 43–63, May 2007.
- [59] S. Gernhardt, *High Precision 3D Localization and Motion Analysis of Persistent Scatterers Using Meter-Resolution Radar Satellite Data*. Munich, Germany: Verlag der Bayerischen Akademie der Wissenschaften, 2012, ser. Reihe C No. 672, p. 130, Deutsche Geodätische Kommission.



Xiao Xiang Zhu (S'10–M'12) received the B.S. degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006 and the M.Sc. and Dr. Ing. degrees from Technische Universität München (TUM), München, Germany, in 2008 and 2011, respectively.

She has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany, since May 2011, where she is the Head of the Team Signal Analysis, and she is also with the Chair of Remote Sensing Technology, TUM. Her main research interests are as follows: advanced InSAR techniques such as high-dimensional tomographic synthetic aperture radar imaging and SqueeSAR; computer vision in remote sensing, including object reconstruction and multidimensional data visualization; and modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multi/hyperspectral image analysis.



Muhammad Shahzad (S'12) received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, and the M.Sc. degree in autonomous systems from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany. Since December 2011, he has been working toward the Ph.D. degree in the Chair of Remote Sensing Technology (LMF), Technische Universität München (TUM), München, Germany. His research topic is the automatic 3-D reconstruction of objects from point clouds retrieved from spaceborne synthetic-aperture-radar image stacks.

B Shahzad, M., Zhu, X. X., 2015. Robust Reconstruction of Building Façades for Large Areas Using Spaceborne TomoSAR Point Clouds. IEEE Transactions on Geoscience and Remote Sensing 53(2): 752-769

<http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6842672>

Robust Reconstruction of Building Facades for Large Areas Using Spaceborne TomoSAR Point Clouds

Muhammad Shahzad, *Student Member, IEEE*, and Xiao Xiang Zhu, *Member, IEEE*

Abstract—With data provided by modern meter-resolution synthetic aperture radar (SAR) sensors and advanced multipass interferometric techniques such as tomographic SAR inversion (TomoSAR), it is now possible to reconstruct the shape and monitor the undergoing motion of urban infrastructures on the scale of centimeters or even millimeters from space in very high level of details. The retrieval of rich information allows us to take a step further toward generation of 4-D (or even higher dimensional) dynamic city models, i.e., city models that can incorporate temporal (motion) behavior along with the 3-D information. Motivated by these opportunities, the authors proposed an approach that first attempts to reconstruct facades from this class of data. The approach works well for small areas containing only a couple of buildings. However, towards automatic reconstruction for the whole city area, a more robust and fully automatic approach is needed. In this paper, we present a complete extended approach for *automatic* (parametric) reconstruction of building facades from 4-D TomoSAR point cloud data and put particular focus on *robust* reconstruction of large areas. The proposed approach is illustrated and validated by examples using TomoSAR point clouds generated from a stack of TerraSAR-X high-resolution spotlight images from ascending orbit covering an approximately 2-km² high-rise area in the city of Las Vegas.

Index Terms—Clustering, facade reconstruction, point density, TerraSAR-X, tomographic SAR (TomoSAR) inversion, 4-D point cloud.

I. INTRODUCTION

AUTOMATIC detection and reconstruction of buildings has been an active research area for at least two decades. Despite an extensive research effort, the topic is still of great interest due to ever increasing growth of urban population which gives rise to a wide range of potential applications in various fields. For instance, building footprints (i.e., 2-D building outline/shape) can be used for urban landscape development [1], urban planning [2], damage assessment [3], disaster management [4], navigation purposes [5], etc. Additionally,

2-D footprint data in combination with height information can generate 3-D building models. Such models are essential for virtual city modeling [6] and 3-D GIS applications (e.g., commercial software such as Google Earth, Apple Maps, etc.). Other possible usages may include analyzing solar potential over building roofs [7], placing and installing telecommunication antenna towers [8], Web-based mapping [9], tourism [6], architecture [10], augmented reality applications [5], [11], and many more.

Spaceborne synthetic aperture radar (SAR) sensors are able to provide day/night global coverage in virtually all weather conditions. Moreover, due to coherent imaging nature, by using acquisitions taken at different times, it is also uniquely capable of imaging the dynamics of the illuminated area in the scale of centimeters or even millimeters. These benefits have motivated many researchers, and therefore, several methods have been developed, which use very high resolution (VHR) spaceborne SAR imagery for detection and reconstruction of man-made structures in particular buildings. For instance, single-aspect SAR-image-based approaches for building reconstruction are proposed in [12]–[14]. Due to the fact that only single SAR images are used, these approaches predominantly perform well mostly only for isolated buildings but not for dense urban areas where the buildings are densely packed and smaller buildings are often occluded by the higher ones [15]. To resolve this, interferometric SAR (InSAR) data, SAR image pairs taken from slightly different viewing angles, are used, e.g., a modified machine vision approach is proposed in [16] to detect and extract buildings. The algorithm is based on local approximation of best fitting planes in the segmented regions of interest. Similarly, Thiele *et al.* [17] also developed a model-based approach to detect and reconstruct building footprints using orthogonal InSAR images. Another automatic approach based on modeling building objects as cuboids using multiaspect polarimetric SAR images is presented in [18]. In data fusion context, the use of optical imagery has also been exploited along with SAR [19] and InSAR [15] datasets, respectively. Despite of the active ongoing research in the area, the problem of building reconstruction still remains challenging due to inherent characteristics with SAR images such as geometrical projection caused by the side-looking geometry [20]. Moreover, complex building structures and high variability of objects appearing in the images make automatic building detection and reconstruction a challenging task. For example, problems posed by occlusion of smaller buildings/objects from the higher ones render difficulties in large area extension. Therefore, prior knowledge is often incorporated with certain regularization (geometric) constraints for realistic and automatic reconstruction. For instance,

Manuscript received December 13, 2013; revised April 7, 2014; accepted May 7, 2014. This work was supported in part by the Helmholtz Association under the framework of the Young Investigators Group “SiPEO.” This work was also part of the project “6.08 4-D City” funded by the IGSSE of Technische Universität München and the German Research Foundation (DFG, Förderkennzeichen BA2033/3-1).

M. Shahzad is with the Lehrstuhl für Methodik der Fernerkundung, Technische Universität München, 80333 Munich, Germany (e-mail: muhammad.shahzad@bv.tum.de).

X. Zhu is with the Lehrstuhl für Methodik der Fernerkundung, Technische Universität München, 80333 Munich, Germany and also with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2014.2327391

facades are often assumed to be vertical [19], building footprint as regular parallelepipeds [14], roofs as polyhedral structures [21], etc.

Modern spaceborne SAR sensors such as TerraSAR-X/TanDEM-X [22] and COSMO-SkyMed [23] can deliver VHR resolution data that fit well to the inherent spatial scales of buildings. Hence, these VHR data are particularly suited for detailed urban mapping [24]–[33]. However, 2-D SAR imaging projects the 3-D scene onto a 2-D image, making it “noninjective” in urban scenarios due to the presence of vertical structures (e.g., building facades or other man-made objects) [25]. The unwelcome effects such as layover and foreshortening seriously handicap the interpretation of SAR images. Advanced interferometric techniques, such as persistent scatterer interferometry (PSI) and SAR tomography (TomoSAR), aim at SAR imaging in 3-D or even higher dimensions. Among them, PSI exploits the coherent pixels, i.e., the bright long-term stable objects (persistent scatterers) [30]. However, it assumes single scatterers in one azimuth-range pixel and therefore does not resolve the layover problem. TomoSAR, on the other hand, aims at real and unambiguous 3-D SAR imaging [25], [34]–[36]. By exploiting multipass SAR images taken from slightly different positions, like PSI does, it builds up a synthetic aperture in the elevation direction that enables retrieval of precise 3-D position of dominant scatterers via spectral analysis within one azimuth-range SAR image pixel [25]. Multiple layovered objects in any pixel are therefore separated from the reconstructed reflectivity profile in elevation direction [36]. Moreover, exploiting the fact that different acquisitions are taken at different times, the synthetic aperture can also be extended in the temporal domain to enable 4-D (space-time) focusing of SAR images. The technique is referred to as D-TomoSAR, which combines the strengths of both TomoSAR and PSI [26], [27], [37]–[40]. It is capable of retrieving elevation and deformation information (linear, seasonal, etc.) even of multiple scatterers inside a single SAR image pixel [25], [29]. Retrieval of rich scatterer information from VHR D-TomoSAR enables generation of 4-D (space-time) point cloud of the illuminated area with point (scatterer) density that is comparable to LiDAR, e.g., experiments using TerraSAR-X high-resolution spotlight data stacks show that the scatterer density retrieved using TomoSAR is on the order of 1 million points/km² [28].

Object reconstruction from these high-quality TomoSAR point clouds can greatly support the reconstruction of dynamic city models that could potentially be used to monitor and visualize the *dynamics* of urban infrastructure in very high level of details. Such models would be immensely helpful to ensure safety/security of growing urban population by monitoring of urban infrastructures against potential threats of damage and structural degradation caused by various factors, e.g., ground subsidence or uplift, bad construction, natural disaster, etc. Motivated by this, the very first results of building facade reconstruction from single-view (ascending stack) and multi-view (fused ascending and descending stacks) perspectives over a small test building area (Bellagio hotel, Las Vegas) were presented in [41] and [42], respectively.

In this paper, we present an approach that allows automatic reconstruction of 3-D building facades using these unstruc-

tured TomoSAR point clouds only. The approach proposes new methods as well as modifications to the previously introduced idea in [42] aimed at finding a more general solution toward automatic reconstruction of the whole city area. The basic idea is to reconstruct 3-D building models via independent modeling of each individual facade to build the overall 2-D shape of the building footprint, followed by its representation in 3-D. The following are the innovative contributions that are specific to the approach proposed in this paper.

- 1) A robust M-estimator-based directional *SD* estimation method is proposed, which provides much better estimates of facade regions compared to the grid-based *SD* estimation proposed in [42] by incorporating the facade geometry. Moreover, instead of rejecting non-facade points via 2-D morphological operations used in [42], robust 3-D surface normal information is utilized. The use of additional-dimensional with the vertical facade assumption helps in better rejecting nonfacade points.
- 2) K-means clustering with a criterion for prior guessing the number of clusters is used in previous works [41], [42]. This technique provides good results for single buildings, but when it comes to larger areas, there are two major concerns: 1) guessing the number of clusters is not always trivial, and 2) a certain shape of clusters is not very well recognized. For this reason, a three-step automatic (unsupervised) clustering approach that combines both the density-based clustering [43] and mean shift algorithm [44], [45] is proposed in this paper. The proposed segmentation approach is able to work directly on bigger areas without requiring any prior knowledge about the number of clusters.
- 3) Facades are modeled using general (first and second order) polynomial equations to cater for a wide variety of building footprint. A detailed methodological description of the modeling procedure is explained, which is able to cater arbitrary (rotated) orientations of building facades. The coefficients of the model are estimated using weighted total least squares (WTLS) method to cope for localization errors of TomoSAR points in both *xy* directions.
- 4) During the reconstruction procedure, the presence of smaller clustered segments occurring at facade transition regions handicaps accurate determination of vertex points from the adjacent facade pair and may cause the reconstruction procedure to fail. To deal with this problem, smaller “conflicting segments” are automatically identified and removed for accurate and robust reconstruction of the adjacent facades.
- 5) Side-looking SAR geometry and complexity of the scene in dense high-rise area of interest can cause occlusions of lower height facades scattered around higher building facade structures. As a consequence, a few or sometimes no data are available for the occluded region, rendering incomplete reconstruction or breaking an individual facade into two or more segments. A partial solution is also presented in this paper, which refines the

reconstructed facade footprints via insertion (of additional segments) and extension operations.

- 6) Finally, this paper presents the first demonstration of automatic large area reconstruction of building facades from this class of data. Moreover, the developed methods are not strictly applicable to TomoSAR point clouds only but are also applicable to work on unstructured 3-D point clouds generated from a different sensor with similar configuration (i.e., oblique geometry) with both low and high point densities.

The aforementioned contributions allow completely *automatic* (but parametric) reconstruction of building facades from TomoSAR point clouds in larger areas.

The remainder of this paper is structured as follows. Section II presents a brief procedural overview of the existing techniques that use 3-D LiDAR point cloud for building reconstruction. Section III presents in detail the proposed approach. In Section IV, the experimental results, obtained from the TomoSAR point cloud generated from a TerraSAR-X high-resolution spotlight data stack (ascending orbit only), are presented and discussed. Finally, in Section V, a conclusion about the proposed approach is drawn, and future perspectives are discussed.

II. RELATED WORK

Most approaches employ airborne LiDAR data for automatic 3-D reconstruction of buildings. Methodologically, the problem is tackled by subdividing the task into two sequential steps, i.e., detection/classification of building points followed by their 3-D modeling and reconstruction.

Detection is generally carried out by first computing the digital terrain model (DTM) by filtering techniques, e.g., morphological filtering [46], gradient analysis [47], or iterative densification of triangular irregular network structures [48]. Nadir-looking LiDAR points essentially give a digital surface model (DSM). Subtracting DSM from the computed DTM provides a normalized DSM (nDSM) which represents the height variation of nonground points. Subsequently, building points are extracted out from nDSM by exploiting geometrical features such as deviations from the surface model, local height measures, roughness, and slope variations. Methods based on building boundary tracing from nDSM [49] or directly from point clouds [50], [51] have also been employed for building detection. With them, finer building boundaries are determined by regularization of the coarsely traced boundaries. All points that lie inside the boundary regions are considered as building points. Building points can also be extracted out by explicitly labeling every point in the data set. For labeling purposes, local neighborhood features such as height, eigenvalue, and plane features have been used in conjunction with supervised [52], semisupervised [21], and unsupervised [53] classification techniques.

Detected building regions or points are, in turn, used for 3-D modeling and reconstruction. Most methods make use of the fact that man-made structures such as buildings usually have either parametric shapes (model driven) or composed of

polyhedral structures only (data driven). The latter is, however, more common in the literature, where local sets of coplanar points are first determined using 3-D Hough transform or RANSAC algorithms and then reconstruction is carried out by surface fitting in the segmented building regions followed by region growing procedure [53] or by building up an adjacency graph [21], [54].

The aforementioned methods and the majority of other techniques present in the literature that make use of 3-D LiDAR data cannot be directly applied to TomoSAR point clouds due to side-looking SAR geometry and different microwave scattering properties of the objects appearing in the scene reflecting different geometrical and material features. Compared to LiDAR, TomoSAR point clouds possess the following peculiarities that should be taken into consideration.

Accuracy and Errors:

- TomoSAR point clouds deliver moderate 3-D positioning accuracy on the order of 1 m [15] as compared to (airborne) LiDAR systems having an accuracy on the order of 0.1 m [21].
- Ghost scatterers [55] may be generated due to multiple scattering that appears as outliers far away from a realistic 3-D position.
- A small number of images and limited orbit speed render the location error of TomoSAR points highly anisotropic, with an elevation error typically one or two orders of magnitude higher than in range and azimuth [25].

Coherent Imaging:

- Due to coherent imaging nature, temporally incoherent objects such as trees cannot be reconstructed from multipass spaceborne SAR image stacks.

Side-Looking Geometry:

- Separation of layover on vertical structures renders geocoded TomoSAR point clouds to possess higher density of points on building facades.
- In order to obtain a full structure of individual buildings from space, multiple views are required [42].

Higher Dimensional Imaging:

- In addition to 3-D spatial information, TomoSAR point clouds also possess the fourth-dimensional information, i.e., temporal or seasonal deformation estimates, making them very attractive for dynamic city modeling.

III. PROCESSING STEPS FOR BUILDING FACADE RECONSTRUCTION

Due to the side-looking SAR geometry, when projected, the TomoSAR point clouds onto ground plane vertical facade regions exhibit higher scatterer (point) density (SD) as compared to nonfacade regions. It is mostly true due to the existence of strong corner reflectors, e.g., window frames on the building facades. Taking this fact into account, in [42], we proposed to extract facade points by projecting all of the points onto the xy grid for estimating SD (rastered image), followed by thresholding and applying morphological dilation operation. This approach works well for high-rise buildings having a much higher point density but limits the extraction of facade points

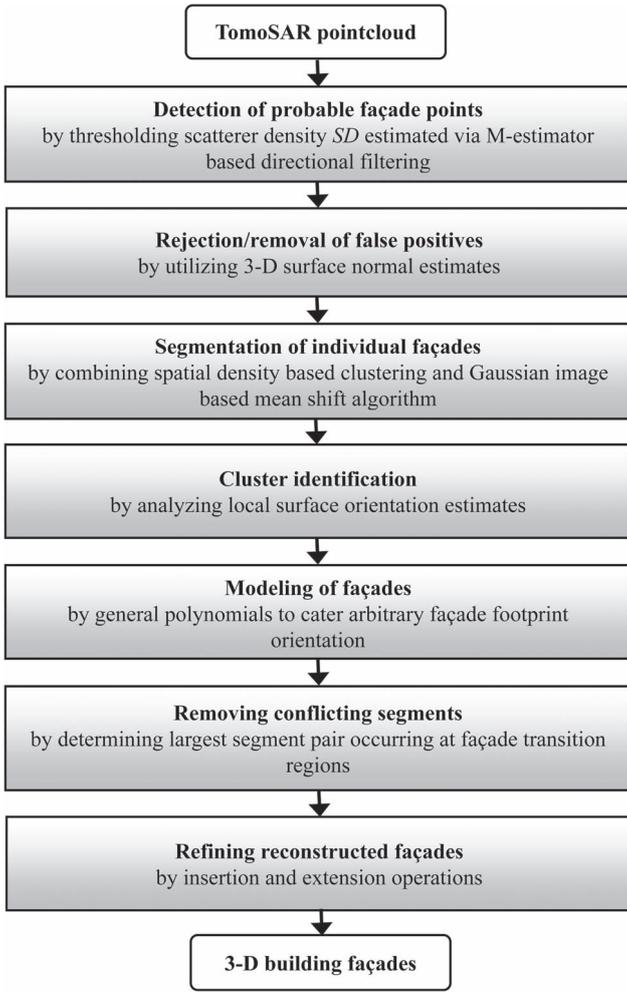


Fig. 1. Block diagram of the proposed method.

from relatively lower buildings. The selection of a particular threshold thus becomes crucial.

To resolve this issue, in this work, a more robust facade extraction approach is proposed, which is based on the directional SD estimation procedure to locally estimate the SD for each point while incorporating the facade geometry [56]. Later, robust 3-D surface normal information is utilized to extract facade points. Automatic segmentation of extracted facade points is obtained by first performing coarse clustering to cluster points belonging to individual buildings. Then, each cluster is further fine-clustered using Gaussian-map-based mean shift clustering algorithm. After that, clusters within clusters are spatially separated. Segmented facades are then classified as flat or curved, and their model parameters are estimated. Subsequently, the geometric primitives such as vertex points are determined from the intersection of adjacent facade pair after removing smaller conflicting segments occurring at transitional regions. Finally, a refinement operation is carried out on the reconstructed facades that remain either incomplete or broken into two or more segments to complete the reconstruction process.

Fig. 1 shows the block diagram of the processing steps involved in the complete methodology. Next, we explain in detail the procedures of the proposed approach in dedicated sections.

A. SD Estimation

For each 3-D TomoSAR point p , points within its local neighborhood v_c are used for SD estimation. v_c includes all of those points that lie inside a vertical cylinder centered at p . To emphasize the building facades, we incorporate facade geometry in estimating SD , i.e., we estimate the direction of the local neighborhood via line fitting using robust M-estimator. The method iteratively reweights the points according to the residuals and computes the so-called M-estimates as follows [57].

- 1) The initial estimates of the line parameters β (e.g., $\beta_1 =$ slope and $\beta_2 =$ offset) are derived from ordinary least squares.
- 2) The weights w_{p_i} of each point $p_i \in v_c$ are then computed using a bisquare function [57], [58]

$$w_{p_i} = \begin{cases} (1 - u^2)^2 & \text{for } \text{abs}(u) < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where } u = \frac{|y_{p_i} - x_{p_i}\beta_1 - \beta_2|}{4.685\hat{\sigma}\sqrt{1-t}} \quad (1)$$

where t is the leverage computed using parameter estimates β of the fitted line and $\hat{\sigma}$ is the estimate for the scale of the error term computed by $\hat{\sigma} = 1.483 * \text{MAD}$, where MAD is the median absolute deviation of the residuals from their median. The term 1.483 is used to make the estimate $\hat{\sigma}$ consistent to the standard deviation at Gaussian distribution [58], [59].

- 3) Updating β using weights w_{p_i} by applying weighted least squares to solve the following objective function:

$$\arg \min_{\beta} \sum_{p_i \in v_c} w_{p_i}(\beta) |y_{p_i} - x_{p_i}\beta_1 - \beta_2|^2 \quad (2)$$

where x_{p_i} and y_{p_i} represent the abscissa and ordinate (i.e., ground coordinates) of the points within v_c .

- 4) Iterating steps 2 and 3 until a fixed number of iterations.

The estimated line describes the main principal axis of the cylindrical footprint of the local neighborhood. The orthogonal distance for every point in v_c is then calculated from the principal axis (shifted to the point p), and the points having distances less than d are taken as “inliers” and used in SD estimation.

SD for each point is thus defined as the number of points within a directional (cylindrical) neighborhood window divided by the area of the window

$$SD = \frac{\text{number of points in } v_d}{\text{Area of } v_d} \quad (3)$$

where $v_d \subseteq v_c$ but includes only those points that lie close to the principal axis of points in v_c .

Fig. 2 shows the graphical representation of the SD estimation procedure.

B. Facade Extraction

Based on the estimated SD , facade points can be extracted. For a large area, both high and low buildings are present. A *hard* threshold, i.e., removing points below a rather high SD value,

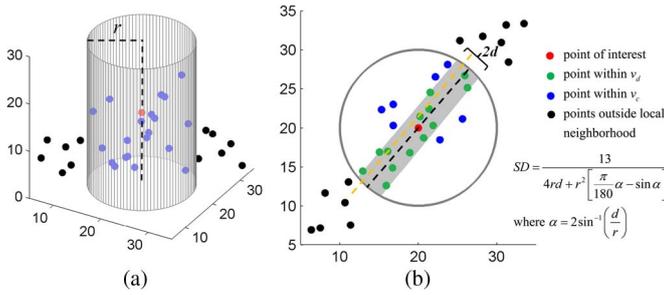


Fig. 2. Illustration of SD estimation. (a) Three-dimensional view of the local cylindrical neighborhood around the point of interest. (b) Top view of (a). The coefficients of the dotted yellow regression line are estimated via M-estimation. The black dotted line shows the shift of the yellow line to the point of interest. The shaded area shows the region of v_d within v_c .

as used in [42], would lead to misdetection of facade points of lower buildings whose SD estimates would be relatively low with respect to high-rise buildings. To avoid such misdetection, we extract facade points in a sequential way. First, we have lowered the SD threshold to detect not only higher buildings but also lower ones. However, a softer threshold would also introduce false positives, i.e., roof points or ground points with a local point density comparable to those of lower buildings. Therefore, we introduce a second step which utilizes 3-D surface normal information by incorporating prior knowledge (i.e., facades are assumed to be vertical surfaces which separate them from nonvertical ground plane and roofs) to eliminate those false positives.

The key issue is then the local surface normal estimation for each 3-D point. Commonly, they are estimated via fitting “best” plane in least squares (LS) sense in the local neighborhood v_c , which is equivalent to performing principal component analysis (PCA) of the points in v_c [60]. This implies that the surface normals can be directly estimated for each 3-D point via eigenvalue/eigenvector analysis of 3-D (i.e., 3×3) covariance matrix Σ_{v_c} . There are two advantages in using eigenvalue/eigenvector analysis of Σ_{v_c} for surface normal estimation.

- 1) First, the eigenvector associated to the smallest eigenvalue of the positive semidefinite matrix Σ_{v_c} is the direct estimate of the local surface normal of the query point [60].
- 2) Second, in addition to giving us the direct solution of estimating local surface normal, it can also help us in determining the dimensionality of each 3-D point [21]. To elaborate, the eigenvectors of Σ_{v_c} essentially give the orthonormal basis for the local neighborhood v_c , with their corresponding eigenvalues representing the magnitude (or variance) of expansion. Analyzing this magnitude implicitly gives us an indication of the beneath surface, e.g., eigenvalue analysis of Σ_{v_c} with all points lying on the plane would ideally return only two nonzero eigenvalues. Similarly, all points lying on a 3-D line would give only one nonzero eigenvalue. Eigenvalue analysis for segmentation and classification of planar points has been exploited in [21], [52], and [53].

Eigenvalue/eigenvector analysis via classical PCA may fail to give precise estimate of the 3-D surface normal using

TomoSAR point cloud due to the presence of outliers and localization errors (see Section II: accuracy and errors). Robust estimation of the covariance matrix Σ_{v_c} is therefore needed. To this end, we estimated Σ_{v_c} using robust minimum covariance determinant (MCD) method [61]. The method finds a subset (fraction) α of the data points $p_i \in v_c$ whose covariance matrix has the lowest determinant. The idea stems from the concept of generalized variance (GV) which is defined to be the determinant of the covariance matrix of any d -D ($d > 1$) random variable [62]. For instance, in case of 2-D (x - y) points, the GV provides a scalar value which measures the overall variability of all points in both x and y dimensions. Points that are clustered tightly together tend to have a smaller GV (i.e., lower determinant of their covariance matrix) as compared to scattered ones. Thus, the subset α of the data points which provides the lowest determinant is taken as the MCD estimate of Σ_{v_c} . If the data points are assumed to have less than 25% outliers, then an appropriate selection of $\alpha = 0.75$ (also used in this work) proposed in [63] provides a good compromise between statistical efficiency and high breakdown value ($\alpha = 0.75$ implies that 75% of the data points has been used in covariance estimation).

The covariance matrix Σ_{v_c} estimated using the MCD method from the local neighboring points $p_i \in v_c$ around (in cylinder) the point of interest $p_o(x_o, y_o, z_o)$ is then used to determine the local 3-D surface normal at p_o . If we denote a plane which robustly fits the neighboring points p_i as $n_x x + n_y y + n_z z + \rho = 0$, with $\rho = -n_x x_o - n_y y_o - n_z z_o$, then $N_o(n_x, n_y, n_z)$ depicts the local 3-D surface normal at p_o . N_o is thus directly estimated from Σ_{v_c} by computing the eigenvector associated to the smallest eigenvalue of Σ_{v_c} (here, v_c includes points in the vicinity of p_o) i.e.,

$$\text{if } \Sigma_{v_c} \cdot v_j = \lambda_j \cdot v_j, \quad j = 1, 2, 3 \text{ (descending order)}$$

then surface normal of the underlying surface at point

$$p_o : N_o(n_x, n_y, n_z) = v_3. \quad (4)$$

From (4), robust 3-D surface normals are computed for each point that is obtained after SD thresholding. Ideally, the direction of surface normal should be parallel to the ground for points on the vertical facades which separate them from non-vertical ground plane and roofs. Taking this fact into account, facade points are extracted out by retaining only those points having normals that are close to the horizontal axis (i.e., parallel to ground for points belonging to a vertical surface). In this manner, the proposed two-step approach allows us to robustly extract facade points over a large area where both high and low buildings are present.

Fig. 3 shows a comparison of the proposed approach with the one presented in [42]. The selected area shown in Fig. 3(a) contains relatively lower height buildings with low and inconsistent density of points on building facades. It can be seen that, in comparison to the SD estimation results from the previous approach depicted in Fig. 3(b), higher and complete density values are obtained for facade regions using the SD estimation method proposed in this paper, shown in Fig. 3(c).

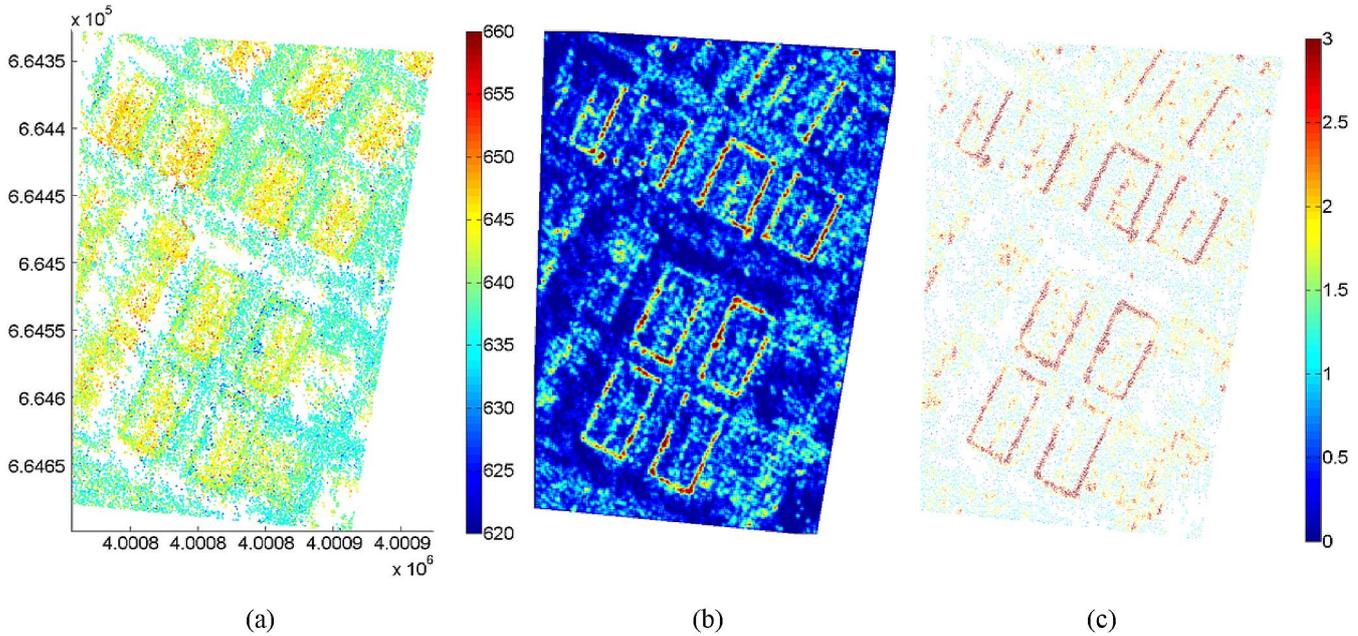


Fig. 3. Comparison of two SD estimation methods using TomoSAR points of an area in Las Vegas containing rectangular shaped buildings. (a) TomoSAR points (top view) generated from ascending stack only. The axes are in UTM coordinates. The height (above sea level) of TomoSAR points is color-coded [unit: meter]. (b) SD image estimated via [42]. (c) SD estimated via M-estimator-based directional filter proposed in this paper. Higher SD regions depict probable facade points. SD is color-coded, with the colorbar representing points/m². (b) and (c) share the same colorbar. Note that the SD estimated in (b) is the rastered image obtained by projecting all points onto the xy grid as compared to (c) where SD is directly computed for each point.

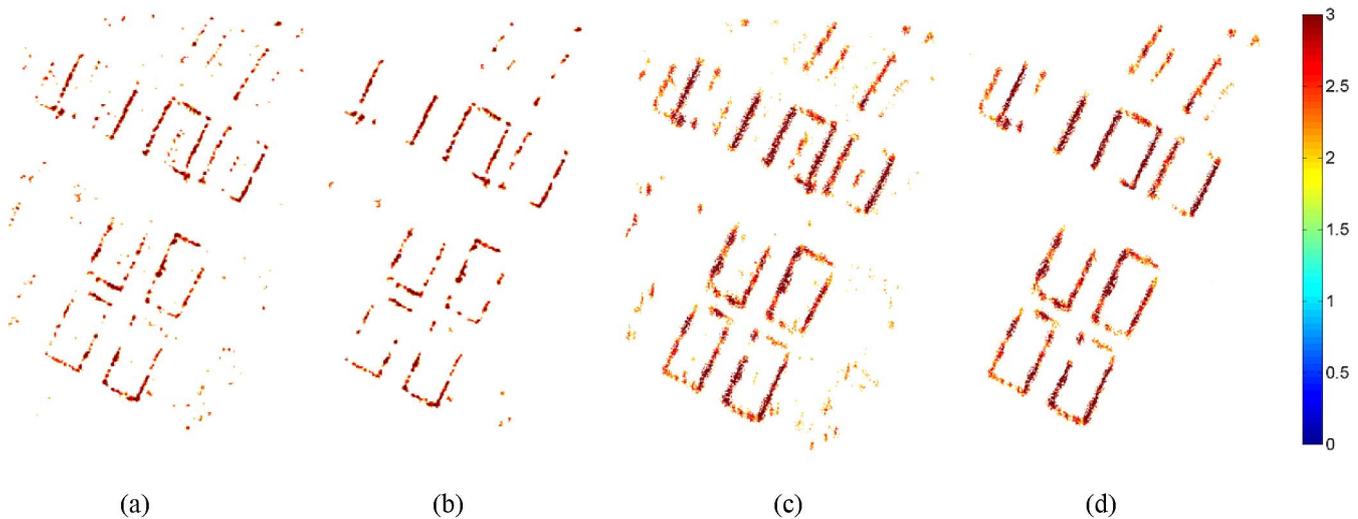


Fig. 4. Extraction of facade regions/points using the SD estimation results from Fig. 3. (a) Building facades obtained by thresholding the rastered SD image of Fig. 3(b) by the TH value. (b) Final extracted facade after 2-D morphological operation on (a) as proposed in [42]. (c) TomoSAR points whose SD estimated in Fig. 3(c) is above TH . (d) Extracted facades from (c) by utilizing robust 3-D surface normal information. The threshold value TH used here is 2 points/m² (empirically found to be optimum in [42]).

Moreover, later use of the third dimension in robust 3-D surface normal estimation provides much better results of extracting facades by rejecting nonfacade points.

Fig. 4 presents the comparison of the facade extraction results obtained using the SD estimates in Fig. 3. Taking the facade point candidates extracted by thresholding SD [Fig. 3(a) and (c)] as inputs, 2-D morphological operations (area opening and dilation) as proposed in [42] and robust 3-D surface normal information as suggested in this paper are performed to reject false positives. The final extracted facades are shown in Fig. 4(b) and (d), respectively.

C. Segmentation of Individual Facades

The extracted facade points belong to different facades. Clustering of points belonging to the same facade is therefore needed. First, a coarse clustering is performed using density-based clustering algorithm [43]. It involves the notion of density connectivity between the points. For example, two points are directly density connected to each other if one is in the neighborhood vicinity of the other point. If the two points are not directly connected to each other, still they can be density connected to each other if there is a chain of points between them such that they all are directly density connected

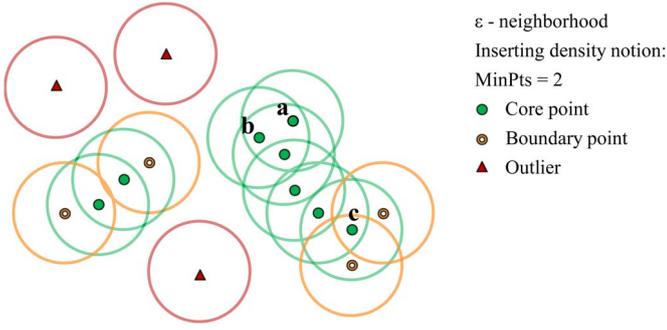


Fig. 5. Density-based clustering. Points a and b are directly density connected to each other, whereas points a and c are density connected to each other since there is a chain of points between them such that they all are directly density connected to each other.

to each other. Two parameters that control the clustering process include the neighborhood parameter ε , i.e., the radius in case of sphere or cylindrical neighborhood, and the minimum number of points MinPts in the ε -neighborhood for any particular point. The resulting clusters K_i thus contain points such that all of the points in any particular cluster are density connected to each other but are not density connected to any other point belonging to another cluster. Moreover, each point inside any particular cluster K_i belongs to one of the three categories (Fig. 5) [43].

- 1) Core points: A point is labeled core point if it contains, within its ε -neighborhood, MinPts number of points.
- 2) Border points: A point is considered border point if it is within ε -neighborhood of any core point but itself is not a core point and does not have MinPts neighbors.
- 3) Outliers: A point neither core point nor border point is termed as an outlier, i.e., any point which do not have density (points) greater than MinPts within its ε -neighborhood and also is not the neighbor of any other point.

Density-connected clusters containing only core and boundary points are used for further processing.

The aforementioned process, however, may merge points of two or more adjacent facade segments into a single cluster. To reconstruct individual facade segments, separation of these segments is therefore necessary. It is done by mapping the facade points in Gaussian image and then employing mean shift clustering.

Let us assume that a coarsely clustered segment K_i consist of one or more vertical adjacent facades F_j , $j = 1, \dots, J$. An image of a map $M: F \rightarrow F^2$ that assigns each point in F to its respective unit surface normal is known as Gaussian image GI of F [64]. Flat F (i.e., planar surface) should ideally be represented by a point in GI . Fig. 6 conceptually illustrates this in an ideal scenario. In practice, surface normals are estimated locally and may fluctuate from one point to another as practical data often contain errors in 3-D positions. However, if the estimation of normals is robust enough, a surface mapped to GI should be represented as a dense cluster of points in GI . The shape of clusters in GI corresponds to the geometry of connected surfaces [44]. The number of clusters in GI tells the number of surfaces in the spatial domain.

If we assume $p_r = 1, \dots, m$ to be 3-D points and n_r as their corresponding 3-D unit normal vectors belonging to one of the

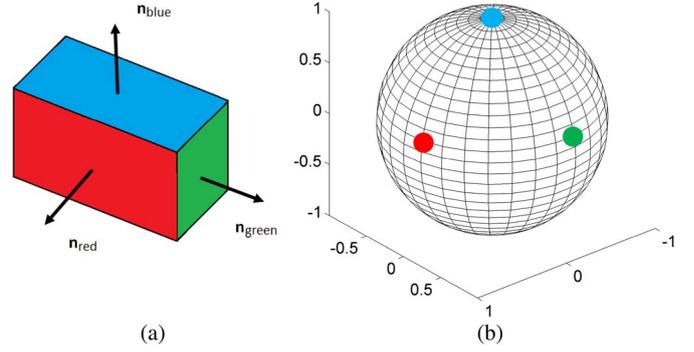


Fig. 6. Gaussian image of three connected planar surfaces. (a) Arrows indicate surface normal vectors ($n_{\text{red}}, n_{\text{green}}, n_{\text{blue}}$) to the respective surfaces. (b) All points belonging to one particular surface are mapped to the same identical point in GI (ideal scenario).

coarsely clustered segments K_i , then the density at any normal point n_q ($q \in r$) in GI (feature space) is defined as [44]

$$D_{n_q} = \frac{c}{mb^3} \sum_{r=1}^m g \left(\left\| \frac{n_q - n_r}{b} \right\|^2 \right) \quad (5)$$

where b is the bandwidth parameter and $g(x)$ is a nonnegative, nonincreasing, and piecewise continuous function with definite integral, i.e., $\int_0^\infty g(x)dx < \infty$. From the concept of kernels [45], the function $g(x)$ is defined as the profile of the radially symmetric kernel $G(x)$ satisfying $G(x) = cg(\|x\|^2)$, where c is a normalization constant ensuring that $G(x)$ integrates to 1. Different kernels, such as the unit flat kernel and the Gaussian kernel, can be used to define the density D_{n_q} . However, the latter with the profile function $\exp(-\|(n_q - n_r)/b\|^2)$ has been used in this work.

Density D_{n_q} is higher for points that belong to planar or parabolic surfaces and lower for points that lie at the transition edges between the surfaces [44]. These higher density points in the GI are identified and clustered using mean shift (MS) clustering procedure. MS is a mode-seeking procedure and works iteratively by shifting every data point toward the weighted mean of points within its neighborhood. The shift vector $m(n_q)$ always points toward the direction of the maximum increase in the density D_{n_q} [65] and is computed as

$$m(n_q) = \frac{\sum_{r=1}^m n_r \exp \left(-\frac{\|n_q - n_r\|^2}{b^2} \right)}{\sum_{r=1}^m \exp \left(-\frac{\|n_q - n_r\|^2}{b^2} \right)} - n_q. \quad (6)$$

Applying MS in GI produces clusters whose corresponding points in spatial domain represent different facades. However, it is also possible that spatial points corresponding to any one particular normal cluster in GI may belong to two or more different facades. This can happen if points of two or more facades that are “nearly” parallel to each other (i.e., having close normal directions) are present in K_i . To resolve this, density-based clustering is again performed in the resulting clusters for spatial separation of parallel facade points clustered into one group. Finally, clusters with very few points are removed from further processing for robust reconstruction.

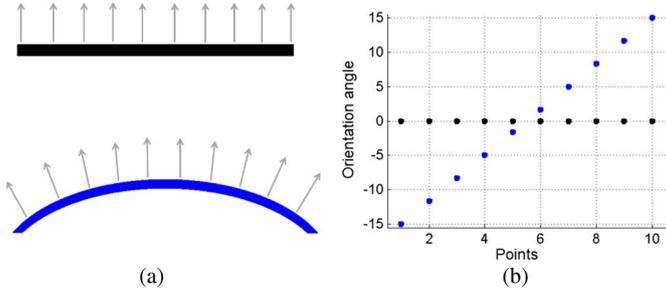


Fig. 7. Illustration of orientation angle for flat and curved vertical footprints (top view). (a) Arrows indicate pattern of change in orientation (azimuthal) angles of ten points on each vertical surface. (b) Plots their respective orientation angles.

D. Cluster Identification

Each cluster is further classified into flat or curved surface by analyzing derivatives of the local orientation angle θ . θ for each 3-D point is equal to the azimuthal angle of the corresponding computed surface normal

$$\theta = \arctan\left(\frac{\lambda_{3y}}{\lambda_{3x}}\right) \quad (7)$$

where λ_{3x} and λ_{3y} represent the x and y components of the surface normal λ_3 of any 3-D point. Ideally, the flat surfaces should have constant orientations, i.e., zero derivatives compared to the curved surfaces that have gradually changing orientations (see Fig. 7). We exploit this fact and compute the first derivative θ' of the orientation angle for each facade footprint. Since the original orientation derivatives θ' are usually noisy, all of the points are first projected to the first principal axis, and polynomial fitting is later applied for denoising. Based on the behavior of θ' , facade footprints are classified as flat or curved.

E. Modeling of Facades

The identified facade clusters in xy plane are then modeled using the following general polynomial equation [42]:

$$f_p(x, y) = \sum_{q=1}^p a_q x^i y^j \quad i + j \leq q \quad (8)$$

where i and j are permuted accordingly, p is the order of the polynomial, and the number of terms in the aforementioned polynomial is equal to $(p+1)(p+2)/2$. Cross terms are introduced in the model in case of the rotated local coordinate system. To solve (8), we restrict ourselves to the first and second orders (i.e., *flat* with $\max(i, j) = 1$ and *curved* with $\max(i, j) = 2$). The coefficients a_q are estimated using the WTLS method, where the total least squares is utilized to cope for localization errors of TomoSAR points in both xy directions and the weight of each point is assigned equal to its corresponding SD . The weighted polynomial fitting (residual) error f_{err} is minimum for the case where we have the unrotated local coordinate system reducing the right-hand side of (8) to $\sum_{i=0}^p a_i x^i$ (i.e., with no cross terms). In case of the rotated local coordinate system (which is often the case), we perform

the following steps to obtain consistent parameter estimates of all facades in a global coordinate system.

- 1) Rotate the points by rotation angle ω , and compute the polynomial fitting error f_{err} by applying the WTLS method.
- 2) Consider coefficients computed with ω_{min} that gives the minimum polynomial fitting error f_{err} as polynomial terms depicting *unrotated* points in the global coordinate system. ω_{min} is computed by using an unconstrained nonlinear optimization procedure to find the minimum of the error function f_{err} by varying ω over $0-360^\circ$ range via the Nelder–Mead simplex algorithm [66].
- 3) Rotate the computed polynomial by replacing the unrotated $(x-, y-)$ axis terms by their rotation counterparts $(x \cos \omega + y \sin \omega, -x \sin \omega + y \cos \omega)$ to yield polynomial terms a_q in global coordinates.

F. Removing Conflicting Segments

After estimation of model parameters, the next step is to describe the overall shape of the building footprint by further identifying adjacent facade pairs and determining the intersection of the facade surfaces. The adjacency of facades is usually described by an adjacency matrix AM that is built up via connectivity analysis [21], [42]. Identified adjacent facade segments are used to determine the vertex points (i.e., facade intersection lines in 3-D) by computing the intersection points between any adjacent facade pair.

Determination of these intersection points can sometimes become difficult if the transition points are segmented as isolated small clusters (also referred to as conflicting segments) rather than part of the corresponding adjacent facade segments. As a consequence, it gets complicated to find a legitimate adjacent facade pair from which intersection points should be computed. To resolve this issue, conflicting segments must be removed prior to vertex point computation. To illustrate how they are removed in an automatic manner, an example is shown in Fig. 8. The labeled line segments indicate the reconstructed facade segments of two different buildings A and B. The endpoints of each segment are denoted as “ Δ ” and “ \bullet .” AM represents the built adjacency matrix, where “1” and “0” denote the adjacent and nonadjacent conditions, respectively. Among the labeled segments, segments [7 4 1 3 6] are “valid” facades, while segments [2 5 8] are the conflicting segments.

The following steps are performed for automatic removal of these conflicting segments.

- 1) The connected series matrix ConnSeg is determined from AM such that rows of ConnSeg represent a set of distinct series of adjacently connected segments, e.g., the i th row of ConnSeg $Seg_i = \{s_j \mid j = 1, \dots, n\}$ represents n segments (i.e., $s_1 \sim s_n$) that are adjacently connected. In Fig. 8, since there exist only two series of adjacently connected facade segments (i.e., belonging to two buildings), ConnSeg therefore consists of two rows only, where the first row contains facade segments $Seg_1 = [2 \ 4 \ 5 \ 7]$ while the second row comprises of segments $Seg_2 = [1 \ 3 \ 6 \ 8]$.
- 2) For each segment, the largest segment that is connected to each endpoint can be identified. Their indices are

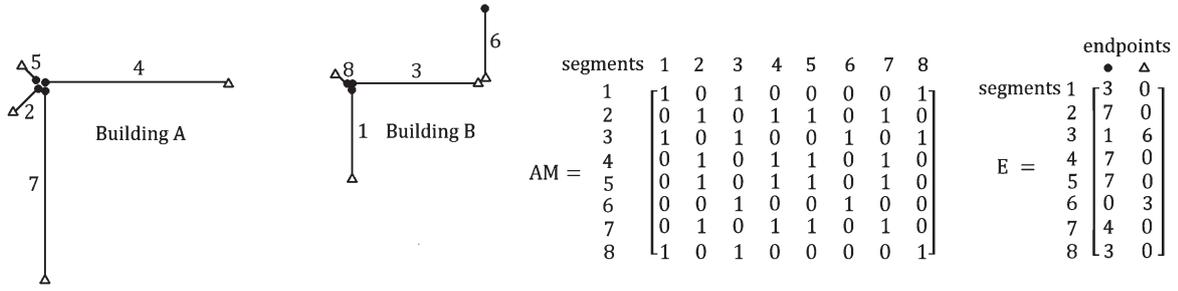


Fig. 8. Example illustrating the removal of conflicting segments.

TABLE I
PROCEDURE TO REMOVE CONFLICTING SEGMENTS

Given: Endpoints matrix EndPts containing endpoints of the reconstructed façade segments &

Adjacency matrix $AM = \{a_{vw} | v, w = 1, \dots, k\}$ such that $a_{v=w} = 1$ & $a_{v \neq w} = 1$ only for conditions where façade

Segments v and w are adjacent to each other, otherwise $a_{v=w} = 0$. k is the total number of façades reconstructed.

1. Initialize: RemoveSeg := \emptyset
2. **for** $i = 1$ to number of rows in ConnSeg
3. Select the adjacently connected series of segments $Seg_i = \text{ConnSeg}(i, :)$
4. **if** number of elements in $Seg_i > 2$
5. **for** $j = 1$ to n (number of segments in Seg_i)
6. Extract adjacent segments of s_j in Seg_i from AM.
7. Extract endpoints of all the segments in Seg_i from EndPts.
8. Compute the Euclidean distance of both endpoints in s_j from end points of all other segments s_k ($j \neq k$) in Seg_i .
9. Find out the largest adjacent segment l_1 connected to one endpoint. Similarly find the largest adjacent segment l_2 connected to the other endpoint of s_j .
10. Store l_1 and l_2 in a two column matrix $E(j, 1:2)$. In case of no adjacent segment to any of the endpoints, insert 0 in the corresponding position in E.
11. **end for**
12. Separately apply union operation to both column of E and store the unique result after excluding 0s in the row matrix RetainSeg.
13. Compare Seg_i and RetainSeg and insert segments of Seg_i that are not members of RetainSeg in the matrix RemoveSeg($i, :$) for removal.
14. **end if**
15. **end for**

recorded in a two column matrix E that captures such an “endpoint”–“largest segment” relationship, e.g., the “•” endpoint of segment 1 shown in Fig. 8 is connected to two segments 8 and 3. Since segment 3 has a larger length than segment 8, therefore 3 is assigned to this endpoint of segment 1 in E . The endpoint matrix E for both buildings is depicted on the right side of Fig. 8. Zeros in E represent the condition of no adjacent facade at that endpoint.

- 3) Applying union operation to all elements in E results in a matrix RetainSeg whose elements contain all building facades that should be retained. Conflicting facades, i.e., the ones that are not part of RetainSeg, are removed. For the example shown in Fig. 8, the union of elements in E gives the RetainSeg [1 3 4 6 7] (zeros are not considered). Subsequently, the segments that are not part of RetainSeg, namely, [2 5 8], are removed.

The pseudocode for the aforementioned procedure is given in Table I.

After removing the conflicting segments, the vertex points are computed from the intersection of valid adjacent segments to complete the reconstruction process.

G. Refining Reconstructed Facades

Sometimes the reconstructed facades remain either incomplete or are broken into more than one segment due to the following reasons: 1) Higher building structures present nearby can partly (or fully) occlude the facades of lower buildings, and 2) due to the geometrical shape, only very few points are available at some parts of building facades. In order to overcome this issue, in this section, we propose a procedure that tries to refine the reconstructed facades by inserting additional

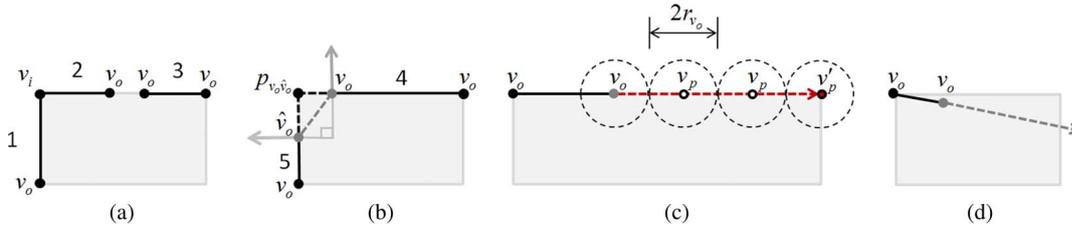


Fig. 9. Vertices for refinement. Gray rectangles depict the 2-D building footprint from the top. (a) Shows the total of five vertices, out of which four are open endpoint vertices and one is an intersection vertex computed from the intersection of segments 1 and 2. (b) Depicts the situation where ignoring condition $C3$ would yield false segment addition shown as a gray dotted line. The gray arrows indicate the local orientation angle θ at open vertices v_o and \hat{v}_o . Two black dotted lines represent the two correct inserted segments between v_o and \hat{v}_o . (c) Depicts the recursive extension procedure of the open vertex v_o . v_p represents the intermediate extension points, whereas v_p' denotes the final point. (d) Illustrates the direction of extension of the gray open vertex over the roof region. This can happen if we only consider h_{\max} and ignore the local standard deviation of height h_σ .

segments between the broken regions and extend those facades that remain incomplete.

Vertex points computed from the previous section are separated into two types: The first type consists of vertices that are computed from the intersection of two adjacent facades, while the second type consists of the other vertices representing “open” endpoints. For each series Seg_i made up of n segments (i.e., the i th updated row of the ConnSeg matrix after removing conflicting segments), there exist two open vertices and $(n - 1)$ intersection vertices. Refinement operations including insertion of additional segment to connect broken facades and extension of incomplete facades are carried out only on the second type, i.e., open endpoint vertices.

If we denote an open end vertex as v_o and an intersection vertex as v_i [see Fig. 9(a)], then the refinement procedure for any one particular adjacently connected series Seg_i having two open vertices is described in Table II.

In Table II, steps 2–4 tries to cope with the broken facades, while steps 5–7 deals with the incomplete facades. Conditions $C1$ and $C2$ in step 4 imply that the two segments are considered part of the same (broken) building facade if both segments are not far enough from each other and at the same time possess data points in between that have close maximum height values. h_{\max} is taken as the mean of at least ten maximum height values (i.e., if there are less than ten points available, then h_{\max} is taken as the mean of all of those point). If conditions $C1$ and $C2$ in step 4 are met, the algorithm then checks condition $C3$. If the two segments belong to the same facade, a segment with vertices (v_o, \hat{v}_o) is inserted, which fills the empty (i.e., broken) regions of the facade. On the contrary, if the open vertex pair $v_o \hat{v}_o$ is not part of the same facade but rather belongs to two different facade segments (determined via the difference in the local orientation angle $> 45^\circ$), then point of intersection $p_{v_o \hat{v}_o}$ is computed, and instead of inserting one segment, two segments with vertices $(v_o, p_{v_o \hat{v}_o})$ and $(\hat{v}_o, p_{v_o \hat{v}_o})$ are inserted. Fig. 9(b) graphically depicts such a situation where the gray open vertices of segments 4 and 5 are (assumed to be) within 2ε distance but have a difference in the local orientation angle of 90° . The gray dotted line shows the addition of new segment without checking condition $C3$. When $C3$ is taken into account, two segments shown in black dotted line are inserted.

In contrast, if any of the conditions $C1$ or $C2$ fail, then the algorithm tries to extend the open vertex point v_o by imposing constraints $C4$ and $C5$ present in step 7. Similar to $C2$, the condition $C4$ ensures that the extended point has the closer

maximum height value. The condition $C5$ ensures that the local 3-D points have certain standard deviation. It is necessary to make sure that the extension is not carried out in the direction that deviates from the facade footprint, i.e., it avoids the extension if the local 3-D points around v_p belong to other nonfacade objects, e.g., roofs, etc. The problem is illustrated in Fig. 9(d), where the gray open vertex can potentially extend over the roof region if the condition $C5$ in step 7 is ignored. Thus, imposing this constraint helps in limiting this false extension.

Finally, the computed vertex points (i.e., the intersection vertices and the open vertices before and after refinement) along with their estimated model parameters are used to reconstruct the 3-D model of the building facades.

IV. EXPERIMENTAL RESULTS AND VALIDATION

A. Data Set

To validate our approach, we tested the algorithm on TomoSAR point clouds generated from a stack of 25 TerraSAR-X high spotlight images from ascending orbit only using the Tomo-GENESIS software developed at the German Aerospace Center [67]. The test area covers approximately 2 km^2 in the high-rise part of the city of Las Vegas. The number of TomoSAR points in the area of interest is about 1.2 million. Fig. 10(a) shows the optical image of our test area, while Fig. 10(b) shows the corresponding TomoSAR point cloud in Universal Transverse Mercator (UTM) coordinates.

B. Results—Extraction of Facade Points

The result of applying the SD estimation procedure is illustrated in Fig. 11(a). The two parameters r (radius of the neighborhood cylinder) and d are empirically set to 5 and 0.9 m, respectively, according to the point density of the data set. One can observe that the TH value influences the number of extracted facade points. A lower TH value results in higher completeness but lower correctness. In [42], we showed the results of estimating SD with varying area sizes and found that a kernel window of size $3 \times 3 \text{ m}^2$ and threshold TH value of about 2 points/ m^2 results in the best tradeoff in terms of completeness and correctness with this class of data. Here, 2 points/ m^2 works well for high-rise buildings but might ignore relatively smaller facades. Therefore, to extract lower facades (and also to automate the procedure), we set the TH

TABLE II
REFINEMENT PROCEDURE

Given: Any one particular adjacently connected series Seg_i

1. Select one of the two open vertex point v_o .
2. Find the nearest open vertex point \hat{v}_o from v_o that belongs to another adjacently connected series $Seg_j (j \neq i)$ and compute their midpoint mid_{v_o} .
3. Locally compute maximum height value h_{max} and height standard deviation h_σ of 3-D points within neighborhood vicinity of v_o , \hat{v}_o and mid_{v_o} . Additionally, compute local orientation θ at v_o and \hat{v}_o .
4. Check following three conditions (C1, C2, C3):

$$C1: \|\hat{v}_o - v_o\|_2 < 2\varepsilon$$

$$C2: \text{abs}(h_{\text{max at } v_o} - h_{\text{max at } mid_{v_o}}) < T_h \ \& \ \text{abs}(h_{\text{max at } \hat{v}_o} - h_{\text{max at } mid_{v_o}}) < T_h$$

$$C3: \text{abs}(\theta_{v_o} - \theta_{\hat{v}_o}) > 45^\circ$$

If conditions C1 & C2 are true, then check condition C3.

If C3 is also true

Insert two additional segments with vertices $(v_o, P_{v_o\hat{v}_o})$ and $(\hat{v}_o, P_{v_o\hat{v}_o})$ and then go to

step 1 ($P_{v_o\hat{v}_o}$ is the point of intersection between two segments).

else

Insert an additional segment with vertices v_o and \hat{v}_o and go to step 1

end if

else

proceed to the next step 5.

end if

5. Determine point v_p at a distance $2r_o$ away from v_o using the estimated model parameters (r_o is the radius used for defining local neighborhood around v_o)
6. Similarly as step 3, locally compute h_{max} and h_σ of 3-D points within neighborhood vicinity of v_p .
7. Check following two conditions (C4, C5)

$$C4: \text{abs}(h_{\text{max at } v_p} - h_{\text{max at } v_o}) < T_h$$

$$C5: \text{abs}(h_{\sigma \text{ at } v_p} - h_{\sigma \text{ at } v_o}) < T_\sigma$$

if both C4 and C5 are true, then recursively extend the vertex v_o towards v_p (see Figure 9(c)) and finally add a new segment with vertices (v_o, v_p')

8. Continue steps 1 to 7 for the second open vertex point of the series Seg_i

to the maximum of SD histogram value. This, as described in Section III, includes not only the facade points but additionally also some nonfacade points with relatively high SD , e.g., roof points. To reject these points from the set of extracted points after SD thresholding, surface normal information is utilized. Fig. 11(b) shows the extracted facade points by retaining only those points having normals between $\pm 15^\circ$ from the horizontal axis (or equivalently $\pm 90^\circ$ from the vertical axis).

C. Results—Automatic Clustering of Extracted Facade Points

Once the facade points are extracted out, the next step is to cluster them into segments, where each segment corresponds to an individual facade. For this, we apply the clustering procedure using the cylindrical neighborhood definition and cluster all of

the points with parameter settings: $\varepsilon = r = 5$ m and $\text{MinPts} = 2$. Here, an important point to notice is that two buildings are considered distinct only in a case when points belonging to facades of two different buildings are separated by ε . Setting ε too small can cause points belonging to a single cluster (i.e., corresponding to an individual facade) to break into more than one cluster. On the other hand, larger values of ε tend to merge points of the nearby facades into one cluster. The value of ε is therefore empirically chosen according to the length and distance among the buildings in the area of interest and implicitly indicates the assumption that two individual facades that belong to different buildings are farther apart than the 5-m radius.

Setting parameter MinPts equal to 2 implies that points are connected to one cluster even if there is a single neighboring point among them. This parameter helps in removing outliers

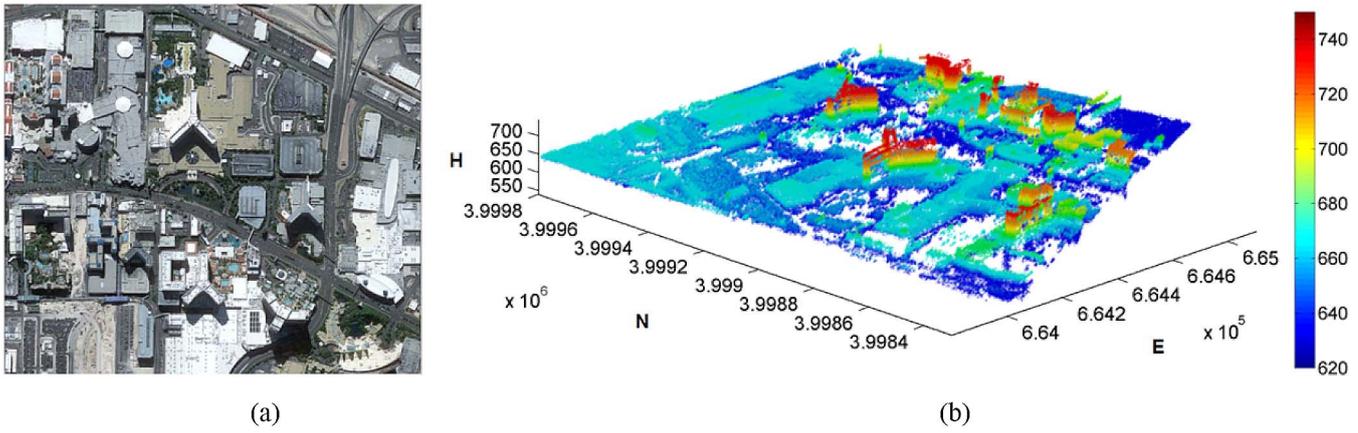


Fig. 10. Dataset. (a) Optical image of the test area in Las Vegas. Copyright Google. (b) TomoSAR points in UTM coordinates of the corresponding test image. The height of TomoSAR points is color-coded [unit: meter].

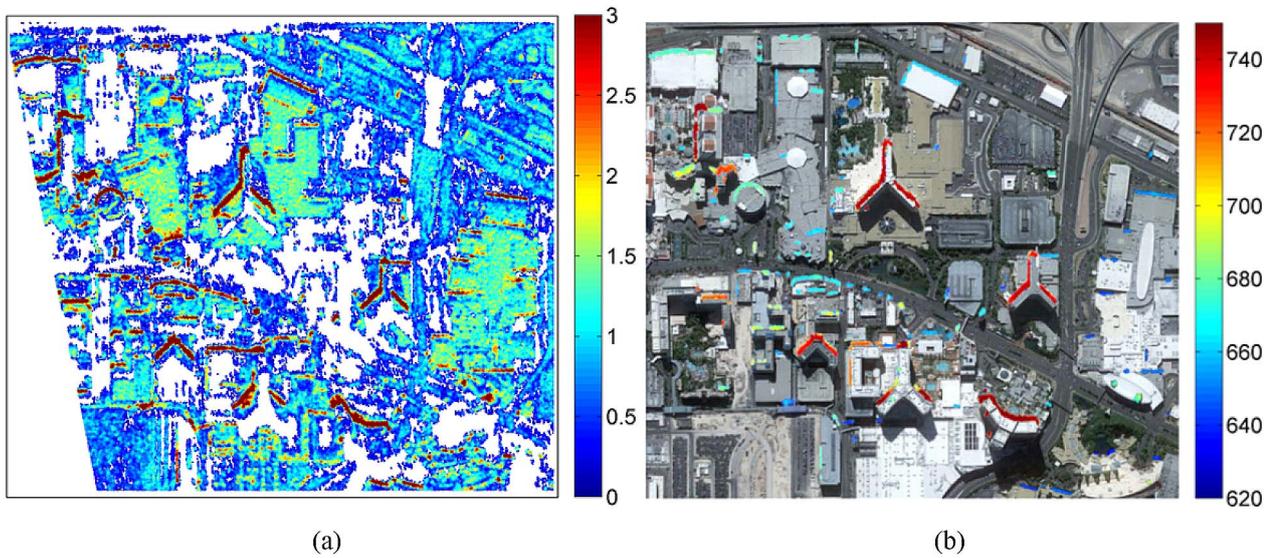


Fig. 11. Facade point extraction. (a) Scatterer (point) density with radius $r = 5$ m and inliers $d = 0.9$ m. (b) Extracted building facade points. Colobar indicates SD and height in meters in (a) and (b), respectively.

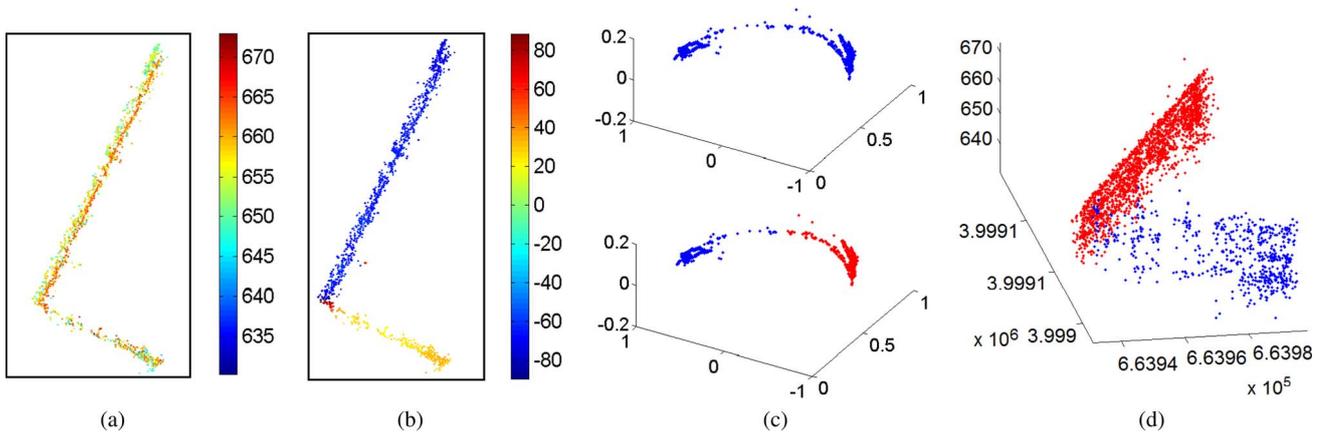


Fig. 12. Fine clustering results after applying mean shift clustering using Gaussian kernel with bandwidth $b = 0.4$ to the coarsely clustered segments in their normal feature space (in GI domain). (a) TomoSAR points of one particular density-connected cluster (top view). The colorbar indicates height in meters. (b) Corresponding orientation angle in degrees. (c) Nonclustered (top) and clustered (bottom) points in the Gaussian image of points in (a). (d) Resulting clustered points in 3-D.

that do not have any neighboring point and produce clusters similar to the clusters obtained from the dendrogram cut at ε in case of hierarchical clustering using single link metric [43].

Increasing MinPts can help in retaining more stable core points but, on the other hand, can also break the clusters into two or more clusters. This property is sometimes useful in cases when

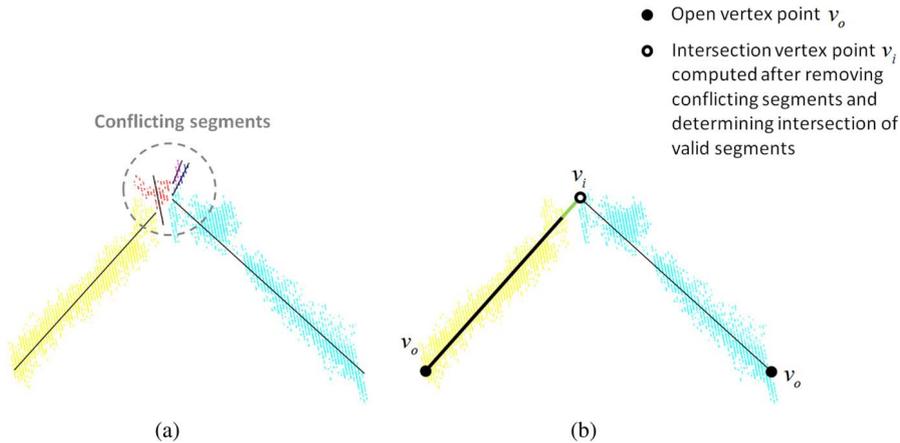


Fig. 13. Removal of conflicting segments and vertex point computation. (a) Shows the case of five clustered point segments belonging to one particular building. At the transitional regions, there are conflicting (smaller) clusters, shown in gray dotted circle, that behave as noisy segments and should be removed before computing the vertex points. (b) Shows the retained segments and their corresponding vertices after removing conflicting segments.

different clusters are merged together by a thin line of points. Estimating the exact value of MinPts is, however, very much dependent on the data set, and certain heuristics based on the “thinnest” cluster in the data set, e.g., k -distance graph, can be employed [43].

In order to reconstruct individual facades, these density-based coarse clusters need to be further clustered. To this end, mean shift clustering has been applied to the coarsely clustered segments in their normal feature space (in GI domain). Fig. 12(b) shows the estimated orientation angle θ for extracted facade points from a single building shown in Fig. 12(a). The variation in orientation angle is quite evident and allows mean shift to cluster points having similar orientations together. Further separation of points in the spatial domain is also required in some cases where the spatially separated points are clustered into one segment. This happens when these points belonging to different facades have similar normals and are spatially closer. Density-based clustering is therefore again applied for spatial separation of the clusters within clusters.

D. Results—Reconstructing Facades

Prior to reconstruction, the segmented facades are first classified to flat and curved surfaces by analyzing the derivatives of the local orientation angle θ . A slope value θ' of 0.3 ($\approx 17^\circ$) is set by empirically testing the computed orientation angles of all of the buildings in the area of interest to distinguish flat and curved surfaces.

After identification, appropriate model parameters are estimated from the core points of the individual clusters. Vertex points are then determined by computing intersections of the adjacent segment pairs. However, in doing so, smaller clusters occurring at facade transition regions behave as noisy segments in the reconstruction procedure. A practical example of these so-called conflicting segments is shown in Fig. 13(a). Following the procedure explained in Table I (see Section III-F), the conflicting (reconstructed) segments occurring at the transitional regions of individual buildings are removed prior to the vertex point computation as exemplified in Fig. 13(b).

Once these transitional clusters are removed, the intersection vertices are determined by computing the intersection point of the two adjacent facades. Refinement operation is then carried out on the open vertices to insert additional segments between the broken facade regions followed by extension of incomplete reconstructed facades.

Fig. 14(a) and (b) depicts the reconstructed facade models of the area of interest before and after refinement, respectively. Green lines show reconstructed facade footprint before refinement. The blue lines indicate additional segments that are added between the vertices of those broken facades that meet the conditions present in step 4 in Section III-G, while the red lines are subsequent extensions of the open vertices after filling the break regions.

E. Results—Validation

The actual ground truth data are missing for exact qualitative evaluation of the approach. In order to provide some quantitative measures of the algorithm performance, we manually counted the actual number of facades that were to be reconstructed. A total of 141 facades are present in the data set, out of which 7 are curved facades and the remaining 134 are flat. Prior to the refinement operation, the algorithm reconstructed a total of 176 facades, i.e., higher than the actual facades present in the data set. As already stated in Section III-G, this is because some individual facades have been broken down into two or more segments due to discontinuity in the number of points available in the data set. After refinement, 29 insertion segments (27 single and 2 double based on the condition $C3$ in Table II) are added between the broken facade regions, whereas 43 facades have been extended. In the final reconstruction, we obtain 147 reconstructed facades, i.e., all 141 facades are successfully reconstructed; among them, five facades remain broken (counted as additional five facades), and there is one case of false alarm which will be explained later. Besides the five cases, we also find seven facades that are not extended and therefore remain incomplete. This is, however, due to the inadequate number of points available in the data.



(a)



- Reconstructed facades without refinement
- Additional facades added between the broken facade regions
- Extension of facades after adding additional segments

(b)

Fig. 14. Reconstructed facades. (a) Two-dimensional view of the facade footprints overlaid onto the optical image prior to refinement. (b) Two-dimensional view of the facade footprints overlaid onto the optical image after refining with parameter settings $T_h = 5$ m and $T_\sigma = 2.5$ m.

As mentioned earlier, there is also one case, shown in Fig. 15, which is considered as false positive (i.e., a facade not actually present but reconstructed by the algorithm). As can

be seen in Fig. 15(c), the reconstructed segment is actually a bridge for pedestrian crossing. Higher number of scatterers is retrieved over the bridge due to its apparently metallic structure.

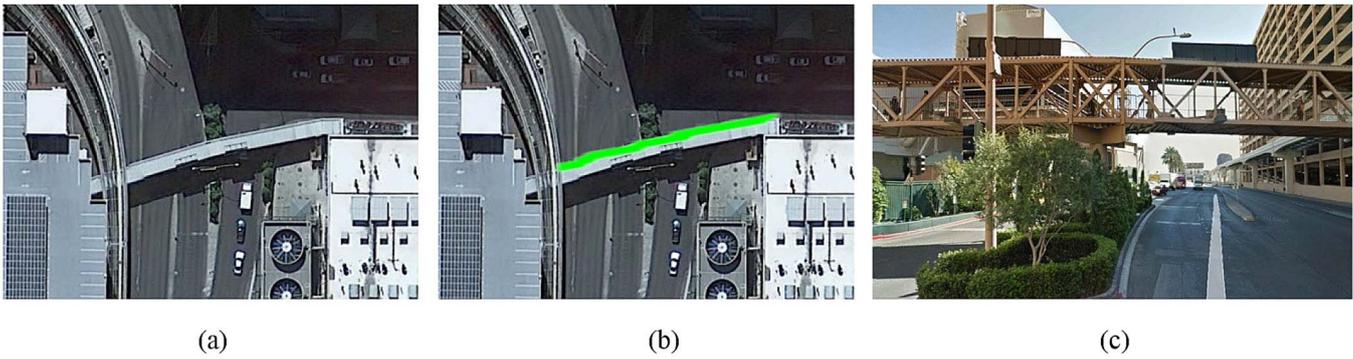


Fig. 15. Case depicting wrong reconstruction of a pedestrian bridge. (a) Optical image of the bridge. (b) Overlays the reconstructed segment in green onto the optical image of (a). (c) Side view of the bridge (copyright Google Street View).

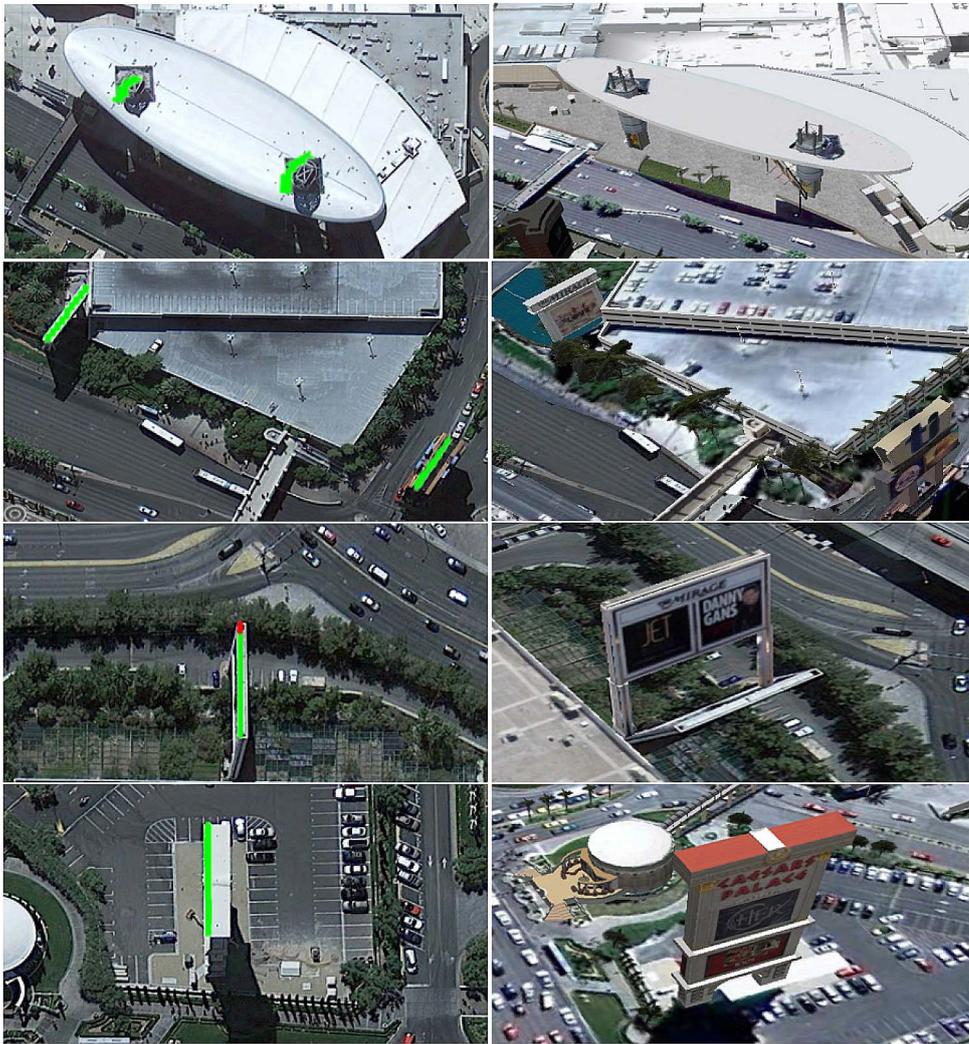


Fig. 16. Reconstructed facades on the left. Their 3-D view on the right (copyright Google Earth).

Moreover, the bridge is also covered from the top, and therefore, scatterers are obtained at the top and bottom and on the metallic rods connecting the upper and lower surfaces of the bridge. The estimated surface normal of these scatterers thus gives a higher horizontal component, and as a consequent, these scatterers are wrongly classified as facade points by satisfying

both extraction constraints: higher SD and higher horizontal component of the surface normals.

It is also interesting here to mention that, in Fig. 14, some small vertical structures on roofs of the buildings or on ground are very well reconstructed. Fig. 16 shows some examples of such objects that might visually appear (or interpreted) as false

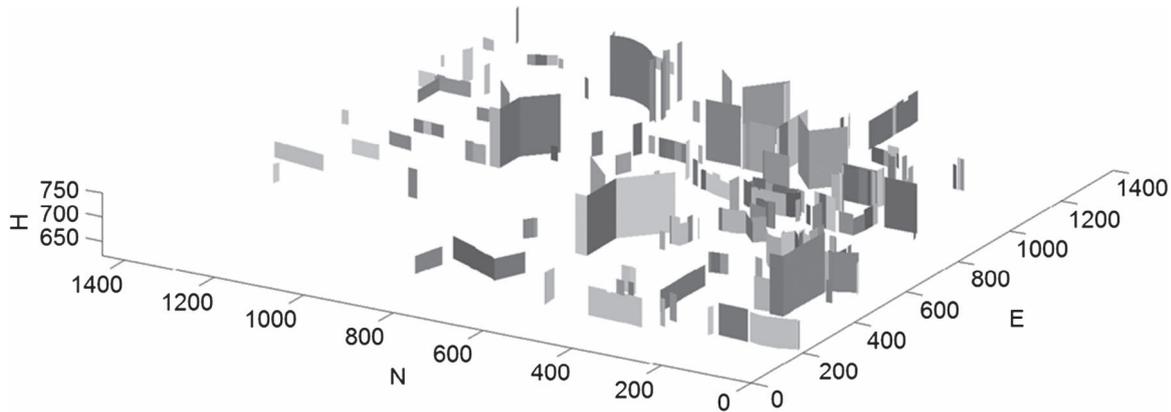


Fig. 17. Three-dimensional view of the final facade reconstruction. The axis is in meter range and has been translated to the origin for better metric clarity by subtracting the UTM easting and northing values by their respective minimum values present in the reconstructed vertices.

reconstructions in Fig. 14 but are actually vertical structures (e.g., advertisement boards, monuments, etc.).

Finally, in Fig. 17, we present the final reconstructed facades in 3-D. As depicted in [42], the shown reconstructed facade model can be used to refine the elevation estimates of the raw TomoSAR points. Moreover, with known deformation estimates of the scatterers, such a model can also lead to the reconstruction of dynamic city models that could potentially be used to monitor and visualize the dynamics of urban infrastructure in very high level of details.

V. OUTLOOK AND CONCLUSION

In this paper, we have presented an automatic (parametric) approach for robust facade reconstruction for large areas using TomoSAR point clouds. The approach is modular and works directly on unstructured 3-D points. It allows for a robust reconstruction of both higher facades and lower height structures, and hence, it is well suited for urban monitoring of larger areas from space. A few points, however, need to be addressed.

- 1) During *SD* estimation, the continuity of an individual facade can be broken due to a limited number of available points. This may result into two or more segments of the same facade. We attempted to cope with this problem by refining the reconstructed facade footprints via insertion and extension operations. Still, the lack of measurements prevents the complete resolution of this problem. The use of 2-D ground plans or cadastral maps can be helpful in this case.
- 2) Since the satellite orbits are bound to pass close to the poles of Earth, we may fail to reconstruct building facades facing North or South due to the missing of measurements. One way to rectify this is by using fused point clouds (i.e., both ascending and descending) and/or inserting new segments by simply connecting the endpoints of the missing facades if they match a certain criteria to get the complete shape of the building footprint.
- 3) The presented approach is a much better option to detect the shape of the building when dense points on the facades are available. However, in cases (usually for lower height buildings) when no or few facade points are avail-

able, one can try to extract roof points and reconstruct the 2-D footprint. This could help in resolving the problems related to the visibility of facades mainly pointing toward the azimuth direction.

In the future, we will work over these considerations and will extend the algorithm toward object-based TomoSAR point cloud fusion and automatic building roof reconstruction.

REFERENCES

- [1] J. Hu, S. You, and U. Neumann, "Approaches to large-scale urban modeling," *IEEE Comput. Graph. Appl.*, vol. 23, no. 6, pp. 62–69, Nov./Dec. 2003.
- [2] J. Benner, A. Geiger, and K. Leinemann, "Flexible generation of semantic 3-D building models," in *Proc. Int. ISPRS/EuroSDR/DGPF-Workshop Next Gen. 3-D City Models*, K. Gröger, Ed., Bonn, Germany, 2005, pp. 18–22, EuroSDR Publication No. 49.
- [3] J. Lee and S. Zlatanova, "A 3-D data model and topological analyses for emergency response in urban areas," in *Geospatial Information Technology for Emergency Response*, L. Zlatanova, Ed. New York, NY, USA: Taylor & Francis, 2008.
- [4] S. Kemec, S. Duzgun, S. Zlatanova, D. I. Dilmen, and A. C. Yalciner, "Selecting 3-D urban visualization models for disaster management: Fethiye tsunami inundation case," in *Proc. 3rd Int. Conf. Cartogr. GIS*, Nessebar, Bulgaria, Jun. 15–20, 2010, pp. 1–9.
- [5] K. Lee, Y. Kim, S. I. Cho, and K. Choi, "Building detection in augmented reality based navigation system," in *Proc. 13th Int. Conf. Multimedia Model.—Volume Part II*, 2006, pp. 544–551.
- [6] J. Döllner, T. H. Kolbe, F. Liecke, T. Sgouros, and K. Teichmann, "The virtual 3-D city model of Berlin—Managing, integrating, and communicating complex urban information," in *Proc. 25th UDMS*, Aalborg, Denmark, May 15–17, 2006, pp. 1–12.
- [7] A. Jochem, B. Hoeffle, M. Rutzinger, and N. Pfeifer, "Automatic roof plane detection and analysis in airborne LiDAR point clouds for solar potential assessment," *Sensors*, vol. 9, no. 7, pp. 5241–5262, 2009.
- [8] J. Kirtner, "Using LiDAR data in wireless communication system design," in *Proc. Am. Soc. Photogramm. Remote Sens. Conf.*, Washington, DC, USA, 2000, pp. 1–8.
- [9] M. Over, A. Schilling, S. Neubauer, and A. Zipf, "Generating Web-based 3-D city models from OpenStreetMap: The current situation in Germany," *Comput., Environ. Urban Syst.*, vol. 34, no. 6, pp. 496–507, Nov. 2010.
- [10] X. Yin, P. Wonka, and A. Razdan, "Generating 3-D building models from architectural drawings: A survey," *IEEE Comput. Graph. Appl.*, vol. 29, no. 1, pp. 20–30, Jan./Feb. 2009.
- [11] First App for LiDAR Data Visualisation on iPad, Accessed Mar. 21, 2014. [Online]. Available: <http://geoawesomeness.com/first-app-lidar-data-visualisation-ipad/>
- [12] M. Quartulli and M. Datcu, "Stochastic geometrical modeling for built-up area understanding from a single SAR intensity image with meter resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1996–2003, Sep. 2004.

- [13] A. Ferro, D. Brunner, and L. Bruzzone, "An advanced technique for building detection in VHR SAR images," in *Proc. SPIE Eur. Remote Sens.*, Berlin, Germany, 2009, vol. 7477, pp. 74770V-1–74770V-12.
- [14] A. Ferro, D. Brunner, and L. Bruzzone, "Automatic detection and reconstruction of building radar footprints from single VHR SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 935–952, Feb. 2013.
- [15] J. D. Wegner, U. Soergel, and A. Thiele, "Building extraction in urban scenes from high resolution InSAR data and optical imagery," in *Proc. URBAN*, 2009, pp. 1–6.
- [16] P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 611–617, Jan. 2000.
- [17] A. Thiele, E. Cadario, K. Schulz, U. Thoennessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007.
- [18] F. Xu and Y.-Q. Jin, "Automatic reconstruction of building objects from multi-aspect meter-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2336–2353, Jul. 2007.
- [19] H. Sportouche, F. Tupin, and L. Denise, "Extraction and three-dimensional reconstruction of isolated buildings in urban scenes from high-resolution optical and SAR spaceborne images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3932–3946, Oct. 2011.
- [20] U. Stilla, U. Soergel, and U. Thoennessen, "Potential and limits of InSAR data for building reconstruction in built-up areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 1/2, pp. 113–123, Jun. 2003.
- [21] A. Sampath and J. Shan, "Segmentation and reconstruction of polyhedral building roofs from aerial LiDAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1554–1567, Mar. 2010.
- [22] R. Bamler, M. Eineder, N. Adam, S. Gernhardt, and X. Zhu, "Interferometric potential of high resolution spaceborne SAR," in *Proc. PGF*, 2009, pp. 407–419.
- [23] S. Mezzasoma *et al.*, "COSMO-SkyMed system commissioning: End-to-end system performance verification," in *Proc. IEEE Radar Conf.*, Rome, Italy, May 2008, pp. 1–5.
- [24] N. Adam, X. Zhu, and R. Bamler, "Coherent stacking with TerraSAR-X imagery in urban areas," in *Proc. JURSE*, Shanghai, China, 2009, pp. 1–6.
- [25] G. Fornaro, D. Reale, A. Pauciuolo, X. Zhu, and R. Bamler, "SAR tomography: An advanced tool for spaceborne 4-D radar scanning with application to imaging and monitoring of cities and single buildings," *IEEE Geosci. Remote Sens. Newsl.*, Dec. 2012.
- [26] X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
- [27] X. Zhu and R. Bamler, "Let's do the time warp: Multi-component nonlinear motion estimation in differential SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, Jul. 2011.
- [28] X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.
- [29] D. Reale, G. Fornaro, A. Pauciuolo, X. Zhu, and R. Bamler, "Tomographic imaging and monitoring of buildings with very high resolution SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 661–665, Jul. 2011.
- [30] S. Gernhardt and R. Bamler, "Deformation monitoring of single buildings using meter-resolution SAR data in PSI," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 68–79, Sep. 2012.
- [31] Y. Wang, X. Zhu, and R. Bamler, "An efficient tomographic inversion approach for urban mapping using meter resolution SAR image stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, p. 1, Jul. 2012.
- [32] D. Reale, G. Fornaro, and A. Pauciuolo, "Extension of 4-D SAR imaging to the monitoring of thermally dilating scatterers," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 12, pp. 5296–5306, Dec. 2013.
- [33] X. Zhu and R. Bamler, "Super-resolving SAR tomography for multi-dimensional imaging of urban areas," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 51–58, Jul. 2014.
- [34] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.
- [35] P. Pasquali, C. Prati, F. Rocca, and M. Seymour, "A 3-D SAR experiment with EMSL data," in *Proc. IGARSS*, 1995, pp. 784–786.
- [36] G. Fornaro, F. Serafino, and F. Lombardini, "Three-dimensional multipass SAR focusing: Experiments with long-term spaceborne data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 702–714, Apr. 2005.
- [37] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 34–44, Jan. 2005.
- [38] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 224–237, Jan. 2009.
- [39] G. Fornaro, F. Serafino, and D. Reale, "4-D SAR imaging: The case study of Rome," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 236–240, Apr. 2010.
- [40] F. Lombardini, *Diff-Tomo opening of the urban SAR pixel: Single-look 4-D and nonuniform motion 3-D extensions*. Presented at FRINGE Workshop, 2011. [Online]. Available: http://earth.eo.esa.int/workshops/fringe2011/files/Lombardini_FRINGE2011.pdf
- [41] M. Shahzad, X. Zhu, and R. Bamler, "Façade structure reconstruction using spaceborne TomoSAR point clouds," in *Proc. IEEE IGARSS*, Munich, Germany, 2012, pp. 467–470.
- [42] X. Zhu and M. Shahzad, "Façade reconstruction using multi-view spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.
- [43] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. KDD*, Portland, OR, USA, 1996, pp. 226–231.
- [44] Y. Liu and Y. Xiong, "Automatic segmentation of unorganized noisy point clouds based on the Gaussian map," *Comput.-Aided Design*, vol. 40, no. 5, pp. 576–594, May 2008.
- [45] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
- [46] G. Sithole and G. Vosselman, "Experimental comparison of filtering algorithms for bare-earth extraction from airborne laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, no. 1/2, pp. 85–101, Aug. 2004.
- [47] G. Vosselman, "Slope based filtering of laser altimetry data," *Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 33, pt. B3, pp. 935–942, 2000.
- [48] G. Sohn and I. Dowman, "Terrain surface reconstruction by the use of tetrahedron model with the MDL criterion," *Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 34, pt. 3A, pp. 336–344, 2002.
- [49] H. Gross, U. Thoennessen, and W. Hansen, "3-D-modeling of urban structures," in *Proc. Int. Archives Photogramm. Remote Sens.*, Vienna, Austria, 2005, vol. XXXVI/3-W24, pp. 137–142.
- [50] F. Rottensteiner and C. Briese, "A new method for building extraction in urban areas from high-resolution LiDAR data," in *Proc. Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, 2002, vol. 34, pp. 295–301, Part 3A.
- [51] A. Sampath and J. Shan, "Building boundary tracing and regularization from airborne LiDAR point clouds," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 7, pp. 805–812, Jul. 2007.
- [52] C. Mallet, F. Bretar, M. Roux, U. Soergel, and C. Heipke, "Relevance assessment of full-waveform LiDAR data for urban area classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. S71–S84, Dec. 2011.
- [53] P. Dorninger and N. Pfeifer, "A comprehensive automated 3-D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds," *Sensors*, vol. 8, no. 11, pp. 7323–7343, 2008.
- [54] G. Forlani, C. Nardinocchi, M. Scaioni, and P. Zingaretti, "Complete classification of raw LiDAR data and 3-D reconstruction of buildings," *Pattern Anal. Appl.*, vol. 8, no. 4, pp. 357–374, Feb. 2006.
- [55] S. Auer, S. Gernhardt, and R. Bamler, "Ghost persistent scatterers related to multiple signal reflections," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 919–923, Sep. 2006.
- [56] Y. Wang, X. Zhu, R. Bamler, and S. Gernhardt, "Towards TerraSAR-X street view: Creating city point cloud from multi-aspect data stacks," in *Proc. JURSE*, Sao Paulo, Brasil, 2013, pp. 198–201.
- [57] P. Holland and R. Welsh, "Robust regression using iteratively reweighted least-squares," *Commun. Statist., Theory Methods*, vol. 6, no. 9, pp. 813–827, 1977.
- [58] J. O. Street, R. J. Carroll, and D. Ruppert, "A note on computing robust regression estimates via iteratively reweighted least squares," *Amer. Statist.*, vol. 42, no. 9, pp. 152–154, 1988.
- [59] P. J. Huber, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 1981.
- [60] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *Comput. Graph.*, vol. 26, no. 2, pp. 71–78, Jul. 1992.
- [61] M. Hubert, P. Rousseeuw, and K. V. Branden, "ROBPCA: A new approach to robust principal components analysis," *Technometrics*, vol. 47, no. 1, pp. 64–79, Feb. 2005.

- [62] R. R. Wilcox, *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*, 2nd ed. New York, NY, USA: Springer-Verlag, 2010.
- [63] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
- [64] M. P. D. Carmo, *Differential Geometry of Curves and Surfaces*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1976.
- [65] D. Comaniciu and P. Meer, "Mean shift: A robust approach towards feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [66] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.
- [67] X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's tomographic SAR processing system," in *Proc. JURSE*, Sao Paulo, Brazil, 2013, pp. 308–313.



Muhammad Shahzad (S'12) received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, and the M.Sc. degree in autonomous systems from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany. He has been working toward the Ph.D. degree in the Chair of Remote Sensing Technology (LMF), Technische Universität München (TUM), München, Germany, since December 2011.

His research topic is the automatic 3-D reconstruction of objects from point clouds retrieved from spaceborne synthetic aperture radar image stacks.



Xiao Xiang Zhu (S'10–M'12) received the Bachelor's degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006 and the M.Sc. degree, the Dr.-Ing. degree, and the "Habilitation" in the field of signal processing from Technische Universität München (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

In October/November 2009, she was a Guest Scientist with the Italian National Research Council (CNR)—Institute for Electromagnetic Sensing of the Environment (IREA), Naples, Italy. Since May 2011, she has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany, where she is the Head of the Team Signal Analysis and with the Chair of Remote Sensing Technology, TUM. Since September 2013, she has also been a Helmholtz Young Investigator Group Leader and appointed as a TUM junior fellow. Her main research interests are advanced InSAR techniques such as high-dimensional tomographic SAR imaging and SqueeSAR, computer vision in remote sensing including object reconstruction and multidimensional data visualization, and modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multi-/hyperspectral image analysis.

C Shahzad, M., Zhu, X. X., 2015. Automatic detection and reconstruction of 2-D/3-D building shapes from spaceborne TomoSAR point clouds. IEEE Transactions on Geoscience and Remote Sensing PP(99): 1-19

<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7283615>

Automatic Detection and Reconstruction of 2-D/3-D Building Shapes From Spaceborne TomoSAR Point Clouds

Muhammad Shahzad, *Student Member, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

Abstract—Modern spaceborne synthetic aperture radar (SAR) sensors, such as TerraSAR-X/TanDEM-X and COSMO-SkyMed, can deliver very high resolution (VHR) data beyond the inherent spatial scales of buildings. Processing these VHR data with advanced interferometric techniques, such as SAR tomography (TomoSAR), allows for the generation of four-dimensional point clouds, containing not only the 3-D positions of the scatterer location but also the estimates of seasonal/temporal deformation on the scale of centimeters or even millimeters, making them very attractive for generating dynamic city models from space. Motivated by these chances, the authors have earlier proposed approaches that demonstrated first attempts toward reconstruction of building facades from this class of data. The approaches work well when high density of facade points exists, and the full shape of the building could be reconstructed if data are available from multiple views, e.g., from both ascending and descending orbits. However, there are cases when no or only few facade points are available. This usually happens for lower height buildings and renders the detection of facade points/regions very challenging. Moreover, problems related to the visibility of facades mainly facing toward the azimuth direction (i.e., facades orthogonally oriented to the flight direction) can also cause difficulties in deriving the complete structure of individual buildings. These problems motivated us to reconstruct full 2-D/3-D shapes of buildings via exploitation of roof points. In this paper, we present a novel and complete data-driven framework for the automatic (parametric) reconstruction of 2-D/3-D building shapes (or footprints) using unstructured TomoSAR point clouds particularly generated from one viewing angle only. The proposed approach is illustrated and validated by examples using TomoSAR point clouds generated using TerraSAR-X high-resolution spotlight data stacks acquired from ascending orbit covering two different test areas, with one containing simple moderate-sized buildings in Las Vegas, USA and the other containing relatively complex building structures in Berlin, Germany.

Index Terms—Building footprint, building reconstruction, clustering, dynamic city models, TerraSAR-X, tomographic SAR (TomoSAR) inversion, four-dimensional (4-D) point cloud.

Manuscript received May 13, 2015; revised July 24, 2015; accepted September 3, 2015. This work was supported in part by the Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018) and in part by the International Graduate School of Science and Engineering of the Technische Universität München through Project “6.08 4-D City.” (Corresponding author: Xiao Xiang Zhu.)

M. Shahzad is with the Signal Processing in Earth Observation (SiPEO), Technische Universität München, 80333 Munich, Germany (e-mail: muhammad.shahzad@tum.de).

X. X. Zhu is with the German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), 82234 Wessling, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2015.2477429

I. INTRODUCTION

AUTOMATED methods for the detection and reconstruction of buildings are becoming increasingly important due to the ever-increasing growth of urban population, giving rise to a wide range of potential applications in numerous fields. For instance, 2-D building footprints (or outlines) can be used in urban energy modeling [1], disaster management [2], and devising emergency responses [3]. Additionally, extruding these footprints using height information leads to the development of 3-D building models. Such models are valuable for various applications, e.g., virtual city modeling [4], urban planning [5], and analyzing electromagnetic reflections for placement of telecommunication antennas [6], etc.

Modern spaceborne synthetic aperture radar (SAR) sensors, such as TerraSAR-X/TanDEM-X and COSMO-SkyMed, are able to provide very high resolution (VHR) data beyond the inherent spatial scales (on the order of 1 m) of buildings, constituting an invaluable data source for detailed urban mapping on a global scale. Moreover, due to the coherent imaging nature of SAR, these sensors are able to provide day/night global coverage in virtually all weather conditions. Moreover, with image acquisitions taken at different times, they are also uniquely capable of imaging the dynamics of the illuminated area down to the scale of centimeters or even millimeters from space. These encouraging attributes have drawn the attention of many remote sensing analysts, and therefore, various techniques have been developed, which make use of VHR SAR imagery for 2-D/3-D reconstruction of man-made structures in particular buildings [7]–[11].

Although much research effort has been put in the area, the problem of building reconstruction remains a challenging task due to the inherent characteristics of SAR images, such as speckle, layover/foreshortening, etc. Moreover, complex building structures and high variability of objects appearing in SAR images also render difficulties in the detection of buildings, particularly in dense urban areas where buildings are compactly packed, and smaller buildings are often occluded by higher buildings. Therefore, prior knowledge with certain regularization (geometric) constraints (e.g., vertical facades [10] and rectangular building shapes [9]) is often incorporated for realistic and automatic reconstruction.

SAR tomography (TomoSAR) is an advanced interferometric technique that aims for real 3-D SAR imaging [12]–[16]. It resolves multiple/dominant scatterers within one resolution element (i.e., able to retrieve more than one scatterer per

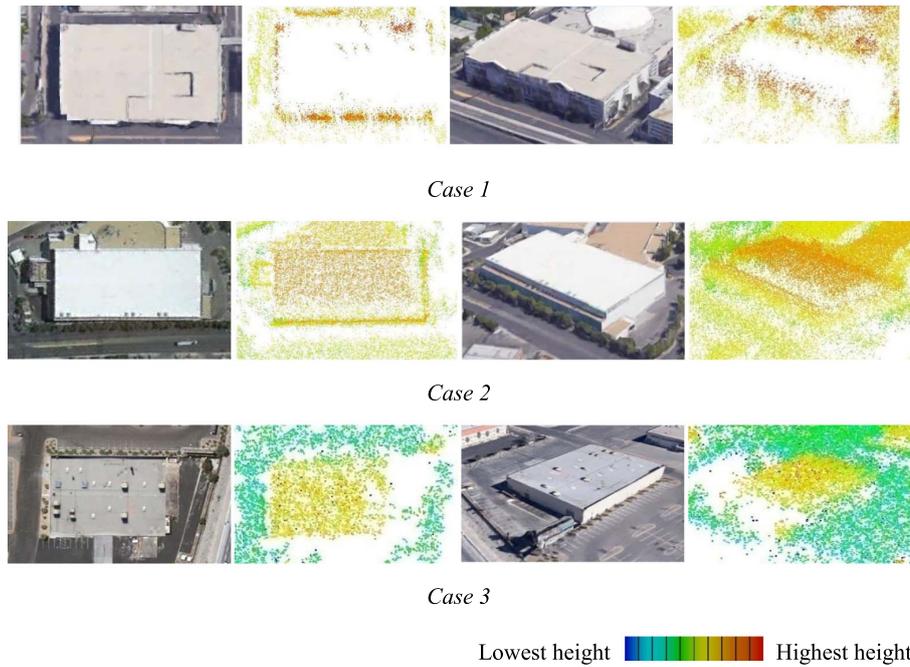


Fig. 1. Three different cases over simple buildings commonly occurring in TomoSAR point clouds. Heights of the points are color coded. (Optical images: Google).

pixel) via spectral analysis by building up a synthetic aperture in the elevation direction to localize each scatterer in 3-D [15], [16]. Likewise, building up a synthetic aperture in the temporal domain enables 4-D space–time) focusing of SAR images [17]–[19]. The technique is referred to as D-TomoSAR, which combines the strengths of both TomoSAR and persistent scatterer interferometry [17]–[21]. It is capable of retrieving elevation and deformation information (linear, seasonal, etc.) even of multiple scatterers inside a single SAR image pixel [15], [18], [22], [23]. Geocoding high density of scatterers, retrieved from TomoSAR, into world coordinates produces high-quality TomoSAR point clouds, containing not only the 3-D positions of the scatterer location but also the estimates of seasonal/temporal deformation, making them very attractive for generating 4-D city models from space.

Object reconstruction from spaceborne TomoSAR point clouds was started not until recently [24], [25]. These point clouds have point density in the range of 600 000–1 000 000 points/km² using TerraSAR-X meter resolution data [16], [27]. Similar statistics using CosmoSkyMed data are reported in [27]. TomoSAR point clouds, however, are associated with some special considerations that are worth to mention [24], [25].

- 1) They deliver moderate 3-D positioning accuracy on the order of 1 m [28].
- 2) Few number of images and limited orbit spread render the location error of TomoSAR points highly anisotropic, with an elevation error typically at least one order of magnitude higher than in range and azimuth.
- 3) Due to the coherent imaging nature, temporally incoherent objects such as trees cannot be reconstructed from multipass spaceborne SAR image stacks.

- 4) Ghost scatterers may be generated due to multiple scattering that appears as outliers far away from a realistic 3-D position [29].
- 5) Side-looking SAR geometry enables strong scatterer reflections from dihedral and trihedral structures or metallic parts resulting in higher density of points on man-made objects, particularly building facades and bridges. Naturally, this urges for the systematic reconstruction of building footprints via facade point analysis.

As depicted over smaller [25] and larger areas [24], facade reconstruction turns out to be an appropriate first step in detecting and reconstructing building shapes from these point clouds when dense points on the facade are available. In particular, when data from multiple views, e.g., from both ascending and descending orbits, are available, the full shape of buildings can be reconstructed using extracted facade points. However, there are cases when no or only few facade points are available. This usually happens for lower height buildings and renders the detection of facade points/regions very challenging. Moreover, problems related to the visibility of facades mainly facing the azimuth direction (i.e., having normals pointing toward the flight direction) can also cause difficulties in deriving the complete structure of an individual building. These problems motivate us to reconstruct full 2-D/3-D shapes of building footprints via facade–roof analysis. In this regard, based on different object contents illuminated by side-looking SAR, the following three cases could be derived using data acquired from one incidence angle, e.g., in this case, image stacks from the ascending orbit only (see Fig. 1).

Case 1) *Higher density of facade points present with no or very few roof points*: In this case, the complete

2-D/3-D building shapes could only be fully reconstructed by adding points from multiple incidence angles. The solution to this case is demonstrated in [24], where 3-D facade models have been reconstructed for high-rise buildings using one incidence angle only, and in [25], where the full shape of the building was derived by prior fusion of two point clouds [30] from ascending and descending stacks.

Case 2) *Higher density of facade points present together with the existence of relatively higher density of roof points*: This case allows for the reconstruction of the full shape of the building footprints from a single data stack by making use of both facade and roof points. Thus, the side of the building visible to the sensor could be reconstructed as the first step using facade points, and later, the other side of the building could be completed by exploiting the available roof points.

Case 3) *No or very few facade points available but enough roof points exist*: This case particularly appears for lower height buildings, rendering the detection of facade points/regions very challenging. This motivates us to obtain the complete 2-D/3-D footprint of these buildings via roof point analysis only using conventional techniques as applied by the light detection and ranging (LiDAR) community. Although these techniques are very much developed, still, their adaptation to TomoSAR point clouds is not straightforward due to different object contents illuminated by side-looking SAR together with problems related to less and varying point density and much less positioning accuracies of TomoSAR point clouds in comparison to airborne LiDAR.

In this paper, we propose a novel data-driven approach that systematically allows for the automatic reconstruction of 2-D/3-D building shapes (or footprints) using unstructured TomoSAR point clouds generated from one incidence angle only. The approach proposes new methods and aims at finding a more general and systematic solution toward the automatic reconstruction of the whole city area. This paper essentially presents solutions for the latter two cases (i.e., case 2 and case 3) by extending (or utilizing) the solution provided for case 1 in [24]. The innovative contributions specific to the approach proposed in this paper are as follows.

- 1) A hybrid approach based on the region-growing procedure and the energy minimization framework is proposed to automatically extract building roof/facade points directly from 3-D TomoSAR points. Seed points are extracted using the reconstructed facade information, and later, a minimum-height-constraint surface-normal-based region-growing approach is adopted to detect probable building regions in case 2. The problem of extracting roof points in case 3 (i.e., having no facade information) from the remaining set of points is then resolved by computing local height and planar features and formulating the whole detection problem into an energy minimization

framework. Graph cuts are later employed to globally extract roof/building points.

- 2) A recursive angular-deviation-based approach is presented to smooth/refine the initial coarse building polygons obtained using alpha shapes (generalization of convex hulls). The smoothed boundaries yield better visually appealing building shapes.
- 3) A novel facade–roof fusion procedure is proposed in this paper. The developed method is robust and fuses the legitimate *facade-Polygon* pair together by interpreting the refined/smoothed polygon of each building as a graph. A series of operations taking part in the reduction of graph (or smoothed polygon) includes identification of points on the graph that are nearest to all the associated/paired reconstructed facades, computation of longest and shortest paths on the graph using these identified points, simplification of shortest paths by representation as line segments, and rotation/projection of certain line segments (simplified shortest paths that are identified as positive paths) to match all the reconstructed facades belonging to the same graph (i.e., smoothed building polygon).
- 4) An effective and robust procedure is developed for rectilinear identification of building polygons. To this end, the dominant direction of each building is first determined based on the weighted-line-segment approach. Angular deviations at each adjacent or consecutive vertex of the polygon are computed from the dominant direction, and histogram analysis is then carried out for rectilinear building footprint identification. Later, rectilinear constraints are added to the identified building polygons.
- 5) The approach presented in [24] may fail to reconstruct building facades facing north or south because of missing measurements caused by the high inclination angle of the TerraSAR-X orbit, i.e., near-polar orbit (see Fig. 2). This paper inherently provides a solution to this problem by exploiting roof points in determining the complete shape/footprint of the building (see Fig. 3).
- 6) Finally, this paper presents the first demonstration of automatic reconstruction of 2-D/3-D building shapes from this class of data. Moreover, the developed methods are applicable not only to TomoSAR point clouds, as demonstrated in this paper, but also to unstructured 3-D point clouds generated from a different sensor with a similar configuration (i.e., oblique geometry) and with both low and high point densities.

The aforementioned contributions allow for a completely automatic (but parametric) reconstruction of 2-D/3-D building shapes from TomoSAR point clouds in larger areas.

The remainder of this paper is structured as follows. Section II provides the basic idea of the whole approach. Section III presents the proposed methodology for detection/extraction of building points. Section IV then presents in detail the processing steps that are employed for reconstruction of building footprints. In Section V, the experimental results obtained from the TomoSAR point cloud generated from a TerraSARX

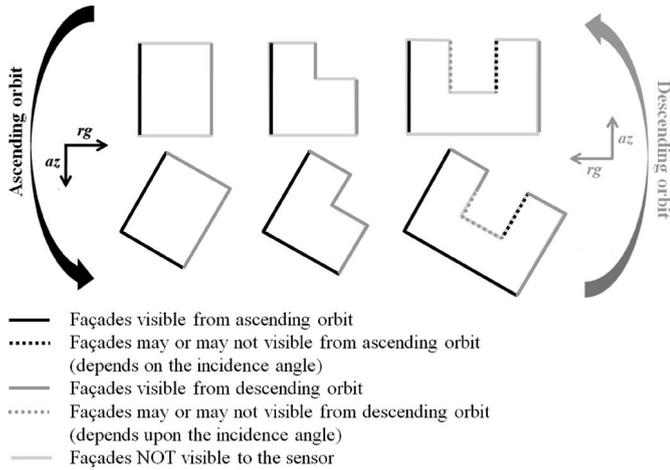


Fig. 2. Illustration of facade visibility problems in data available from single- and/or multiview perspective (i.e., data from both ascending and descending orbits). Six polygons depict the 2-D building footprints or outlines (top view) of three buildings in two different orientations. Due to side-looking geometry, not all facades are visible to the sensor. If we consider data available from the ascending orbit only, it can be easily inferred that there is no possibility to reconstruct the shape of the footprint for any building as oriented in the first row. In the second row, only the first building could be reconstructed if we impose certain strict constraints (e.g., symmetric or rectangular-shaped footprint assumption). However, such a constraint is not always feasible due to a variety of building shapes, and therefore, for the latter two buildings (for which two facades are visible), the complete shape of the footprint could only be reconstructed if either the data from multiple views are available or enough roof points exist, which could be exploited together with the reconstructed facades (see Fig. 3). If data from the descending orbit are also incorporated, it could be seen that the building footprints in the second row could be almost completely reconstructed. However, even in this case, it is still difficult to reconstruct building footprints in the first row due to lack of data on facades that are not visible to the sensor.

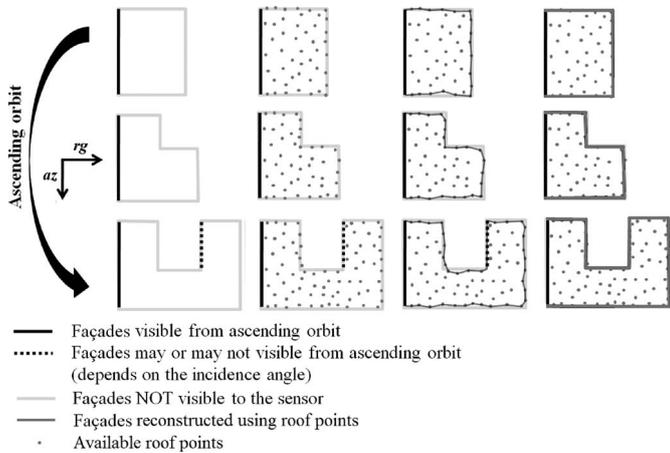


Fig. 3. Illustration of the solution to the facade visibility problems from the single-view perspective, as proposed in this paper. The first column contains the same three buildings oriented as in the first row of Fig. 2. Assuming that enough facade points available, the facades shown by black polylines could be reconstructed. Moreover, if enough roof points, which are represented by gray dots in the second column, are available, the complete shape of the building footprints could be reconstructed by fusing the rough shape obtained by connecting roof points occurring at the building boundary together with the reconstructed facades, as shown in the third column. Finally, if the building shape is rectilinear, constraints could be added to yield geometrically correct and visually aesthetic building shapes (fourth column).

high-resolution spotlight data stack (ascending orbit only) are presented and validated. Finally, in Section VI, conclusions are drawn, and future perspectives are discussed.

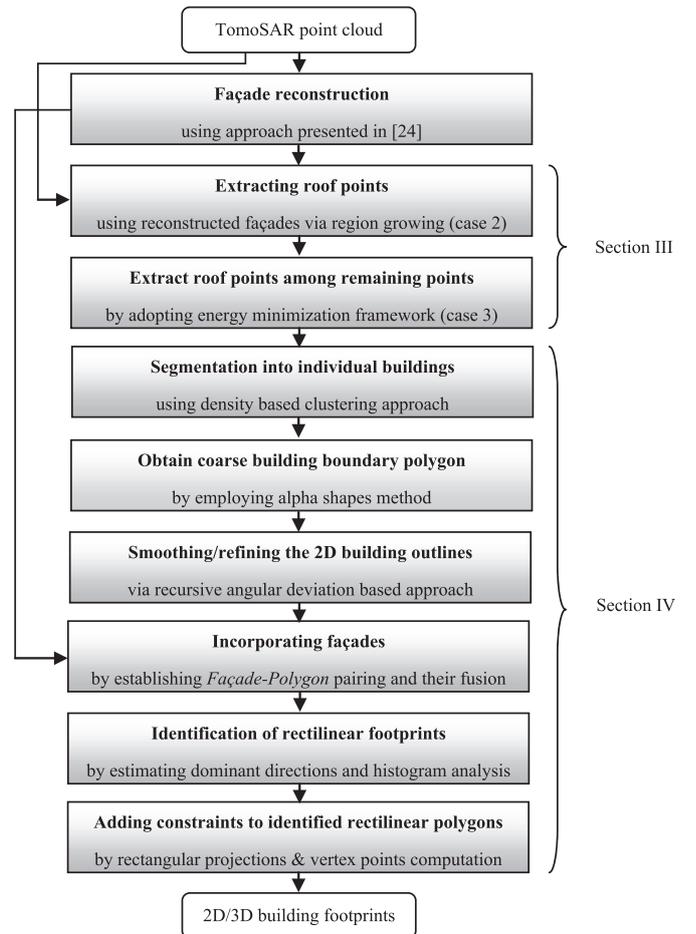


Fig. 4. Block diagram of the proposed approach.

II. METHODOLOGICAL OVERVIEW OF THE RECONSTRUCTION PROCEDURE

Fig. 4 shows the block diagram of the processing steps involved in the complete methodology. The whole processing begins by first identifying the probable building regions. This is accomplished by incorporating information pertaining to facades as prior knowledge (i.e., regions corresponding to higher point density indicate probable facade regions). Thus, building facade points are first extracted, segmented to points belonging to individual facades, and further reconstructed. Detailed processing schemes are described in [24] and [25]. The reconstructed facades are used to select seed points, from which a minimum-height-constraint surface-normal-based region-growing procedure is adopted to extract available roof points. Then, roof points without the support of facade points (i.e., case 3) are further extracted from the remaining points by formulating the extraction problem into an energy minimization framework. Automatic segmentation of detected/extracted building points is then obtained by clustering points belonging to individual buildings. Later, boundary/outline polygons (or footprints) are reconstructed and refined/smoothed for each individually segmented building cluster. Afterward, robust fusion of legitimate *facade-Polygon* pairs is carried out to improve the geometrical accuracy of the refined footprints. Finally, after the identification of rectilinear

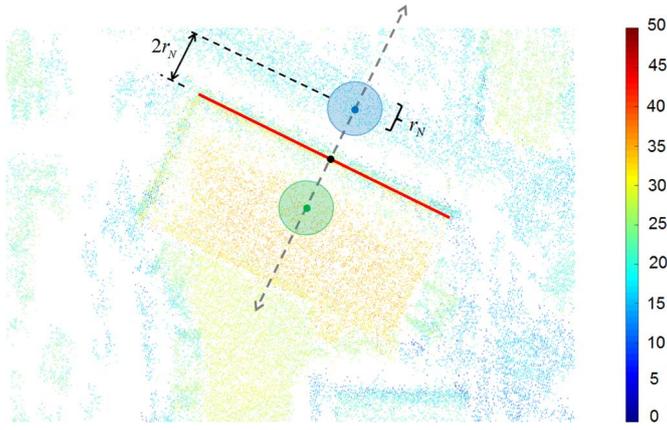


Fig. 5. Graphical illustration of the seed point selection procedure. The top view of 3-D TomoSAR points of a small region containing one building is depicted. The height (in meters) of each TomoSAR point is color coded. The red line shows the reconstructed facade segment with the black point as its midpoint. The green and blue points show the selected points in two opposite orthogonal directions with respect to the reconstructed facade (depicted as gray dotted lines). r_N is the cylindrical radius used to determine local neighbors (i.e., points within shaded circles in respective colors). The local mean height of the blue and green points is computed, denoted as m_b and m_g , respectively. Seed point s_p is chosen as the maximum height point among the local neighbors of blue points if $m_b > m_g$ or among the neighbors of green points if $m_g > m_b$, where the latter is true in the depicted case.

footprints, rectangular constraints are inserted to yield geometrically correct and visually aesthetic building shapes.

Next, we explain in detail the procedures of the proposed approach in the following two dedicated sections.

III. BUILDING DETECTION

A. Selection of Seed Points

The reconstructed facades provide direct hints for the desired building regions. The idea is then to search in the nearby vicinity of the reconstructed facade to select seed points and later employ a surface-normal-based region-growing algorithm to extract existing roof points. In this regard, the seed points are selected by first computing the midpoint of the reconstructed facade segment and then analyzing the local height statistics in the two opposite directions perpendicular to the reconstructed segment. Fig. 5 provides the graphical illustration of the seed selection procedure.

B. Region-Growing Procedure

Extracted seed points from the above procedure are then used in the region-growing procedure to extract existing roof points based on the similarity of their surface normals (robustly estimated [31]), i.e., points having an angular difference between their surface normals less than $\theta_{normals}$ are added to the grown cluster. However, using surface normals only could also add ground points during the growing procedure. An example of such a case will be where both the roof and the neighboring ground are flat. A minimum height constraint is therefore needed and is incorporated in the growing procedure to restrict the addition of points below a certain height h_{min} , which is adaptively computed for each seed point as $m_p + (\text{height of } s_p - m_p) \cdot fac$, where $m_p = m_b$ if $m_g > m_b$ or

$m_p = m_g$ if $m_b > m_g$ (see Fig. 5). The factor fac (empirically set to 0.55 in this work) can be adjusted between 0 and 1 to adjust h_{min} .

The above procedure extracts the existing roof points only for those buildings whose facades can be determined/reconstructed. However, as already mentioned, still, there exist cases when no or very few points are available on the building facades, rendering the detection of facade points/regions very difficult. Moreover, if data are acquired from one viewing angle, e.g., ascending orbit only, the facades of lower buildings could get fully (or partly) occluded due to the presence of nearby higher building structures. As a consequence, the corresponding roof points will be misdetections. To cope with the aforementioned issue, we sequentially formulate the building detection problem among the remaining points into a simple energy minimization framework to extract those building/roof points where no information pertaining to facades could be exploited (i.e., resolve case 3).

C. Formulation of Energy Equation

The detection/extraction of building roof points from the remaining point cloud can be formulated in terms of an energy minimization problem as follows [32]–[35]:

$$E(l) = \sum_{p_i \in \mathbf{P}} D_{p_i}(l_{p_i}) + \sum_{(p_i, p_j) \in \mathbf{N}} w_{p_{ij}} \delta(l_{p_i} \neq l_{p_j}) \quad (1)$$

where $\mathbf{P} = \{p_i | i = 1, \dots, n\}$ denotes the set of remaining n 3-D points; \mathbf{N} denotes the set containing pairs of neighboring points (p_i, p_j) ; l_{p_i} represents the label assigned to point p_i , i.e., roof or nonroof point; $\delta(\cdot)$ is an indicator function; and $w_{p_{ij}}$ is the weight of each pair of neighboring points (p_i, p_j) and is defined as $w_{p_{ij}} = \exp(-\|\mathbf{p}_i - \mathbf{p}_j\|)$. The second summation term in (1) ensures spatial smoothness, i.e., it favors consistent labeling between the neighboring point pairs, whereas the first term in (1) denotes the data discrepancy term that measures the suitability/unsuitability of a particular labeling l assigned to the set of points \mathbf{P} . It is defined as [32]

$$D_{p_i}(l_{p_i}) = \begin{cases} (1 - h_{p_i}) + \eta \cdot r_{p_i}, & \text{if } l_{p_i} = \text{building roof point} \\ h_{p_i} + \eta \cdot (1 - r_{p_i}), & \text{if } l_{p_i} \neq \text{building roof point} \end{cases} \quad (2)$$

where r_{p_i} is the plane residual distance feature computed for each point p_i by locally fitting the robust RANSAC plane among its local neighbors, and h_{p_i} is the differential height of point p_i obtained after subtracting the height of the underlying terrain, which is approximated by fitting a cubic polynomial surface to the nonbuilding/ground points via the robust least absolute residual (LAR) method, as explained in the following section. Both features h_{p_i} and r_{p_i} are normalized to the scale of 0–1 by adopting the forms [32]: $h_{p_i} = \min(1, h_{p_i}/\varepsilon)$ and $r_{p_i} = \min(1, r_{p_i}/r_N)$, where ε is the tuning factor adjusting the sensitivity of the height feature [i.e., it ensures that all points having relative heights greater than ε provide a minimum data discrepancy term in (1)], and r_N is the radius size used to extract local neighbors v_c . η is the relative importance factor for the feature r_{p_i} , i.e., η defines the relative importance of r_{p_i}

with respect to h_{p_i} . The value of η less than 1 (e.g., $\eta = 0.5$ used in this work) means more importance is given to the height feature. This is reasonable since r_{p_i} essentially depicts the degree of planarity and is lower for both building roofs and flat ground surfaces. Thus, to differentiate between points on a flat planar ground segment, e.g., parking area and on the flat roof, their heights become the only discriminative factor in the labeling process and, therefore, should be given relatively higher importance. Moreover, planar objects with higher height are more probable to be part of the building structure than to the ground surface.

D. Approximation of Terrain Height

The remaining point cloud \mathbf{P} mentioned in Section III-C may contain roof points belonging to buildings for which no or very few facade points are available. To adopt the above energy-based formulation, the height of the underlying terrain surface required in (2) is approximated by fitting a cubic polynomial surface to the nonbuilding/ground points via the robust LAR method. Ground points are extracted via successive reduction of nonground points in the remaining point cloud \mathbf{P} . This is done by adopting the following sequence of steps.

- Determine the local height difference of each point by taking the difference between the maximum and minimum height of points among its neighbors.
- Identify those points whose local height jump is greater than 5 m. These higher jump points are referred to as transition points.
- Cluster these transition points and, for each transition cluster having at least ten points, begin a region-growing procedure (similar to that explained in Section III-B).
- Probable ground points are then extracted by removing all the grown regions from the set of remaining points \mathbf{P} .

It is worth mentioning here that the grown regions can also be incorporated into the set of extracted roof points, as depicted in [36]. However, due to gaps in the data and localization errors of TomoSAR, it is still possible that few buildings remain undetected. Formulating the problem into the energy minimization framework helps us to detect these buildings.

E. Minimization via Graph Cuts

The above energy formulation in (1) is solved (minimized) via graph-cut-based optimization library using the $\alpha\beta$ -swap move algorithm [34], [35], [37], [38]. The minimum energy corresponds to the labeling l such that higher planar points are detected as building roof points. Combining them with the set of roof points extracted in Section III-B via facade information completes the extraction procedure.

IV. BUILDING SHAPE/FOOTPRINT RECONSTRUCTION

A. Segmentation Into Individual Buildings

The extracted building points are segmented such that each cluster represents points from an individual building. This is done by means of density connectivity [39], i.e., two points

are considered to be directly density connected to each other if one point is in the neighborhood vicinity of the other point. If the two points are not directly connected to each other, still, they can be density connected to each other if there is a chain of points between them such that they are all directly density connected. Thus, starting from a point, all points that are density connected to each other are clustered into a single group representing an individual building. These clustered points are then removed, and the procedure is repeated for the remaining points until all the points are assigned to a particular cluster.

B. Coarse Building Footprint

The reconstruction of building shapes is initially obtained by employing alpha shapes (or α -shape, i.e., generalization of convex hull) around each segmented building [40]. This results in vertices that describe the coarse 2-D polygonal boundary of the building footprint. The reconstructed shape depends on a particular value of α , which has to be carefully chosen since it controls the model complexity. For instance, an overlarge α could make it difficult to follow concave polygonal shapes, e.g., an L-shaped building. In [41], it is recommended that α be chosen as twice the mean Euclidean point distance among the building roof points to produce a reliable building shape, including smaller structures.

C. Refinement of Alpha Shape Vertices by Recursively Analyzing Angular Deviations

The alpha shape method provides good initial estimates of building outlines. However, due to lesser point density of TomoSAR points, alpha shapes only define the coarse outline of an individual building. The resulting polygons are therefore irregular and need to be refined/regularized.

If we denote $V_{alpha} = \{\mathbf{V}_{i=1,\dots,N}\}$ as a set containing N matrices of building polygons returned by the alpha shape algorithm and \mathbf{V}_j with ($j \in i$) as the matrix containing 2-D vertices of the initial alpha polygon of the j th building, then inspired by the work in [41] (although different), the recursive procedure provided in Table I is adopted to refine the coarse reconstructed building footprints returned by the alpha shape algorithm.

The procedure in Table I begins by computing the angular deviations at each vertex point of the alpha polygon as

$$\beta_j = \begin{cases} \theta_j, & \text{if } \theta_j \leq 90^\circ \\ 180 - \theta_j, & \text{if } \theta_j > 90^\circ \end{cases} \text{ with } \theta_j = \cos^{-1} \left(\frac{\mathbf{d}\mathbf{v}_j \cdot \mathbf{d}\mathbf{v}_{j+1}}{\|\mathbf{d}\mathbf{v}_j\| \|\mathbf{d}\mathbf{v}_{j+1}\|} \right) \quad (3)$$

where “ \cdot ” denotes the dot product, and $\mathbf{d}\mathbf{v}_j$ is the direction vector computed at each edge formed by connecting two consecutive vertices \mathbf{v}_j and \mathbf{v}_{j+1} of the polygon \mathbf{V}_{prev} (initialized to \mathbf{V}_j). Steps 4 and 5 ensure that all those vertices (or edges) having angular deviations less than the threshold θ_{ang} are removed. \mathbf{V}_{new} and \mathbf{V}_{prev} are then compared, and the process repeats itself if any vertex is removed in the current recursive iteration, i.e., \mathbf{V}_{new} and \mathbf{V}_{prev} do not contain the same number of elements. Finally, the process terminates when there is no further removal of vertices.

TABLE I
REFINEMENT OF ALPHA SHAPE POLYGONS

Given: Alpha shape polygon of j th building V_j .

1. Initialize: $V_{prev} = V_j$
2. **while** (true)
3. Compute angular deviation matrix β
4. Identify indices in β greater than θ_{ang}
5. Extract vertices that corresponds to the identified indices in V_{prev} and assign them to V_{new}
6. **if** number of elements in V_{prev} equals number of elements in V_{new}
7. **break**
8. **else**
9. $V_{prev} = V_{new}$
10. **end if**
11. **end loop**
12. $V_{final} = V_{new}$

D. Incorporating Reconstructed facades

To improve the geometrical accuracy of the footprints, the reconstructed facades are fused with the refined building polygons. For this purpose, the facade associated to each refined building polygon is categorized into the following two types:

- type-I facade: facade fully or partly inside the refined polygon;
- type-II facade: facade lying completely outside but associated to the refined polygon.

The aforementioned two facade types are fused with the refined building polygon in slightly different manners, as will be explained later.

a) Identification of legitimate facade-polygon pair (facade-polygon pairing): To achieve fusion of reconstructed facades with the refined building polygons, the foremost task is to identify the association of each facade to its respective building polygon.

Identification of type-I facades is easily achieved by checking if the endpoints of the reconstructed facades lie inside the polygon. Thus, if both or at least one of the facade endpoints lie inside the building polygon, it is categorized as a type-I facade.

To identify facades of type II, the following procedure is adopted.

- 1) First, the midpoint of the reconstructed facade is computed, and two points are chosen in opposite directions orthogonal to the reconstructed facade at a distance d from the midpoint.
- 2) Compute intersections of $line_1$ and $line_2$ with all the building polygons. Here, $line_1$ denotes the line segment formed by connecting the midpoint to one of the chosen points, and similarly, $line_2$ is the line segment formed by connecting the midpoint to the other opposite point.
- 3) If there exists an intersection of $line_1$ or $line_2$ with any of the building polygons, the facade is assigned to the polygon with which the intersection occurs. In case there

are more than one line–polygon intersections or both $line_1$ and $line_2$ intersect with different polygons, the facade is assigned to the polygon having the intersection point nearest to it.

Implementation-wise, steps 1–3 are performed in a recursive manner. That is, d is initialized to 1 m, and steps 1–3 are carried out. If there exists no line–polygon intersection (i.e., the facade is not assigned to any polygon), the procedure repeats itself, but this time, d is incremented by 1 m. The recursion stops if either the facade is assigned to any polygon or distance d exceeds a certain threshold, which is set to a fixed value of 20 m in this work. Thus, a facade is only associated/paired to any building polygon if it lies at a distance less than 20 m; otherwise, it is regarded to have no polygon associated to it (i.e., categorized to case 1).

b) Fusion of reconstructed facades: Similar to earlier notation, let us denote $V_{refined} = \{\tilde{V}_{i=1,\dots,N}\}$ as a set containing N matrices of refined building polygons, with $\tilde{V}_j = \{\mathbf{v}_{k=1,\dots,m}\}$ ($j \in i$) being the matrix containing 2-D vertices of the j th refined polygon having m vertices and $f_{r=1,\dots,s}$ being the corresponding s number of (paired) reconstructed facades. Now, if the building polygon, formed by connecting vertices of \tilde{V}_j , is interpreted as a graph, then we may define a path P_t for any particular facade f_t as a path consisting of a polygonal chain of vertices that connect two points \mathbf{v}_{ta} and \mathbf{v}_{tb} lying on the graph/polygon. That is, the polygonal segment comprising of all the points of polygon \tilde{V}_j within the interval $[\mathbf{v}_{ta}, \mathbf{v}_{tb}]$ defines path P_t . \mathbf{v}_{ta} and \mathbf{v}_{tb} denote points on the building polygon, which are nearest to the two endpoints of the particular reconstructed facade f_t . Since, in our case, the polygon is nonintersecting (or simple), it thus renders only two distinct paths to exist, which are referred to as $P_{t_shortest}$ and $P_{t_longest}$ [see Fig. 6(b)]. If the path length of P_t is denoted as $PL_{\mathbf{v}_{ta}\mathbf{v}_{tb}}$, then P_t is $P_{t_shortest}$ only if $PL_{\mathbf{v}_{ta}\mathbf{v}_{tb}} < TL/2$, where TL is the total path length (i.e., perimeter) of the polygon.

$P_{t_shortest}$ is further classified into two types: positive path P^+ and negative path P^- . If we denote the set of points on the polygons that are nearest to the endpoints of all facades of the same building other than f_t as \mathbf{K}_t (i.e., \mathbf{K}_t contains points nearest to the endpoints of facades $f_{r'=1,\dots,t-1,t+1,\dots,s}$ such that $r' \cup t = s$), then $P_{t_shortest}$ of the reconstructed facade f_t is defined to be P^+ if the set of points $\mathbf{P}_{t_shortest}$ belonging to path $P_{t_shortest}$ does not contain any element of \mathbf{K}_t , i.e., $\mathbf{P}_{t_shortest} \cap \mathbf{K}_t = \emptyset$. Thus, all facades whose paths are identified as positives are incorporated in the fusion process, whereas facades having negative paths are not considered any further. A sample illustration of the concept of a positive path is presented in Fig. 6(c).

Table II provides the complete procedure to incorporate facades of both types with the refined building polygon.

Fig. 7 practically illustrates the procedure of fusing both types of facades. Note that there lie some differences in the computation of \mathbf{v}'_{ta} and \mathbf{v}'_{tb} for type-I and type-II facades (steps 8, 9, and 17 in Table II). The reason for this is due to the fact that point density on building roofs is quite varying and can contain gaps in between. This could lead to *under reconstruct*

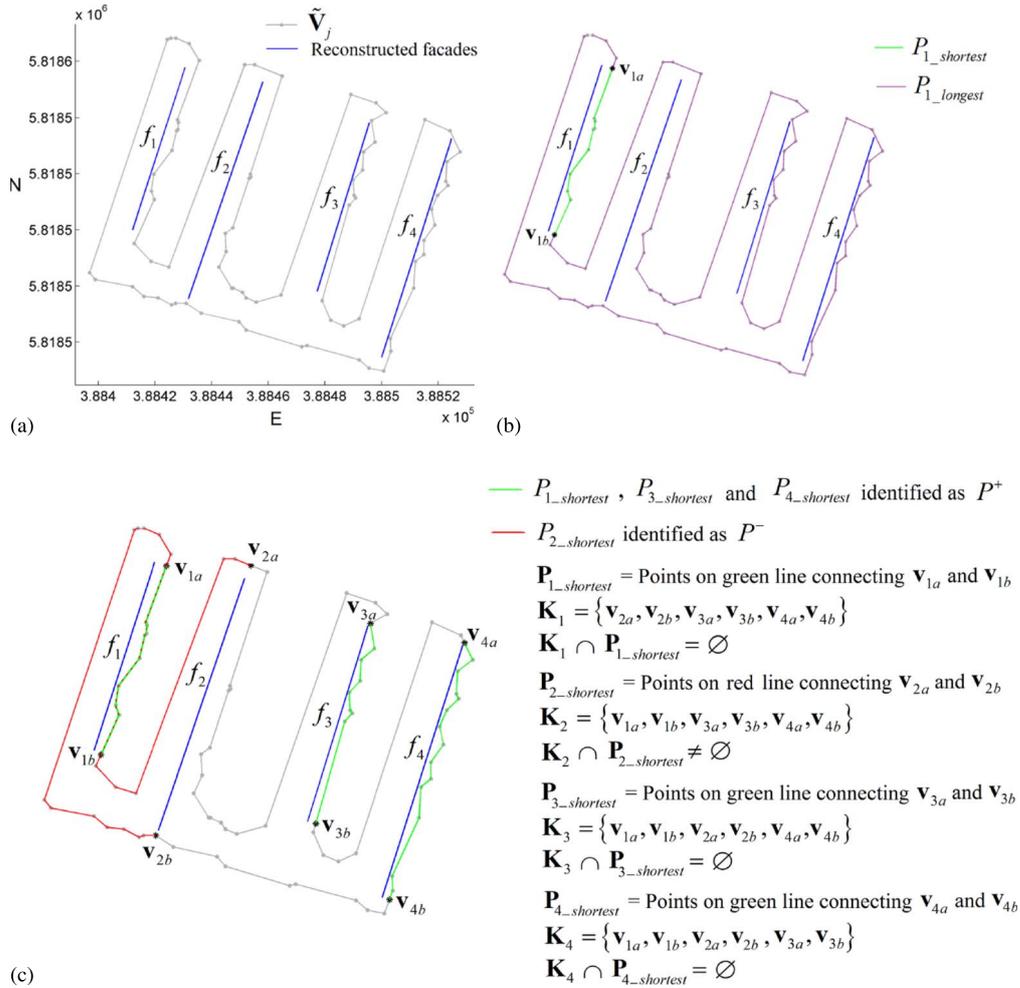


Fig. 6. Procedure describing the concept of $P_{t_shortest}$ and P^+ . (a) Refined polygon of one particular j th building $\tilde{\mathbf{V}}_j$ having vertices $\mathbf{v}_{k=1,\dots,m}$ with four reconstructed facades $f_1, f_2, f_3,$ and f_4 that are to be incorporated. (b) Sample illustration of the concept of shortest and longest paths associated for a particular facade f_1 . \mathbf{v}_{1a} and \mathbf{v}_{1b} denote the closest points on the polygon/graph $\tilde{\mathbf{V}}_j$ to the two endpoints of facade f_1 , respectively. (c) Sample illustration of the concept of positive path P^+ . $P_{2_shortest}$ is identified as P^- as there exist points in \mathbf{K}_2 that are also present in $\mathbf{P}_{2_shortest}$.

of the building footprint, i.e., part of the building roof region could not be reconstructed due to unavailability of points. The presence of type-II facades implicitly validates this plausible phenomenon, and therefore, fusion of refined polygons by fully incorporating the reconstructed facades (of type II only) results in improved overall accuracy of reconstruction. Doing the same for type-I facades, on the other hand, may affect the footprint polygon in the presence of facades belonging to inner building structures. Thus, only the orientation of type-I facades is essentially incorporated by the proposed procedure (steps 8 and 9 in Table II). In addition, steps 12–15 in Table II also pose a condition $C1$ for type-I facades such that they do not take part in the fusion process if the change in area of the polygon after incorporating the particular facade is greater than the certain fraction a_f (fixed to 0.15 in this work) of the previous polygonal area. Thus, using condition $C1$ together with the method of type-I facade fusion, it is ensured that facades belonging to the inner structures of the building do not interfere during the fusion procedure, or in other words, only facades that are exterior and define the building outlines are utilized.

E. Identification of Rectilinear Footprints

The next step in the reconstruction procedure is to identify if the building is composed of two or more than two dominant directions. If the building polygon is composed of only two dominant directions orthogonal to each other, rectilinear constraints are added to derive geometrically correct and better visually looking building shapes.

a) *Estimation of principal direction:* The decision of identifying a rectilinear building is based on its estimated dominant/principal direction. For this purpose, building polygons belonging to case 2 (i.e., having one or more reconstructed facades associated to each polygon), the principal direction is easily determined by assigning it to the direction vector computed by subtracting the endpoints of the longest reconstructed facade paired to it. For case-3 building polygons, the principal direction is directly estimated from the polygon itself. Since no facade is associated to them, a weighted method based on polygonal edge lengths is employed to estimate the two orthogonal principal directions of the building. The basic idea is to give weight to each edge of the polygon according to its

TABLE II
PROCEDURE TO FUSE FACADES OF BOTH TYPES

Given: Refined polygon of j th building \tilde{V}_j & s reconstructed façades $f_{r=1, \dots, s}$ belonging to the same j th building

1. $V_{prev} = \tilde{V}_j$
2. **for** $t = 1$ to s (i.e., total number of reconstructed façades associated to this building polygon)
3. Determine points v_{ia} and v_{ib} by computing points on the polygon that are nearest to the two endpoints of the façade f_t
4. Build a matrix K_t specific to the façade f_t that contains points on the polygon that are nearest to the endpoints of all other façades
5. Determine $P_{t_shortest}$
6. **if** $P_{t_shortest}$ is also P^*
7. **if** façade is identified as type I
8. Compute midpoint of the line segment L_t formed by connecting the two points v_{ia} and v_{ib}
9. Determine two new points v'_{ia} and v'_{ib} by projecting v_{ia} and v_{ib} onto another line parallel to the respective façade f_t but passing through the midpoint of L_t (see Figure 7(c))
10. Replace vertices of V_{prev} within $P_{t_shortest}$ (i.e., points in $P_{t_shortest}$) by v'_{ia} and v'_{ib} and store the result in the matrix V_{new}
11. Compute area of old V_{prev} and new V_{new} polygons denoted as $A_{V_{prev}}$ and $A_{V_{new}}$ respectively
12. Check the condition C1: $|A_{V_{prev}} - A_{V_{new}}| \leq \alpha_f A_{V_{prev}}$
13. **if** C1 is satisfied
14. $V_{prev} = V_{new}$
15. **end if**
16. **else**
17. Assign v'_{ia} and v'_{ib} to the respective endpoints of the reconstructed façade (see Figure 7(c))
18. Replace vertices of V_{prev} within $P_{t_shortest}$ (i.e., points in $P_{t_shortest}$) by v'_{ia} and v'_{ib} & store the result in the matrix V_{new}
19. $V_{prev} = V_{new}$
20. **end**
21. **end if**
22. **end loop**

relative length (with respect to total polygon length) and the angular deviation it makes with a particular direction vector $\mathbf{d}\mathbf{v}_q$. $\mathbf{d}\mathbf{v}_q$ is a candidate for one of the two orthogonal principal directions and is rotated within a certain interval to minimize the following objective function [42]:

$$PD = \sum_{i=1}^n g_1(l_i) g_2(\varphi_i(\beta_i, \phi_{\mathbf{d}\mathbf{v}_q})) \quad (4)$$

where n is the total number of vertices of the polygon, and $\beta_i (0 \leq \beta_i \leq 90^\circ)$ is the angular deviation of each edge l_i with respect to the direction vector $\mathbf{d}\mathbf{v}_q$. β_i is computed similar to (3) with the difference that the two direction vectors are $\mathbf{d}\mathbf{v}_j$ and $\mathbf{d}\mathbf{v}_q$ instead of direction vectors of consecutive edges $\mathbf{d}\mathbf{v}_j$ and $\mathbf{d}\mathbf{v}_{j+1}$. $\phi_{\mathbf{d}\mathbf{v}_q}$ is the anticlockwise rotation angle that $\mathbf{d}\mathbf{v}_q$ makes with the unrotated coordinate system. $\varphi_i(\cdot)$ is a function that maps the angular deviations β_i to one of the two orthogonal directions (or axes) as defined by direction vector $\mathbf{d}\mathbf{v}_q$ and its corresponding normal vector. It is computed as

$$\varphi_i(\beta_i, \phi_{\mathbf{d}\mathbf{v}_q}) = \begin{cases} \beta_i, & \text{if } \beta_i \leq 45^\circ \\ 90 - \beta_i, & \text{if } \beta_i > 45^\circ. \end{cases} \quad (5)$$

Both $g_1(\cdot)$ and $g_2(\cdot)$ are the weighting functions. $g_1(\cdot)$ assigns weight to each edge based on its relative length with

respect to the overall length of the polygon edges. It is constructed such that edges with longer lengths contribute less in (4) as compared with shorter edge lengths. The following linear function is used to describe $g_1(\cdot)$:

$$g_1(l_i) = 1 - \frac{l_i}{\sum_{i=1}^n l_i}. \quad (6)$$

Similarly, $g_2(\cdot)$ assign weights to each edge based on its φ_i value. The assignment of weight is directly proportional to φ_i , i.e., lower weight is given to an edge with lower φ_i , inferring that edges close to one of the two orthogonal directions are given less weight as compared with those that are deviating. Since the span of φ_i for each edge is defined to be within the interval $[0, 45^\circ]$, $g_2(\cdot)$ is thus computed by adopting the following linear function:

$$g_2(\varphi_i(\beta_i, \phi_{\mathbf{d}\mathbf{v}_q})) = \frac{\varphi_i}{45}. \quad (7)$$

The solution of (4) is obtained by rotating $\mathbf{d}\mathbf{v}_q$ within the interval $[0, 90^\circ]$. An optimum (or minimum) $\hat{\phi}_{\mathbf{d}\mathbf{v}_q}$ is found by comparing PD for each $\hat{\phi}_{\mathbf{d}\mathbf{v}_q}$ value. The direction vector and its corresponding normal vector associated to the optimum (or minimum) $\hat{\phi}_{\mathbf{d}\mathbf{v}_q} (= \arg \min_{\phi_{\mathbf{d}\mathbf{v}_q}} (PD))$ thus describe the two orthogonal principal directions.

b) Identification procedure: Once the principal/dominant directions are determined, the following procedure is adopted for identification of rectilinear buildings.

- Determine angular difference $\beta_i (0 \leq \beta_i \leq 90^\circ)$ of all the edges of the polygon with respect to the dominant/principal directions.
- Compute the histogram of these angular differences.
- Find the edges whose angular differences are within the bin intervals $[0^\circ, 20^\circ]$ and $[70^\circ, 90^\circ]$.
- Identify the polygon to be rectilinear if the total sum of the lengths of these edges is more than a certain fraction L_f of the total polygonal length TL . (L_f is fixed to 0.75 in our work, i.e., 75% of the total polygonal length TL).

F. Addition of Rectilinear Constraints

Subsequently, rectilinear constraints are added to the identified building polygons to yield much better (visually appealing) geometric building shapes. The following steps are performed to obtain rectilinear building footprint.

- Classify each edge of the building polygon such that it belongs to one of the two orthogonal principal axes based on its angular deviation (i.e., an edge is associated to that principal direction with whom the angular difference is less).
- Merge all adjacent edges that share the same class, i.e., associated to the same principal direction.

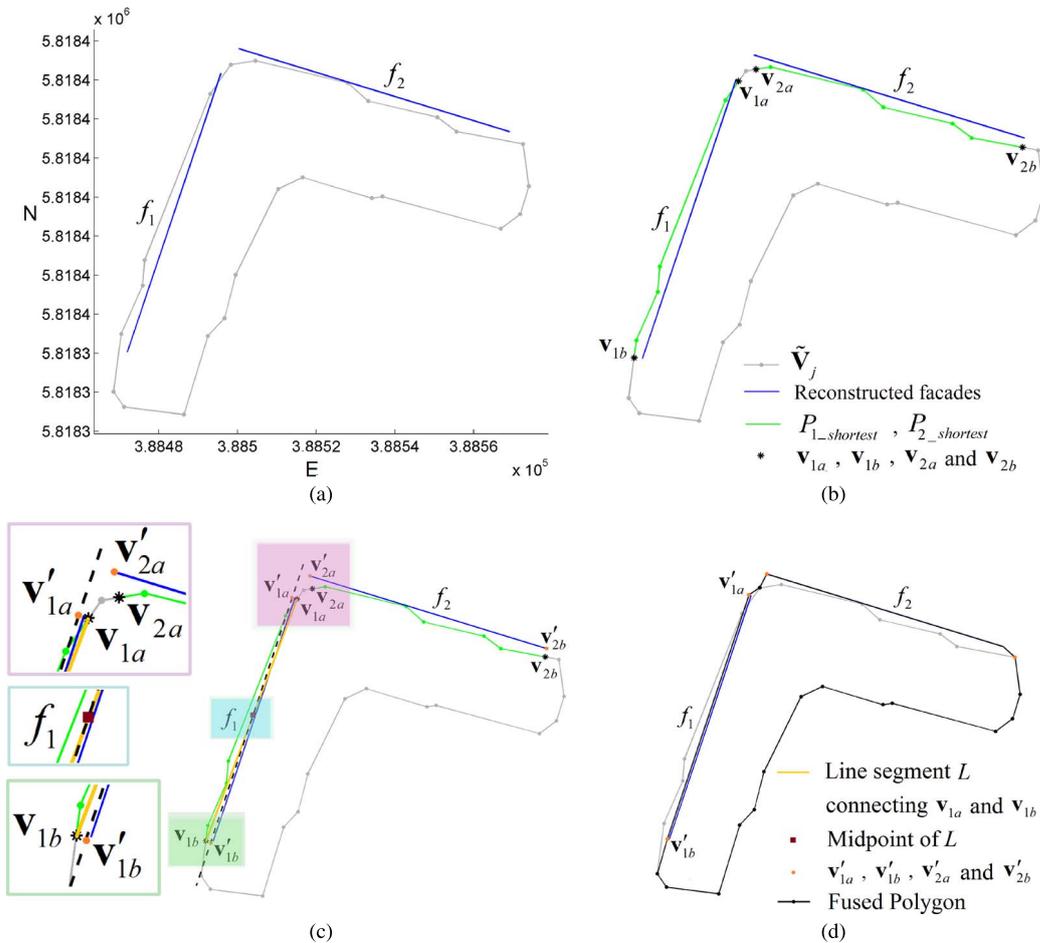


Fig. 7. Fusion of both type-I and type-II facades. (a) Refined polygon of one particular j th building \tilde{V}_j with two reconstructed facades f_1 and f_2 with types I and II, respectively. (b) Shortest paths (also P^+) determined for both facades. (c) Procedure of computing points v'_{ta} and v'_{tb} . Note that for the type-I facade, v'_{1a} and v'_{1b} are computed by projecting v_{1a} and v_{1b} onto the black dotted line, which is parallel to the reconstructed facade f_1 and passes through the midpoint of line segment L formed by connecting points v_{1a} and v_{1b} (steps 8 and 9 in Table II), whereas for the other facade having type II, v'_{2a} and v'_{2b} are simply the endpoints of facade f_2 (step 17 in Table II). All vertices of \tilde{V}_j on $P_{1_shortest}$ and $P_{2_shortest}$ (i.e., points in $\mathbf{P}_{1_shortest}$ and $\mathbf{P}_{2_shortest}$) are replaced by vertices v'_{1a} , v'_{1b} and v'_{2a} , v'_{2b} , respectively. (d) Resulting polygon in black after fusing facades f_1 and f_2 with types I and II, respectively.

- Apply rectilinear transformation to every merged polygonal edge by projecting it onto its corresponding principal axis/vector.
- Computing intersection (or vertex) points between the consecutive vertices.

Fig. 8 illustrates the mechanism of practically adding rectilinear constraints to the building polygon using the aforementioned procedure. Note that the depicted polygon contains no associated facade and, therefore, belongs to case 3. For case-2 polygons, the associated facades are first fused to the refined polygon, and if identified as a rectilinear, constraints are added to complete the reconstruction procedure.

V. EXPERIMENTAL RESULTS AND VALIDATION

A. Data Sets

To validate our approach, we tested the proposed algorithms on two different data sets. One is composed of TomoSAR point clouds generated from a stack of 25 TerraSAR-X high-resolution spotlight images covering approximately

($900 \times 600 \approx$) 0.54 km^2 area in the city of Las Vegas, USA. It contains 0.48 million points and consists of moderate-sized buildings with relatively simple geometry. To test the generality of the proposed algorithms, the other data set is composed of TomoSAR point clouds of complex building structures produced from a stack of 102 TerraSAR-X high-resolution spotlight images covering around ($1750 \times 900 \approx$) 1.5 km^2 area in the city of Berlin, Germany. The number of points in the Berlin data set is approximately 0.52 million. Both TomoSAR point cloud data sets are generated from images taken from an ascending orbit using the TomoGENESIS software developed in the German Aerospace Center (DLR) [43], [44].

B. Reference Footprints

For the area of interest in Las Vegas, the reference footprints for the Las Vegas data set were acquired from CyberCity3D [45]. These footprints are highly precise with positional accuracy up to $\pm 15 \text{ cm}$ and are generated using automated and

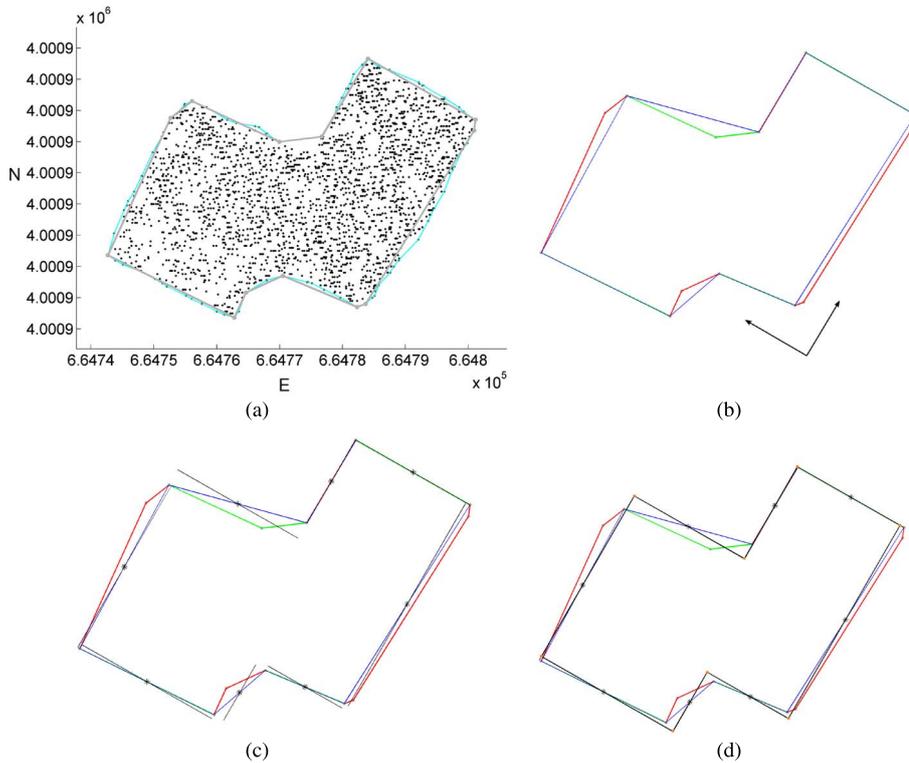


Fig. 8. Procedure to add rectilinear constraints to any one particular building polygon. (a) Black dots indicate the extracted building points, the cyan polygon represents the initial coarse outline obtained using the alpha shape algorithm, and the gray polygon depicts the refined/smoothed polygon to which rectilinear constraints are added. (b) Each edge of the refined gray polygon of (a) is segmented to belong to one of the two estimated dominant directions indicated by black arrows based on their angular deviation (i.e., an edge is associated to that principal direction with whom the angular difference is less). Segmented edges are shown in red and green. The blue polygon then depicts the reduced refined polygon of (a) by merging all adjacent edges that belong to the same principal axis. (c) Each edge of the blue polygon is then rotated/projected around its midpoint onto its corresponding principal axis. (d) Finally, the intersection (vertex) points among the adjacent edges of the projected blue polygon edges of (c) are computed for rectilinearization.

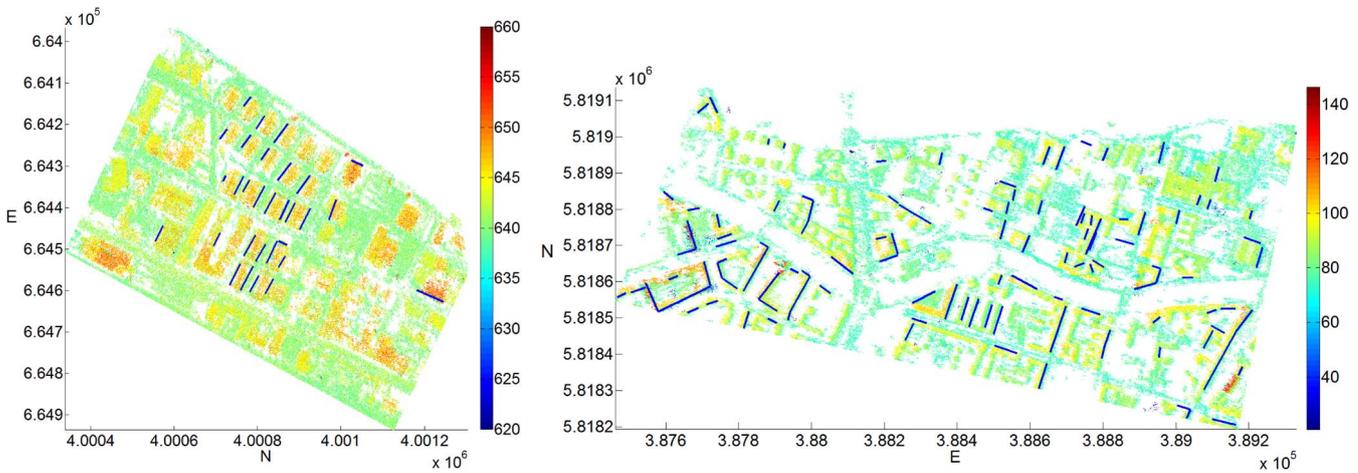


Fig. 9. Data sets. Top view of the 3-D TomoSAR points in UTM coordinates of the area of interest in (left) Las Vegas, USA and (right) Berlin, Germany. Blue lines depict the reconstructed facade segments (longer than 10 m). The height of TomoSAR points is color coded [unit: m].

semiautomated photogrammetry-based techniques with data source derived from aerial, oblique, or satellite stereo imagery.

For the Berlin data set, we compared our building extraction results to reference polygons downloaded from OpenStreetMap (OSM) [46]. Based on the concept of crowd sourcing, which involves crowd or community to effectively and efficiently fulfill a task at hand, OSM with around two million registered users (as of today and also rapidly growing) is considered to

be the most successful Volunteered Geographic Information (VGI) project [47], [48]. The OSM database contains multitude of building footprints represented as polygons with an ordered list of nodes/vertices (i.e., pairs of UTM or Latitude/Longitude coordinates according to the WGS 84 coordinate system) and is updated every day. The data are free to download and comes under the open license Open Data Commons Database License (ODbL). Since it is a VGI project, the data quality

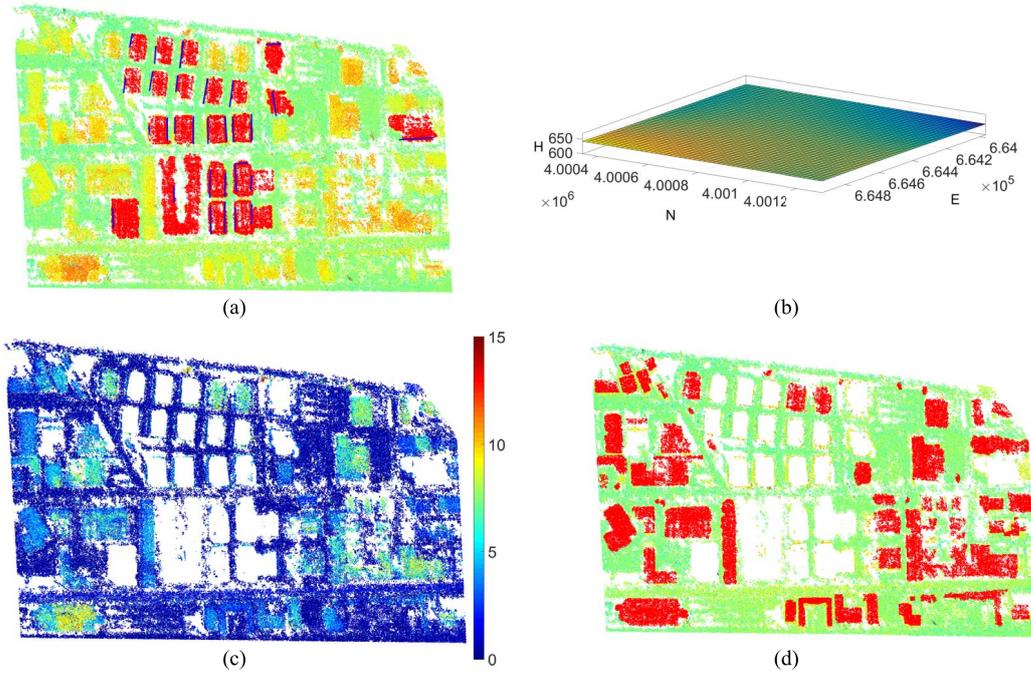


Fig. 10. Building extraction procedure. (a) Red points depict building points extracted via utilizing reconstructed facades. (b) Approximated terrain surface from the remaining set of points via successive reduction of building points and polynomial surface fitting, as explained in Section III-D. (c) Differential height computed via subtracting the terrain height from the height of each point. (d) Further building points (in red) are then extracted using the energy minimization framework with tuning parameters: $\varepsilon = 10$ m, $r_N = 5$ m.

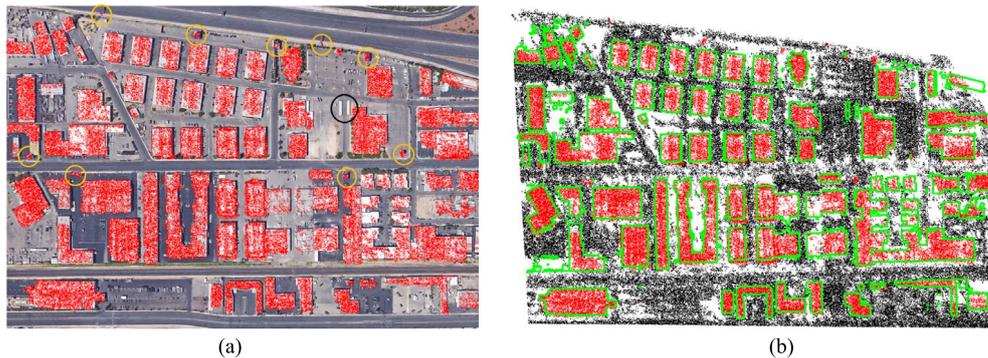


Fig. 11. Results of building extraction: Las Vegas. (a) Extracted building points in red from Fig. 10(a) and (d) are overlaid onto the optical image (Google) of the area of interest. Yellow circles indicate extracted points originating from sources such as advertisement boards and monuments. Large black circle encloses two buildings that remain undetected due to lower relative heights. (b) Finally, reference footprints (in green) overlaid onto the extracted building points. Red points are building points, whereas black points are nonbuilding points.

may vary from region to region. To this end, the first investigations regarding OSM data quality were carried out for roads [48], followed by an assessment of other attributes present in the database, e.g., lines [50] and polygonal objects [51]. Recently, the building footprints have also been evaluated for their completeness [52] and correctness [49] for various cities in Germany. The analysis of OSM data with surveying data sets reveals fairly precise positioning accuracy varying within 4 m [47], [49].

C. Results: Extraction of Building Points

Fig. 9 shows the result of applying facade reconstruction procedure over both data sets. Seed points are selected from each reconstructed facade, and the region is grown using a

surface-normal-based similarity measure with $\theta_{normals}$ set to 15° . Fig. 10(a) shows the grown region using seeds from the reconstructed facades. Later, among remaining points, terrain is approximated, and building points are extracted by adopting an energy minimization procedure, as previously explained in Section III-C [see Fig. 10(b)–(d)]. Figs. 11 and 12 show the final extracted building points in both data sets overlaid onto the optical images.

D. Validation: Extraction of Building Points

Figs. 11(b) and 12(b) show the reference polygons overlaid onto the extracted building points of the test area in Las Vegas and Berlin, respectively. It can be visually seen that the extracted building points fit very well to these reference polygons.



Fig. 12. Results of building extraction: Berlin. (a) Extracted roof points in red are overlaid onto the optical image (Google) of the area of interest. (b) Red and black points depict building and nonbuilding points, respectively. The overlaid green polygons are reference buildings downloaded from OSM [46]. Blue polygons are manually extracted buildings not present in OSM data. Gray polygons are newly constructed buildings that are not present in our data set, whereas magenta polygons are buildings that do not actually exist but are present in OSM data. Both gray and magenta polygons are not included in the evaluation.

For the Berlin data set, we found out that few buildings are missing in the OSM data set, and therefore, by analyzing the detected buildings from TomoSAR point clouds and validating using optical data, we completed few missing buildings in the OSM data set; polygons are depicted as blue polygons in Fig. 12(b). The performance of the (detection) extraction procedure in both data sets is then assessed by employing the evaluation metrics [53], [54] given in

$$\left. \begin{aligned} \text{Completeness (\%)} : comp &= 100 \times \left(\frac{TP}{TP+FN} \right) \\ \text{Correctness (\%)} : corr &= 100 \times \left(\frac{TP}{TP+FP} \right) \\ \text{Quality (\%)} : Q &= \frac{comp \times corr}{comp + corr - comp \times corr} = \frac{TP}{TP+FP+FN} \end{aligned} \right\} \quad (8)$$

where

- TP (true positives) represents the number of detected building points inside the reference building polygons;
- FN (false negatives) represents the number of detected nonbuilding points inside the reference building polygons;
- FP (false positives) represents the number of detected building points outside the reference building polygons; and
- TN (true negatives) represents the number of detected nonbuilding points outside the reference building polygons.

TABLE III
EVALUATION STATISTICS OF THE DETECTION ALGORITHM

Datasets	Las Vegas	Berlin
TP	168065	295367
FN	10874	16269
FP	26966	50834
TN	228551	154420
Completeness (%) - $comp$	93.923	94.779
Correctness (%) - $corr$	86.173	85.316
Quality (%) - Q	81.622	81.487

The aforementioned metrics assess the overall performance of the building extraction algorithm. Completeness tells up to what percentage the algorithm has detected the roof points, whereas correctness provides a measure of correct classification. Quality combines both completeness and correctness metrics to provide an overall measure of the algorithm performance. Results of the evaluation statistics for both data sets are provided in Table III.

In both data sets, all of the buildings having relative heights (with respect to ground) of more than 5 m are detected by the extraction procedure. However, an example of two smaller buildings, for which no points could be extracted, is highlighted by a black circle in Fig. 11(a). These buildings have a footprint

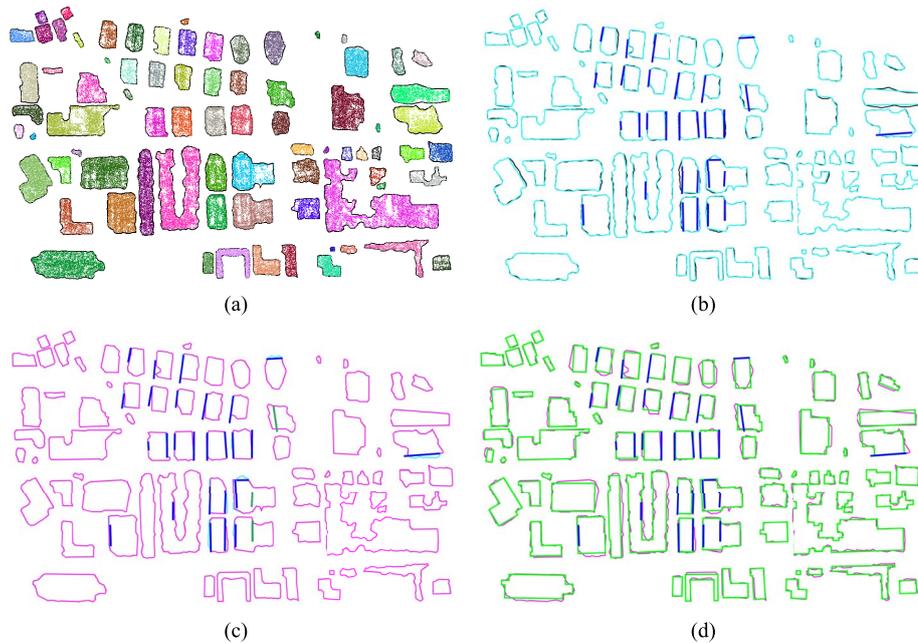


Fig. 13. Reconstruction results: Las Vegas. (a) Clustered (or segmented) building points. Black polygons V_{alpha} surrounding individual segmented building points are the initial coarse boundary/outline obtained using the alpha shape algorithm. (b) Refined (cyan) polygons $V_{refined}$ obtained after applying a recursive angular deviation approach together with 2-D reconstructed facades (depicted in blue, are overlaid onto alpha shape polygons). (c) facades are then incorporated into the refined polygons from (b) depicted in magenta, symbolized as $V_{facade\ fused}$. facades either identified as P^- or having condition $C1$ in Table II not satisfied are depicted in green and are not utilized during the *facade-Polygon* fusion process. (d) Final rectilinearized polygons V_{final} obtained after adding rectilinear constraints.

area of approximately $(28 \times 5 =) 140 \text{ m}^2$ but possess relative heights of only 2 m. Since during the extraction procedure, seed points are chosen based on local height jumps of 5 m, no seed point could thus be chosen for these buildings. As a consequence, they remain undetected.

In terms of false alarms, it is worth mentioning that it might happen that points belonging to some small vertical structures on ground (e.g., advertisement boards and monuments) also get detected during our extraction procedure. Few examples of such cases are also highlighted by yellow circles in Fig. 11(a). The reason for this occurrence of false positives is our implicit definition (or assumption) pertaining to buildings, i.e., higher points with a higher degree of planarity are detected as buildings by the extraction procedure.

E. Results: Reconstruction of Building Footprints

Extracted building points are then spatially segmented such that each cluster represents an individual building. Figs. 13(a) and 14(a) depict the result of spatially clustering points into individual buildings in the Las Vegas and Berlin data sets, respectively. The initial coarse outline of each cluster is then determined using the alpha shape algorithm. Black polygons in Figs. 13(a) and 14(a) surrounding each individual segmented building cluster depict its corresponding alpha polygon.

Refinement of the initial coarse alpha vertices is then carried out by computing the angular deviation at each vertex point. The threshold value $\theta_{ang} = 20^\circ$ is used, which consequently removes all vertices having angular deviations less than 20° from their adjacent neighboring vertices. Refined or smoothed alpha polygons are then fused with the reconstructed facades. Later,

rectilinear constraints are added to the building polygons that are identified to be rectilinear. Figs. 13(b)–(d) and 14(b)–(d) depict the results of building footprint reconstruction on the Las Vegas and Berlin data sets, respectively.

F. Validation: Reconstruction of Building Footprints

To evaluate the reconstruction results, we rasterized both the reconstructed and reference polygonal footprints onto an image with pixel resolution of 1 m (i.e., one pixel corresponds to 1 m^2 spatial area). A difference image created by subtracting the reconstructed footprint image from the reference footprint image is then used to compute the commission and omission errors as follows:

$$\begin{aligned} \text{Commission error}(\%) &= \frac{F_N}{A_{ref}} \times 100 \\ \text{Omission error}(\%) &= \frac{F_P}{A_{ref}} \times 100 \end{aligned} \quad (9)$$

where A_{ref} is the area of the reference polygons, whereas F_N and F_P are the number of pixels in the difference image having values of -1 and 1 , respectively.

Fig. 15 presents the common (top row) and difference (bottom row) images. The red pixels in the difference images indicate the building regions that are not reconstructed by the proposed algorithm contributing to the omission errors, whereas blue pixels are over reconstructed regions, i.e., pixels not part of the reference footprint image but present in the reconstructed image.

Table IV lists the commission and omission errors obtained for the reconstructed footprints.

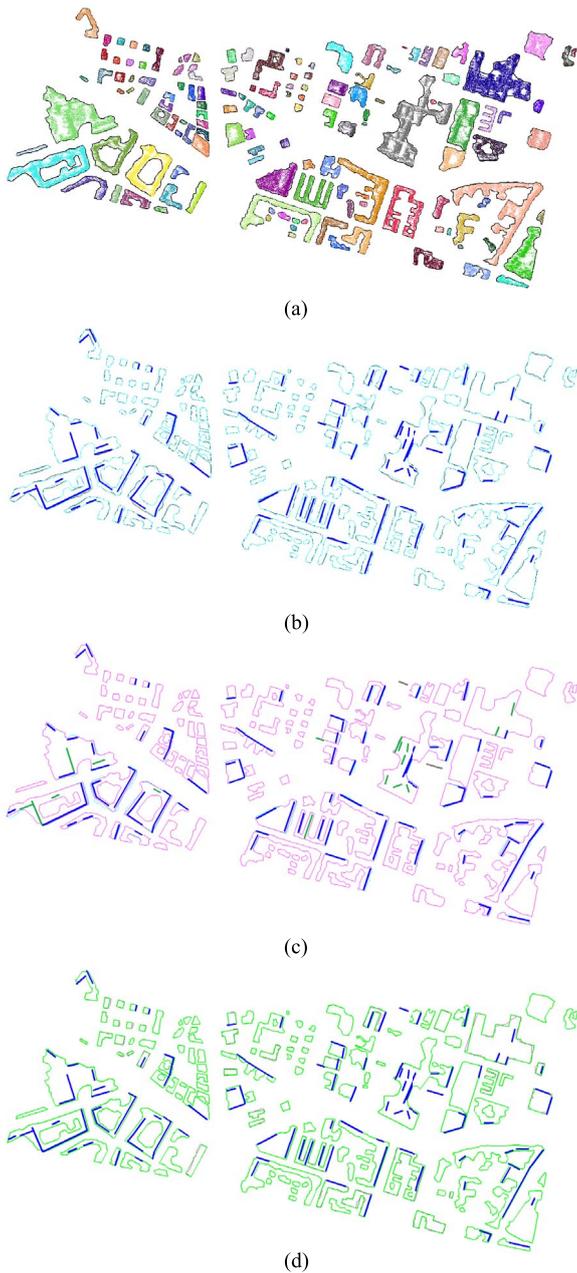


Fig. 14. Reconstruction results: Berlin. (a) Clustered (or segmented) building points. Black polygons V_{α} surrounding individual segmented building points are the initial coarse boundary/outline obtained using the alpha shape algorithm. (b) Refined (cyan) polygons $V_{refined}$ obtained after applying a recursive angular deviation approach together with 2-D reconstructed facades (depicted in blue, are overlaid onto alpha shape polygons). (c) facades are then incorporated into the refined polygons from (b) depicted in magenta, symbolized as $V_{facade\ fused}$. facades either identified as P^- or having condition $C1$ in Table II not satisfied are depicted in green, whereas facades not associated to any building polygon (i.e., case 1) are depicted in gray. Both red and gray facades are not utilized during the *facade-Polygon* fusion process. (d) Final rectilinearized polygons V_{final} obtained after adding rectilinear constraints.

Hypothetically, the reconstruction results will be improved with higher density of TomoSAR points because more points would be available for parameter estimation. Numerical experiments also demonstrated that reconstruction accuracy is better for buildings with higher density of roof points. For low-density roof regions, the reconstruction accuracy is, however, restricted

by the number of available points, which consequently reduces omission errors. A further improved-model-based approach might be helpful in this regard.

Additionally, the reconstruction errors between the final and coarse 2-D topologies (polygons) are also varying. Thus, there is a tradeoff in achieving geometrically correct footprints while simultaneously keeping the commission and omission errors in control. As evident, the best tradeoff can be obtained by incorporating facades to the coarser building polygons. However, more visually appealing results are produced by introducing rectangular constraints to the rectilinear buildings.

Finally, in Fig. 16, we present the final reconstructed building shapes/footprints in 3-D. As depicted in [25], the shown reconstructed building model can be used to refine the elevation estimates of the raw TomoSAR points. Moreover, with known deformation estimates of the scatterers, such a model can also lead to the reconstruction of dynamic city models [55] that could potentially be used to monitor and visualize the dynamics of urban infrastructure in very high level of details.

G. Discussion on Parameter Selection

Although the parameter values have been tuned according to the investigated scenes, some parameters are not strictly related to the scene under consideration and, therefore, can be easily set *a priori* by using some general rules or constraints. Moreover, most of the parameters involved have a clear physical meaning associated to them, which aids the user in incorporating it as prior knowledge on a particular scene. Based on aforementioned guidelines provided in the previous two sections, in the following, some more insights regarding the selection of parameters and their effects on the results are discussed in detail.

a) *Extraction of building points*: Building points are extracted by following a two-step procedure: First, the building points are extracted using a region-growing procedure. Later, among the remaining points, an energy minimization formulation is adopted to extract points belonging to lower height buildings for which no facade information was available. Parameters that control the extraction procedure are fac (controlling the minimum height constraint value h_{min}), threshold on angular difference $\theta_{normals}$, the tuning factor adjusting the sensitivity of the height feature ε , the radius size r_N , and the parametric value of local height jumps.

- fac adaptively sets the value of h_{min} for each seed point during the region-growing procedure. Theoretically, the value of fac should be close to 1. The reason for this is because the height of facades that are part of the building boundary is either lower than or equal to the height of the building roof (i.e., it is lower for polyhedral roofs and equal for flat roofs with the assumption of no overhanging roof structure). Thus, setting fac equal to 1 would set h_{min} (almost) equal to the height of the facade, and therefore, only those points that are planar and have height values greater than h_{min} would be clustered for a particular seed point during the region-growing procedure. Lower fac values, on the other hand, may cause flat (or planar) ground points to be added in the growing procedure. As

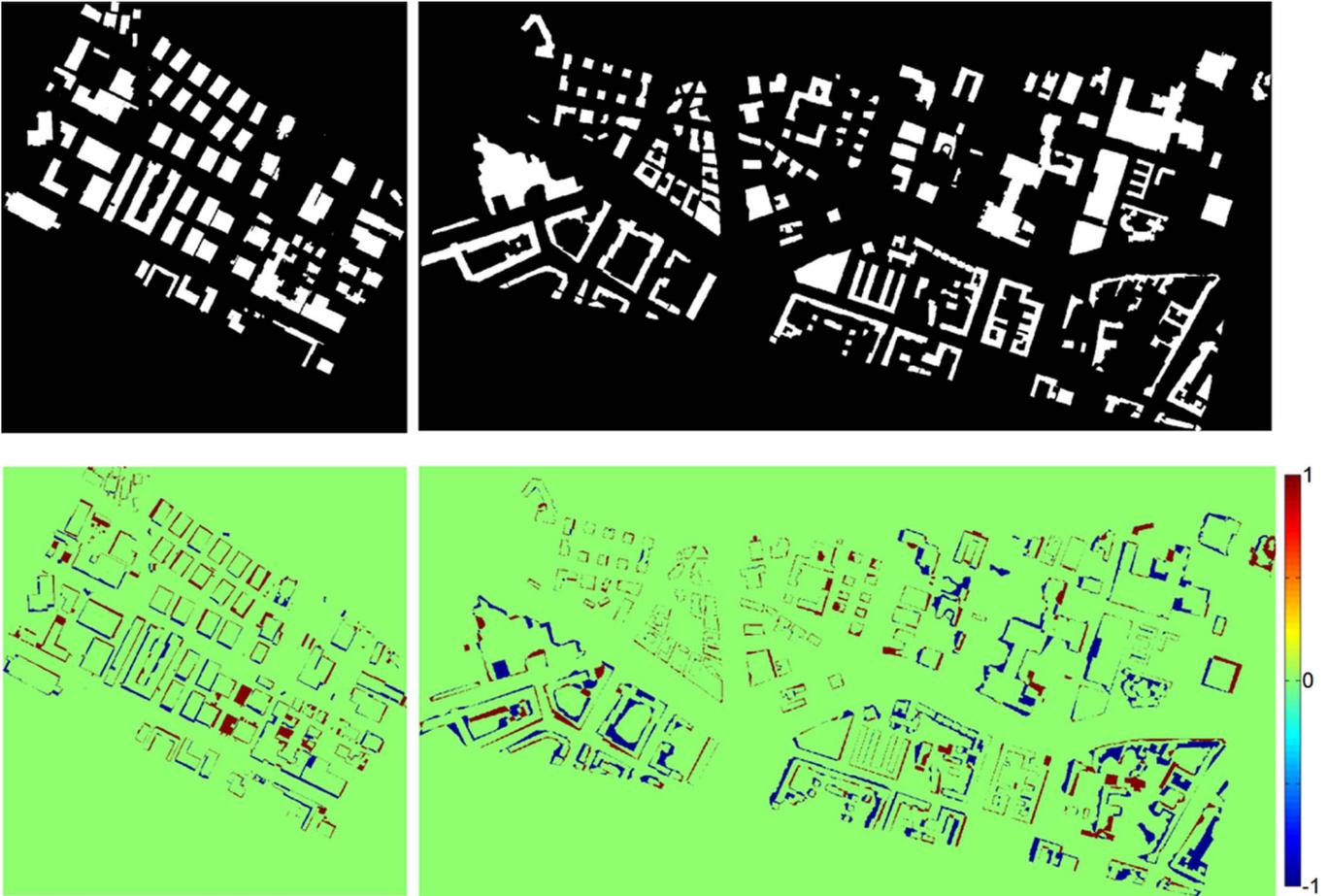


Fig. 15. Common (top row) and difference (bottom) images computed using the reference footprint image and final reconstructed footprints. The first row depicts the common regions, whereas the second row shows the difference image computed by subtracting the final reconstructed footprint image from the reference image. Note that green pixels indicate no difference between reconstructed and reference rastered images.

previously mentioned in Section I, since TomoSAR point clouds have a rather higher elevation error, the suggested range of values for fac based on empirical evaluation for different scenes is between 0.5 and 0.7. However, for other sensors giving better positioning accuracy, a higher fac value may be used.

- The angular difference threshold value $\theta_{normals}$ is mainly dependent on noise in the data. If the normals are robustly estimated, this parameter is more or less independent of both the scene and sensor configurations. The possible range of values for this parameter should be set between 10° and 20° .
- ε can be easily set based on *a priori* knowledge related to the average height of the buildings in the scene. For instance, with $\varepsilon = 10$ m (used in this work), it is ensured that any planar point above 10 m provides a minimum data discrepancy term in (1) and is therefore regarded as belonging to the building structure. ε is a scene- and data-dependent parameter. For data with relatively higher density and positioning accuracy, a lower value for ε may be used. For instance, a reasonable value for ε equal to the height of two floors (i.e., 6 m) has been proposed in [32].
- The role of radius size parameter r_N is twofold: First, it is used to compute local neighbors for feature computation,

TABLE IV
FOOTPRINT RECONSTRUCTION STATISTICS

Reconstructed footprints	V_{α}	$V_{refined}$	$V_{facalified}$	V_{final}
Errors(Las Vegas / Berlin)				
Commission error (%)	14.01 / 20.20	13.56 / 20.43	13.13 / 19.13	10.79 / 19.43
Omission error (%)	12.49 / 12.31	13.39 / 12.84	13.14 / 13.24	17.20 / 14.57

and second, it is used to spatially cluster extracted points such that they belong to individual buildings for subsequent reconstruction. Based on experimental evaluation over TomoSAR point clouds, a good choice for radius size between 5 and 10 m providing reasonable results for feature computation have been utilized [24], [56]. The use of a particular radius size, however, puts a constraint on the minimum distance between the buildings. That is, setting $r_N = 5$ m allows the algorithm to separate two buildings only if they are at least farther than 5 m from each other. Otherwise, the algorithm will merge them into one single cluster.

- Apart from these four parameters controlling the whole extraction procedure, the parametric value of local height

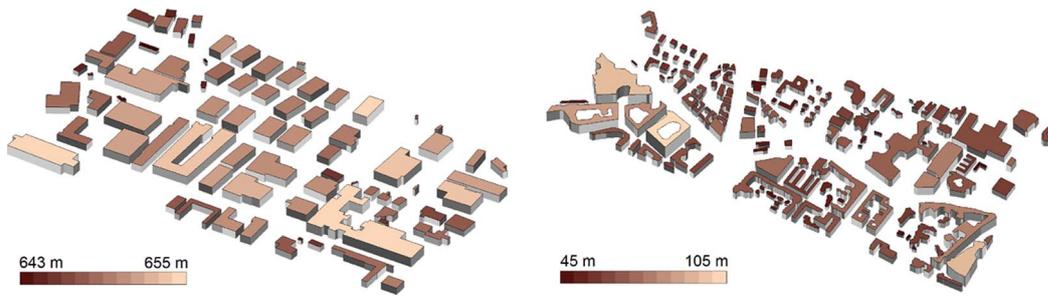


Fig. 16. Final reconstructed 3-D building footprints. (Left) Las Vegas data set. (Right) Berlin data set.

jumps that is used to select seed points while approximating the terrain surface also puts a constraint on minimum building height that could be detected. This parameter is merely data dependent and, therefore, has been empirically fixed in this work to 5 m. With a different sensor of higher resolution and positioning accuracy, this value may be lowered.

b) Reconstruction of building footprints: The whole reconstruction procedure is essentially governed by only two parameters: α and θ_{ang} .

- α provides good initial estimates of building outlines. However, the value of α effects the shape of the initial coarse polygon. Setting a larger α restricts in obtaining concave boundaries, whereas lower values may result in more smaller boundary polygons that are actually present. In addition, with smaller α , it is also possible that the outer and inner polygons share one (or more) common vertex and, hence, leads to improper footprint geometry. The use of one particular value for α may not be feasible for scenes containing arbitrary-sized buildings. Thus, to adaptively select an appropriate value of α , we initialize $\alpha = 5$ m (reasonable tradeoff for our data), which is recursively incremented by 1 m if resulting polygons share common vertices or the minimum area of any resulting polygon is less than 50 m².
- θ_{ang} is used to refine the initial (coarse) building boundary obtained using the alpha shape algorithm. $\theta_{ang} = 0$ results in no refinement or regularization, i.e., the original alpha polygons are returned. Setting too high a value for θ_{ang} may, however, result in overrefinement/smoothing.

Both α and θ_{ang} are very stable parameters for a variety of different input scenes with arbitrary building sizes. For α , the point density of the data plays an important role. For lower density data, the value of α should be higher, whereas for high-density data, lower α values may be used. For θ_{ang} , relatively stable values for data from different scenes (or sensors) range from 5° to 25°.

Although all of the remaining subsequent operations, including the *facade-Polygon* fusion, estimating the principal direction, and rectilinearization, are completely parameter free, one fixed threshold value, i.e., $L_f = 0.75$, providing a careful balance for our data, has been used to identify the rectilinear

polygons. With lower values of L_f , it is possible to wrongly classify more polygons as rectilinear, whereas higher values of L_f may increase the chance of a miss hit. This parameter also depends on the point density of the input data, and therefore, with higher density point clouds, values of L_f close to 1 may be employed.

To summarize, in total, eight parameters control the complete processing chain. Although the proposed approach is able to work with unstructured 3-D point clouds generated from any other sensor with a similar configuration (i.e., oblique geometry), the point density and positioning accuracy plays an important role in tuning these parameters. For spaceborne TomoSAR point clouds utilized in our experiments, Table V provides the values used in this work together with their possible (stable) range, which has been empirically evaluated for a variety of different input scenes. Additionally, categorization of each parameter according to its dependency type has also been presented.

VI. OUTLOOK AND CONCLUSION

In this paper, we have presented an automatic (parametric) approach that utilized unstructured spaceborne TomoSAR point clouds from one viewing angle only to detect and reconstruct 2-D/3-D building shapes/footprints. The approach is modular and allows for a robust detection of both tall and low buildings and, hence, is well suited for urban monitoring of larger areas from space. The approach is completely data driven and, therefore, imposes no restrictions on the shape of the building, i.e., any arbitrarily shaped footprint could be reconstructed. Moreover, the presented approach utilizes roof points in determining the complete shape of the buildings and, therefore, resolves problems, as mentioned in [24], related to the visibility of facades mainly facing toward the azimuth direction.

TABLE V
LIST OF PARAMETERS TOGETHER WITH THEIR VALUES USED AND DEPENDENCY TYPE

Module	Parameters	Dependency type	Value used (stable range)
Building extraction & segmentation	fac	Neither scene nor data dependent	0.55 (0.5 ~ 0.7)
	θ_{normal}	Data dependent	15° (10° ~ 20°)
	ε	Scene dependent	10 m (10 ~ 20m)
segmentation	r_s	Both scene and data dependent	5 m (5 ~ 10m for features computation)
	local height jump	Both scene and data dependent	5 m
Reconstruction	α	Data dependent	Initial value of 5 (3 ~ 15)
	θ_{ang}	Data dependent	20° (5° ~ 25°)
	L_f	Data dependent	0.75 (0.7 ~ 0.85)

In the following, several aspects in further extension/improvement of the proposed approach are mentioned.

- The approach presented in [24] is a much better option for detecting the shape of the building when dense points on the facades are available. However, there are cases occurring particularly for lower height buildings or buildings having a smoothed facade structure where no or only very few scatterers are detected. The approach proposed in this paper presents nice solutions to such cases by exploiting roof points. However, the availability of roof points is somewhat dependent on the structure and geometry (area) of the roofs. For the part of the roof that is smooth or completely invisible to the sensor, no points would be obtained, which may result in the underreconstruction of the footprint. Such situations can be rectified by using points from other viewing angles (e.g., use of TomoSAR points generated from both ascending and descending orbits) and/or by incorporating some prior model knowledge of the shape to obtain complete building footprints.
- Building points are extracted based on two assumptions, namely, planarity and higher relative heights. This assumption also enables the detection/reconstruction of other man-made vertical structures, e.g., advertisement boards and monuments. Imposing dimensional constraints on the reconstructed footprints may help in further distinguishing them from buildings.
- Unknown absolute height values of different reference points, which are chosen independently while processing individual VHR SAR data stacks, are a hindrance to the direct fusion of TomoSAR point clouds from multiple views. To this end, 2-D/3-D footprints are independently produced from each single aspect, and TomoSAR point cloud could be matched/correlated to develop a more precise object-based geometric TomoSAR fusion method compared with two existing ground-point-based [57] and feature-based [30] approaches. An alternative approach is geodetic TomoSAR [58], where the absolute 3-D coordinates of the reference points are retrieved using imaging Geodesy [59] and stereoSAR [60].

In the future, we will also explore the potential of extending the algorithm toward the generation of automatically reconstructed complete watertight prismatic (or polyhedral) 3-D/4-D building models from space.

REFERENCES

- [1] P. Sehrawat and K. Kensek, "Urban energy modeling: GIS as an alternative to BIM," 2014 ASHRAE/IBPSA-USA Building Simulation Conf., Atlanta, GA, USA, 2014.
- [2] B. Tomaszewski, *Geographic Information Systems (GIS) for Disaster Management*. Boca Raton, FL, USA: CRC Press, 2014.
- [3] T. H. Kolbe, G. Gröger, and L. Plümer, "CityGML-3D city models and their potential for emergency response," in *Geospatial Information Technology Emergency Response*, vol. 257. London, U.K.: Taylor & Francis, 2008.
- [4] J. Döllner, T. H. Kolbe, F. Liecke, T. Sgouros, and K. Teichmann, "The virtual 3D city model of Berlin—Managing, integrating, and communicating complex urban information," in *Proc. 25th UDMS*, 2006, pp. 15–17.
- [5] A. Köninger and S. Bartel, "3D-GIS for urban purposes," *Geoinformatica*, vol. 2, no. 1, pp. 79–103, Mar. 1998.
- [6] J. Kirtner, "Using LIDAR data in wireless communication system design," in *Proc. Amer. Soc. Photogramm. Remote Sens. Conf.*, Washington, DC, USA, 2000, pp. 1–8.
- [7] R. Bamler, M. Eineder, N. Adam, X. Zhu, and S. Gernhardt, "Interferometric recognition of high resolution spaceborne SAR," *Photogramm.-Fernerkund.-Geoinformation*, vol. 2009, no. 5, pp. 407–419, Nov. 2009.
- [8] A. Ferro, D. Brunner, and L. Bruzzone, "Automatic detection and reconstruction of building radar footprints from single VHR SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 935–952, Feb. 2013.
- [9] A. Thiele, E. Cadario, K. Schulz, U. Thonnessen, and U. Soergel, "Building recognition from multi-aspect high-resolution InSAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3583–3593, Nov. 2007.
- [10] Y.-Q. Jin and F. Xu, "Automatic Reconstruction of Building Objects from Multi-Aspect SAR Images," in *Polarimetric Scattering and SAR Information Retrieval*. Hoboken, NJ, USA: Wiley, 2013, pp. 235–274.
- [11] L. Zhao, X. Zhou, and G. Kuang, "Building detection from urban SAR image using building characteristics and contextual information," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 56, Mar. 2013.
- [12] G. Fornaro, F. Serafino, and F. Soldovieri, "Three-dimensional focusing with multipass SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 507–517, Mar. 2003.
- [13] P. Pasquali et al., "A 3-D SAR experiment with EMSL data," in *Proc. IGARSS*, 1995, vol. 1, pp. 784–786.
- [14] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.
- [15] G. Fornaro, D. Reale, A. Pauciuillo, X. Zhu, and R. Bamler, "SAR tomography: An advanced tool for spaceborne 4-D radar scanning with application to imaging and monitoring of cities and single buildings," *IEEE Geosci. Remote Sens. Soc. Newslett.*, pp. 9–17, Dec. 2012.
- [16] X. X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
- [17] F. Lombardini, "Differential tomography: A new framework for SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 37–44, Jan. 2005.
- [18] X. X. Zhu and R. Bamler, "Let's do the time warp: Multicomponent nonlinear motion estimation in differential SAR tomography," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 735–739, Jul. 2011.
- [19] G. Fornaro, D. Reale, and F. Serafino, "Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 224–237, Jan. 2009.
- [20] S. Gernhardt and R. Bamler, "Deformation monitoring of single buildings using meter-resolution SAR data in PSI," *ISPRS J. Photogramm. Remote Sens.*, vol. 73, pp. 68–79, Sep. 2012.
- [21] A. Ferretti, C. Prati, and F. Rocca, "Permanent scatterers in SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 8–20, Jan. 2001.
- [22] D. Reale, G. Fornaro, A. Pauciuillo, X. Zhu, and R. Bamler, "Tomographic imaging and monitoring of buildings with very high resolution SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 661–665, Jul. 2011.
- [23] G. Fornaro, D. Reale, and S. Verde, "Bridge thermal dilation monitoring with millimeter sensitivity via multidimensional SAR imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 677–681, Jul. 2013.
- [24] M. Shahzad and X. X. Zhu, "Robust reconstruction of building facades for large areas using spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 752–769, Feb. 2015.
- [25] X. X. Zhu and M. Shahzad, "Facade reconstruction using multiview spaceborne TomoSAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.
- [26] X. X. Zhu and R. Bamler, "Demonstration of super-resolution for tomographic SAR imaging in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.
- [27] G. Fornaro, A. Pauciuillo, D. Reale, and S. Verde, "Multilook SAR tomography for 3-D reconstruction and monitoring of single structures applied to COSMO-SKYMED data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2776–2785, Jul. 2014.
- [28] X. X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, Jan. 2012.

- [29] S. Auer, S. Gernhardt, and R. Bamler, "Ghost persistent scatterers related to multiple signal reflections," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 919–923, Sep. 2011.
- [30] Y. Wang and X. X. Zhu, "Automated feature-based geometric fusion of multiview TomoSAR point clouds in urban area," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 3, pp. 953–965, Mar. 2015.
- [31] M. Hubert, P. J. Rousseeuw, and K. V. Branden, "ROBPCA: A new approach to robust principal component analysis," *Technometrics*, vol. 47, no. 1, pp. 64–79, Feb. 2005.
- [32] F. Lafarge and C. Mallet, "Creating large-scale city models from 3D-point clouds: A robust approach with hybrid representation," *Int. J. Comput. Vis.*, vol. 99, no. 1, pp. 69–85, Aug. 2012.
- [33] J. Yan, J. Shan, and W. Jiang, "A global optimization approach to roof segmentation from airborne lidar point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 94, pp. 183–193, Aug. 2014.
- [34] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [35] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [36] M. Shahzad and X. X. Zhu, "Reconstructing 2-D/3-D building shapes from spaceborne tomographic synthetic aperture radar data," in *Proc. ISPRS Spatial Inf. Sci.*, Aug. 2014, vol. XL-3, pp. 313–320.
- [37] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 670–677.
- [38] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.
- [39] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Data Mining Knowl. Discov.*, vol. 96, no. 2, pp. 226–231, Jun. 1996.
- [40] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Trans. Graph.*, vol. 13, no. 1, pp. 43–72, Jan. 1994.
- [41] P. Dorninger and N. Pfeifer, "A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds," *Sensors*, vol. 8, no. 11, pp. 7323–7343, Nov. 2008.
- [42] K. Zhang, J. Yan, and S.-C. Chen, "Automatic construction of building footprints from airborne LIDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2523–2533, Sep. 2006.
- [43] X. Zhu, *Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring - A Sparse and Nonlinear Tour*, C. Reihe Ed. Munich, Germany: Deutsche Geodätische Kommission, Verlag der Bayerischen Akademie der Wissenschaften, 2011, No. 666, ISBN 978-3-7696-5078-5.
- [44] X. X. Zhu, Y. Wang, S. Gernhardt, and R. Bamler, "Tomo-GENESIS: DLR's tomographic SAR processing system," in *Proc. JURSE*, 2013, pp. 159–162.
- [45] CyberCity3D, Inc. [Accessed: 07-Apr-2015]. [Online]. Available: <http://www.cybercity3d.com/>
- [46] GEOFABRIK," 2015. [Accessed: 07-Apr-2015]. [Online]. Available: <http://www.geofabrik.de/data/download.html>
- [47] Stats-OpenStreetMap Wiki. [Accessed: 22-Jan-2015]. [Online]. Available: <http://wiki.openstreetmap.org/wiki/Statistics>
- [48] H. Fan, A. Zipf, Q. Fu, and P. Neis, "Quality assessment for building footprints data on OpenStreetMap," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 4, pp. 700–719, Apr. 2014.
- [49] M. Haklay, "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," *Environ. Plan. B, Plan. Des.*, vol. 37, no. 4, pp. 682–703, 2010.
- [50] M. Helbich, C. Amelunxen, and P. Neis, "Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata," in *Proc. AGILE*, 2011, pp. 24–33.
- [51] P. Mooney, P. Corcoran, and A. C. Winstanley, "Towards quality metrics for OpenStreetMap," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, New York, NY, USA, 2010, pp. 514–517.
- [52] R. Hecht, C. Kunze, and S. Hahmann, "Measuring completeness of building footprints in OpenStreetMap over space and time," *ISPRS Int. J. Geo-Inf.*, vol. 2, no. 4, pp. 1066–1091, Nov. 2013.
- [53] M. Rutzinger, F. Rottensteiner, and N. Pfeifer, "A comparison of evaluation techniques for building extraction from airborne laser scanning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 11–20, Mar. 2009.

- [54] G. Sohn and I. Dowman, "Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 1, pp. 43–63, May 2007.
- [55] L. Ding, X. Zhu, and L. Meng, "Scientific visualization for 4-D building deformation data analysis," presented at the 26th Int. Cartographic Conf., Dresden, Germany, 2013.
- [56] O. D'Hondt, S. Guillaso, and O. Hellwich, "Geometric primitive extraction for 3D reconstruction of urban areas from tomographic SAR data," in *Proc. JURSE*, 2013, pp. 206–209.
- [57] S. Gernhardt, X. Cong, M. Eineder, S. Hinze, and R. Bamler, "Geometrical fusion of multitrack PS point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 1, pp. 38–42, Jan. 2012.
- [58] X. X. Zhu, S. Montazeri, C. Gisinger, R. Hanssen, and R. Bamler, "Geodetic SAR Tomography," in *IEEE Trans. Geosci. Remote Sens.*, Jul. 2015, pp. 1–18, to be published.
- [59] M. Eineder, C. Minet, P. Steigenberger, X. Cong, and T. Fritz, "Imaging geodesy—Toward centimeter-level ranging accuracy with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 661–671, Feb. 2011.
- [60] C. Gisinger *et al.*, "Precise three-dimensional stereo localization of corner reflectors and persistent scatterers with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1782–1802, Apr. 2015.



Muhammad Shahzad (S'12) received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, and the M.Sc. degree in autonomous systems from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany. He is currently working toward the Ph.D. degree at Signal Processing in Earth Observation (SiPEO), Technische Universität München, Munich, Germany.

His research interests include the automatic 3-D reconstruction of objects from point clouds retrieved from spaceborne synthetic aperture radar image stacks.



Xiao Xiang Zhu (S'10–M'12–SM'14) received the Bachelor's degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006 and the M.Sc., Dr.-Ing., and "Habilitation" degrees in signal processing from the Technische Universität München (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy; Fudan University, Shanghai, China; and the University of Tokyo, Tokyo, Japan, in 2009, 2014, and 2015, respectively. Since 2011, she has been a Scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany, where she is also the Head of Team Signal Analysis. Since 2013, she has been a Helmholtz Young Investigator Group Leader and appointed as a TUM Junior Fellow. In 2015, she was appointed as a Professor for signal processing in Earth observation with TUM. Her main research interests include advanced interferometric synthetic aperture radar techniques, such as high-dimensional tomographic synthetic aperture radar imaging and SqueeSAR; computer vision in remote sensing, including object reconstruction and multi-dimensional data visualization; and modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multispectral/hyperspectral image analysis.

Dr. Zhu is an Associate Editor of the IEEE TRANSACTIONS ON GEOGRAPHICS AND REMOTE SENSING.

D Zhu, X. X., Ge, N., Shahzad, M., 2015. Joint Sparsity in SAR Tomography for Urban Mapping. IEEE Journal of Selected Topics in Signal Processing 9(8): 1498 - 1509

<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7208791>

Joint Sparsity in SAR Tomography for Urban Mapping

Xiao Xiang Zhu, *Senior Member, IEEE*, Nan Ge, Muhammad Shahzad, *Student Member, IEEE*,

Abstract—With meter-resolution images delivered by modern SAR satellites like TerraSAR-X and TanDEM-X, it is now possible to map urban areas from space in very high level of detail using advanced interferometric techniques such as PSI and SAR Tomography (TomoSAR), whereas these multi-pass techniques are based on a great number of images. We aim at precise TomoSAR reconstruction while significantly reducing the required number of images by incorporating building a priori knowledge to the estimation. In the paper, we propose a novel workflow that marries the freely available 2-D building footprint GIS data and the joint sparsity concept for TomoSAR inversion. Experiments on bistatic TanDEM-X data stacks demonstrate the great potential of the proposed approach, e.g., highly accurate tomographic reconstruction is achieved using six interferograms only.

Index Terms—Synthetic aperture radar, SAR tomography, compressive sensing, joint sparsity, GIS, TanDEM-X.

I. INTRODUCTION

MODERN spaceborne synthetic aperture radar (SAR) sensors, such as TerraSAR-X, TanDEM-X and COSMO-SkyMed, deliver SAR data with a very high spatial resolution (VHR) of up to 1 m. With these meter resolution data, advanced multi-pass interferometric techniques such as persistent scatterer interferometry (PSI) and tomographic SAR inversion (TomoSAR) allow retrieving not only the 3-D geometrical shape but also the undergoing temporal motion in the scale of millimeter of individual building [1][2][3][4][5][6][7]. In particular, sparse reconstruction based methods [8][9], like SLIMMER [10], give robust TomoSAR inversion with very high elevation resolution, and can offer so far ultimate 3-D, 4-D and 5-D SAR imaging [11][12].

The downside of advanced repeat-pass InSAR techniques [13][14][15][16][17][18][19][20], like PSI and TomoSAR, are their high demand on the data, i.e., typically a stack of 20–100 images over the illuminated area are required. For instance, it is demonstrated in [10] that even using most efficient algorithms, like non-linear least squares (NLS) and SLIMMER, a

minimum number of 11 acquisitions are required to achieve a reasonable reconstruction in the interesting parameter range of spaceborne SAR. "Reasonable" in this context means that given an average signal-to-noise ratio (SNR) of 6 dB, the detection rate of double scatterers with an elevation distance of one Rayleigh resolution unit reaches at least 90%. However, if we can extract certain detailed features or patterns of high-rise buildings in SAR images, the required number of images can be significantly reduced by incorporating such features as prior for a joint estimation.

For this purpose, we propose a novel workflow marrying the globally available (2-D building footprint) GIS data and the joint sparsity concept for TomoSAR inversion, both of which have not yet been addressed in the community so far. Within this workflow, our main contributions are as follows:

- A robust procedure is proposed to use online freely assessable 2-D building footprints for extracting detailed high-rise building features including building masks, orientation, and iso-height lines (defined in [21]) in SAR image stacks (see Section III);
- The M-SLIMMER algorithm is proposed to promote joint sparsity for tomographic inversion of the identified iso-height pixel groups (see Section IV);
- The performance of M-SLIMMER is systematically evaluated using simulated data in terms of elevation estimation accuracy, detection rate and false alarm rate of the overlaid scatterers' separation, and its super-resolution capability. Compared to the single-snapshot sparsity model, as used in SLIMMER, the superior performance of the proposed joint sparsity approach is evident for all above mentioned quantitative metrics (see Section V);
- The first tomographic reconstruction using bistatic TanDEM-X data stacks is presented. The superior performance of M-SLIMMER is demonstrated in practice, e.g., highly accurate tomographic reconstruction is achieved using six interferograms only (see Section VI).

II. DATA SET

We work with 21 bistatic interferograms acquired by the German SAR satellites TerraSAR-X and TanDEM-X, with cross-track baselines ranging between approximately ± 200 [m]. The single-pass characteristic renders atmospheric effects very small and deformation negligible. For this reason these datasets are ideal to test our proposed methodology. An optical image of the test area is shown in Figure 1(a) while the corresponding SAR mean intensity image is shown in Figure 1(b).

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

X. X. Zhu is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany and Signal Processing in Earth Observation (SiPEO), Technische Universität München, 80333 Munich, Germany (E-mail: xiao.zhu@dlr.de).

N. Ge is with Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany (E-mail: nan.ge@dlr.de).

M. Shahzad is with Signal Processing in Earth Observation (SiPEO), Technische Universität München, 80333 Munich, Germany (E-mail: muhammad.shahzad@tum.de).

final manuscript submitted on August 2015.

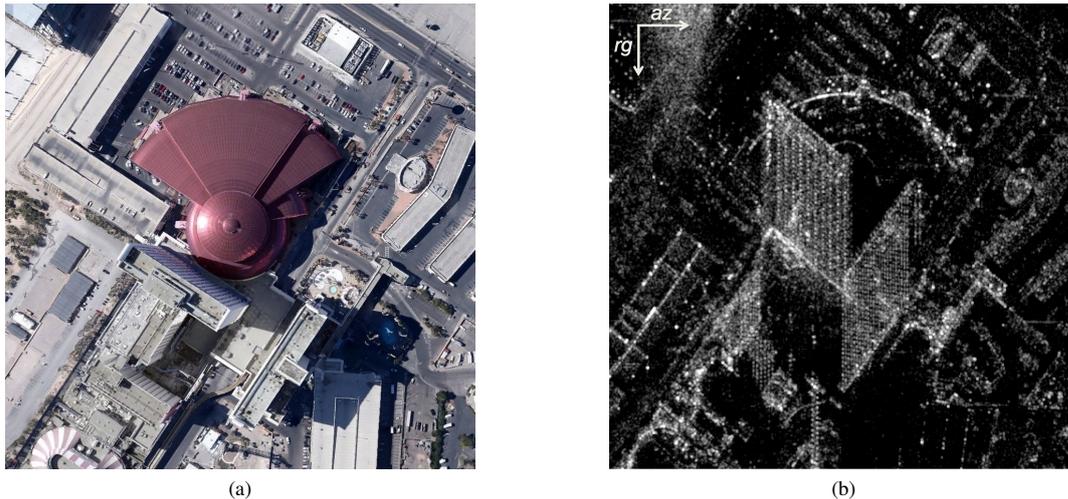


Fig. 1. Test Area: (a) Optical image of the test area ©Google; (b) Corresponding SAR intensity map (rg and az refer to the range and azimuth coordinates respectively).

III. PRIOR KNOWLEDGE RETRIEVAL

In order to retrieve prior information pertaining to building regions, the 2-D building footprints are downloaded from OpenStreetMap (OSM). Based on the concept of crowd sourcing that involve crowd or community to effectively and efficiently fulfill a task at hand, OSM with around 2 million registered users (as of today and also rapidly growing) is considered to be the most successful Volunteered Geographic Information (VGI) project [22][23]. The OSM database contains multitude of building footprints represented as polygons with ordered list of nodes/vertices (i.e., pairs of UTM or latitude/longitude coordinates according to WGS 84 coordinate system) and is updated every day. The data are free to download and comes under the open license Open Data Commons Database License (ODbL). Since it is a VGI project, the data quality may vary from region to region. To this end, the first investigations regarding OSM data quality were carried out for roads [24] followed by assessment of other attributes present in the database e.g., lines [25], polygonal objects [26] etc. Recently, the building footprints have also been evaluated for their completeness [27] and correctness [23]. The analysis of OSM data with surveying datasets reveals fairly precise positioning accuracies varying within 4 meters [23][24]. The completeness percentage is already very high for many cities in Europe and US and is consistently increasing with time. Available 2-D footprints of the buildings in the city of Las Vegas are shown in Figure 2 to give the reader an insight of the existing database. The high availability of such type of data triggers us to change our perspective of thinking, namely, instead of using Earth observation (EO) satellite data to build-up sources of geo-information for open users, we can explore the knowledge provided by social media to support information retrieval from EO data. In this regard, one mission of this work is to demonstrate this concept in tomographic SAR reconstruction.

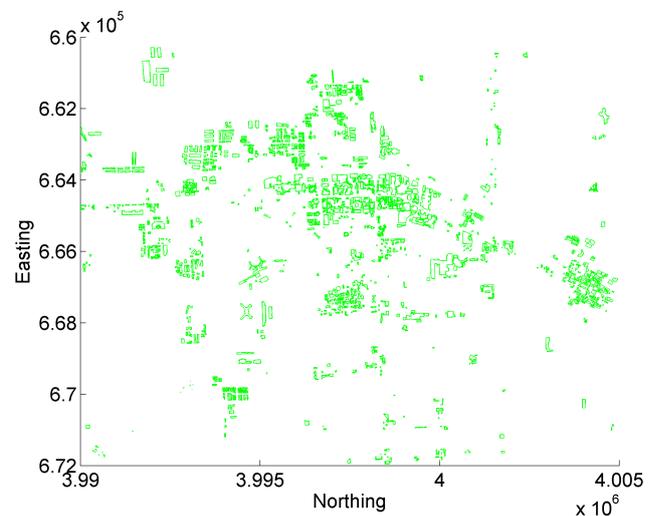


Fig. 2. GIS data (2-D building footprints) of Las Vegas from OSM.

A. Automatic extraction of building mask in SAR image

The key idea is to make use of the aforementioned online freely assessable 2-D building footprints to extract detailed high rise building features including building masks, orientations, and the iso-height lines in SAR image data stacks. The extracted information can be further incorporated as a prior knowledge into the estimation for a more accurate tomographic SAR inversion. For this purpose, in this section we propose a sophisticated approach that is tolerable to moderate errors in the input GIS data for automatic extraction of aforementioned high rise building features in the SAR image data stacks:

- First the available building footprints from OpenStreetMap in world (latitude/longitude) coordinates are transformed/geo-coded into SAR (azimuth/range) imaging coordinate system. Figure 3(a) shows the resulting projected reference polygons overlaid onto the buildings

of interest in the corresponding SAR image shown in Figure 1;

- Secondly, due to the side looking geometry, SAR illuminates only one side of the building. Therefore, the complete building footprint of individual buildings is further segmented into two parts by means of a simple 2-D visibility test: 1) the part illuminated by the sensor which will be further used for iso-height pixel extraction (as depicted by red polylines in Figure 3(a)); 2) the part in the shadow area not visible to the sensor which will not be used in further processing (as depicted by green polylines in Figure 3(a));
- Finally, errors in the identified red polylines, caused by inaccuracies of the input GIS data in both orientation and translation, are compensated and the mask of individual buildings is further generated by iteratively shifting the corrected polylines towards the sensor.

In this regard, the approach depicted in Algorithm 1 is adopted. After transforming the available building footprints from world coordinates to SAR imaging coordinate system, we identify the side of the building footprint facing the SAR sensor as follows. If we assume that $v_{i=1,\dots,n}$ denote the indices of ordered 2-D footprint vertices of one particular building. Then any vertex $v_k (k \in n)$ belongs to the side facing the sensor if and only if its projection onto the line at zeroth range axis (i.e., line defined as $rg = 0$ with zero azimuth slope) does not self-intersect the reference polygon. The range of total number of vertices belonging to the side visible to the sensor in any footprint is m where $1 < m \leq n$. The inequality that $m > 1$ depicts that, if not occluded, at least one side or two vertices of the building are always visible to the side looking SAR sensor.

Once the vertices facing to the sensor are identified, the step 3 in Algorithm 1 compensates for any positioning inaccuracy in the OSM footprint of the building in the area of interest. Possible error in OSM footprint is compensated by adopting the following sequence of steps:

- 1) Shift/translate the identified polyline in 2-D sliding window fashion within the intervals: range shift $[-10 \ 10]$ and azimuth shift $[-5 \ 5]$;
- 2) Within each shift, rotate the polyline between interval $[-7.5 \ 7.5]$ degrees and compute median of intensities along the rotated polyline (similar to steps 6-7 in Algorithm 1);
- 3) The polyline is rotated and shifted with the rotation angle and the azimuth-range shifts which give the maximum of computed median intensities (from the previous step);
- 4) Finally, the rotation causes the change in the azimuthal length of the polyline which needs to be adjusted (see Figure 4(a)). This is accomplished by first slightly extending the polyline and later adjusting the lengths of the outer (first and the last) edges of the polyline by analysing their (interpolated) intensities (same as step 2 except that only the first and last edge of the polyline is used). Figure 4(b) graphically illustrates the adjustment procedure.

Algorithm 1 Procedure to automatically generate mask (or ROI) of an individual building

Require: 2-D polygonal footprint vertices $v_{i=1,\dots,n}$ in SAR coordinates of one particular building & SAR image of the scene.

- 1: Initialize: $MaxRgShift := 595$ and $d := 1$
 - 2: Identify the polyline comprising of m out of n vertices belonging to the building side facing the sensor
 - 3: Apply orientation correction and range-azimuth shift compensation to the identified polyline in order to cope for any positioning inaccuracy in the OSM footprint of the building
 - 4: **while** (1) **do**
 - 5: Shift/translate (in range direction) the polyline after orientation correction and azimuth-range compensation towards the sensor by distance d
 - 6: Compute the intensity values along the shifted polyline. This is accomplished by selecting equally spaced points along the shifted polyline, and then using nearest neighbor interpolation to find the intensity value for each point
 - 7: Take the median of computed intensity values along the shifted polyline and store the result in a column matrix $C(d, 1)$
 - 8: **if** $d == MaxRgShift$ **then**
 - 9: **break**
 - 10: **else**
 - 11: $d := d + 1$
 - 12: **end if**
 - 13: **end while**
 - 14: Take approximate derivative of C (i.e., calculate differences between adjacent elements of C), and store the result in matrix D
 - 15: Compute $\arg \max_{RgShift} (D_{RgShift})$ where $RgShift (= 1, \dots, d - 1)$ denotes the maximum change point in D
 - 16: Use $RgShift$ and m vertices of the identified (compensated) polyline facing the sensor to extract $poly_{ROI}$ of the particular building
-

In our experiment, the shifting intervals used in compensation are based on already mentioned 4m inaccuracy of OSM data i.e., considering 4m inaccuracy, approx. range and azimuth resolution of 0.588m and 1.1m requires range and azimuth shifts of atmost $(4/0.588 \approx) 6.8$ pixels and $(4/1.1 \approx) 3.6$ pixels.

After compensating any orientation and/or shifting inaccuracies, steps 4 to 13 in Algorithm 1 iteratively shifts the compensated polyline towards the sensor (in range direction only). Since the tallest building in the city of Las Vegas, the Stratosphere Tower, is around 350 m, $MaxRgShift$ in Algorithm 1 is set to 595 i.e., maximum building size along elevation appearing in the SAR image of Las Vegas city, computed as $350/0.588 (\approx 595)$ pixels) where 0.588m is the approx. range resolution. Thus the polyline is shifted till $MaxRgShift$ and the column vector C stores the median

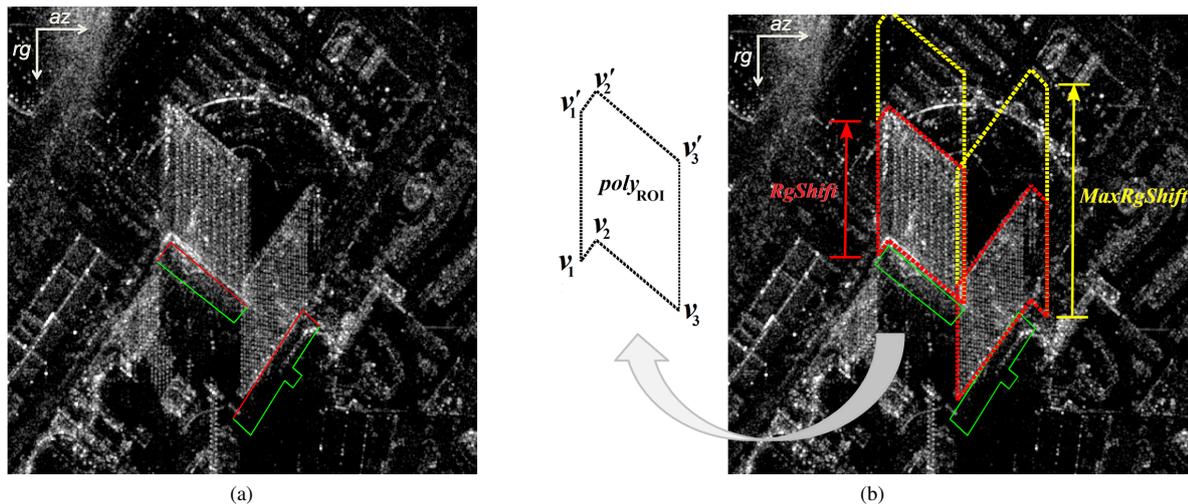


Fig. 3. Building mask extraction: (a) Reference polygons (shown in red and green polylines) of two buildings in the area of interest overlaid onto the SAR intensity map after geocoding. Side of the buildings facing the sensor are shown in red while the other side not visible to the sensor in green; (b) After rotation and range-azimuth shift compensation, the red polylines in (a) are shifted towards the sensor. The yellow dotted lines indicate the maximum range shift of $MaxRgShift$ ($= 595\text{m}$) where as the red dotted lines indicate the $RgShift$ obtained by the proposed procedure in Algorithm 1.

of computed intensity values along each range shift. Steps 14 and 15 in Algorithm 1 then computes the maximum change point $RgShift$ in the approximate derivative of C . $RgShift$ is then used in step 16 to determine $polyROI$ which describes the polygon surrounding the overlaid pixels of the same building in the SAR image. To elaborate how $polyROI$ is computed, consider a building having three adjacently connected vertices $v_1 - v_2 - v_3$ of the polyline facing the sensor where $-$ denotes the adjacency (i.e., v_2 is adjacently connected to v_1 and v_3 , and so on). Assuming that the polyline has been compensated for rotation and range-azimuth shifts, the $polyROI$ is then simply formed as $v_1 - v_2 - v_3 - v'_3 - v'_2 - v'_1 - v_1$ where $\mathbf{v}'_j(az, rg) = \mathbf{v}_j(az, rg - RgShift)$ with $j = 1, 2, 3$ (see Figure 3(b)). Finally, $polyROI$ is used to generate the building mask of an individual building.

B. Pixel grouping

Based on the extracted masks of individual buildings, pixels sharing similar heights are then grouped together. This procedure is done in three steps:

- 1) Iso-height lines will be reproduced by translating the adjusted polyline towards both ends of the building mask with sub-pixel increments;
- 2) The distance between each pixel and its adjacent iso-height lines will be calculated;
- 3) Each pixel will be assigned to the closest iso-height line.

The distance between a pixel and an iso-height line is defined as the minimum absolute amount of translation (in pixels) towards or away from sensor. Figure 5 illustrates one exemplary iso-height line in the cropped intensity image, as well as the final results of pixel grouping with each group of pixels plotted with a random color. Note that the color-coding already gives a rough idea about monotonic height change of the investigated façades.

IV. JOINT SPARSITY IN TOMOSAR

In this section, we first revisit a data model commonly used in TomoSAR, as well as the SLIMMER algorithm. Following this, we extend the SLIMMER algorithm for the multiple-snapshot case. The extended version exploits joint sparsity and is named as M-SLIMMER.

A. TomoSAR system model

For a single SAR image, information along the third dimension, the so-called elevation axis s , which is perpendicular to the azimuth-range ($x-r$) plane, is integrated (see Figure 6). I.e., echoes from, e.g., tree crown, building roof, or double-bounce effects on a balcony sharing the same distance to the sensor, are mapped onto one single pixel. To reconstruct reflectivity along s and to further separate those different contributions, TomoSAR utilizes scenes acquired from slightly different viewing angles to synthesize an elevation aperture Δb (cf. aperture along x created by steering the radar beam) for full 3-D SAR imaging [28]. A well-established model, which can be found, e.g., in [29], approximates each pixel value g_n as follows:

$$g_n = \int_{\Delta s} \gamma(s) \exp(-j2\pi\xi_n s) ds, \quad (1)$$

which is essentially the Fourier transform of the reflectivity function $\gamma(s)$ sampled at the spatial (elevation) frequency $\xi_n = -2b_n/(\lambda r)$, Δs is the elevation extent and λ is wavelength. Note that a possible motion term has been neglected here without loss of generality. For differential SAR tomography that takes into account the motion component, the readers are recommended to consult [17][30][31].

Discretizing the continuous function in Eq. (1) along elevation s into L elevation positions $s_l \forall l \in \{1, \dots, L\}$, replacing integral by sum and taking into account measurement noise yield the following discrete system model:

$$\mathbf{g} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2)$$

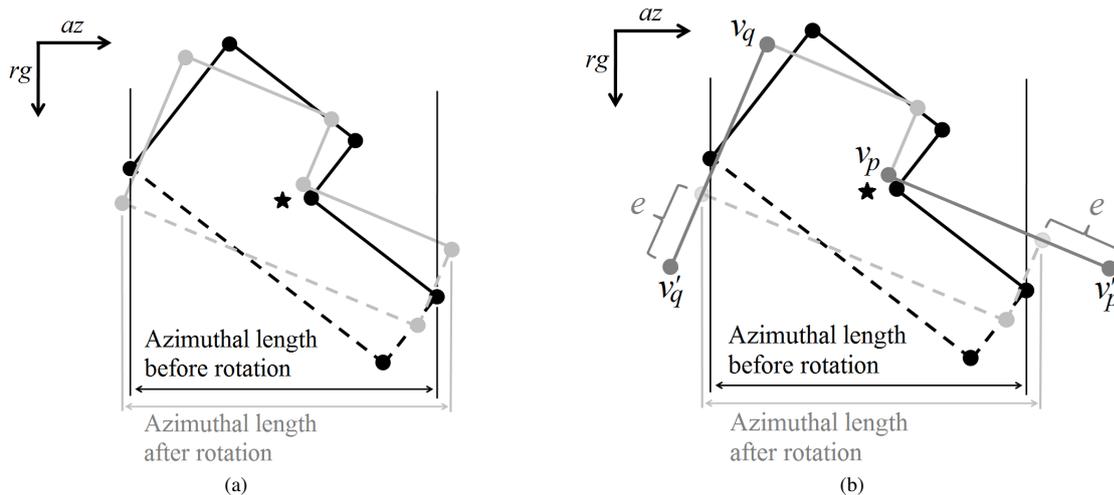


Fig. 4. Graphical illustration of adjusting polyline length: (a) Black and gray polygons indicate polygons before and after rotation (around centroid depicted as black star) respectively with circles representing the corresponding vertices. The dotted polylines represent building side not visible to the sensor. It is shown that after rotation the azimuthal length is changed; (b) The length of the outer (first and last edge depicted in dark gray) edges of the polyline facing the sensor is slightly extended by distance e ($= 5\text{m}$ in this work). Intensities (interpolated) over these extended edges are analysed and first and last extended points (i.e., v'_p and v'_q) are respectively replaced by the points on the edges $v_p v'_p$ and $v_q v'_q$ where the approximate derivative (or change in intensities) is maximum.

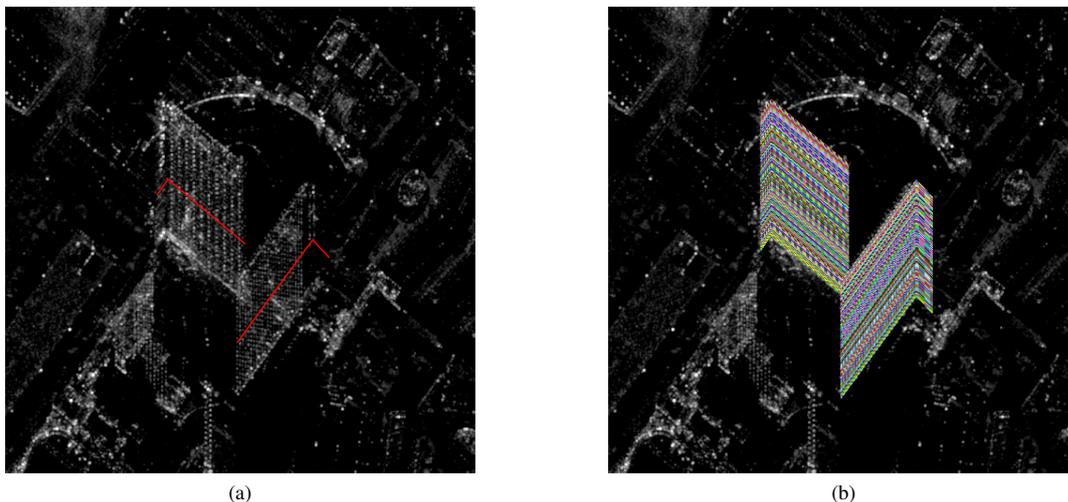


Fig. 5. Illustration of pixel grouping with (a) exemplary iso-height lines, and (b) grouped iso-height pixels color-coded with group indices.

where $\mathbf{g} \in \mathbb{C}^{N \times 1}$ is the measurement vector with $g_n \forall n \in \{1, \dots, N\}$, $\mathbf{R} \in \mathbb{C}^{N \times 1}$ is an irregularly sampled Fourier transform matrix with $R_{nl} = \exp(-j2\pi\xi_n s_l)$, $\gamma \in \mathbb{C}^{L \times 1}$ is the discretized reflectivity vector $\gamma_l = \gamma(s_l) \forall l \in \{1, \dots, L\}$, and $\varepsilon \in \mathbb{C}^{N \times 1}$ is additive noise which can be modeled as a zero-mean circular Gaussian random process. Typically we have $N \ll L$, which renders Eq. (2) underdetermined.

Similar to the resolution in azimuth, the Rayleigh resolution ρ_s is inversely proportional to the aperture size [2]

$$\rho_s = \frac{\lambda r}{2\Delta b}. \quad (3)$$

For high resolution spotlight data of TerraSAR-X/TanDEM-X, ρ_s is much worse than azimuth and range resolution (approx. 1.10 and 0.588 m, respectively) due to tight orbit control and amounts to about 24.9 m for our test data set.

B. The SLIMMER algorithm

To solve Eq. (2), an algorithm called SLIMMER, which stands for Scale-down by L_1 norm Minimization, Model selection, and Estimation Reconstruction, has been proposed to achieve promising super-resolution power while guaranteeing the efficiency[8][10]. SLIMMER has been originally designed for TomoSAR in urban areas, under the assumption that there are only a few dominant scatterers (phase centers) along elevation axis within each azimuth-range pixel [2]. I.e., γ has merely K non-zero entries where typically $K \leq 4$. As its name suggests, this algorithm consists of the following three main steps.

1) *Scale-down by L_1 norm minimization*: To exploit the sparse prior on γ , we solve the following L_1 -regularized least

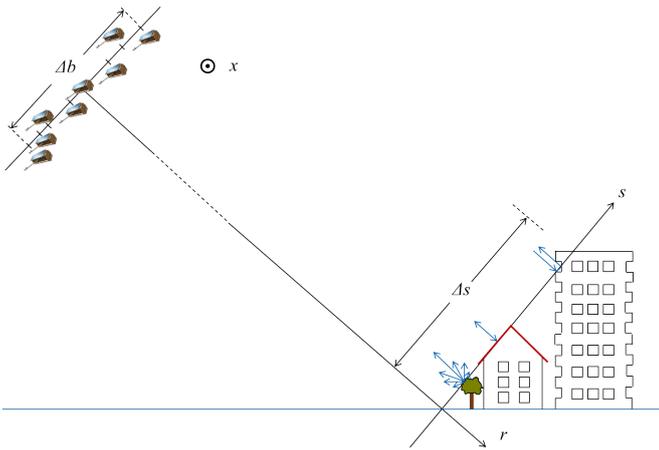


Fig. 6. TomoSAR imaging geometry with an artistic view of TerraSAR-X/TanDEM-X ©DLR. The satellite flies into the plane and looks to its right.

squares problem

$$\hat{\gamma} = \arg \min_{\gamma} \left\{ \frac{1}{2} \|\mathbf{g} - \mathbf{R}\gamma\|_2^2 + \lambda_K \|\gamma\|_1 \right\}, \quad (4)$$

where λ_K is a hyperparameter balancing model error and the sparsity of γ . Eq. 4 is known to deliver robust elevation estimates \hat{s}_l of dominant scatterers. Therefore, by identifying the most significant entries in $\hat{\gamma}$ and choosing certain columns of \mathbf{R} accordingly, the dimension of the original problem in Eq. (2) can be downscaled by a large factor. However, solving 4 is prone to amplitude bias due to the L_1 norm relaxation. Moreover, outliers might appear when the required mathematical conditions of \mathbf{R} are not fully fulfilled as most of the engineering problems do [32]. These make the next two steps necessary.

2) *Model selection*: The initial estimate $\hat{\gamma}$ from Eq. (4) may contain artifacts, which falsifies its sparsity level. In order to detect and remove them, the goodness of fit of a model should be penalized by its complexity, so that overfitting of data can be avoided. Model selection can be regarded as the following optimization problem

$$\hat{K} = \arg \min_K \left\{ -2 \ln p(\mathbf{g} | \hat{\theta}(K), K) + 2C(K) \right\}, \quad (5)$$

where $p(\mathbf{g} | \hat{\theta}(K), K)$ is the likelihood function of \mathbf{g} given the estimates of unknown $\theta(K)$ and K , $C(K)$ is the penalty term for model complexity. Various alternatives of $C(K)$ have been devised for different needs, e.g., Bayesian information criterion, Akaike information criterion, minimum description length, to name a few [33]. By choosing one specific criterion suitable for the given datasets, Eq. (5) is then solved as combinatorial problem over a pre-defined range of K . Likewise, the most likely positions \hat{s} of non-zero elements in $\hat{\gamma}$ will hereby be estimated, which further shrinks \mathbf{R} . This leaves only one last step to correct amplitude bias.

3) *Parameter estimation*: At this stage, we have a much slimmer sensing matrix $\mathbf{R}(\hat{s}) \in \mathbb{C}^{N \times \hat{K}}$. This renders Eq. (2) to

$$\mathbf{g} = \mathbf{R}(\hat{s}) \gamma(\hat{s}) + \mathbf{e}, \quad (6)$$

where $\gamma(\hat{s}) \in \mathbb{C}^{\hat{K} \times 1}$, and $\mathbf{e} \in \mathbb{C}^{N \times 1}$ is the sum of measurement noise and the error introduced by model selection. Since Eq. (6) is now overdetermined, it can be solved with ordinary least squares (OLS)

$$\hat{\gamma} = \mathbf{R}^+(\hat{s}) \mathbf{g}, \quad (7)$$

where $(\cdot)^+$ denotes pseudo inverse.

Within the framework of SLIMMER, the merits from both sparse regression and OLS have been joined as a whole, namely, robust identification of scatterers' elevation positions, as well as accurate amplitude estimation. Other advantages of SLIMMER over conventional parametric and non-parametric methods have been discussed in [8] and its theoretical limits in terms of estimation accuracy, super-resolution power and the required minimum number of acquisitions for a reasonable reconstruction have been investigated in [10].

C. The M-SLIMMER algorithm

We extend the SLIMMER method to M-SLIMMER, i.e., the multiple-snapshot case. Assume that by applying the method described in Section III, we have already detected M pixels along an iso-height line. We further assume that within each pixel, there is a dominant scatterer located on the considered building façade. Hence, those M scatterers should reside at the same height or elevation position. For each pixel, we have, similar to Eq. (2),

$$\mathbf{g}_m = \mathbf{R}_m \gamma_m + \varepsilon_m, \quad (8)$$

$\forall m \in \{1, \dots, M\}$. If the iso-height line stretches principally in azimuth direction, we expect ξ_n to vary little among all concerned pixels. For this reason, we define $\mathbf{R} := \mathbf{R}_1 \cong \mathbf{R}_2 \cong \dots \cong \mathbf{R}_M$. By using the identical degree of discretization along elevation axis, we can rewrite Eq. (8) as

$$\mathbf{G} = \mathbf{R}\mathbf{\Gamma} + \mathbf{E}, \quad (9)$$

where $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_M]$ is the observation matrix with M measurements vectors, $\mathbf{\Gamma} = [\gamma_1, \dots, \gamma_M]$ is the unknown discretized reflectivity matrix, and \mathbf{E} accounts for both additive noise and possible model error. Eq. (9) is again an under-determined system with $N \ll L$ which has infinitely many solutions. Since we assume that all the measurement vectors have one contribution from the same height on a façade, the non-zero entry positions in the columns of $\mathbf{\Gamma}$ are aligned in a row-wise fashion. This property of signals is also referred to as joint sparsity. Indeed, there can be more non-zero rows related to ground, lower infrastructures, building roof, etc. Still, the number of non-zeros rows of $\mathbf{\Gamma}$ is very limited. To solve Eq. (9) while incorporating this prior, $\hat{\mathbf{\Gamma}}$ can be estimated by solving $L_{1,2}$ -regularized least squares problem [34],

$$\hat{\mathbf{\Gamma}} = \arg \min_{\mathbf{\Gamma}} \left\{ \frac{1}{2} \|\mathbf{G} - \mathbf{R}\mathbf{\Gamma}\|_F^2 + \lambda_K \|\mathbf{\Gamma}\|_{1,2} \right\}, \quad (10)$$

where F denotes the Frobenius norm, and the mixed norm $\|\mathbf{\Gamma}\|_{1,2} = \sum_{l=1}^L (\|\gamma^l\|_2)$, where γ^l is the l^{th} , promotes joint sparsity. It has been shown in [35] that the probability of successful recovery increases with the number of snapshots.

Note that different polarimetric channels or neighboring pixels were used in a similar way in [36][37].

After the downscaling step based on the estimate in Eq. (10), model selection and parameter estimation will be performed individually for each pixel as the SLIMMER algorithm does.

D. Cramér Rao Lower Bounds (CRLB) for elevation estimates

The Cramér Rao lower bound (CRLB) for elevation estimates \hat{s} for the single-scatterer case has been derived in [38] as

$$\sigma_{\hat{s},0} = \frac{\lambda r}{4\pi\sqrt{N} \cdot \sqrt{2SNR} \cdot \sigma_b}, \quad (11)$$

where σ_b is the standard deviation of b_n . Given an SNR of 3 dB, the CRLB is approx. 1.11 m with all 21 acquisitions.

In urban environment, due to the side-looking geometry of SAR, multiple scatterers are often mapped into one azimuth-range pixel. The fact that closely spaced scatterers will interfere with each other renders a degraded estimation accuracy of individual scatterers [12]. In the interest of super-resolution, the CRLB for elevation estimate of the q^{th} ($q = 1, 2$) scatterer has been derived in [10] as

$$\sigma_{s_q} = c_0 \cdot \sigma_{s_q,0}, \quad (12)$$

where

$$c_0 \approx \max \left\{ \sqrt{2.57(\alpha^{-1.5} - 0.11)^2 + 0.62}, 1 \right\} \quad (13)$$

is the interference factor depending on α which is the distance between two scatterers normalized w.r.t. the Rayleigh resolution unit [10]. c_0 is equal to one (no interference) when two scatterers are far apart, i.e., $\alpha \gg 1$, greater than one since the two scatterers are closely spaced ($\alpha < 1.5$) and increasing with decreasing α .

V. PERFORMANCE EVALUATION USING SIMULATED DATA

In general, as an extension of SLIMMER, M-SLIMMER has the same basic principle. However, instead of exploiting sparsity, M-SLIMMER uses multiple snapshots of iso-height pixels identified in SAR images (with the help of supporting OSM data). Since M-SLIMMER makes use of multiple snapshots and thus more observations than SLIMMER, we naturally expect it to achieve better performance.

In this section, the performance of the proposed M-SLIMMER algorithm, including elevation estimation accuracy, detection rate and false alarm rate in separating overlaid scatterers, and its super-resolution capability, is evaluated using simulated data.

We simulate façade-ground interaction of two scatterers spaced by decreasing elevation distances, which is a well-known TomoSAR benchmark test [2][8]. Note that we only work in the super-resolution regime, i.e., the elevation distance between façade and ground is no larger than the Rayleigh resolution ρ_s . Four scenarios are taken into account with the number of measurements $N \in \{6, 11\}$ and $SNR \in \{3, 10\}$ in [dB] because:

- As mentioned above, eleven is the minimum required number of measurements for a reasonable reconstruction

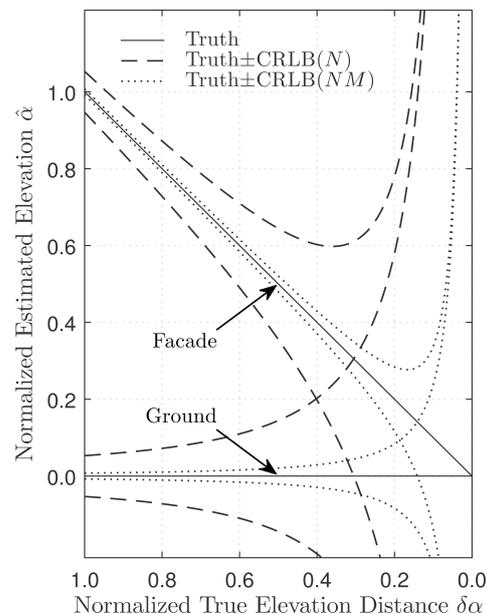


Fig. 7. Normalized true elevation α of simulated façade and ground (solid lines), as well as the CRLB of normalized elevation estimates, both w.r.t. normalized true elevation distance $\delta\alpha$. Dashed lines: CRLB with N baselines; dotted lines: CRLB with NM baselines.

in the interesting SNR range of spaceborne SAR if SLIMMER is used [10];

- In case of two scatterers, six is the number of unknowns, namely the amplitude, phase and elevation position of each scatterer;
- SNR of 3 dB and 10 dB are usually considered as the lower and upper bound of persistent scatterers, respectively [39].

For each façade-ground interaction with a given elevation distance, we independently generate $M = 48$ snapshots, which is an average case for the test buildings in Figure 5. The true elevation of simulated façade and ground is shown as two solid line segments w.r.t. their normalized elevation distance $\delta\alpha$ in Figure 7, respectively. In addition, we show for the case $N = 11$ the evolution of CRLB, which increases with decreasing elevation distance of two interfering scatterers in the super-resolution regime, as implied by 12 and 13. The dashed lines mark true elevation $\pm 1 \times \text{CRLB}$ with N interferograms, denoted as $\text{CRLB}(N)$, while we plot true elevation $\pm 1 \times \text{CRLB}$ with NM interferograms, denoted as $\text{CRLB}(NM)$, as dotted lines. We will show that M-SLIMMER using N interferograms and M snapshots approaches the latter bound, which can be achieved by SLIMMER given NM interferograms.

We solve the L_1 - and $L_{1,2}$ -regularized least squares problems independently, and then follow the SLIMMER procedures to perform model selection and parameter estimation. λ_K is chosen adaptively, which depends on N , M and the noise level of observations [34]. The results are shown in Figure 8 and Figure 9 with elevation estimates $\hat{\alpha}$ of façade and ground plotted w.r.t. their normalized true elevation difference $\delta\alpha$. Each dot depicts mean value of all estimates, with error bar indicating its standard deviation. In each subplot, the two solid line segments mark the true elevation for façade

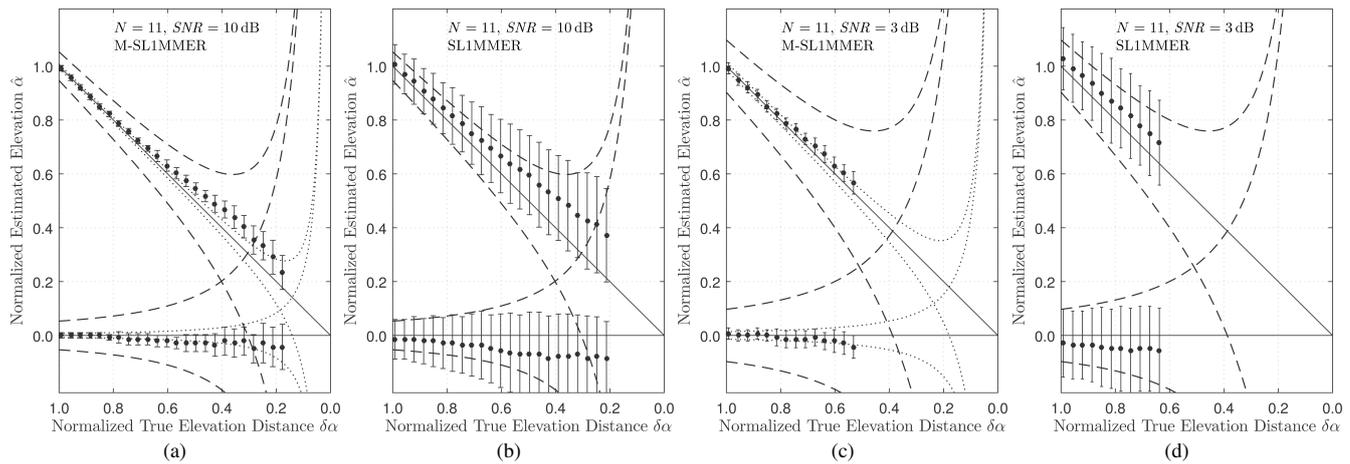


Fig. 8. Reconstructed elevation of simulated façade and ground with $M = 48$, $N = 11$; (a) $SNR = 10$ dB with M-SLIMMER (b) $SNR = 10$ dB with SLIMMER, (c) $SNR = 3$ dB with M-SLIMMER and (d) $SNR = 3$ dB with SLIMMER respectively. Each dot has the sample mean of all estimates as its y value and the corresponding standard deviation as error bar.

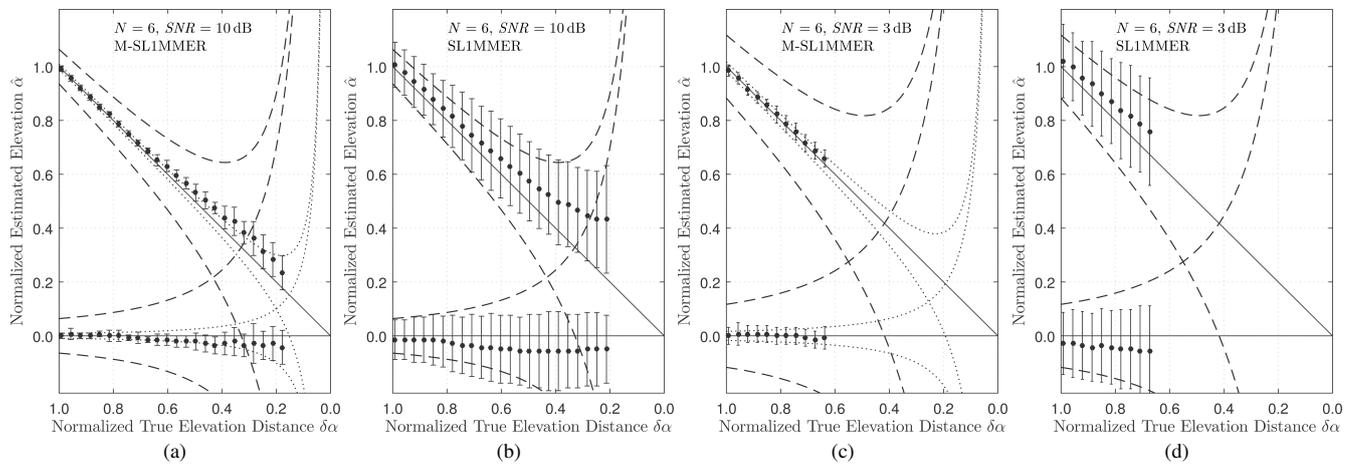


Fig. 9. Reconstructed elevation of simulated façade and ground with $M = 48$, $N = 6$, (a) $SNR = 10$ dB with M-SLIMMER, (b) $SNR = 10$ dB with SLIMMER, (c) $SNR = 3$ dB with M-SLIMMER, and (d) $SNR = 3$ dB with SLIMMER, respectively. Each dot has the sample mean of all estimates as its y value and the corresponding standard deviation as error bar.

and ground, respectively, while the dashed and dotted lines denote true elevation $\pm 1 \times \text{CRLB}$, which is the same as in Figure 7. Missing points suggest that detection rate is below 25%. Note that we call detection when not only two scatterers are separated, but also their estimates should be bounded by $\pm 3 \times \text{CRLB}(N)$ of their true elevation. For $N = 11$, the elevation estimates with both methods are still somewhat comparable, despite the fact that joint sparsity model leads to much smaller variance and slightly better super-resolution capability. SLIMMER performs in particular worse with smaller N and lower SNR. On the contrary, even for the case $N = 6$, reasonable elevation has been reconstructed with M-SLIMMER. In particular, M-SLIMMER with N interferograms and M snapshots is equivalent to SLIMMER with NM measurements in relation to the standard deviation of elevation estimates, although the mean values slowly drift away from the true elevation with increasing interference between two scatterers.

In Figure 10, the detection rate P_D is provided for the

case $N = 11$ w.r.t. normalized true elevation distance $\delta\alpha$. The red and blue colors denote M-SLIMMER and SLIMMER, respectively. The solid and dashed lines illustrate the results with $SNR = 10$ dB and 3 dB, respectively. If we define elevation resolution to be the minimum distance between façade and ground required to achieve at least 50% detection rate, then the resolution of M-SLIMMER is approximately one tenth of Rayleigh resolution better than the one of SLIMMER, given an SNR of 10dB before. We also analyzed P_D w.r.t. SNR for two typical elevation distances $\alpha = 1$ or 0.4, i.e., when two scatterers are spaced by one or two fifths of Rayleigh resolution. The results are shown in Figure 11(a). Moreover, false alarm rate P_F is illustrated in Figure 11(b) as a function of SNR for M-SLIMMER (red) and SLIMMER (blue), respectively. In this context, we simulate only one scatterer and define false alarm for the case when two scatterers are detected. These analyses confirm the fact that M-SLIMMER outperforms SLIMMER in all respects significantly. For the case $N = 6$, the gain of using multiple snapshots regarding

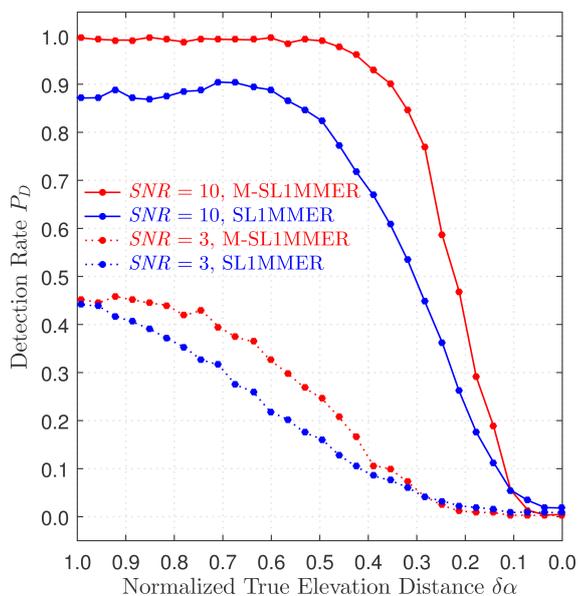


Fig. 10. Detection rate P_D w.r.t. normalized true elevation distance $\delta\alpha$ between façade and ground, for the case $N = 11$. Red: $SNR = 10$ dB with M-SLIMMER, orange: $SNR = 10$ dB with SLIMMER, violet: $SNR = 3$ dB with M-SLIMMER, and $SNR = 3$ dB with SLIMMER, respectively.

P_D and P_F is comparable to the case of $N = 11$.

VI. PRACTICAL DEMONSTRATION USING TANDEM-X DATA

In this section, M-SLIMMER is applied to the TanDEM-X data mentioned in Section II. The results are compared to those obtained using SLIMMER. Figure 12–14 show the reconstructed and color-coded elevation of the two test buildings in Figure 5, overlaid with intensity. From Figure 12 to 14, 21, 11 and 6 interferograms are used, respectively. The separated superimposed scatterers, from left to right estimated using M-SLIMMER and SLIMMER and from top to down of first and second layer, are illustrated respectively.

On top of the test buildings, reflections from building roof and façade are overlaid. In these practical examples, dominating scattering from roof (dark red) can be seen in the first layer, whereas the corresponding parts of façade (light red) are visible in the second layer. We do not expect many reflections from lower structures though, due to the large slope of the shell-like roof in front of the test buildings. It is evident that M-SLIMMER (left) using joint sparsity model significantly outperforms SLIMMER (right). In particular, when $N = 6$, i.e., using extremely small number of scenes, the second layer estimated using SLIMMER (lower right plot of Figure 14) is deteriorated by false alarms while M-SLIMMER still achieves reasonable results (lower left plot of Figure 14).

In particular, due to the significantly improved estimation accuracy, M-SLIMMER reconstructs some interesting details which was not accessible so far. For a practical demonstration, we calculated elevation distance between first and second layer for the double-scatterer case, which is shown in Figure 15. The red parallelogram marks the area where facade and roof are overlaid, cf. Figure 16. At the far-range side of this area,

the elevation distance amounts to approximately 22.60 [m] (cyan). Accordingly, the width of the roof can be calculated to be 18.27 [m], which agrees, up to the decimeter level, with what we estimated from the 3-D building model of Google Earth. Besides, the yellow parallelogram circumscribes the area where two neighboring windows in the diagonal direction, exemplified as S_1 and S_2 in Figure 16, are superimposed. Thus their elevation distance is more or less constant. This can be easily verified by comparing the SAR amplitude image to the optical one in Figure 16. As can be counted from Figure 16, the number of windows per floor adds up to 20, whereas there are only 10 extraordinarily bright pixels in the SAR amplitude image. In this area, the results with M-SLIMMER are much more homogeneous, which indicates a more robust reconstruction.

VII. CONCLUDING REMARKS

In this paper, a novel framework is proposed which can achieve precise TomoSAR reconstruction while significantly reducing the required number of images. The core idea is the exploitation of joint sparsity in iso-height SAR pixel groups that can be identified with the support of online available GIS data—2D building footprints. Experiments using bistatic TanDEM-X data stacks demonstrate the great potential of the proposed approach.

A few additional remarks might be helpful for further use of our results:

- The approach we proposed for grouping iso-height pixels can also be used for all other estimators that support multiple-snapshot (also referred to as multi-look) estimation, like NLS, MUSIC, etc.;
- Our experiments are performed over Las Vegas where moderate quality GIS data are available. For test areas with relatively poor 2-D footprint data, the benefit of the proposed approach that refine the iso-height models of each building using, e.g., SAR intensity map, could be even more evident;
- In the cities where the GIS data are not available, one can use the 2-D footprint reconstructed using a preliminarily retrieved TomoSAR point cloud [40] to alternately improve the TomoSAR estimation procedure;
- The proposed M-SLIMMER is a general spectral estimator, even though we applied it here for tomographic reconstruction;
- In our work, the mix $L_{1,2}$ norm is introduced to promote the joint sparsity prior. Depending on the applications, this constraint can be relaxed by replacing $L_{1,2}$ with $L_{p,2}$ norm with $p > 1$;

In the future, we will extend the proposed M-SLIMMER for higher dimensional spectral estimation problems, e.g., differential tomographic SAR reconstruction using mixed single- and multi-pass monostatic data stacks.

ACKNOWLEDGMENT

This work is supported by the Helmholtz Association under the framework of the Young Investigators Group

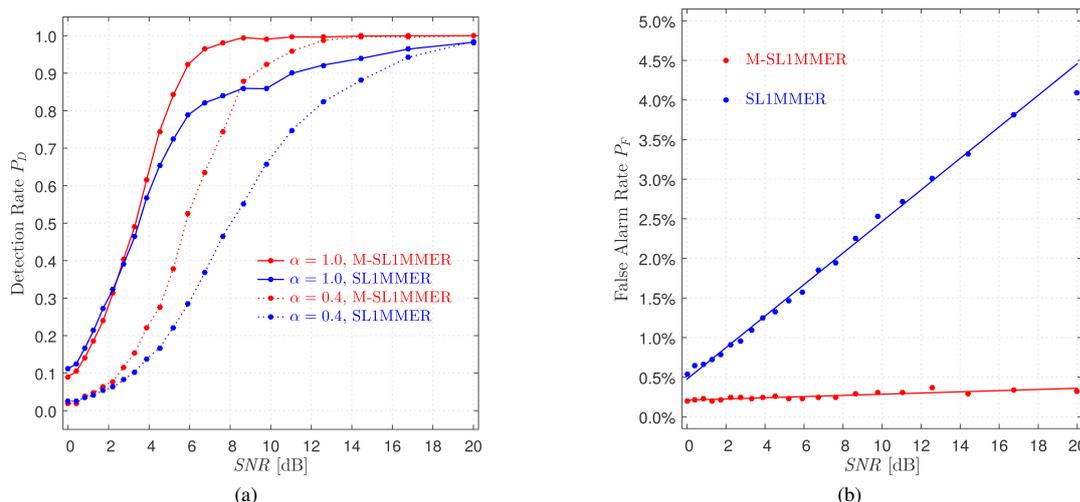


Fig. 11. Detection rate P_D and false alarm rate P_F w.r.t. SNR , for the case $N = 11$: (a) P_D for $\alpha = 1.0$ with M-SLIMMER (red), $\alpha = 1.0$ with SLIMMER (orange), $\alpha = 0.4$ with M-SLIMMER (violet), and $\alpha = 0.4$ with SLIMMER (blue), respectively. (b) P_F with M-SLIMMER (red), and SLIMMER (orange), respectively.

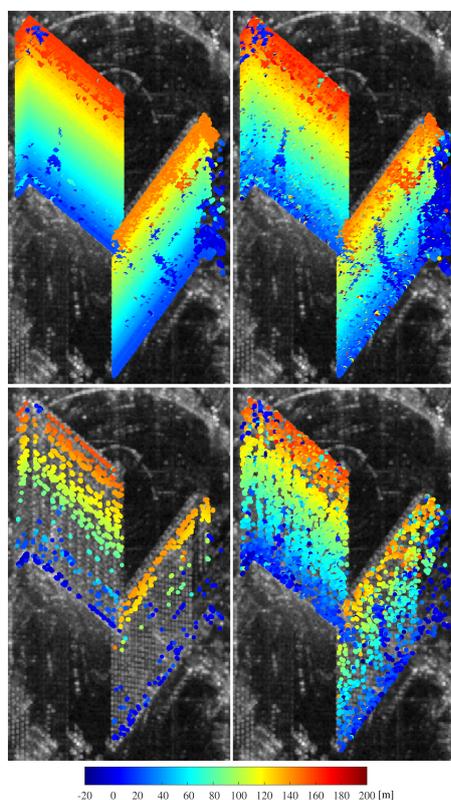


Fig. 12. Reconstructed and color-coded elevation of the two test buildings using 21 interferograms, visualized in two layers, overlaid with intensity. From top to down: first and second layer, respectively; from left to right: M-SLIMMER and SLIMMER, respectively.

”SiPEO” (VH-NG-1018, www.sipeo.bgu.tum.de) and International Graduate School of Science and Engineering, Technische Universität München (Project 6.08: ”4D City”).

REFERENCES

- [1] S. Gernhardt, N. Adam, M. Eineder, and R. Bamler, ”Potential of Very High Resolution SAR for Persistent Scatterer Interferometry in Urban Areas,” *Annals of GIS*, vol. 16, no. 2010-06, pp. 103–111, 2010.
- [2] X. X. Zhu and R. Bamler, ”Very High Resolution Spaceborne SAR Tomography in Urban Environment,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.
- [3] D. Reale, G. Fornaro, A. Pauciuolo, X. Zhu, and R. Bamler, ”Tomographic Imaging and Monitoring of Buildings With Very High Resolution SAR Data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 661–665, Jul. 2011.
- [4] G. Fornaro, A. Pauciuolo, D. Reale, X. Zhu, and R. Bamler, ”Sar tomography: an advanced tool for 4d spaceborne radar scanning with application to imaging and monitoring of cities and single buildings,” *IEEE Geosci. Remote Sensing Newslett*, pp. 10–18, 2012.
- [5] Y. Wang, X. X. Zhu, Y. Shi, and R. Bamler, ”Operational tomosar processing using terrasars-x high resolution spotlight stacks from multiple view angles,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, 2012, pp. 7047–7050.
- [6] X. X. Zhu and M. Shahzad, ”Facade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3541–3552, Jun. 2014.
- [7] G. Fornaro, F. Lombardini, A. Pauciuolo, D. Reale, and F. Viviani, ”Tomographic Processing of Interferometric SAR Data: Developments, applications, and future research perspectives,” *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 41–50, Jul. 2014.
- [8] X. X. Zhu and R. Bamler, ”Tomographic SAR Inversion by L1-Norm Regularization: The Compressive Sensing Approach,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839–3846, Oct. 2010.
- [9] A. Budillon, A. Evangelista, and G. Schirinzi, ”Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 488–499, Jan. 2011.
- [10] X. X. Zhu and R. Bamler, ”Super-Resolution Power and Robustness of Compressive Sensing for Spectral Estimation With Application to Spaceborne Tomographic SAR,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 247–258, Jan. 2012.
- [11] —, ”Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 8, pp. 3150–3157, Aug. 2012.
- [12] —, ”Superresolving SAR Tomography for Multidimensional Imaging of Urban Areas: Compressive sensing-based TomoSAR inversion,” *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 51–58, Jul. 2014.
- [13] A. Ferretti, C. Prati, and F. Rocca, ”Analysis of Permanent Scatterers in SAR interferometry,” in *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, vol. 2, 2000, pp. 761–763 vol.2.

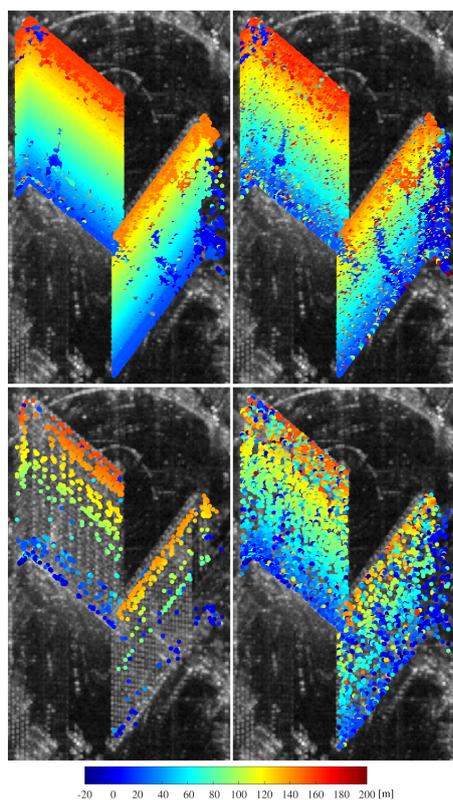


Fig. 13. The same results as Figure 12, but using 11 interferograms.

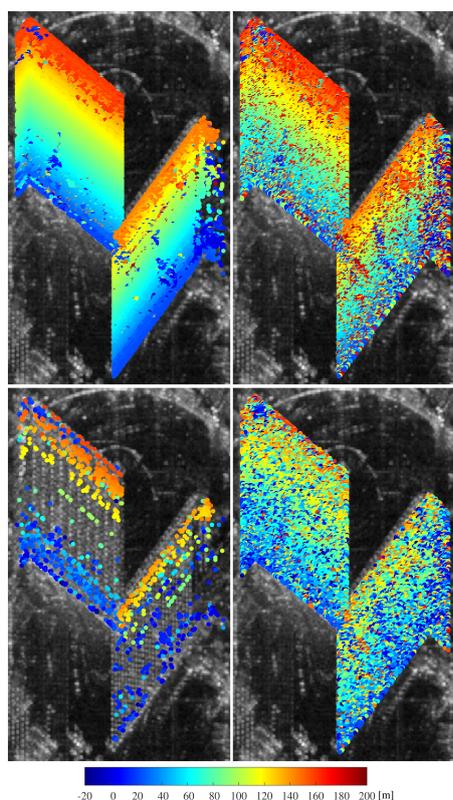


Fig. 14. The same results as Figure 12, but using 6 interferograms only.

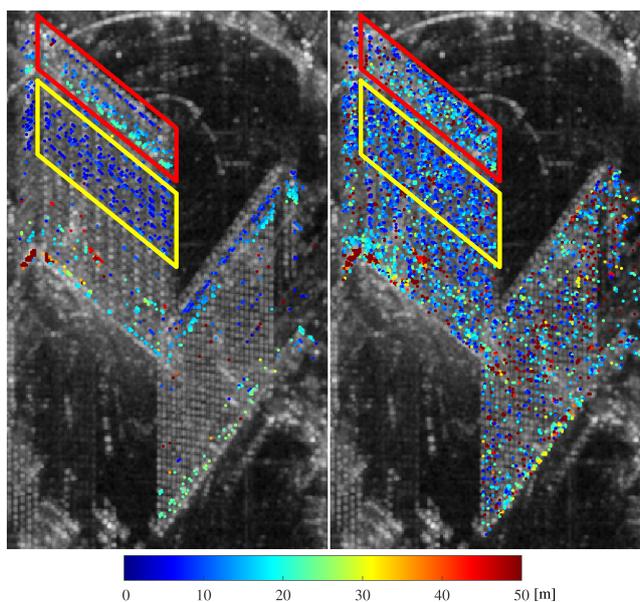


Fig. 15. Elevation distance between first and second layer with $N = 21$. Left: M-SLIMMER; right: SLIMMER.

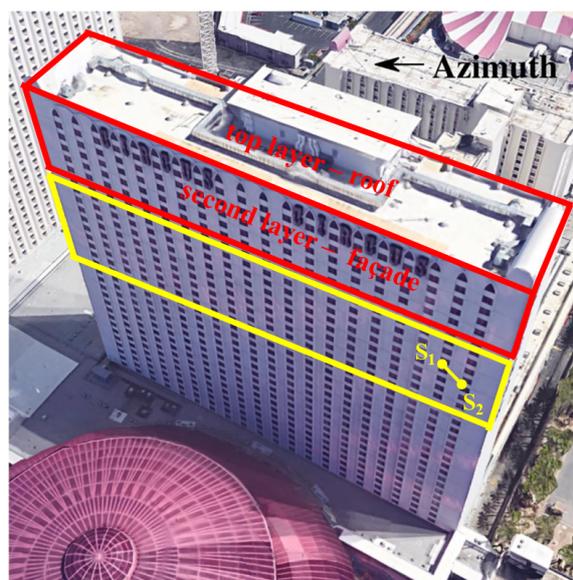


Fig. 16. Optical image of the left test building in Figure 15 ©Google.

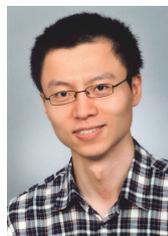
- [14] B. M. Kampes, *Radar Interferometry: Persistent Scatterer Technique*. Springer Science & Business Media, Sep. 2006.
- [15] N. Adam, B. Kampes, M. Eineder, J. Worawattanamateekul, and M. Kircher, "The development of a scientific permanent scatterer system," in *ISPRS Workshop High Resolution Mapping from Space, Hannover, Germany*, 2003.
- [16] G. Fornaro, F. Lombardini, and F. Serafino, "Three-dimensional multi-pass SAR focusing: experiments with long-term spaceborne data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 702–714, Apr. 2005.
- [17] F. Lombardini, "Differential tomography: a new framework for SAR interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 1, pp. 37–44, Jan. 2005.
- [18] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "A new algorithm for processing interferometric data-stacks: Squeesar,"

- Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 9, pp. 3460–3470, Sept 2011.
- [19] Y. Wang, X. X. Zhu, and R. Bamler, “Retrieval of phase history parameters from distributed scatterers in urban areas using very high resolution sar data,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 73, pp. 89 – 99, 2012.
- [20] M. Schmitt and U. Stilla, “Adaptive multilooking of airborne single-pass multi-baseline insar stacks,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 305–312, Jan 2014.
- [21] R. Guo and X. X. Zhu, “High-Rise Building Feature Extraction Using High Resolution Spotlight TanDEM-X Data,” in *EUSAR 2014; 10th European Conference on Synthetic Aperture Radar; Proceedings of*, Jun. 2014, pp. 1–4.
- [22] “Stats - OpenStreetMap Wiki.” [Online]. Available: <http://wiki.openstreetmap.org/wiki/Statistics>
- [23] H. Fan, A. Zipf, Q. Fu, and P. Neis, “Quality Assessment for Building Footprints Data on OpenStreetMap,” *Int. J. Geogr. Inf. Sci.*, vol. 28, no. 4, pp. 700–719, Apr. 2014.
- [24] M. Haklay, “How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets,” *Environment and Planning B: Planning and Design*, vol. 37, no. 4, pp. 682–703, 2010.
- [25] M. Helbich, C. Amelunxen, and P. Neis, “Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata,” in *AGILE 2011*, 2011, gIScience, Department of Geography, University of Heidelberg, Berliner Strasse 48, 69120 Heidelberg, Germany.
- [26] P. Mooney, P. Corcoran, and A. C. Winstanley, “Towards Quality Metrics for OpenStreetMap,” in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 514–517.
- [27] C. K. Robert Hecht, “Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time,” *ISPRS International Journal of Geo-Information*, vol. 2, pp. 1066–1091, 2013.
- [28] A. Reigber and A. Moreira, “First demonstration of airborne sar tomography using multibaseline l-band data,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 38, no. 5, pp. 2142–2152, Sep 2000.
- [29] G. Fornaro, F. Serafino, and F. Soldovieri, “Three-dimensional focusing with multipass SAR data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 507–517, Mar. 2003.
- [30] G. Fornaro, D. Reale, and F. Serafino, “Four-Dimensional SAR Imaging for Height Estimation and Monitoring of Single and Double Scatterers,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1, pp. 224–237, Jan. 2009.
- [31] X. X. Zhu and R. Bamler, “Let’s Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography,” *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 735–739, Jul. 2011.
- [32] X. X. Zhu, “Very high resolution tomographic sar inversion for urban infrastructure monitoring: a sparse and nonlinear tour,” Ph.D. dissertation, Technische Universität München, 2011.
- [33] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer, Dec. 2003.
- [34] D. Malioutov, M. Cetin, and A. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
- [35] Y. Eldar and H. Rauhut, “Average Case Analysis of Multichannel Sparse Recovery Using Convex Relaxation,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.
- [36] E. Aguilera, M. Nannini, and A. Reigber, “Multi-signal compressed sensing for polarimetric SAR tomography,” in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, Jul. 2011, pp. 1369–1372.
- [37] M. Schmitt and U. Stilla, “Compressive Sensing Based Layover Separation in Airborne Single-Pass Multi-Baseline InSAR Data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 313–317, Mar. 2013.
- [38] R. Bamler, M. Eineder, N. Adam, X. Zhu, and S. Gernhardt, “Interferometric Potential of High Resolution Spaceborne SAR,” *Photogrammetrie - Fernerkundung - Geoinformation*, vol. 2009, no. 5, pp. 407–419, Nov. 2009.
- [39] N. Adam, R. Bamler, M. Eineder, and B. Kampes, “Parametric estimation and model selection based on amplitude-only data in ps-interferometry,” in *ESA FRINGE Workshop*, 2005.
- [40] M. Shahzad and X. X. Zhu, “Robust reconstruction of building facades for large areas using spaceborne tomosar point clouds,” *Geoscience and*

Remote Sensing, IEEE Transactions on, vol. 53, no. 2, pp. 752–769, Feb 2015.



Xiao Xiang Zhu (S'10-M'12-SM'14) received the bachelor degree in space engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2006. She received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her Habilitation in the field of signal processing from Technische Universität München (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively. Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China and the University of Tokyo, Tokyo, Japan in 2009, 2014 and 2015, respectively. Since 2011, she is a scientist with the Remote Sensing Technology Institute at the German Aerospace Center (DLR), Oberpfaffenhofen, where she is the head of the Team Signal Analysis. Since 2013, she is also a Helmholtz Young Investigator Group Leader and appointed as TUM junior fellow. In 2015, she is appointed as the Professor for Signal Processing in Earth Observation at TUM. Prof. Zhu is an associate Editor of *IEEE Transactions on Geoscience and Remote Sensing*. Her main research interests are: advanced InSAR techniques such as high dimensional tomographic SAR imaging and SqueeSAR; computer vision in remote sensing including object reconstruction and multi-dimensional data visualization; and modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multi/hyperspectral image analysis.



Nan Ge received the bachelor and the master degree in geoenvironmental engineering at TU Clausthal, Germany. He has been pursuing his doctor's degree since 2013 in the Helmholtz Young Investigators Group (SiPEO) at the Remote Sensing Technology Institute of German Aerospace Center (DLR), Germany. His work mainly involves developing advanced tomographic techniques for very high resolution SAR systems.



Muhammad Shahzad (S'12) received the B.E. degree in electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, and the M.Sc. degree in autonomous systems from the Bonn Rhein Sieg University of Applied Sciences, Sankt Augustin, Germany. Since December 2011, he has been working towards his Ph.D. degree at Signal Processing in Earth Observation (SiPEO), Technische Universität München (TUM), Munich, Germany. His research topic is the automatic 3-D reconstruction of objects from point clouds retrieved from spaceborne synthetic-aperture-radar image stacks.

E Schmitt, M., Shahzad, M., Zhu, X. X., 2015. Reconstruction of individual trees from multi-aspect TomoSAR data. Remote Sensing of Environment 165: 175-185

<http://www.sciencedirect.com/science/article/pii/S0034425715300110>



Reconstruction of individual trees from multi-aspect TomoSAR data



Michael Schmitt*, Muhammad Shahzad, Xiao Xiang Zhu*

Helmholtz Young Investigators Group "SiPEO", Technische Universität München (TUM), Arcisstr. 21, 80333 Munich, Germany

ARTICLE INFO

Article history:

Received 9 January 2015

Received in revised form 8 May 2015

Accepted 16 May 2015

Available online xxxx

Keywords:

Synthetic aperture radar (SAR)

Multi-aspect

SAR tomography

Trees

3D reconstruction

Forested areas

Point cloud segmentation

ABSTRACT

The localization and reconstruction of individual trees as well as the extraction of their geometrical parameters is an important field of research in both forestry and remote sensing. While the current state-of-the-art mostly focuses on the exploitation of optical imagery and airborne LiDAR data, modern SAR sensors have not yet met the interest of the research community in that regard. This paper presents a prototypical processing chain for the reconstruction of individual deciduous trees: First, single-pass multi-baseline InSAR data acquired from multiple aspect angles are used for the generation of a layover- and shadow-free 3D point cloud by tomographic SAR processing. The resulting point cloud is then segmented by unsupervised mean shift clustering, before ellipsoid models are fitted to the points of each cluster. From these 3D ellipsoids the relevant geometrical tree parameters are extracted. Evaluation with respect to a manually derived reference dataset prove that almost 74% of all trees are successfully segmented and reconstructed, thus providing a promising perspective for further research toward individual tree recognition from SAR data.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The analysis of individual trees in remote sensing data has been a widely studied field of research for quite some years now. This is mainly caused by the fact that it is an important topic for a variety of application fields, among which the most relevant certainly is the task of sustainable forest management: In many countries, single-tree related parameters are used as a basis for forest inventory, e.g. tree species, mean tree height or timber volume. Still, most of these variables are collected manually by measuring sample plots in staff-, time-, and thus cost-intensive field surveys (Shiver & Borders, 1996), although remote sensing-based methods have been investigated for some decades now (Fagan & DeFries, 2009). Another exemplary application is the extraction of individual urban trees in order to generate city tree cadastres or to provide additional objects for geoinformation systems (GIS) and 3D city models (Straub & Heipke, 2001).

Since the mid-1990s an abundance of literature has been published on the detection and localization of individual trees (Chang, Eo, Kim, & Kim, 2013; Chen, Baldocchi, Gong, & Kelly, 2006; Leckie et al., 2005; Pollock, 1996; Wulder, Niemann, & Goodenough, 2000) as well as the delineation of their tree crowns (Culvenor, 2002; Erikson, 2003; Jing, Hu, Li, & Noland, 2012; Koch, Heyder, & Weinacker, 2006; Pouliot, King, Bell, & Pitt, 2002) both from aerial images and LiDAR-derived canopy height models. In contrast to that, the analysis of forested

areas on the single-tree level by means of synthetic aperture radar (SAR) remote sensing has not yet met the interest of the community, although modern sensors have reached sub-meter resolutions down to the decimeter-range in recent years, enabling a detailed mapping of natural and urban scenes (Brenner, 2012; Mittermayer, Wollstadt, Prats-Iraola, & Scheiber, 2014). Until now, most of the research exploiting SAR for forest mapping purposes has focused on large-scale forest classification (Perko, Raggam, Deutscher, Gutjahr, & Schardt, 2011) or canopy height model (CHM) reconstruction using X-band sensors (Izzawati, Wallington, & Woodhouse, 2006). In addition, another research direction based on longer-wavelength, canopy-penetrating SAR (e.g. L-band or P-band) is aiming at volume and biomass retrieval using allometric equations (Le Toan, Beaudoin, Riou, & Guyon, 1992), regression models (Englhart, Keuck, & Siegert, 2011), polarimetric SAR interferometry (Neumann, Ferro-Famil, & Reigber, 2010) or SAR tomography (TomoSAR) (Frey, Morsdorf, & Meier, 2008). Only recently, the very first investigations toward single-tree recognition in airborne SAR data have been introduced (Schmitt, Brück, Schönberger, & Stilla, 2013).

Considering the advantages of radar remote sensing, i.e. the active sensor principle, the independence of sunlight and weather conditions, and the potential to map wide swaths in short time, this article intends to further ignite the discussion about the SAR-based mapping of forests on the individual-tree level by proposing a method for the reconstruction of trees from tomographic single-pass SAR data. Exploiting an airborne millimeterwave system equipped with four simultaneously receiving, spatially separated antennas (and thus, multiple independent baselines), first a three-dimensional point cloud is generated by TomoSAR processing. In analogy to the approaches based on 3D LiDAR point clouds, in the

* Corresponding authors.

E-mail addresses: m.schmitt@tum.de (M. Schmitt), muhhammad.shahzad@tum.de (M. Shahzad), xiao.zhu@dlr.de (X.X. Zhu).

URL: <http://www.sipeo.bgu.tum.de> (X.X. Zhu).

second step a clustering algorithm is applied to the TomoSAR point cloud in order to segment the individual trees in 3D space. While the studies of Morsdorf et al. (2004) or Gupta, Weinacker, and Koch (2010) are based on k -means clustering, which needs the number of expected clusters and an initialization of their centers as a priori knowledge, in the presented work the unsupervised mean-shift clustering algorithm is used, which enables a fully automatic procedure (Comaniciu & Meer, 2002) and has already successfully been applied to the reconstruction of building facades (Shahzad & Zhu, 2015) and roofs (Shahzad & Zhu, 2014) in TomoSAR point clouds. Finally, rotational ellipsoids are used to model the segmented clusters in order to approximate the tree crown shapes. From these ellipsoids the tree positions, heights and crown diameters can be extracted. This tree reconstruction strategy is evaluated using airborne millimeterwave InSAR data acquired from multiple aspect angles. The purpose of this study is to demonstrate the potential of millimeterwave SAR remote sensing for the reconstruction of individual trees including a first estimate of the achievable accuracy level.

2. Generation of point clouds from airborne tomographic SAR data

2.1. The principle of SAR tomography

The first step of the methodology presented in this paper is the generation of a point cloud of the forested area under investigation from single-pass multi-baseline InSAR data. In single-pass InSAR configurations, two (or more) coherent SAR images are acquired simultaneously by sensors separated by a certain spatial baseline. This simultaneity is necessary because of temporal decorrelation effects, which especially occur over vegetated areas due to wind or growth effects or other, potentially more rapid changes of the scene (Ahmed, Siqueira, Hensley, Chapman, & Bergen, 2011). Since only recently satellite missions offering spaceborne single-pass InSAR data have been realized (Moreira et al., 2004), airborne multi-antenna systems have been the only possibility to acquire such data so far.

As discussed by Hoekman and Varekamp (2001) already, unfortunately the side-looking SAR imaging geometry leads to severe height, displacement and occlusion errors for individual trees due to the well-known layover and shadowing effects (cf. Fig. 1). While the shadowed scene parts can be filled with information by multi-aspect data fusion, the layover-affected resolution cells need to be treated by tomographic SAR inversion: Basically, SAR tomography aims at creating

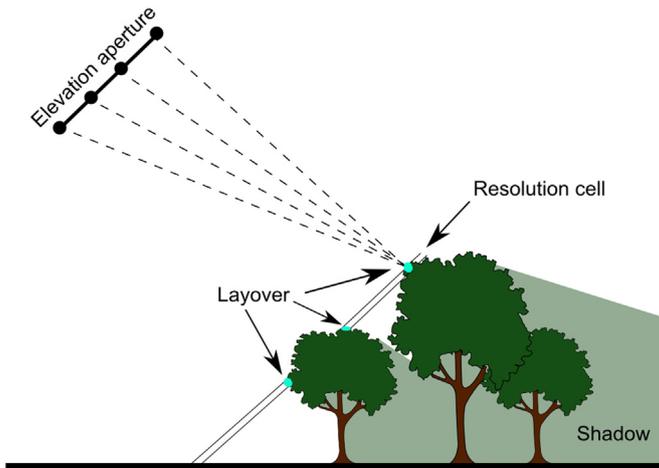


Fig. 1. Side-looking SAR viewing geometry for multi-baseline system illuminating a forest scene. While several scattering components are overlaid in one resolution cell, some trees are fully or partially shadowed by larger trees in front of them.

a second synthetic aperture along elevation direction. However, in opposition to the high number of densely and regularly spaced azimuth samples, for the elevation aperture comparably few observations per resolution cell are available in the form of a stack of N coregistered single look complex (SLC) SAR images acquired from slightly different antenna positions (Gini & Lombardini, 2005; Reigber & Moreira, 2000; Zhu & Bamler, 2010). The complex measurement stored in a pixel of the n th acquisition with a baseline B_n between the respective slave antenna n and the master antenna is the integral of the reflected signal weighted by a linear phase term:

$$z_n = \int_{s_{min}}^{s_{max}} x(s) \exp(j \cdot \phi_n(s)) ds, \quad (1)$$

where $x(s)$ is the reflectivity function along elevation, $[s_{min}; s_{max}]$ defines the relevant part of the elevation profile, and j is the imaginary unit. $\phi(s)$ is the phase related to system parameters and two-way distance between the scattering center at elevation s and the receiving antenna.

Using SAR sensors operating with rather long wavelengths (e.g. L- or P-band), this enables a fully three-dimensional, canopy-penetrating imaging of vegetation volumes. If, however, shorter wavelengths are used, where canopy penetration is less likely, Eq. (1) can also be approximated by discretizing and sparsifying the reflectivity profile. The measurement vector is then formulated as

$$\mathbf{z} = \sum_{k=1}^K x_k \mathbf{a}(s) + \mathbf{n} = \mathbf{A}(\mathbf{s}) \mathbf{x} + \mathbf{n}. \quad (2)$$

In this notation, $\mathbf{x} = [x_1, \dots, x_K]^T$ is the source signal vector containing the complex reflectivities of the K discrete scattering contributions, and \mathbf{n} represents complex circularly symmetric Gaussian noise. $\mathbf{A}(\mathbf{s}) = [\mathbf{a}(s_1), \dots, \mathbf{a}(s_K)]$ is the $N \times K$ system matrix of K concatenated steering vectors each of which corresponding to one backscattering source. The task of SAR tomography basically is to invert this imaging model in order to reconstruct either a continuous reflectivity profile in the case of volumetric scenes or a discrete reflectivity profile in case of scenes containing mostly discrete scatterers. Since the tree reconstruction approach presented in this paper is based on point cloud segmentation, here a discrete TomoSAR model is chosen aiming at sparse reflectivity profiles. The utilized approach was first presented in Schmitt & Stilla (2014b) and tested for urban scenes, but has been shown to work well for decimeter-resolution millimeterwave SAR data of forested areas as well (Schmitt & Stilla, 2014a). Its core algorithm is quickly recapitulated in the following section. Due to the low level of canopy penetration provided by millimeterwave SAR, of course the situation would be a little different for very dense forests: In this case it would be unlikely that tree crowns overlap with each other or with any ground contribution in the SAR viewing direction, unless they differ strongly in height, or, e.g., if forest clearing appears. For dense forests it could therefore be advisable to only use standard SAR interferometry for height reconstruction followed by the multi-aspect data fusion step described in Section 2.4 of this paper in order to avoid unnecessary computational costs and methodical complexity.

2.2. Layover separation and height reconstruction

A flowchart of the employed TomoSAR algorithm is shown in Fig. 2, whereas for the scope of this paper only the part enclosed by the rectangle is described. For the remaining pre- and postprocessing steps, the reader is referred to Schmitt and Stilla (2014b). Therefore, here the starting situation is the sample coherence matrix for every pixel in the stack of coregistered SAR images, as well as an estimate \hat{K} of the model order K , i.e. of the expected number of scattering contributions in the sparse reflectivity profile. The sample covariance matrix is the

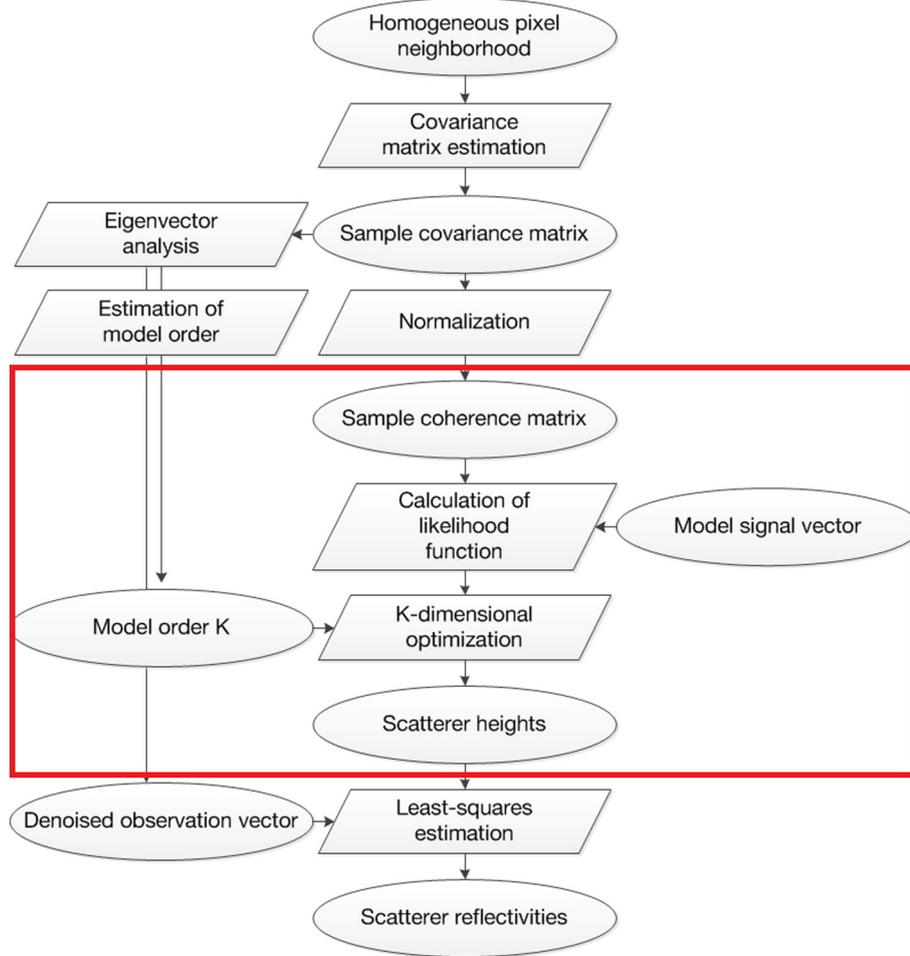


Fig. 2. Flowchart of the TomoSAR algorithm used in the presented work.

result of normalizing the observed sample covariance matrix and is defined as

$$\hat{\mathbf{\Gamma}} = \begin{bmatrix} 1 & |\hat{\gamma}_{12}| \exp(j\hat{\phi}_{12}) & \dots & |\hat{\gamma}_{1N}| \exp(j\hat{\phi}_{1N}) \\ |\hat{\gamma}_{21}| \exp(j\hat{\phi}_{21}) & 1 & \dots & |\hat{\gamma}_{2N}| \exp(j\hat{\phi}_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ |\hat{\gamma}_{N1}| \exp(j\hat{\phi}_{N1}) & |\hat{\gamma}_{N2}| \exp(j\hat{\phi}_{N2}) & \dots & 1 \end{bmatrix}, \quad (3)$$

which means it contains all the interferometric, i.e. phase-related information of the respective pixel: $|\hat{\gamma}_{ij}|$ is the magnitude of coherence between acquisitions i and j , while $\hat{\phi}_{ij}$ is the corresponding interferometric phase. Based on the assumption of circularly-symmetric zero-mean Gaussian SAR pixels, this sample coherence matrix can be used as input to the likelihood function

$$\mathcal{L}(\hat{\mathbf{\Gamma}}; \tilde{\mathbf{z}}(\mathbf{s})) = \frac{1}{\pi^N \det(\hat{\mathbf{\Gamma}})} \exp(-\tilde{\mathbf{z}}^H(\mathbf{s}) \hat{\mathbf{\Gamma}}^{-1} \tilde{\mathbf{z}}(\mathbf{s})), \quad (4)$$

where $\tilde{\mathbf{z}}(\mathbf{s})$ resembles the steering vector described in Section 2.1, but contains a hypothetical mixture of \hat{K} scattering contributions:

$$\tilde{\mathbf{z}}(\mathbf{s}) = [\tilde{z}_1(\mathbf{s}), \dots, \tilde{z}_N(\mathbf{s})]^T, \quad (5)$$

where

$$\tilde{z}_n(\mathbf{s}) = \sum_{k=1}^{\hat{K}} \exp(j\tilde{\varphi}_n(s_k)) \quad (6)$$

and $\tilde{\varphi}_n(s_k)$ is the absolute signal phase caused by scatterer k at elevation s_k minus any “flat earth” phase component potentially measured for the reference surface at elevation $s = 0$. Using the elevation-to-height conversion $h = s \cdot \sin(\theta)$, where θ is the off-nadir angle of the radar line-of-sight, Eq. (4) can be maximized with respect to the unknown scatterer heights $\hat{\mathbf{h}} = [\hat{h}_1, \dots, \hat{h}_k]$. Since the denominator of Eq. (4) stays constant during the optimization, the simplified estimator finally becomes:

$$\hat{\mathbf{h}} = \arg \max_{\mathbf{h} \in [h_{min}, h_{max}]} \exp(-\tilde{\mathbf{z}}(\mathbf{h})^H \hat{\mathbf{\Gamma}}^{-1} \tilde{\mathbf{z}}(\mathbf{h})). \quad (7)$$

This maximization can be solved by a simple \hat{K} -dimensional grid search: For this, the objective function in Eq. (7) is calculated for \hat{K} hypothetical scattering heights taken from a the pre-defined discretized search interval $[h_{min}; h_{max}]$. Then the maximum is found in the \hat{K} -dimensional search space aggregated from the \hat{K} search intervals.

2.3. Geocoding

The result of the TomoSAR inversion described in Section 2.2 is a three-dimensional point cloud in radar geometry, i.e. the data are still

stored in the coordinate system of the SAR imagery, although now for every resolution cell \hat{K} scattering heights are available. This point cloud must then be transformed from the so-called slant range plane to a geodetic reference frame. For this task, the non-linear equation system

$$R - \|\mathbf{p} - \mathbf{s}_M(t)\| = 0 \quad (8)$$

$$\mathbf{v}_M(\mathbf{p} - \mathbf{s}_M(t)) = 0 \quad (9)$$

$$\frac{p_x^2 + p_y^2}{(a+h)^2} + \frac{p_z^2}{(b+h)^2} - 1 = 0 \quad (10)$$

has to be solved for the unknown 3D-coordinates $\mathbf{p} = [p_x, p_y, p_z]^T$ in world geometry (Schwäbisch, 1998). In this context, R is the slant range distance between \mathbf{p} and the master antenna position $\mathbf{s}_M(t) = [s_{MX}(t), s_{MY}(t), s_{MZ}(t)]^T$ at time t . It has to be mentioned that the sensor velocity $\mathbf{v}_M = [v_{MX}, v_{MY}, v_{MZ}]^T$ in Eq. (9) is modeled as constant and referring to a linearized flight trajectory, while the data are assumed to be zero-Doppler processed. Although this is a simplification not valid for arbitrary mission designs, the extension to non-linear trajectories or non-zero-Doppler processed data is straight-forward. a and b in Eq. (10) denote the semi-major and semi-minor axes of the reference Earth ellipsoid, respectively, while h is the previously reconstructed height of the corresponding scatterer above the ellipsoid's surface.

2.4. Fusion of point clouds from multiple aspects

As described in Section 2.1, the side-looking SAR imaging geometry causes not only layover, but also shadowing effects that lead to scene parts that don't contain any exploitable measurements behind elevated objects. For the analysis of urban areas, multi-aspect data fusion has been demonstrated as a viable solution for this problem (e.g. Schmitt, 2015). Therefore, the same approach is used for fusing point clouds of multiple viewing directions in the context of this work. If several independent point clouds of the forested scene under investigation can be generated by the procedure described in the last sections, they may need to either be registered via their corresponding flight navigation data a priori (Schmitt, Maksymiuk, Magnard, & Stilla, 2013) or using point cloud-based registration methods *a posteriori* (e.g. Makadia, A. P., & Daniilidis, 2006; or Gernhardt, Cong, Eineder, Hinz, & Bamler, 2012). If the data then are aligned in the same common reference system this way, an additional fusion step was proved to be beneficial: By combining the individual point clouds intelligently, not only the accuracy of the 3D data is improved, but also the number of points in the point cloud can be significantly reduced without any information loss. This greatly benefits any further processing due to reduction of computational costs and memory requirements (Schmitt, 2015). This fusion step can be implemented by a voxel-space-based strategy: In order to impose a voxel-space onto the registered, unstructured 3D point clouds, a regular point cloud containing the centers of the desired voxel space has to be created. This is done via a k -d tree of dimension $k = 3$. A k -d tree is a multidimensional binary search tree, which serves as a space-partitioning data structure for storage of information to be retrieved by associative searches in a k -dimensional space (Bentley, 1975): If data are represented as a k -d tree, then each data point is stored as a node in the tree. Every non-leaf node implicitly generates a splitting hyperplane that divides the space into two parts (i.e. half-spaces). Points to the left of this hyperplane are then represented by the left subtree of that node, and points on the right are represented by the right subtree. Since every node in the tree is associated with one of the k dimensions the hyperplane is chosen perpendicular to that dimension's axis. Each split can be denoted using the dimension number and split value, whereas the splits are arranged in order to balance the tree, i.e. its maximum depth is kept as small as possible. If points are queried, the k -d tree search first locates the

respective point in its appropriate node and then searches nearby leaves in the tree until it can guarantee that the correct point has been found. Due to their efficient storage structure, k -d trees are particularly useful for multi-dimensional search operations such as range searches or nearest neighbor searches. Besides, an abundance of ready-to-use implementations in different programming languages is available online, many of them open source.

After creating the voxel-space defining point cloud this way, the irregularly sampled 3D TomoSAR point clouds are examined with the established k -d tree by applying a range search based on the Chebychev distance. This distance metric is defined for two points $\mathbf{p} = [p_1, \dots, p_n]^T$ and $\mathbf{q} = [q_1, \dots, q_n]^T$ as

$$\lim_{k \rightarrow \infty} \left(\sum_{i=1}^n |p_i - q_i|^k \right)^{\frac{1}{k}} = \max_{i \in \{1, \dots, n\}} (|p_i - q_i|), \quad (11)$$

which means it defines the distance between \mathbf{p} and \mathbf{q} as the greatest of their differences along any coordinate dimension. For three-dimensional vectors and a Chebyshev distance of $\frac{d}{2}$, this corresponds to a cube with an edge length of d . Therefore, the distance query results in a list of voxels and the corresponding points from the unstructured point clouds contained in that voxel (see Fig. 3). The final point cloud fusion is then realized by returning the mean value of the coordinates of all points contained in each voxel. The core parameter of this procedure is the edge length d of the voxel cubes, which basically defines both the resolution of the fused point cloud as well as the degree of data and noise reduction. As has been shown in Schmitt (2015), while the combination of multi-aspect data already helps to fill previously shadowed scene parts, this fusion strategy additionally provides a measurable improvement of the 3D accuracy of the reconstructed 3D points, as well as a significant reduction of the number of points.

3. Segmentation and reconstruction of individual trees

3.1. Point cloud segmentation

The 3D points resulting from the TomoSAR processing explained in Section 2 are clustered by the mean shift algorithm as described by

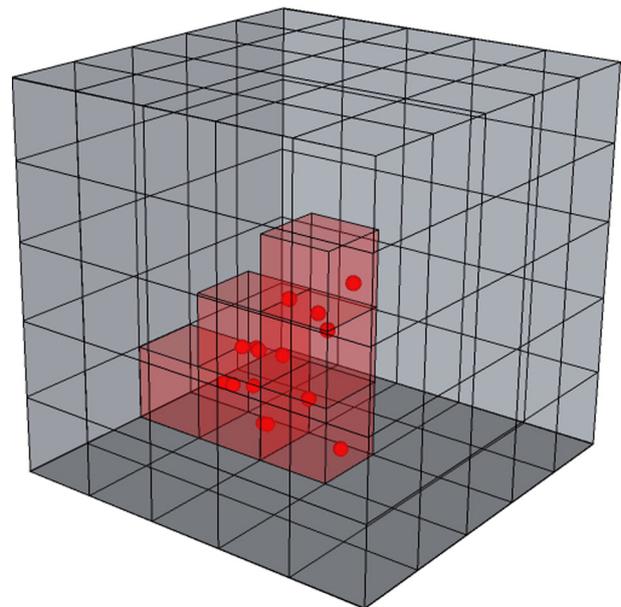


Fig. 3. Discretization of the scene by imposing a voxel space. The red cubes indicate voxels containing at least one 3D TomoSAR point. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

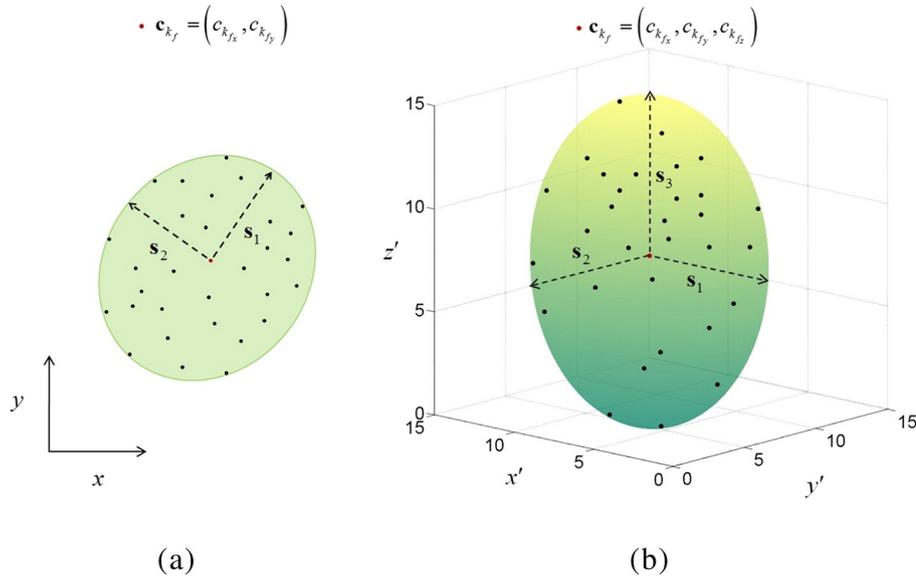


Fig. 4. Illustration of the ellipsoid modeling: (a) MVEE computed using 3D points projected onto the xy -plane; s_1 and s_2 are the computed semi-axes of the MVEE. (b) The MVEE of (a) is extruded in z -direction both upwards and downwards forming a 3D ellipsoid with a third semi-axis denoted as s_3 . x' , y' and z' in (b) represent axes of the local coordinate system aligned to the ellipsoid semi-axes. The red points in both (a) and (b) represent the ellipsoid centers \mathbf{c}_{k_f} . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Comaniciu and Meer (2002) in order to extract individual trees. Since the tree crowns generally show a comparably high point density when projected into the xy -plane, the points are clustered in 2D space, i.e. the feature space consists of xy -coordinates only. The kernel density estimate at any point \mathbf{p}_i of the n points is given by the expression

$$D_{\mathbf{p}_i} = \frac{c}{nb^3} \sum_{j=1}^n g\left(\left\|\frac{\mathbf{p}_i - \mathbf{p}_j}{b}\right\|^2\right), \quad (12)$$

where b is the bandwidth parameter and $g(x)$ is a non-negative, non-increasing, piecewise continuous function with definite integral, i.e. $\int_0^\infty g(x)dx < \infty$. Based on the concept of kernels discussed by Cheng (1995) as well as Comaniciu and Meer (2002), the function $g(x)$ is defined as the profile of the radially symmetric kernel $G(x)$ satisfying

$$G(x) = cg\left(\|x\|^2\right), \quad (13)$$

where c is a normalization constant ensuring that $G(x)$ integrates to 1. Different kernels, such as the Epanechnikov kernel and the Gaussian kernel can be used to define the density $D_{\mathbf{p}_i}$. Mean shift clustering essentially seeks modes of the kernel density estimates and works iteratively by shifting every data point toward the weighted mean of points within its neighborhood (defined to be cylindrical in the presented case). The shift vector $\mathbf{m}(\mathbf{p}_i)$ always points toward the direction of the maximum increase in the density $D_{\mathbf{p}_i}$, and is computed as

$$\mathbf{m}(\mathbf{p}_i) = \frac{\sum_{j=1}^n \mathbf{p}_j \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{b^2}\right)}{\sum_{j=1}^n \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{b^2}\right)} - \mathbf{p}_i. \quad (14)$$

The iteration process continues until there is no or only little shift in $\mathbf{m}(\mathbf{p}_i)$ anymore, i.e. the length of the shift vector $\mathbf{m}(\mathbf{p}_i)$ is close to 0. Due

to the gradient ascent nature, the mean shift algorithm returns clusters using the concept *attraction of basin*, i.e. those points whose trajectories lead to the same mode form the basin of attraction for that mode and are clustered into one group. The clustering procedure is repeated until all points are assigned to their respective modes.

Clustering via mean shift is a non-parametric procedure in the sense that it does not require the number of clusters a priori, nor does it need any pre-defined model for the shape of the resulting clusters. Nevertheless, it still does require a bandwidth parameter (corresponding to the radius of the kernel), which affects the number of clusters, i.e. the number of modes, that are returned by the algorithm. However, unlike other clustering algorithms such as k -means, fuzzy c -means, expectation maximization etc., the bandwidth parameter has some physical meaning and can be set based on prior knowledge such as the expected average diameter of the tree crowns in the scene.

3.2. Ellipsoid modeling

Once the clustering is done, the individual tree clusters are modeled in three dimensions using generalized tri-axial ellipsoids that are aligned to the z -axis. For this purpose, parameters of an arbitrarily oriented minimum volume enclosing ellipse (MVEE) are estimated by first projecting points belonging to individual tree clusters onto the

Table 1
MEMPHIS sensor parameters used during the 2013 measurement campaign.

Sensor	MEMPHIS
Carrier frequency	35 GHz (Ka-band)
Wavelength	8.5 mm
Range bandwidth	900 MHz
Nominal depression angle	35°
Resolution	
Azimuth	8.2 cm
Range	16.7 cm
Pixel spacing	
Azimuth	5.3 cm
Range	16.7 cm
Available baselines	5.5 cm, 11 cm, 16.5 cm, 22 cm, 27.5 cm
Approximate ambiguity heights	180 m, 90 m, 60 m, 45 m, 36 m

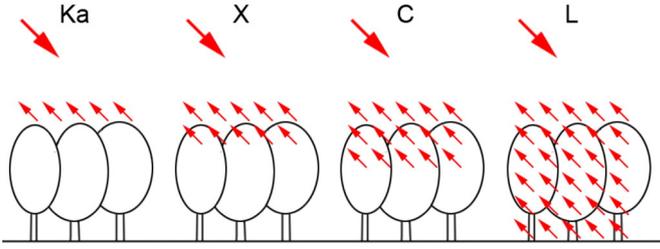


Fig. 5. Expected canopy penetration of common SAR wavelengths.

xy -plane followed by extruding the 2D xy -ellipse in z -direction to form a 3D ellipsoid. The motivation for expanding the ellipsoid along the z -axis is based on geometrical considerations: It is assumed that correct tree models may have an arbitrary orientation in the xy -plane, but remain upright or vertical with respect to the ground (cf. Fig. 4). This is based on the assumption that tree trunks are vertical to the ground surface.

3.2.1. Computation of the MVEE

If $\mathbf{K} = \{k_i | i = 1, \dots, m\}$ denotes m clusters returned by the mean shift algorithm, and $\mathbf{Q} = \{\mathbf{q}_u | u = 1, \dots, r\}$ denotes the set of r points \mathbf{q}_u belonging to a particular cluster k_f ($f \in i$), then any arbitrarily oriented ellipse ε can be a candidate for the MVEE(\mathbf{Q}), if and only if all points in \mathbf{Q} lie on or inside its boundary, i.e. if the following condition is satisfied (Kumar & Yildirim, 2005):

$$(\mathbf{q}_u - \mathbf{c}_{k_f})^T \mathbf{A} (\mathbf{q}_u - \mathbf{c}_{k_f}) \leq 1 \quad \text{for } u = 1, \dots, m. \quad (15)$$

In this equation, \mathbf{A} is a $d \times d$ positive definite matrix, where $d = 2$ in the presented case, and \mathbf{c}_{k_f} is the center of the ellipse surrounding the clustered points \mathbf{Q} . The semi-axes \mathbf{s}_i of such an ellipse are given as

$$\mathbf{s}_i = \lambda_i^{-\frac{1}{2}} \mathbf{v}_i, \quad (16)$$

where \mathbf{v}_i denote the eigenvectors of \mathbf{A} , which correspond to the directions of the semi-axes. λ_i denotes the eigenvalues of \mathbf{A} , which are related to the length of these axes: The length of each axis is equal to $\frac{1}{\sqrt{\lambda_i}}$. The area of an ellipse or volume of an ellipsoid, respectively, is thus directly proportional to $\det\left(\frac{1}{\sqrt{\mathbf{A}}}\right)$.

Therefore, in order to obtain an MVEE(\mathbf{Q}), $\det\left(\frac{1}{\sqrt{\mathbf{A}}}\right)$ has to be minimized such that Eq. (15) is satisfied in conjunction with \mathbf{A} being positive definite. In order to solve this minimization, Khachyan's first order algorithm is used, which formulates the problem as optimization using Lagrangian duality (Khachyan, 1996).

The computed MVEE(\mathbf{Q}) is extended to the third dimension by extruding it in z -axis in order to form a 3D ellipsoid. The z -coordinate of the ellipsoid center and its semi-axis length s_3 in z -direction are estimated by

$$\mathbf{c}_{k_f z} = \text{median}\left(h_{\min, k_f, i}\right) + s_3, \quad (17)$$

where

$$s_3 = \frac{1}{2} \left(\text{median}\left(h_{\max, k_f, i}\right) - \text{median}\left(h_{\min, k_f, i}\right) \right). \quad (18)$$

In this context, $h_{\min, k_f, i}$ and $h_{\max, k_f, i}$ ($i = 1, \dots, N$) are the N lowest heights and the N largest heights of all points in the cluster k_f , respectively.

Once this modeling is complete, the tree parameters *tree height*, *crown diameter*, and *trunk location* can directly be extracted from the ellipsoid model: The tree height is the maximum height of the ellipsoid in z -direction, the tree crown radii are calculated by taking the geometric mean of the x - and y -semi-axes of the ellipsoid, and the xy -coordinates of the ellipsoid center point provide the location of the tree trunk. Of course, this is a simplifying model only valid for deciduous trees of approximately ellipsoidal shape, but an extension toward a more general tree model as, e.g., described by Sheng, Gong, and Biging (2001) basically seems possible.

4. Utilized test data

4.1. Airborne millimeterwave sensor

The sensor used for the experimental considerations in this paper is the German MEMPHIS system created by the Fraunhofer Institute for High Frequency Physics and Radar Techniques (Schimpf, Essen, Boehmsdorff, & Brehm, 2002). Of its different modes and configurations, in this work only the basic airborne side-looking configuration with a carrier frequency of 35 GHz (Ka-band) and a bandwidth of 900 MHz is employed. This leads to a slant range resolution of 16.7 cm, whereas the azimuth resolution is 8.2 cm. Since MEMPHIS is equipped with four receiving antennas, it is able to provide multi-baseline InSAR datasets from just a single pass over the scene of interest. Additional parameters of the sensor can be found in Table 1.

For the task of tree reconstruction, this system configuration provides two key advantages: First, it is a single-pass InSAR system, which means the four receiving antennas acquire four images of the same scene simultaneously, thus leading to highly coherent data even for vegetation areas. Second, the system uses radar signals of the millimeterwave domain, which can be expected to provide much less canopy penetration than longer wavelengths (cf. Fig. 5). Therefore, it is possible to use the TomoSAR model based on sparse reflectivity profiles as described in Section 2.

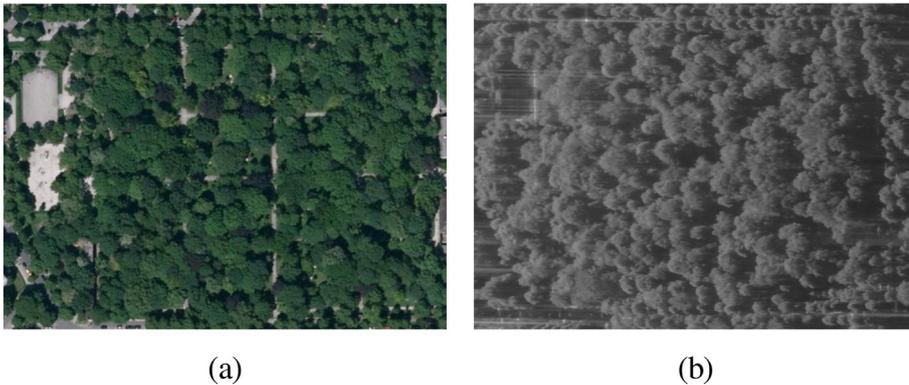


Fig. 6. Test scene "Nordfriedhof" in Munich, Germany: (a) Optical image, (b) MEMPHIS intensity image, range direction from left to right.

Table 2
Parameters of the MEMPHIS test data used in the experiments.

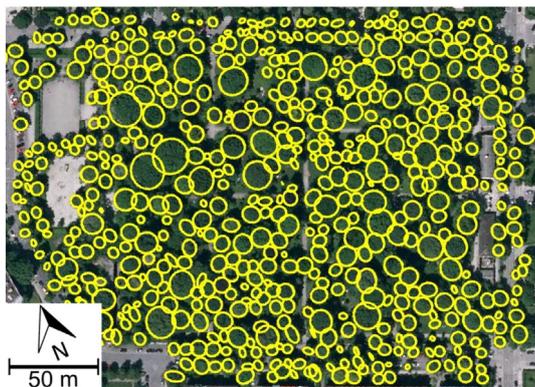
Aspect 1	
Flying altitude	760 m
Heading angle	20°
Number of reconstructed points	1.78 million
Aspect 2	
Flying altitude	760 m
Heading angle	200°
Number of reconstructed points	1.89 million

4.2. Test scene

The available experimental MEMPHIS data were acquired during a campaign over Munich, Germany, in June 2013. The test scene contains the “Alter Nordfriedhof”, an abandoned cemetery, which is used as a public park today, with the target coordinates 48°09′13″N, 11°34′13″E. As can be seen in Fig. 6(a), this approximately 5 ha large areal is characterized by a light planting of deciduous trees (mainly beeches), resembling a grove or little wood. A corresponding SAR intensity image is shown in Fig. 6(b). For the investigations in this paper, multi-baseline InSAR data acquired from two opposing aspects (flight heading angles 20° and 200°, respectively) were processed to a fused 3D point cloud as described in Section 2. The spacing of the height search interval of the TomoSAR optimization step as well as the edge length of the voxel space used for point cloud fusion were both chosen as 0.5 m. A summary of the data parameters can be found in Table 2. In order to emphasize the scope of this study, which is to prove the capability of millimeterwave



(a)



(b)

Fig. 7. Reference data of the test scene, created from a LiDAR point cloud and a co-registered orthophoto. (a) Texturized LiDAR point cloud. (b) Every circle indicates one manually extracted reference tree.

Table 3
Statistics of the reference trees.

Number of trees	570
Tree crown radii	
Min/max	0.39 m/10.09 m
Mean/median	3.48 m/3.25 m
Tree heights	
Min/max	2.81 m/26.97 m
Mean/median	15.37 m/15.43 m

SAR for individual tree parameter reconstruction rather than tree detection, non-tree points are manually removed from the dataset before any further processing.

5. Experiments and reconstruction results

The results of the experimental assessment of the proposed tree reconstruction procedure are summarized in the following sections. First the results of the segmentation are shown in order to provide an evaluation of the effectiveness of unsupervised mean shift clustering for 3D TomoSAR point clouds. In the second part, the results of the ellipsoid fitting process including the derivation of tree parameters are compiled.

For evaluation purposes, the results have been compared to a manually generated reference dataset that is displayed in Fig. 7. This reference dataset is based on the combination of a helicopter-borne LiDAR point cloud containing approximately 0.16 million points (i.e. 3 points/m²) and a co-registered orthophoto. From these data, the reference trees were extracted by a human operator as follows:

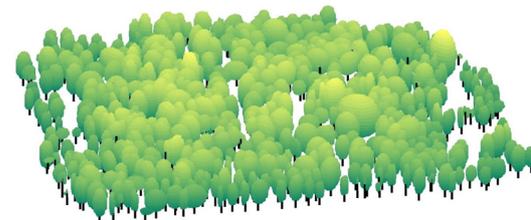
- Identification of individual trees by simultaneous visual comparison of LiDAR point cloud and orthophoto



(a)



(b)



(c)

Fig. 8. The scene shown in different processing stages: (a) The 3D point cloud as derived by multi-aspect TomoSAR data fusion; (b) the clustered point cloud; (c) the reconstructed tree models.

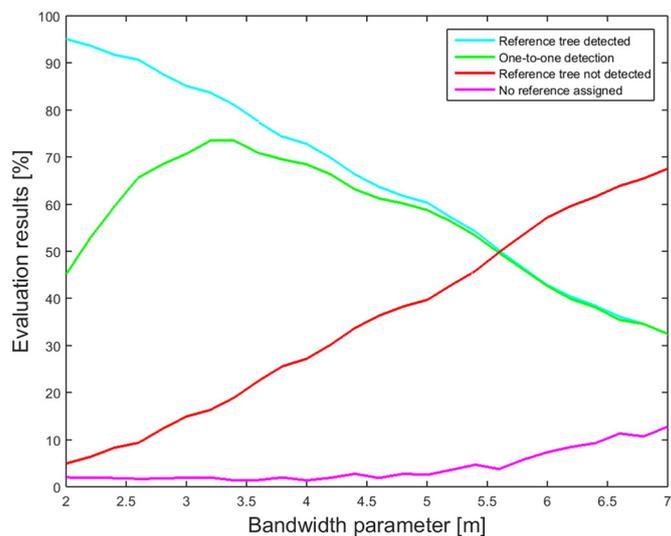


Fig. 9. Determination of the optimal bandwidth parameter by analysis of tree segmentation accuracy for all 570 reference trees. Green indicates perfect matches of one cluster to one reference tree, red missed reference trees, and purple tree clusters that cannot be assigned to any reference tree. In addition, cyan summarizes the overall detected trees, including oversegmented trees. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- Manual measurement of approximate tree crown diameters
- Determination of highest LiDAR point of each tree crown
- Parameterization of each tree by a circle, plus LiDAR-derived tree height
- Modeling of ellipsoids to each tree using the method described in Section 3.2, using only the LiDAR points located within these preliminary tree circles

This way, a reliable set of reference tree models was created, consisting of reference tree locations, heights, and tree crown radii. The statistics of this reference dataset are compiled in Table 3.

5.1. Point cloud generation results

The original TomoSAR point cloud as derived from the test dataset by the method described in Section 2 is displayed in Fig. 8(a). After application of the voxel-space-based fusion procedure (applied with a voxel edge length of 0.5 m), the number of points in the fused point cloud is approximately 1.66 million (corresponding to a point density of about 22 points/m²). That means that due to redundant observations

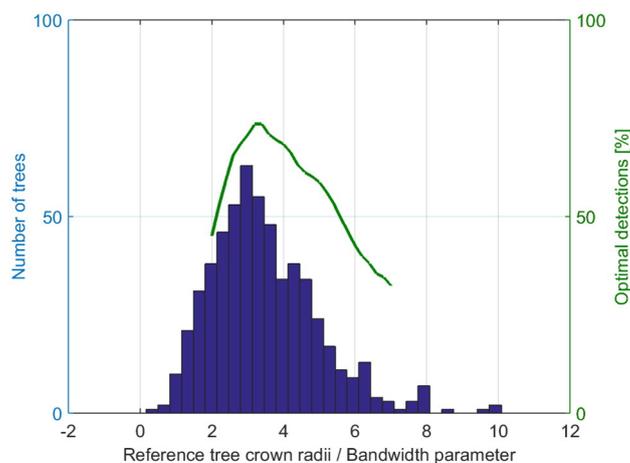


Fig. 10. Correlation of bandwidth parameter and distribution of reference tree radii.

Table 4
Statistics of the tree segmentation for the optimal bandwidth parameter.

Measure	Value
Producer accuracy	73.5%
User accuracy	74.0%
Commission error	1.9%
Omission error	16.3%

and a slight reduction of the spatial resolution, a significant reduction of the data volume (considering more than 3.5 million points existing in the original point clouds) has been achieved. Compared to the LiDAR point cloud mentioned above, the 3D accuracy of the TomoSAR point cloud lies in the sub-meter domain, depending on the evaluation strategy. It has to be mentioned, however, that any evaluation of a TomoSAR point cloud with respect to LiDAR reference data only gives a rough estimate for the 3D accuracy level due to strongly different point densities. Apart from this fact, the accuracy can always further be improved either from the data side, i.e. by providing additional aspects, or from the processing side, i.e. by using a finer spacing during TomoSAR height reconstruction and voxel-space-based fusion.

5.2. Optimal bandwidth parameter selection

Since the only parameter of the mean shift clustering is the bandwidth that is used for generating the clustering kernel, the first set of experiments aimed at determining the optimal bandwidth parameter. Fig. 9 shows a summary of the segmentation results for all 570 reference trees by distinguishing four potential cases: Of course, the desired result of the clustering would be that each reference tree corresponds to one particular cluster, which is referred to as one-to-one detection of this tree. Together with all reference trees, which are oversegmented, i.e. more than one cluster is assigned to them, this adds to the overall rate of detected trees. Then there is the class of missed trees, i.e. reference trees that were not detected as no cluster center falls within their 2D crown outline. Finally, some clusters cannot be assigned to any reference tree at all, which is often referred to as false positives. In the context of this study, it is important to mention that all remaining evaluations are based on one-to-one detections only, whereas oversegmented detections are counted as non-detected trees in order to provide a fair assessment. Analyzing Fig. 9, it becomes obvious that the optimal bandwidth parameter is about 3.2 m, giving an optimal detection rate of 73.5% of the trees, plus oversegmented detections at the rate of 10.2%. Thus, in total 83.7% of all reference trees are discovered, only 16.3% are missed, and 1.9% are false positives.

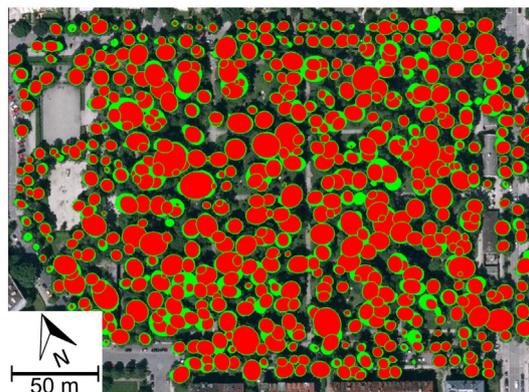


Fig. 11. Ellipsoid models projected onto the 2D reference dataset for one-to-one comparison. The green circles indicate the reference trees, the circles filled in red correspond to the reconstructed trees. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

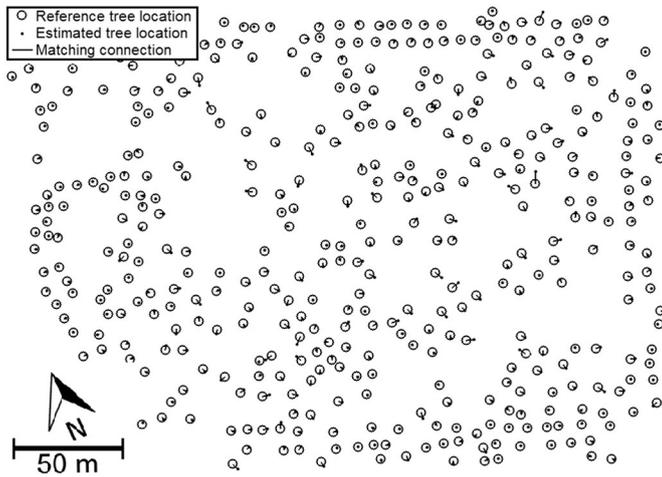


Fig. 12. Localization errors of reconstructed trees and reference trees.

In this context, it is interesting to note that the median tree radius of the reference trees is 3.25 m. The correlation between tree radius distribution and bandwidth parameter is further illustrated in Fig. 10. It can nicely be seen that the optimal bandwidth parameter for the mean shift clustering approximately corresponds to the peak of the tree radius histogram. That means that only light prior knowledge about the expected tree radii of the scene of interest is sufficient to tune the clustering process, while keeping it otherwise fully unsupervised.

5.3. Segmentation results

The result of the mean shift clustering of the point cloud with optimal bandwidth parameter of 3.2 m is displayed in Fig. 8(b). The points have been segmented into 566 clusters, which already resemble individual trees by visual impression. A statistical evaluation of the segmentation based on the one-to-one detections only is summarized in Table 4.

5.4. Ellipsoid modeling results

The final result of the ellipsoid modeling process can be assessed in Fig. 8(c), including tree crowns of different shape and hypothetical stem positions. A projection of the ellipsoids onto the 2D reference data is shown in Fig. 11, while the reference tree positions and the estimated tree locations corresponding to the ellipsoid centers projected into the xy -plane are opposed in Fig. 12. A summary of the tree parameter reconstruction errors is given in Table 5. In addition, the error distributions for tree heights and crown radii are shown in Fig. 13. Again it has to be noted that for this quantitative analysis only the optimally segmented trees are used, whereas the oversegmented ones are discarded.

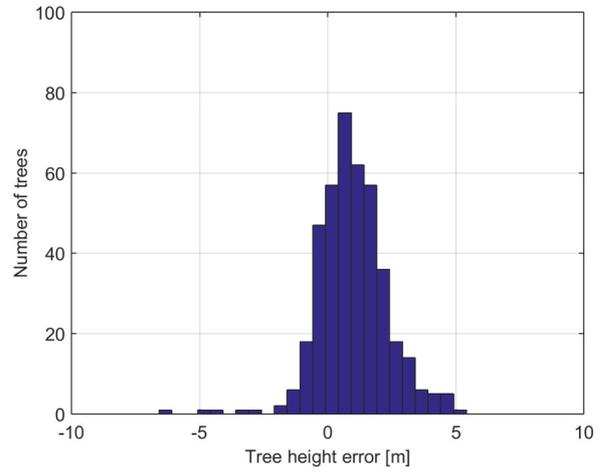
6. Discussion

The experiments presented in this paper show a variety of things:

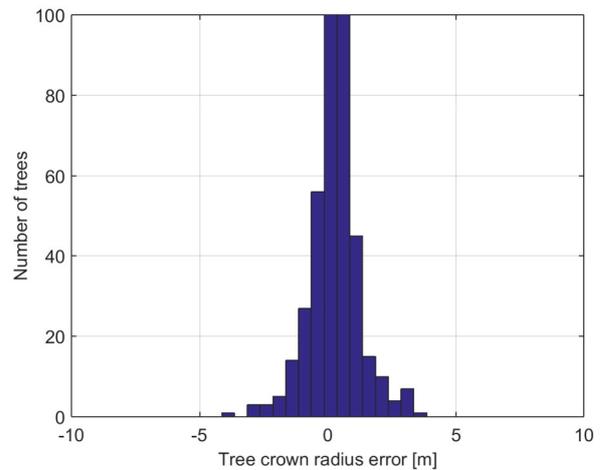
- It is possible to generate 3D point clouds of forested areas from airborne multi-aspect TomoSAR data if single-pass interferometers

Table 5
Error statistics of the reconstructed tree parameters.

Errors	Mean	Standard deviation
Height	0.93 m	1.92 m
Radius	0.28 m	0.96 m
x -Localization	0.50 m	1.05 m
y -Localization	−0.29 m	1.20 m



(a)



(b)

Fig. 13. Distributions of (a) the tree height errors and (b) the crown radii errors. It can be seen that both tree heights and crown radii tend to be slightly overestimated.

and millimeterwaves are employed.

- Depending on the sampling sizes of the SAR pixels, the TomoSAR height reconstruction step and the voxel-space-based fusion, 3D accuracies of the TomoSAR points reach the decimeter domain at very high density.
- Unsupervised clustering of such a point cloud is possible by mean shift clustering if an expectation value of the typical tree crown radii in the investigated scene is available.
- Ellipsoids can be used to model the shapes of deciduous trees, thus providing approximate estimates for core tree parameters such as location, height and diameter.

Although the results of this study are already very promising, there is still room for further improvement. First of all, it is obvious that the clustering is highly dependent on proper tuning of the bandwidth parameter. Although this can be handled by only light prior knowledge, an adaptive setting of the bandwidth parameter could possibly enhance the segmentation accuracy, in particular concerning oversegmented trees. Secondly, the ellipsoid model of course is only a coarse approximation of real-life tree crowns, and only useful for deciduous trees at that. Here, e.g. a generalized ellipsoid model also accounting for varying crown curvature could help to create a more universal approach and more detailed results. In addition, a more robust estimation of the tree

heights and the crown radii is expected to reduce the over-estimation bias in these parameters significantly.

Concerning the number of missed trees, there is unfortunately always the sensor-inherent limitation: If a small tree is surrounded by large trees on all sides, not even multi-aspect SAR data will help to avoid missing that tree due to the side-looking nature of the SAR imaging process. In such a case, only approaches based on volume tomography might provide a viable solution.

Last but not least, it has to be mentioned that the reference data also provides some potential for erroneous modeling, since no analysis of any kind of data can replace in-situ observations. For example, the smallest tree in the reference data is only about 80 cm wide, i.e. in a real ground truth dataset, it would possibly not have been included at all.

7. Summary and conclusion

In this article, an unsupervised approach for segmentation and reconstruction of individual trees from multi-aspect millimeterwave TomoSAR data has been presented. Starting from at least two stacks of single-pass multi-baseline SAR images, first 3D point clouds are generated by TomoSAR height reconstruction. After geocoding, the point clouds derived from multiple aspects are fused to form a single homogeneous 3D point cloud mostly free from any layover or shadow parts. This point cloud is then segmented by unsupervised mean shift clustering, and for every cluster a three-dimensional ellipsoid is modeled to the contained points. Since these ellipsoids are supposed to serve as satisfying approximations of deciduous tree crowns, three important tree parameters are extracted from each ellipsoid: tree location, tree height and tree crown diameter. Experiments based on an airborne millimeterwave dataset of two opposing aspects acquired over a cemetery in the city of Munich, Germany, have shown that about 74% of all trees are properly segmented and reconstructed by the presented technology. Although the side-looking SAR imaging geometry serves as a system-inherent limitation and leads to the fact that particularly small trees fully surrounded by large trees will always be missed, it is expected that an adaptive selection of the kernel bandwidth during mean shift clustering will further enhance the results. This extension of the mean shift algorithm will be part of future research activities.

In any case, the results presented in this paper are expected to further stimulate the research interest in exploiting SAR imagery for forest remote sensing on the individual tree level.

Acknowledgments

The authors would like to thank Thorsten Brehm and Dr. Stephan Stanko (Fraunhofer Institute for High Frequency Physics and Radar Technology FHR) for providing the MEMPHIS test data. They would also like to acknowledge the effort of Christophe Magnard and Dr. Erich Meier (Remote Sensing Laboratories at the University of Zurich) had focusing the raw SAR data.

This work is supported by the Helmholtz Association under the framework of the Young Investigators Group “SiPEO” (VH-NG-1018, www.sipeco.bgu.tum.de) and the International Graduate School of Science and Engineering, Technische Universität München (Project 6.08: “4D City”).

References

- Ahmed, R., Siqueira, P., Hensley, S., Chapman, B., & Bergen, K. (2011). A survey of temporal decorrelation from spaceborne L-band repeat-pass InSAR. *Remote Sensing of Environment*, 115(11), 2887–2896.
- Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9), 509–517.
- Brenner, A. R. (2012). Ultra-high resolution airborne SAR imaging of vegetation and man-made objects based on 40. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium* (pp. 7397–7400).
- Chang, A., Eo, Y., Kim, Y., & Kim, Y. (2013). Identification of individual tree crowns from LiDAR data using a circle fitting algorithm with local maxima and minima filtering. *Remote Sensing Letters*, 4(1), 30–38.
- Chen, Q., Baldocchi, D., Gong, P., & Kelly, M. (2006). Isolating individual trees in a savanna woodland using small footprint LiDAR data. *Photogrammetric Engineering & Remote Sensing*, 72(8), 923–932.
- Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8), 790–799.
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.
- Culvenor, D. S. (2002). TIDA: An algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences*, 28(1), 33–44.
- Englhart, S., Keuck, V., & Siegert, F. (2011). Aboveground biomass retrieval in tropical forests – The potential of combined X- and L-band SAR data use. *Remote Sensing of Environment*, 115(5), 1260–1271.
- Erikson, M. (2003). Segmentation of individual tree crowns in colour aerial photographs using region growing supported by fuzzy rules. *Canadian Journal of Forest Research*, 33(8), 1557–1563.
- Fagan, M., & DeFries, R. (2009). Measurement and monitoring of the world's forests. A review and summary of remote sensing technical capability 2009–2015. *Tech. rep. Resources for the Future*.
- Frey, O., Morsdorf, F., & Meier, E. (2008). Tomographic imaging of a forested area by airborne multi-baseline P-band SAR. *Sensors*, 8(9), 5884–5896.
- Gernhardt, S., Cong, X., Eineder, M., Hinz, S., & Bamler, R. (2012). Geometrical fusion of multitrack PS point clouds. *IEEE Geoscience and Remote Sensing Letters*, 9(1), 38–42.
- Gini, F., & Lombardini, F. (2005). Multibaseline cross-track SAR interferometry: A signal processing perspective. *IEEE Aerospace and Electronic Systems Magazine*, 20(8), 71–93.
- Gupta, S., Weinacker, H., & Koch, B. (2010). Comparative analysis of clustering-based approaches for 3-D single tree detection using airborne fullwave LiDAR data. *Remote Sensing*, 2(4), 968–989.
- Hoekman, D. H., & Verekamp, C. (2001). Observation of tropical rain forest trees by airborne high-resolution interferometric radar. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3), 584–594.
- Izzawati, Wallington, E. D., & Woodhouse, I. H. (2006). Forest height retrieval from commercial X-band SAR products. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4), 863–870.
- Jing, L., Hu, B., Li, J., & Noland, T. (2012). Automated delineation of individual tree crowns from LiDAR data by multi-scale analysis and segmentation. *Photogrammetric Engineering & Remote Sensing*, 78(12), 1275–1284.
- Khachiyan, L. (1996). Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21(2), 307–320.
- Koch, B., Heyder, U., & Weinacker, H. (2006). Detection of individual tree crowns in airborne LiDAR data. *Photogrammetric Engineering & Remote Sensing*, 72(4), 357–363.
- Kumar, P., & Yildirim, E. (2005). Minimum volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications*, 126(1), 1–21.
- Le Toan, T., Beaudoin, A., Riou, J., & Guyon, D. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 403–411.
- Leckie, D. G., Gougeon, F. A., Tinis, S., Nelson, T., Burnett, C. N., & Paradine, D. (2005). Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sensing of Environment*, 94(3), 311–326.
- Makadia, A., Patterson, A., & Daniilidis, K. (2006). Fully automatic registration of 3D point clouds. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1297–1304).
- Mittermayer, J., Wollstadt, S., Prats-Iraola, P., & Scheiber, R. (2014). The TerraSAR-X staring spotlight mode concept. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6), 3695–3706.
- Moreira, A., Krieger, G., Hajnsek, I., Hounam, D., Werner, M., Riegger, S., et al. (2004). TANDEM-X: A TerraSAR-X add-on satellite for single-pass SAR interferometry. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium* (pp. 1000–1003).
- Morsdorf, F., Meier, E., Kötz, B., Itten, K. I., Dobbertin, M., & Allgöwer, B. (2004). LiDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment*, 92(3), 353–362.
- Neumann, M., Ferro-Famil, L., & Reigber, A. (2010). Estimation of forest structure, ground, and canopy layer characteristics from multi-baseline polarimetric interferometric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3), 1086–1104.
- Perko, R., Raggam, H., Deutscher, J., Gutjahr, K., & Schardt, M. (2011). Forest assessment using high resolution SAR data in X-band. *Remote Sensing*, 3(4), 792–815.
- Pollock, R. J. (1996). *The automatic recognition of individual trees in aerial images of forests based on a synthetic tree crown image model*. Ph.D. thesis The University of British Columbia.
- Pouliot, D., King, D., Bell, F., & Pitt, D. (2002). Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, 82(2–3), 322–334.
- Reigber, A., & Moreira, A. (2000). First demonstration of airborne SAR tomography using multi-baseline L-band data. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5), 2142–2152.
- Schimpf, H., Essen, H., Boehmsdorff, S., & Brehm, T. (2002). MEMPHIS – A fully polarimetric experimental radar. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium* (pp. 1714–1716).
- Schmitt, M. (2015). Three-dimensional reconstruction of urban areas by multi-aspect TomoSAR data fusion. *Proceedings of Joint Urban Remote Sensing Event* (On CD).
- Schmitt, M., Brück, A., Schönberger, J., & Stilla, U. (2013). Potential of airborne single-pass millimeterwave InSAR data for individual tree recognition. *Proceedings of 33. Wissenschaftlich-Technische Jahrestagung der DGPF* (pp. 427–436).

- Schmitt, M., Maksymiuk, O., Magnard, C., & Stilla, U. (2013). Radar-grammetric registration of airborne multi-aspect SAR data of urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86, 11–20.
- Schmitt, M., & Stilla, U. (2014 aa). Generating point clouds of forested areas from airborne millimeterwave InSAR data. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium* (pp. 1–4).
- Schmitt, M., & Stilla, U. (2014 bb). Maximum-likelihood-based approach for single-pass synthetic aperture radar tomography over urban areas. *IET Radar, Sonar & Navigation*, 8(9), 1145–1153.
- Schwäbisch, M. (1998). A fast and efficient technique for SAR interferogram geocoding. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium Proceedings* (pp. 1100–1102).
- Shahzad, M., & Zhu, X. (2014). Reconstructing 2-D/3-D building shapes from spaceborne tomographic SAR point clouds. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 15. (pp. 313–320).
- Shahzad, M., & Zhu, X. (2015). Robust reconstruction of building facades for large areas using spaceborne TomoSAR point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2), 752–769.
- Sheng, Y., Gong, P., & Biging, G. S. (2001). Model-based conifer-crown surface reconstruction from high-resolution aerial images. *Photogrammetric Engineering & Remote Sensing*, 67(8), 957–965.
- Shiver, B. D., & Borders, B. E. (1996). *Sampling techniques for forest resource inventory*. Hoboken, NJ: Wiley.
- Straub, B. -M., & Heipke, C. (2001). Automatic extraction of trees for 3D-city models from images and height data. In E. Baltsavias, A. Gruen, & L. V. Gool (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images, III*. (pp. 267–277).
- Wulder, M., Niemann, K. O., & Goodenough, D. G. (2000). Local maximum filtering. *Remote Sensing of Environment*, 73(1), 103–114.
- Zhu, X., & Bamler, R. (2010). Tomographic SAR inversion by L1-norm regularization – The compressive sensing approach. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10), 3839–3846.

List of Abbreviations

VHR very high resolution	1
SAR synthetic aperture radar	1
2-D two-dimensional	1
3-D three-dimensional	1
InSAR SAR interferometry	1
PSI persistent scatterer interferometry	1
TomoSAR SAR tomography	1
4-D four-dimensional	2
D-TomoSAR differential SAR tomography	2
LiDAR Light Detection and Ranging	2
PS persistent scatterer	3
LOS line-of-sight	5
MAP maximum a posteriori	6
PolSAR polarimetric SAR	7
SLC single-look-complex	8
DEM digital elevation model	8
D-InSAR differential InSAR	8
SLIMMER Scale-down by L_1 norm Minimization, Model selection, and Estimation Reconstruction	10
CRF conditional random field	13
SVM support vector machine	13
GIS geographic information system	16
DSM digital surface model	16
MRF markov random field	16
M-SLIMMER multi-snapshot SLIMMER	17
ALS airborne laser scanning	19

DTM digital terrain model	19
TIN triangular irregular network.....	19
nDSM normalized DSM.....	21
PCA principal component analysis	22
RANSAC random sample consensus.....	24
DP Douglas-Peucker	27
MBR minimum bounding rectangle.....	28
AM adjacency matrix.....	31
CHM canopy height model.....	33
SD scatterer (or point) density	36
WTLS weighted total least squares	36
MCD minimum covariance determinant	38
GI Gaussian image.....	39
Tomo-GENESIS TomoSAR module of the GENERic System for Interferometric SAR.....	44
DLR German Aerospace Center	44
UTM universal transverse mercator	44
OSM OpenStreetMap	60
MVEE minimum volume enclosing ellipse	78

List of Figures

- 1 Two dimensional resolution element of a conventional SAR image. ρ_r depends on the duration of transmitted pulse and is determined from chirp bandwidth f_W while ρ_a is approximately half of the physical antenna length L . x represents flight direction and is normally called azimuth or along track, r refers to the LOS direction of the sensor normally called range or slant range, θ denotes the elevation (or incidence) angle which actually defines the LOS direction, y refers to the ground range and z denotes the height direction that is perpendicular to the x - y plane. All scatterers lying within the yellow curve are mapped into the same azimuth-range pixel. 6

- 2 Geometrical distortions appearing in SAR images. Three objects are shown in gray shades. θ refers to the incidence (or look) angle of the sensor while β represents the angle of the object surface with respect to the horizontal ground terrain (i.e., slope of the surface). The flight direction or azimuth axis x is perpendicular to the y - z plane i.e., coming out of the page in the depicted case. 7

- 3 Schematic illustration of the TomoSAR imaging geometry. The elevation aperture is built by exploiting multi-pass/multi-baselines (six in the depicted case) from slightly different viewing angles. It is shown that the backscattering contribution from the edge of two buildings and small portion of ground are mapped into a single azimuth-range SAR image pixel. TomoSAR aims to estimate the depicted reflectivity profile $\hat{\gamma}(s)$ for discretized (pink region) elevation extent Δs . Typically, the discretization factor is much higher i.e., $N \ll Q$ which renders (9) to be undetermined (i.e., more equations than unknowns). s denote the elevation axis which is actually a curve but is usually approximated as a straight line due to large range distances. 11

- 4 Geometric illustration of multi- (double-) bounce and shadow effect. The light gray line depicts the double-bounce phenomenon that is frequently observed in urban regions particularly between vertical structures and ground. These lines give strong hint towards building detection. Shadow region (depicted as black) indicates the part not visible to the sensor. By knowing key parameters of the imaging radar e.g., θ , ρ_r , ρ_a etc., individual building height based on dimensions of shadow and/or layover could be retrieved. 12

- 5 Density based clustering algorithm (Ester et al., 1996). Points a and b are directly density connected to each other whereas points a and c are density connected to each other since there is a chain of points between them such that they all are directly density connected to each other. Two parameters that control the clustering process include the neighborhood parameter ε , i.e., the radius in case of sphere or cylindrical neighborhood, and the minimum number of points $MinPts$ in the ε -neighborhood for any particular point. The resulting clusters thus contains points such that all the points in any particular cluster are density connected to each other but are not density connected to any other point belonging to another cluster. Moreover, each point inside any particular cluster belongs to one of the three categories: 1) A point is labeled core point if it contains, within its ε -neighborhood, $MinPts$ number of points; 2) A point is considered border point if it is within ε -neighborhood of any core point but itself is not a core point and does not have $MinPts$ neighbors; 3) A point neither core point nor border point is termed as an outliers i.e., any point which do not have density (number of points) greater than $MinPts$ within its ε -neighborhood and also is not the neighbor of any other point. 23
- 6 Typical workflow related to building footprint reconstruction. The first column represents the segmentation process that separates the extracted building points into respective individual buildings (black and gray points belonging to the two buildings in the depicted case). Coarse shape of the building footprints are reconstructed by connecting the boundary points (second column). Different methods to determine this coarse shape are presented in Section 3.1.2.1. The initial shapes are smoothed using methods presented in Section 3.1.2.2 (third column). Finally constraints are added to yield geometrically correct and visually appealing building shapes via methods provided in Section 3.1.2.3 (fourth column). Each module is elaborated in the corresponding subsections. 26
- 7 Coarse building footprint reconstruction using alpha shapes method with varying α such that $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < \alpha_5 < \alpha_6$. It is demonstrated that with increasing α , the algorithm loses the ability to reconstruct concave building footprints and eventually with very large value of α or when $\alpha \approx \inf$ returns a convex hull. 27
- 8 Example depicting failure of MBR to correctly determine the dominant direction of the building footprint: (a) 2-D points of buildings (blue color) are depicted. Convex hull with 14 edges around these points are drawn as red polygon. Grey rectangle is the minimum area bounded rectangle computed via rotating callipers method. It can be seen that the dominant directions estimated via grey MBR do not correctly represent the true dominant orientation of the building points. The desired MBR that correctly represent the dominant orientation is shown as black dotted rectangle; (b) the areas of the MBR by rotating MBR around its center of gravity at angles equal to every edge of the convex hull. The grey circle shows the minimum area bounded rectangle corresponding to grey MBR in (a) (Shahzad and Zhu, 2014). 29

- 9 Procedure to add rectilinear constraints to any one particular building polygon: (a) Black dots indicate the extracted building points, cyan polygon represents the initial coarse outline obtained using alpha shapes algorithm and gray polygon depicts the refined/smoothed polygon to which rectilinear constraints are added; (b) Each edge of the refined gray polygon of (a) is segmented to belong to one of the two estimated dominant directions indicated by black arrows. The segmentation is based on their angular deviation (i.e., an edge is associated to that principal direction with whom the angular difference is less). Segmented edges are shown in red and green colors. The blue polygon then depicts the reduced refined polygon of (a) by merging all adjacent edges that belong to the same principal axis. (c) Each edge of the blue polygon is then rotated/projected around its midpoint onto its corresponding principal axis; (d) Finally the intersection (vertex) points among the adjacent edges of the projected blue polygon edges of (c) are computed for rectilinearization. Note that the depicted polygon contains no associated façade and therefore belongs to case 3. For case 2 polygons, the associated façades are first fused to the refined polygon and, if identified as rectilinear, constraints are added to complete the reconstruction procedure. 30
- 10 Spaceborne TomoSAR point clouds of high rise area in Las Vegas processed by DLR's Tomo-GENESIS system (Zhu, 2011) (Zhu et al., 2013). Height is color-coded. It can be seen that the side-looking geometry render rich amount of points/scatterers on building façades. Also due to the coherent imaging nature of SAR and the repeat-pass data-takes, temporally incoherent objects such as trees or vegetation cannot be reconstructed, depicted in black. 34
- 11 Illustration of SD estimation: (a) 3-D view of the local cylindrical neighborhood around the point of interest; (b) Top view of (a). The coefficients of the dotted yellow regression line are estimated via M estimation. Black dotted line shows the shift of yellow line to the point of interest. Shaded area shows the region of v_d within v_c . 38
- 12 Gaussian image of three connected planar surfaces: (a) Arrows indicate surface normal vectors ($n_{red}, n_{green}, n_{blue}$) to the respective surfaces; (b) All points belonging to one particular surface are mapped to same identical point in GI (ideal scenario). 40
- 13 Illustration of orientation angle for flat and curved vertical footprints (top view); (a) Arrows indicate pattern of change in orientation (azimuthal) angles of ten points on each vertical surface; (b) plots their respective orientation angles. 41
- 14 Removal of conflicting segments and vertex point computation: (a) shows the case of five clustered point segments belonging to one particular building. At the transitional regions, there are conflicting (smaller) clusters, shown in grey dotted circle, that behave as noisy segments and should be removed before computing the vertex points; (b) shows the retained segments and their corresponding vertices after removing conflicting segments. 42

- 15 Vertices for refinement. Grey rectangles depicts the 2-D building footprint from the top: (a) shows the total of 5 vertices out of which 4 are open endpoint vertices and one is intersection vertex computed from the intersection of segments 1 and 2; (b) depicts the situation where ignoring condition C3 would yield false segment addition shown as grey dotted line. The grey arrows indicate the local orientation angle θ at open vertices v_o and \hat{v}_o . Two black dotted lines represent the two correct inserted segments between v_o and \hat{v}_o ; (c) depicts the recursive extension procedure of the open vertex v_o . v_p represent the intermediate extension points where as the v'_p denotes the final point; (d) illustrates the direction of extension of grey open vertex over the roof region. This can happen if we only consider the h_{\max} and ignore the local standard deviation of height h_σ . 44
- 16 Dataset: (a) Optical image of the test area in Las Vegas. Copyright Google. (b) TomoSAR points in UTM coordinates of the corresponding area in Las Vegas generated from a stack acquired from ascending orbit only. The height is color-coded. 45
- 17 SD estimated with radius $r = 5\text{m}$ and inliers $d = 0.9\text{m}$. Colorbar indicates computed SD. 46
- 18 Automatic selection of TH value. SD histogram for the TomoSAR points of the corresponding test area depicted in Figure 16. The maximum SD histogram value used for thresholding is 1.132. Number of bins is set to bins = min(SD) : (max(SD) – min(SD))/100 : max(SD). It is worth mentioning here that for a rural or sub-urban area, it is possible that the maximum of histogram occurs at SD = 0 which would consequently render all points to be retained. However for urban regions, this usually never happens due to existence of higher point densities over man-made structures. 46
- 19 Extracted façade points from TomoSAR point cloud depicted in Figure 16 overlaid onto the corresponding optical image (©Google). 47
- 20 Fine clustering results after applying mean shift clustering using Gaussian kernel with bandwidth $b = 0.4$ to the coarsely clustered segments in their normal feature space (in GI domain): (a) TomoSAR points of one particular density connected cluster (top view). Colorbar indicates height in meters; (b) Corresponding orientation angle in degrees; (c) Non clustered (top) and clustered (bottom) points in the Gaussian image of points in (a); (d) Resulting clustered points in 3-D. 47
- 21 Reconstructed façades: (a) shows the 2-D view of the façade footprints overlaid onto the optical image prior to refinement; (b) shows the 2-D view of the façade footprints overlaid onto the optical image after refining with parameters settings $T_h = 5\text{m}$ and $T_\sigma = 2.5\text{m}$. 48
- 22 A case depicting wrong reconstruction of a pedestrian bridge: (a) shows the optical image of the bridge; (b) overlays the reconstructed segment in green onto the optical image of (a); (c) side view of the bridge (© Google Street View). 49
- 23 3-D view of the final façade reconstruction. The axis is in meters range and has been translated to the origin for better metric clarity by subtracting UTM easting and northing values by their respective minimum values present in the reconstructed vertices. 49

- 24 Depicting three different cases over simple buildings commonly occurring in TomoSAR point clouds. Heights of the points are color-coded. Optical images © Google. 51
- 25 Graphical illustration of the seed point selection procedure. Top view of 3-D TomoSAR points of a small region containing one building is depicted. The height in meters of each TomoSAR point is color-coded. The red line shows the reconstructed façade segment with black point as its midpoint. The green and blue points show the selected points in two opposite orthogonal directions with respect to the reconstructed façade (depicted as gray dotted lines). r_N is cylindrical radius used to determine local neighbors (i.e., points within shaded circles in respective colors). Local mean height of the blue and green points are computed, denoted as m_b and m_g respectively. Seed point s_p is chosen as the maximum height point among the local neighbors of blue points if $m_b > m_g$ or among the neighbors of green points if $m_g > m_b$ where latter is true in the depicted case. 52
- 26 Procedure describing the concept of $P_{t_shortest}$ and P^+ : (a) Refined polygon of one particular j th building \tilde{V}_j having vertices $\mathbf{v}_{k=1,\dots,m}$ with four reconstructed façades f_1, f_2, f_3 and f_4 that are to be incorporated; (b) illustrates the concept of shortest and longest paths associated for a particular façade f_1 . \mathbf{v}_{1a} and \mathbf{v}_{1b} denotes the closest points on the polygon/graph \tilde{V}_j to the two endpoints of façade f_1 respectively; (c) Example illustration of the concept of positive path P^+ . $P_{2_shortest}$ is identified as P^- as there exist points in \mathbf{K}_2 that are also present in $\mathbf{P}_{2_shortest}$. 55
- 27 Fusion of both type I and II façades. The refined polygon of one particular j th building \tilde{V}_j with two reconstructed façades f_1 and f_2 with types I and II respectively is shown; (a) depicts the procedure of computing points \mathbf{v}'_{ta} and \mathbf{v}'_{tb} . Note that for type I façade, \mathbf{v}'_{1a} and \mathbf{v}'_{1b} are computed by projecting \mathbf{v}_{1a} and \mathbf{v}_{1b} onto black dotted line which is parallel to the reconstructed façade f_1 and passes through the midpoint of line segment L formed by connecting points \mathbf{v}_{1a} and \mathbf{v}_{1b} while for the other façade having type II, \mathbf{v}'_{2a} and \mathbf{v}'_{2b} are simply the endpoints of the façade f_2 . All vertices of \tilde{V}_j on $P_{1_shortest}$ and $P_{2_shortest}$ (i.e., points in $\mathbf{P}_{1_shortest}$ and $\mathbf{P}_{2_shortest}$) are replaced by vertices \mathbf{v}'_{1a} , \mathbf{v}'_{1b} and $\mathbf{v}'_{2a}, \mathbf{v}'_{2b}$ respectively; (b) Resulting polygon in black after fusing façades f_1 and f_2 with types I and II respectively. 56
- 28 Datasets: Top view of the three dimensional TomoSAR points in UTM coordinates of the area of interest in Las Vegas, USA (a) and Berlin (b). Blue lines depict the reconstructed façade segments (longer than 10 meters). The height of TomoSAR points is color-coded [unit: m]. 59
- 29 Building extraction results – Las Vegas: (a) Extracted building points in red are overlaid onto the optical image (© Google) of the area of interest. Yellow circles indicate extracted points originating from sources like advertisement boards, monuments etc. Large black circle encloses two buildings which remain undetected due to lower relative heights; (b) Finally, reference footprints (in green) overlaid onto the extracted building points. Red points are building points while black points are non-building points. 60

- 30 Results of building extraction – Berlin: (a) Extracted roof points in red are overlaid onto the optical image (© Google) of the area of interest; (b) Red and black points depict building and non-building points, respectively. The overlaid green polygons are reference buildings downloaded from OSM (Geofabrik, 2015). Blue polygons are manually extracted buildings not present in OSM data. Gray polygons are newly constructed buildings that are not present in our dataset whereas magenta colored polygons are buildings that do not actually exist but present in OSM data. Both gray and magenta polygons are not included in the evaluation. 61
- 31 Common images computed using reference footprint image and final reconstructed footprints. 62
- 32 Difference images computed by subtracting the final reconstructed footprint image from the reference image. Red pixels indicate building regions that are not reconstructed by the proposed algorithm while blue pixels are over reconstructed regions i.e., pixels not part of the reference footprint image but present in the reconstructed image. Note that green pixels indicate no difference between reconstructed and reference rastered images. 63
- 33 Final reconstructed 3-D building footprints: Las Vegas (top) and Berlin (bottom) datasets. 64
- 34 Test buildings - Bellagio hotel, Las Vegas: (a) Optical image (© Google); (b) Fused TomoSAR point clouds from both ascending and descending orbits in UTM coordinates. 66
- 35 SD estimates of TomoSAR point cloud depicted in Figure 34(b). 66
- 36 3-D Façade reconstruction results: (a) Extracted façade points; (b) Segmented façade points; (b) Reconstructed 3-D façade model with extracted façade points. For (a), (c) and (d), the height of the points is color-coded from blue (lowest) to red (highest) [unit: m]. 67
- 37 Reconstructed 4-D building façade model, amplitude of seasonal motion is color-coded [unit: mm]. 67
- 38 Reconstructed 3-D SAR image overplotted on the reconstructed façade model. Note that this is not only a projection of the SAR image onto the building models. Rather, the lay-overed brightness contributions from façade and ground have been separated in the tomographic reconstruction step. 68
- 39 The basic principle for the elevation estimates refinement of the TomoSAR points by using their more accurate azimuth and range coordinates and the reconstructed façade surfaces. 69
- 40 Elevation estimates refinement: (a) TerraSAR-X mean intensity map from ascending stacks (the red dots are the analyzed points) along with the projection geometry; Height estimates of the analyzed points before (b) and after (c) refinement. 70
- 41 Test Area: (a) Optical image of the test area © Google; (b) Corresponding SAR intensity map with spatial resolution of $1.1 \text{ m} \times 0.588 \text{ m}$ in azimuth (*az*) and range (*rg*) respectively. 71
- 42 GIS data (2-D building footprints) of Las Vegas from OSM. 71

- 43 Building mask extraction: (a) Reference polygons (shown in red and green polylines) of two buildings in the area of interest overlaid onto the SAR intensity map after geocoding. Side of the buildings facing the sensor are shown in red while the other side not visible to the sensor in green; (b) After rotation and range-azimuth shift compensation, the red polylines in (a) are shifted towards the sensor. The yellow dotted lines indicate the maximum range shift of $MaxRgShift$ ($= 318\text{m}$) where as the red dotted lines indicate the $RgShift$ obtained by taking the maximum of the approximate derivative of C . 72
- 44 Graphical illustration of adjusting polyline length: (a) Black and gray polygons indicate polygons before and after rotation (around centroid depicted as black star) respectively with circles representing corresponding vertices. The dotted polylines represent building side not visible to the sensor. It is shown that after orientation the azimuthal length is changed; (b) Length of the outer (first and last edge depicted in dark gray) edges of the polyline facing the sensor is slightly extended by distance e ($= 5\text{m}$ in this work). Intensities (interpolated) over these extended edges are analyzed and first and last extended points (i.e., v'_p and v'_q) are respectively replaced by the points on the edges $v_p v'_p$ and $v_q v'_q$ where the approximate derivative (or change in intensities) is maximum. 72
- 45 Reconstructed and color-coded elevation of the two test buildings using 6 interferograms, visualized in two layers, overlaid with intensity. From top to down: first and second layer, respectively; from left to right: M-SL1MMER and SL1MMER, respectively (Zhu et al., 2015b). 73
- 46 Test scene “Nordfriedhof” in Munich, Germany: (a) Optical image © Google; (b) MEMPHIS intensity image, range direction from left to right; (c) Corresponding airborne 3-D TomoSAR point cloud. 76
- 47 Illustration of the ellipsoidal modeling: (a) MVEE computed using 3-D points denoted as black dots (projected onto the xy plane). s_1 and s_2 are the computed semi axes of the MVEE; (b) MVEE of (a) is extruded in z direction both upwards and downwards forming a 3-D ellipsoid with third semi axes denoted as s_3 . x' and y' in (a) represents axes of the global coordinate system while x and y denotes the coordinates in local coordinate system. The orientation and axes of the rotated ellipsoid are computed using singular value decomposition of matrix A . Red point in both (a) and (b) represents the ellipsoid center. 77
- 48 Reference data of the test scene, created from a LiDAR point cloud and a co-registered orthophoto. Each yellow ellipse indicates one manually extracted reference tree. 79
- 49 Determination of the optimal bandwidth parameter by analysis of tree segmentation accuracy for all 570 reference trees. Green indicates perfect matches of one cluster to one reference tree (i.e., optimal detection), red missed reference trees, and purple tree clusters that cannot be assigned to any reference tree. In addition, cyan summarizes the overall detected trees, including oversegmented trees. 80
- 50 Correlation of bandwidth parameter and distribution of reference tree radii. 80
- 51 Tree reconstruction results: (a) Clustering results of applying mean shift algorithm with bandwidth 3.2 m over TomoSAR point clouds depicted in Figure 46(c); (c) Reconstructed 3-D ellipsoidal tree models. 81

-
- 52 Ellipsoid models projected onto the 2-D reference dataset for one-to-one comparison. The green circles indicate the reference trees, the circles filled in red correspond to the reconstructed trees. 82
- 53 Localization errors of reconstructed trees and reference trees. 82
- 54 Distributions of (a) the tree height errors and (b) the crown radii errors. It can be seen that both tree heights and crown radii tend to be slightly overestimated. 83

References

- Adam, N., Zhu, X., Bamler, R., 2009. Coherent stacking with TerraSAR-X imagery in urban areas. In: Proceedings of Joint Urban Remote Sensing Event (JURSE), 1–6.
- Alharthy, A., Bethel, J., 2002. Heuristic filtering and 3d feature extraction from LIDAR data. In: ISPRS Commission III Symposium, Photogrammetric and Computer Vision, Graz, 29–34.
- Ardila, J. P., Bijker, W., Tolpekin, V. A., Stein, A., 2012. Multitemporal change detection of urban trees using localized region-based active contours in VHR images. *Remote Sensing of Environment* 124: 413–426.
- Arefi, H., Reinartz, P., 2013. Building Reconstruction Using DSM and Orthorectified Images. *Remote Sensing* 5 (4): 1681–1703.
- Argenti, F., Alparone, L., 2002. Speckle removal from SAR images in the undecimated wavelet domain. *IEEE Transactions on Geoscience and Remote Sensing* 40 (11): 2363–2374.
- Argenti, F., Lapini, A., Bianchi, T., Alparone, L., 2013. A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images. *IEEE Geoscience and Remote Sensing Magazine* 1 (3): 6–35.
- Auer, S., Gernhardt, S., 2014. Linear Signatures in Urban SAR Images—Partly Misinterpreted? *IEEE Geoscience and Remote Sensing Letters* 11 (10): 1762–1766.
- Auer, S., Gernhardt, S., Bamler, R., 2011. Ghost Persistent Scatterers Related to Multiple Signal Reflections. *IEEE Geoscience and Remote Sensing Letters* 8 (5): 919–923.
- Awrangjeb, M., Fraser, C., 2014. Automatic Segmentation of Raw LIDAR Data for Extraction of Building Roofs. *Remote Sensing* 6 (5): 3716–3751.
- Awrangjeb, M., Lu, G., 2014. Automatic Building Footprint Extraction and Regularisation from LIDAR Point Cloud Data. In: Proceedings of IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA), 1–8.
- Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. In: International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII (Part B4), 110–117.
- Ballard, D. H., Brown, C. M., 1982. *Computer Vision*, first edition Edition. Prentice Hall, Englewood Cliffs, N.J.
- Bamler, R., Eineder, M., 2005. Accuracy of differential shift estimation by correlation and split-bandwidth interferometry for wideband and delta-k SAR systems. *IEEE Geoscience and Remote Sensing Letters* 2 (2): 151–155.
- Bamler, R., Hartl, P., 1998. Synthetic aperture radar interferometry. *Inverse problems* 14 (4): R1.
- Barthelet, E., Mercier, G., Denise, L., Reynaud, S., 2012. Feature based maximum likelihood model inversion for three-dimensional building extraction from single high resolution SAR images. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 3600–3603.
- Baselice, F., Ferraioli, G., Reale, D., 2014. Edge Detection Using Real and Imaginary Decomposition of SAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 52 (7): 3833–3842.
- Belgiu, M., Tomljenovic, I., Lampoltshammer, T. J., Blaschke, T., Höfle, B., 2014. Ontology-Based Classification of Building Types Detected from Airborne Laser Scanning Data. *Remote Sensing* 6: 1347–1366.
- Bennett, A. J., Blacknell, D., 2003. The extraction of building dimensions from high resolution SAR imagery. In: Proceedings of the International Radar Conference, 182–187.
- Bolter, R., 2001. Buildings from SAR: Detection and Reconstruction of Buildings from Multiple View High Resolution Interferometric SAR Data. PhD thesis, Graz University of Technology (TUG), Graz, Austria.
- Bolter, R., Leberl, F., 2000. Detection and reconstruction of human scale features from high resolution interferometric SAR data. In: Proceedings of 15th International Conference on Pattern Recognition, Vol. 4, IEEE, 291–294.
- Borrmann, D., Elseberg, J., Lingemann, K., Nüchter, A., 2011. The 3d Hough Transform for plane detection in point clouds: A review and a new accumulator design. *3D Research* 2 (2).
- Botterill, T., Mills, S., Green, R., 2009. New Conditional Sampling Strategies for Speeded-Up

- RANSAC. 33.1–33.11.
- Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (9): 1124–1137.
- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (11): 1222–1239.
- Brandtberg, T., 1999. Automatic individual tree-based analysis of high spatial resolution remotely sensed data. PhD thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden.
- Brédif, M., 2010. 3d Building Modeling: Topology-Aware Kinetic Fitting of Polyhedral Roofs and Automatic Roof Superstructure Reconstruction. PhD thesis, Telecom ParisTech, France.
- Brédif, M., Tournaire, O., Vallet, B., Champion, N., 2013. Extracting polygonal building footprints from digital surface models: A fully-automatic global optimization framework. *ISPRS Journal of Photogrammetry and Remote Sensing* 77: 57–65.
- Brenner, A., Roessing, L., 2008. Radar Imaging of Urban Areas by Means of Very High-Resolution SAR and Interferometric SAR. *IEEE Transactions on Geoscience and Remote Sensing* 46 (10): 2971–2982.
- Brenner, C., 2000. Towards fully automatic generation of city models. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 33 (PART B3), 85–92.
- Brenner, C., Haala, N., 1999. Rapid production of virtual city models. *Geoinformationssysteme* 2: 22–28.
- Brunner, D., Lemoine, G., Bruzzone, L., Greidanus, H., 2010. Building Height Retrieval From VHR SAR Imagery Based on an Iterative Simulation and Matching Technique. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3): 1487–1504.
- Buckreuss, S., Werninghaus, R., Pitz, W., 2008. The German Satellite Mission TerraSAR-X. In: *Proceedings of IEEE Radar Conference*, Rome, Italy, 1–5.
- Cao, Y., Su, C., Yang, G., 2014. Detecting the number of buildings in a single high-resolution SAR image. *European Journal of Remote Sensing* : 513–535.
- Carmo, M. P. D., 1976. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Englewood Cliffs, NJ, USA.
- Cellier, F., Colin, E., 2006. Building Height Estimation using Fine Analysis of Altimetric Mixtures in Layover Areas on Polarimetric Interferometric X-band SAR Images. 4004–4007.
- Cellier, F., Oriot, H., Nicolas, J.-M., 2006. Hypothesis Management for Building Reconstruction from High Resolution InSAR Imagery. 3639–3642.
- Chang, A., Eo, Y., Kim, Y., Kim, Y., 2013. Identification of individual tree crowns from LiDAR data using a circle fitting algorithm with local maxima and minima filtering. *Remote Sensing Letters* 4 (1): 29–37.
- Chehata, N., Guo, L., Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 39-Part 3/W8, 207–212.
- Chen, C., Li, Y., Li, W., Dai, H., 2013. A multiresolution hierarchical classification algorithm for filtering airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 82: 1–9.
- Chen, C.-T., Chen, K.-S., Lee, J.-S., 2003. The use of fully polarimetric information for the fuzzy neural classification of SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 41 (9): 2089–2100.
- Chen, D., Zhang, L., Li, J., Liu, R., 2012a. Urban building roof segmentation from airborne lidar point clouds. *International Journal of Remote Sensing* 33 (20): 6497–6515.
- Chen, H., Zhang, Y., Wang, H., Ding, C., 2012b. Stationary-Wavelet-Based Despeckling of SAR Images Using Two-Sided Generalized Gamma Models. *IEEE Geoscience and Remote Sensing Letters* 9 (6): 1061–1065.
- Chen, Q., 2007. Airborne lidar data processing and information extraction. *Photogrammetric engineering and remote sensing* 73 (2): 109.
- Chen, Q., 2009. Improvement of the edge based morphological method for lidar data filtering. *Inter-*

- national Journal of Remote Sensing 30 (4): 1069–1074.
- Chen, Q., Baldocchi, D., Gong, P., Kelly, M., 2006. Isolating individual trees in a savanna woodland using small footprint lidar data. *Photogrammetric Engineering & Remote Sensing* 72 (8): 923–932.
- Chen, S., Wang, H., Xu, F., Jin, Y.-Q., 2015. Automatic Recognition of Isolated Buildings on Single-Aspect SAR Image Using Range Detector. *IEEE Geoscience and Remote Sensing Letters* 12 (2): 219–223.
- Chen, Y., Cheng, L., Li, M., Wang, J., Tong, L., Yang, K., 2014. Multiscale Grid Method for Detection and Reconstruction of Building Roofs from Airborne LiDAR Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (10): 4081–4094.
- Cheng, Y., 1995. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17 (8): 790–799.
- Chum, O., Matas, J., 2005. Matching with PROSAC - progressive sample consensus. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, 220–226 vol. 1.
- Cloude, S. R., Pottier, E., 1996. A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing* 34 (2): 498–518.
- Cloude, S. R., Pottier, E., 1997. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing* 35 (1): 68–78.
- Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (5): 603–619.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine learning* 20 (3): 273–297.
- Cumming, I. G., Wong, F. H., 2005. *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Artech House Remote Sensing Library. Artech House.
- CyberCity3D, 2015.
- Dai, E., Jin, Y.-Q., Hamasaki, T., Sato, M., 2008. Three-Dimensional Stereo Reconstruction of Buildings Using Polarimetric SAR Images Acquired in Opposite Directions. *IEEE Geoscience and Remote Sensing Letters* 5 (2): 236–240.
- Dai, M., Peng, C., Chan, A., Loguinov, D., 2004. Bayesian wavelet shrinkage with edge detection for SAR image despeckling. *IEEE Transactions on Geoscience and Remote Sensing* 42 (8): 1642–1648.
- d’Alessandro, M. M., Tebaldini, S., 2012. Phenomenology of P-Band Scattering From a Tropical Forest Through Three-Dimensional SAR Tomography. *IEEE Geoscience and Remote Sensing Letters* 9 (3): 442–446.
- de Kluijver, H., Stoter, J., 2003. Noise mapping and GIS: optimising quality and efficiency of noise effect studies. *Computers, Environment and Urban Systems* 27 (1): 85–102.
- Deledalle, C.-A., Denis, L., Tupin, F., 2009. Iterative Weighted Maximum Likelihood Denoising With Probabilistic Patch-Based Weights. *IEEE Transactions on Image Processing* 18 (12): 2661–2672.
- Demantke, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality based scale selection in 3d lidar point clouds. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Laser Scanning*, Vol. 2011.
- Demirhan, A., Gueler, İ., 2011. Combining stationary wavelet transform and self-organizing maps for brain MR image segmentation. *Engineering Applications of Artificial Intelligence* 24 (2): 358–367.
- Deng, L., Wang, C., 2014. Improved Building Extraction With Integrated Decomposition of Time-Frequency and Entropy-Alpha Using Polarimetric SAR Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (10): 4058–4068.
- Denis, L., Tupin, F., Darbon, J., Sigelle, M., 2009. Joint Regularization of Phase and Amplitude of InSAR Data: Application to 3-D Reconstruction. *IEEE Transactions on Geoscience and Remote Sensing* 47 (11): 3774–3785.
- D’Hondt, O., Guillaso, S., Hellwich, O., 2012. Automatic extraction of geometric structures for 3d reconstruction from tomographic SAR data. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 3728–3731.
- D’Hondt, O., Guillaso, S., Hellwich, O., 2013. Geometric primitive extraction for 3d reconstruction of urban areas from tomographic SAR data. In: *Proceedings of Joint Urban Remote Sensing Event*

- (JURSE), 206–209.
- Döllner, J., Kolbe, T. H., Liecke, F., Sgouros, T., Teichmann, K., 2006. The virtual 3d city model of berlin-managing, integrating, and communicating complex urban information. In: Proceedings of the 25th Urban Data Management Symposium UDMS, Vol. 2006.
- Dorninger, P., Pfeifer, N., 2008. A Comprehensive Automated 3d Approach for Building Extraction, Reconstruction, and Regularization from Airborne Laser Scanning Point Clouds. *Sensors* 8 (11): 7323–7343.
- Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization* 10 (2): 112–122.
- Dubois, C., Thiele, A., Hinz, S., 2013. Pixel-based approach for building heights determination by SAR radargrammetry. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W3, ISPRS, Antalya, Turkey, 19–24.
- Dubois, C., Thiele, A., Hinz, S., 2014. Building Detection in TanDEM-X Data. 1–4.
- Duda, R. O., Hart, P. E., 1972. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM* 15 (1): 11–15.
- Edelsbrunner, H., Kirkpatrick, D., Seidel, R., 1983. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory* 29 (4): 551–559.
- Eineder, M., Minet, C., Steigenberger, P., Cong, X., Fritz, T., 2011. Imaging Geodesy - Toward Centimeter-Level Ranging Accuracy With TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing* 49 (2): 661–671.
- Elachi, C., 1988. *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE, New York.
- Elberink, S. O., Vosselman, G., 2009. Building Reconstruction by Target Based Graph Matching on Incomplete Laser Data: Analysis and Limitations. *Sensors* 9 (8): 6101–6118.
- Engel, J., Döllner, J., 2009. Approaches towards visual 3d analysis for digital landscapes and its applications. *Digital Landscape Architecture Proceedings 2009*: 33–41.
- Englhart, S., Keuck, V., Siegert, F., 2011. Aboveground biomass retrieval in tropical forests — The potential of combined X- and L-band SAR data use. *Remote Sensing of Environment* 115 (5): 1260–1271.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Portland, Oregon, USA, 226–231.
- Eysn, L., Hollaus, M., Lindberg, E., Berger, F., Monnet, J.-M., Dalponte, M., Kobal, M., Pellegrini, M., Lingua, E., Mongus, D., Pfeifer, N., 2015. A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space. *Forests* 6 (5): 1721–1747.
- Fan, H., Yao, W., Fu, Q., 2014a. Segmentation of Sloped Roofs from Airborne LiDAR Point Clouds Using Ridge-Based Hierarchical Decomposition. *Remote Sensing* 6 (4): 3284–3301.
- Fan, H., Zipf, A., Fu, Q., Neis, P., 2014b. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science* 28 (4): 700–719.
- Ferraioli, G., 2010. Multichannel InSAR Building Edge Detection. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3): 1224–1231.
- Ferretti, A., Bianchi, M., Prati, C., Rocca, F., 2005. Higher-order permanent scatterers analysis. *EURASIP Journal on Applied Signal Processing* 2005: 3231–3242.
- Ferretti, A., Prati, C., Rocca, F., 2000. Analysis of permanent scatterers in SAR interferometry. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 2, 761–763.
- Ferretti, A., Prati, C., Rocca, F., 2001. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 39 (1): 8–20.
- Ferretti, A., Prati, C., Rocca, F., Guarnieri, A. M., 1997. Multibaseline SAR interferometry for automatic DEM reconstruction. In: *3rd ERS Workshop*, Florence, Italy.
- Ferro, A., Brunner, D., Bruzzone, L., 2013. Automatic Detection and Reconstruction of Building Radar Footprints From Single VHR SAR Images. *IEEE Transactions on Geoscience and Remote*

- Sensing 51 (2): 935–952.
- Fischler, M. A., Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* 24 (6): 381–395.
- Forlani, G., Nardinocchi, C., Scaioni, M., Zingaretti, P., 2006. Complete classification of raw LIDAR data and 3d reconstruction of buildings. *Pattern Analysis and Applications* 8 (4): 357–374.
- Fornaro, G., Lombardini, F., Serafino, F., 2005. Three-dimensional multipass SAR focusing: experiments with long-term spaceborne data. *IEEE Transactions on Geoscience and Remote Sensing* 43 (4): 702–714.
- Fornaro, G., Pauciuolo, A., Reale, D., Verde, S., 2014. Multilook SAR Tomography for 3-D Reconstruction and Monitoring of Single Structures Applied to COSMO-SKYMED Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (7): 2776–2785.
- Fornaro, G., Pauciuolo, A., Reale, D., Zhu, X. X., Bamler, R., 2012. SAR Tomography: An advanced tool for 4f spaceborne radar scanning with application to imaging and monitoring of cities and single buildings. *IEEE Geoscience and Remote Sensing Society Newsletter* : 9–17.
- Fornaro, G., Reale, D., Serafino, F., 2009. Four-Dimensional SAR Imaging for Height Estimation and Monitoring of Single and Double Scatterers. *IEEE Transactions on Geoscience and Remote Sensing* 47 (1): 224–237.
- Fornaro, G., Serafino, F., Reale, D., 2010. 4-D SAR Imaging: The Case Study of Rome. *IEEE Geoscience and Remote Sensing Letters* 7 (2): 236–240.
- Fornaro, G., Serafino, F., Soldovieri, F., 2003. Three-dimensional focusing with multipass SAR data. *IEEE Transactions on Geoscience and Remote Sensing* 41 (3): 507–517.
- Franceschetti, G., Iodice, A., Riccio, D., 2002. A canonical problem in electromagnetic backscattering from buildings. *IEEE Transactions on Geoscience and Remote Sensing* 40 (8): 1787–1801.
- Franceschetti, G., Lanari, R., 1999. Synthetic aperture radar processing. *Electronic engineering systems series*. CRC Press, Boca Raton, FL, USA.
- Freeman, A., Durden, S. L., 1998. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing* 36 (3): 963–973.
- Freeman, H., Shapira, R., 1975. Determining the Minimum-area Encasing Rectangle for an Arbitrary Closed Curve. *Communications of the ACM* 18 (7): 409–413.
- Frey, O., Morsdorf, F., Meier, E., 2008. Tomographic Imaging of a Forested Area By Airborne Multi-Baseline P-Band SAR. *Sensors* 8 (9): 5884–5896.
- Frost, V. S., Stiles, J. A., Shanmugan, K. S., Holtzman, J. C., 1982. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2): 157–166.
- Frulla, L. A., Milovich, J. A., Karszenbaum, H., Gagliardini, D. A., 1998. Radiometric corrections and calibration of SAR images. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 2, 1147–1149.
- Fulkerson, B., Vedaldi, A., Soatto, S., 2009. Class segmentation and object localization with superpixel neighborhoods. In: *Proceedings of IEEE 12th International Conference on Computer Vision (CVPR)*, 2009, 670–677.
- Gabriel, A. K., Goldstein, R. M., Zebker, H. A., 1989. Mapping small elevation changes over large areas: differential radar interferometry. *Journal of Geophysical Research* 94: 9183–9191.
- Galvanin, E. A. d. S., Poz, A. P. D., 2012. Extraction of Building Roof Contours From LiDAR Data Using a Markov-Random-Field-Based Approach. *IEEE Transactions on Geoscience and Remote Sensing* 50 (3): 981–987.
- Gamba, P., Houshmand, B., Saccani, M., 2000. Detection and extraction of buildings from interferometric SAR data. *IEEE Transactions on Geoscience and Remote Sensing* 38 (1): 611–617.
- Geofabrik, 2015.
- Gernhardt, S., Adam, N., Eineder, M., Bamler, R., 2010. Potential of very high resolution SAR for persistent scatterer interferometry in urban areas. *Annals of GIS* 16: 103–111.
- Gernhardt, S., Bamler, R., 2012. Deformation monitoring of single buildings using meter-resolution

- SAR data in PSI. *ISPRS Journal of Photogrammetry and Remote Sensing* 73: 68–79.
- Gernhardt, S., Cong, X., Eineder, M., Hinz, S., Bamler, R., 2012. Geometrical Fusion of Multitrack PS Point Clouds. *IEEE Geoscience and Remote Sensing Letters* 9 (1): 38–42.
- Gini, F., Lombardini, F., 2002. Multilook APES for multibaseline SAR interferometry. *IEEE Transactions on Signal Processing* 50 (7): 1800–1803.
- Gisinger, C., Bals, U., Pail, R., Zhu, X. X., Montazeri, S., Gernhardt, S., Eineder, M., 2015. Precise Three-Dimensional Stereo Localization of Corner Reflectors and Persistent Scatterers With TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing* 53 (4): 1782–1802.
- Goldstein, R., Werner, C., 1998. Radar interferogram filtering for geophysical applications. *Geophysical Research Letters* 25 (21): 4035–4038.
- Goldstein, R. M., Zebker, H. A., Werner, C. L., 1988. Satellite radar interferometry Two-dimensional phase unwrapping. *Radio Science* 23 (4): 713–720.
- Goodman, J., 1975. Statistical properties of laser speckle patterns. Vol. 9 of Dainty J, Ennos A, Françon M, Goodman J, McKechnie T, Parry G, Goodman J (eds) *Laser Speckle and Related Phenomena*. Springer-Verlag, 9–75.
- Gross, H., Thoennesen, U., Hansen, W. v., 2005. 3d-modeling of urban structures. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXVI, Part 3/W24, ISPRS, 137–142.
- Guida, R., Iodice, A., Riccio, D., 2010. Height Retrieval of Isolated Buildings From Single High-Resolution SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 48 (7): 2967–2979.
- Guillaso, S., D’Hondt, O., Hellwich, O., 2012. Extraction of points of interest from SAR tomograms. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 471–474.
- Guillaso, S., D’Hondt, O., Hellwich, O., 2013. Building characterization using polarimetric tomographic SAR data. In: *Proceedings of Joint Urban Remote Sensing Event (JURSE)*, 234–237.
- Guillaso, S., Ferro-Famil, L., Reigber, A., Pottier, E., 2005. Building Characterization Using L-Band Polarimetric Interferometric SAR Data. *IEEE Geoscience and Remote Sensing Letters* 2 (3): 347–351.
- Guillaso, S., Reigber, A., 2005. Scatterer characterisation using polarimetric SAR tomography. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 4, 2685.
- Guo, B., Huang, X., Zhang, F., Sohn, G., 2015. Classification of airborne laser scanning data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing* 100: 71–83.
- Guo, Q., Li, W., Yu, H., Alvarez, O., 2010. Effects of Topographic Variability and Lidar Sampling Density on Several DEM Interpolation Methods. *Photogrammetric Engineering & Remote Sensing* 76 (6): 701–712.
- Gupta, S., 2010. Single tree detection and modeling using airborne laser scanner data. PhD thesis, Faculty of Forest and Environmental Sciences, Albert Ludwigs University of Freiburg, Germany.
- Haala, N., Kada, M., 2010. An update on automatic 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (6): 570–580.
- Haensch, R., 2014. Generic object categorization in PolSAR images - and beyond. PhD thesis, Technical University of Berlin, Berlin, Germany.
- Haensch, R., Hellwich, O., 2010. Object Recognition from Polarimetric SAR Images. In: U. Soergel (Ed), *Radar Remote Sensing of Urban Areas*, 1st Edition. Chapter 5. Springer, 109–132.
- Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design* 37 (4): 682 – 703.
- Hallberg, B., Smith-Jonforsen, G., Ulander, L., 2005. Measurements on individual trees using multiple VHF SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 43 (10): 2261–2269.
- He, W., Jägere, M., Reigber, A., Hellwich, O., 2008. Building Extraction from Polarimetric SAR Data using Mean Shift and Conditional Random Fields. In: *Proceedings of EUSAR 2008*, 439–442.
- He, Y., Zhang, C., Fraser, C. S., 2014. An energy minimization approach to automated extraction of regular building footprints from airborne LiDAR data. In: *ISPRS Annals of the Photogrammetry*,

- Remote Sensing and Spatial Information Sciences, Vol. II-3, ISPRS, Zurich, Switzerland, 65–72.
- Hecht, R., Kunze, C., Hahmann, S., 2013. Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *ISPRS International Journal of Geo-Information* 2 (4): 1066–1091.
- Henn, A., Gröger, G., Stroh, V., Plümer, L., 2013. Model driven reconstruction of roofs from sparse LIDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 76: 17–29.
- Hill, R., Moate, C., Blacknell, D., 2008. Estimating building dimensions from synthetic aperture radar image sequences. *IET Radar, Sonar & Navigation* 2 (3): 189.
- Hirschmugl, M., Ofner, M., Raggam, J., Schardt, M., 2007. Single tree detection in very high resolution remote sensing data. *Remote Sensing of Environment* 110 (4): 533–544.
- Höfle, B., Hollaus, M., Hagenauer, J., 2012. Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 134–147.
- Höfle, B., Hollaus, M., Lehner, H., Pfeifer, N., Wagner, W., 2008. Area-based parameterization of forest structure using full-waveform airborne laser scanning data. In: *Proceedings of Silvilaser 2008*, Edinburgh, Scotland, 227–235.
- Höfle, B., Mücke, W., Dutter, M., Rutzinger, M., Dorninger, P., 2009. Detection of building regions using airborne LiDAR - A new combination of raster and point cloud based GIS methods. In: Car, A., Griesebner, G. & Strobl, J.: *Geospatial Crossroads @ GI Forum '09: Proceedings of the Geoinformatics Forum Salzburg*, Salzburg, Austria, 66–75.
- Hong, S.-H., Kim, H.-O., Wdowinski, S., Feliciano, E., 2015. Evaluation of Polarimetric SAR Decomposition for Classifying Wetland Vegetation Types. *Remote Sensing* 7 (7): 8563–8585.
- Hong, S.-H., Wdowinski, S., 2014. Double-Bounce Component in Cross-Polarimetric SAR From a New Scattering Target Decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 52 (6): 3039–3051.
- Hooper, A., Zebker, H., Segall, P., Kampes, B., 2004. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophysical Research Letters* 31 (23).
- Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W., 1992. Surface reconstruction from unorganized points. *ACM SIGGRAPH Computer Graphics* 26 (2): 71–78.
- Hough, P. V. C., 1962. Method and means for recognizing complex patterns.
- Hron, V., Halounová, L., 2015. Automatic Generation of 3d Building Models from Point Clouds. In: *Geoinformatics for Intelligent Transportation*. Springer, 109–119.
- Hu, J., Guo, R., Zhu, X. X., Baier, G., Wang, Y., 2015. Non-Local Means Filter for Polarimetric Sar Speckle Reduction-Experiments Using TerraSAR-X Data. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W4, Munich, Germany, 71–77.
- Huang, H., Brenner, C., Sester, M., 2013. A generative statistical approach to automatic 3d building roof reconstruction from laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 29–43.
- Huber, P. J., 1981. *Robust Statistics*. John Wiley & Sons, Hoboken, New Jersey, USA.
- Hubert, M., Rousseeuw, P., Branden, K. V., 2005. ROBPCA: A new approach to robust principal components analysis. *Technometrics* 47 (1): 64–79.
- Hulik, R., Spanel, M., Smrz, P., Materna, Z., 2014. Continuous plane detection in point-cloud data based on 3d Hough Transform. *Journal of Visual Communication and Image Representation* 25 (1): 86–97.
- Hyypä, J., Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland* 16: 27–42.
- Hyypä, J., Kelle, O., Lehtikoinen, M., Inkinen, M., 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing* 39 (5): 969–975.
- Izzawati, Wallington, E., Woodhouse, I., 2006. Forest height retrieval from commercial X-band SAR products. *IEEE Transactions on Geoscience and Remote Sensing* 44 (4): 863–870.
- Jarzabek-Rychard, M., 2012. Reconstruction of Building Outlines in Dense Urban Areas Based on

- LIDAR Data and Address Points. In: ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIX-B3, Melbourne, Australia, 121–126.
- Jiang, L.-b., Wang, Z., Yu, W.-x., 2011. Model based building height retrieval from single SAR images. In: Proceedings of 6th IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Vol. 1, 379–384.
- Jiang, M., Lin, Y., 2013. Individual Deciduous Tree Recognition in Leaf-Off Aerial Ultrahigh Spatial Resolution Remotely Sensed Imagery. *IEEE Geoscience and Remote Sensing Letters* 10 (1): 38–42.
- Jin, Y.-Q., Xu, F., 2013. Automatic Reconstruction of Building Objects from Multi-Aspect SAR Images. In: *Polarimetric Scattering and SAR Information Retrieval*. John Wiley & Sons (Asia) Pte Ltd, 235–274.
- Jing, L., Hu, B., Li, J., Noland, T., 2012. Automated delineation of individual tree crowns from LiDAR data by multi-scale analysis and segmentation. *Photogrammetric Engineering & Remote Sensing* 78 (12): 1275–1284.
- Jochem, A., Höfle, B., Rutzinger, M., Pfeifer, N., 2009. Automatic Roof Plane Detection and Analysis in Airborne Lidar Point Clouds for Solar Potential Assessment. *Sensors* 9 (7): 5241–5262.
- Jochem, A., Höfle, B., Wichmann, V., Rutzinger, M., Zipf, A., 2012. Area-wide roof plane segmentation in airborne LiDAR point clouds. *Computers, Environment and Urban Systems* 32: 54–64.
- Jwa, Y., Sohn, G., Tao, V., Cho, W., 2008. An implicit geometric regularization of 3d building shape using airborne lidar data. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Laser Scanning*, Vol. 37, 69–76.
- Kaartinen, H., Hyypä, J., Yu, X., Vastaranta, M., Hyypä, H., Kukko, A., Holopainen, M., Heipke, C., Hirschmugl, M., Morsdorf, F., Næsset, E., Pitkänen, J., Popescu, S., Solberg, S., Wolf, B. M., Wu, J.-C., 2012. An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. *Remote Sensing* 4 (12): 950–974.
- Kada, M., McKinley, L., 2009. 3d building reconstruction from LiDAR based on a cell decomposition approach. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38 (Part 3/W4), Paris, France, 47–52.
- Kampes, B. M., 2006. *Radar Interferometry: Persistent Scatterer Technique*, 1st Edition. 12. Springer Netherlands.
- Kang, X., Liu, J., Lin, X., 2014a. Streaming Progressive TIN Densification Filter for Airborne LiDAR Point Clouds Using Multi-Core Architectures. *Remote Sensing* 6 (8): 7212–7232.
- Kang, Z., Zhang, L., Wang, B., Li, Z., Jia, F., 2014b. An Optimized BaySAC Algorithm for Efficient Fitting of Primitives in Point Clouds. *IEEE Geoscience and Remote Sensing Letters* 11 (6): 1096–1100.
- Khachiyan, L. G., 1996. Rounding of Polytopes in the Real Number Model of Computation. *Mathematics of Operations Research* 21 (2): 307–320.
- Khoshelham, K., Elberink, S. O., 2012. Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data. In: *Proceedings of the International Conference on Geographic Object Based Image Analysis*, Rio de Janeiro, Brazil, 372–377.
- Kim, K., Shan, J., 2011. Building roof modeling from airborne laser scanning data based on level set approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (4): 484–497.
- Kirtner, J., 2000. Using LiDAR data in wireless communication system design. In: *Proceedings of American Society for Photogrammetry & Remote Sensing (ASPRS)*, ASPRS, Washington DC, USA.
- Kobler, A., Pfeifer, N., Ogrinc, P., Todorovski, L., Oštir, K., Džeroski, S., 2007. Repetitive interpolation: A robust algorithm for DTM generation from Aerial Laser Scanner Data in forested terrain. *Remote Sensing of Environment* 108 (1): 9–23.
- Koch, B., Heyder, U., Weinacker, H., 2006. Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering & Remote Sensing* 72 (4): 357–363.
- Koeninger, A., Bartel, S., 1998. 3d-GIS for urban purposes. *Geoinformatica* 2 (1): 79–103.
- Kolbe, T. H., Gröger, G., Plümer, L., 2008. CityGML–3d city models and their potential for emergency response. *Geospatial information technology for emergency response* 257.
- Kolmogorov, V., Zabih, R., 2004. What energy functions can be minimized via graph cuts? *IEEE*

- Transactions on Pattern Analysis and Machine Intelligence 26 (2): 147–159.
- Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 53 (4): 193–203.
- Kumar, P., Yildirim, E. A., 2005. Minimum-Volume Enclosing Ellipsoids and Core Sets. *Journal of Optimization Theory and Applications* 126 (1): 1–21.
- Lach, S. R., Kerekes, J. P., 2008. Robust extraction of exterior building boundaries from topographic LiDAR data. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 2, II–85.
- Lafarge, F., Alliez, P., 2013. Surface Reconstruction through Point Set Structuring. *Computer Graphics Forum* 32 (2pt2): 225–234.
- Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2008. Automatic building extraction from DEMs using an object approach and application to the 3d-city modeling. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (3): 365–381.
- Lafarge, F., Descombes, X., Zerubia, J., Pierrot-Deseilligny, M., 2010. Structural Approach for Building Reconstruction from a Single DSM. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (1): 135–147.
- Lafarge, F., Mallet, C., 2012. Creating Large-Scale City Models from 3d-Point Clouds: A Robust Approach with Hybrid Representation. *International Journal of Computer Vision* 99 (1): 69–85.
- Laprade, G. L., Leonardo, E. S., 1969. ELEVATIONS FROM RADAR IMAGERY. *Photogrammetric Engineering*.
- Leberl, F. W., 1990. *Radargrammetric Image Processing*. Artech House, Norwood, MA.
- Ledoux, H., Meijers, M., 2011. Topologically consistent 3d city models obtained by extrusion. *International Journal of Geographical Information Science* 25 (4): 557–574.
- Lee, H., Slatton, K. C., Roth, B. E., JR, W. P. C., 2010. Adaptive clustering of airborne LiDAR data to segment individual tree crowns in managed pine forests. *International Journal of Remote Sensing* 31 (1): 117–139.
- Lee, J.-S., 1980. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2): 165–168.
- Lee, J.-S., Ainsworth, T. L., 2011. The Effect of Orientation Angle Compensation on Coherency Matrix and Polarimetric Target Decompositions. *IEEE Transactions on Geoscience and Remote Sensing* 49 (1): 53–64.
- Lee, J.-S., Grunes, M. R., Ainsworth, T. L., Du, L.-J., Schuler, D. L., Cloude, S. R., 1999. Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing* 37 (5): 2249–2258.
- Lee, J.-S., Hoppel, K. W., Mango, S., Miller, A. R., 1994. Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 32 (5): 1017–1028.
- Lee, J. S., Pottier, E., 2009. *Polarimetric Radar Imaging: From Basics to Applications*. Optical Science and Engineering. CRC Press.
- Li, F. K., Goldstein, R. M., 1990. Studies of multibaseline spaceborne interferometric synthetic aperture radars. *IEEE Transactions on Geoscience and Remote Sensing* 28 (1): 88–97.
- Li, W., Guo, Q., Jakubowski, M. K., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. *Photogrammetric Engineering & Remote Sensing* 78 (1): 75–84.
- Li, Y., Yong, B., Wu, H., An, R., Xu, H., 2014. An Improved Top-Hat Filter with Sloped Brim for Extracting Ground Points from Airborne Lidar Point Clouds. *Remote Sensing* 6: 12885–12908.
- Lillesand, T. M., Kiefer, R. W., Chipman, J. W., 2004. *Remote Sensing and Image Interpretation*, 5th Edition. John Wiley & Sons, New York, NY, USA.
- Lin, X., Zhang, J., 2014. Segmentation-Based Filtering of Airborne LiDAR Point Clouds by Progressive Densification of Terrain Segments. *Remote Sensing* 6 (2): 1294–1326.
- Lindberg, E., Eysn, L., Hollaus, M., Holmgren, J., Pfeifer, N., 2014. Delineation of Tree Crowns and Tree Species Classification From Full-Waveform Airborne Laser Scanning Data Using 3-D Ellipsoidal Clustering. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote*

- Sensing 7 (7): 3174–3181.
- Lindenberger, J., 1993. Laser-Profilmessungen zur topographischen Geländeaufnahme. PhD thesis, Fakultät für Bauingenieur- und Vermessungswesen der Universität Stuttgart, Stuttgart, Germany.
- Liu, B., Tupin, F., Liu, X., Yu, W., 2013a. Characterization and extraction of building layovers in urban areas using high resolution SAR imagery. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 895–898.
- Liu, C., Shi, B., Yang, X., Li, N., Wu, H., 2013b. Automatic Buildings Extraction From LiDAR Data in Urban Area by Neural Oscillator Network of Visual Cortex. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (4): 2008–2019.
- Liu, W., Yamazaki, F., Adriano, B., Mas, E., Koshimura, S., 2014. Development of Building Height Data in Peru from High-Resolution SAR Imagery. *Journal of Disaster Research* 9 (6): 1042–1049.
- Liu, Y., Xiong, Y., 2008. Automatic segmentation of unorganized noisy point clouds based on the Gaussian map. *Computer-Aided Design* 40 (5): 576–594.
- Lodha, S. K., Fitzpatrick, D. M., Helmbold, D. P., 2007a. Aerial lidar data classification using adaboost. In: Proceeding of Sixth International Conference on 3-D Digital Imaging and Modeling, 3DIM'07, 435–442.
- Lodha, S. K., Fitzpatrick, D. M., Helmbold, D. P., 2007b. Aerial lidar data classification using expectation-maximization. In: *Electronic Imaging 2007*, International Society for Optics and Photonics, 64990L–64990L.
- Lombardini, F., 2005. Differential tomography: a new framework for SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 43 (1): 37–44.
- Lombardini, F., Gini, F., Matteucci, P., 2001. Application of array processing techniques to multibaseline InSAR for layover solution. In: *Proceedings of the IEEE Radar Conference*, 210–215.
- Lombardini, F., Pardini, M., 2012. Superresolution Differential Tomography: Experiments on Identification of Multiple Scatterers in Spaceborne SAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 50 (4): 1117–1129.
- Lombardini, F., Reigber, A., 2003. Adaptive spectral estimation for multibaseline SAR tomography with airborne L-band data. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 3, III–2014.
- Lopes, A., Nezry, E., Touzi, R., Laur, H., 1990. Maximum a posteriori speckle filtering and first order texture models in SAR images. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2409–2412.
- Ma, P., Lin, H., Lan, H., Chen, F., 2015. Multi-dimensional SAR tomography for monitoring the deformation of newly built concrete buildings. *ISPRS Journal of Photogrammetry and Remote Sensing* 106: 118–128.
- Maas, H.-G., Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2–3): 153–163.
- Maksymiuk, O., Schmitt, M., Auer, S., Stilla, U., 2014. Single Tree Detection in Millimeterwave SAR Data by Morphological Attribute Filters. *Proc. 34. Jahrestagung der DGPF*.
- Maksymiuk, O., Stilla, U., 2012. A concept for building reconstruction from airborne multi-aspect SAR data. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 7405–7408.
- Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (6): S71–S84.
- Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., Rabaute, T., 1993. The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364 (6433): 138–142.
- May, N. C., Toth, C. K., 2007. Point positioning accuracy of airborne LiDAR systems: A rigorous analysis. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36 (3/W49B), ISPRS, 107–111.
- Medeiros, E., Velho, L., Lopes, H., 2004. Restricted BPA: applying ball-pivoting on the plane. In:

- Proceedings of 17th Brazilian Symposium on Computer Graphics and Image Processing, 372–379.
- Meng, X., Currit, N., Zhao, K., 2010. Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sensing* 2 (3): 833–860.
- Meng, X., Wang, L., Currit, N., 2009. Morphology-based Building Detection from Airborne Lidar Data. *Photogrammetric Engineering & Remote Sensing* 75 (4): 437–442.
- Mezzasoma, S., Gallon, A., Impagnatiello, F., Angino, G., Fagioli, S., Capuzi, A., Caltagirone, F., Leonardi, A. R., Ziliotto, C. U., 2008. COSMO-SkyMed system commissioning: End-to-end system performance verification. In: *Proceedings of IEEE Radar Conference*, 1–5.
- Michaelsen, E., Soergel, U., Thoennessen, U., 2005. Potential of building extraction from multi-aspect high resolution amplitude SAR data. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 149–154.
- Michaelsen, E., Soergel, U., Thoennessen, U., 2006. Perceptual grouping for automatic detection of man-made structures in high-resolution SAR data. *Pattern Recognition Letters* 27 (4): 218–225.
- Miliaresis, G., Kokkas, N., 2007. Segmentation and object-based classification for the extraction of the building class from LIDAR DEMs. *Computers & Geosciences* 33 (8): 1076–1087.
- Minh, H. T., Toan, T. L., Rocca, F., Tebaldini, S., d’Alessandro, M. M., Villard, L., 2014. Relating P-Band Synthetic Aperture Radar Tomography to Tropical Forest Biomass. *IEEE Transactions on Geoscience and Remote Sensing* 52 (2): 967–979.
- Mitra, N. J., Nguyen, A., 2004. Estimating surface normals in noisy point cloud data. *International Journal of Computational Geometry & Applications* 14 (04n05): 261–276.
- Mongus, D., Lukač, N., Žalik, B., 2014. Ground and building extraction from LiDAR data based on differential morphological profiles and locally fitted surfaces. *ISPRS Journal of Photogrammetry and Remote Sensing* 93: 145–156.
- Mongus, D., Zalik, B., 2012. Parameter-free ground filtering of LiDAR data for automatic DTM generation. *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 1–12.
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K. P., 2013. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine* 1 (1): 6–43.
- Morsdorf, F., Meier, E., Kötz, B., Itten, K. I., Dobbertin, M., Allgöwer, B., 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Environment* 92 (3): 353–362.
- Munoz, D., Bagnell, J. A., Vandapel, N., Hebert, M., 2009. Contextual classification with functional max-margin markov networks. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 975–982.
- Nannini, M., Scheiber, R., Horn, R., 2008. Imaging of targets beneath foliage with SAR tomography. In: *Proceedings of 7th European Conference on Synthetic Aperture Radar (EUSAR)*, VDE, 1–4.
- Nelder, J. A., Mead, R., 1965. A Simplex Method for Function Minimization. *The Computer Journal* 7 (4): 308–313.
- Neumann, M., Ferro-Famil, L., Reigber, A., 2010. Estimation of Forest Structure, Ground, and Canopy Layer Characteristics From Multibaseline Polarimetric Interferometric SAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3): 1086–1104.
- Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 152–165.
- Over, M., Schilling, A., Neubauer, S., Zipf, A., 2010. Generating web-based 3d City Models from OpenStreetMap: The current situation in Germany. *Computers, Environment and Urban Systems* 34 (6): 496–507.
- Palsson, F., Sveinsson, J. R., Ulfarsson, M. O., Benediktsson, J., others, 2012. SAR image denoising using total variation based regularization with sure-based optimization of the regularization parameter. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2160–2163.
- Pasquali, P., Prati, C., Rocca, F., Seymour, M., Fortuny, J., Ohlmer, E., Sieber, A. J., 1995. A 3-D SAR experiment with EMSL data. In: *Proceedings of IEEE International Geoscience and Remote Sensing*

- Symposium (IGARSS), Vol. 1, 784–786.
- Pauly, M., Keiser, R., Gross, M., 2003. Multi-scale Feature Extraction on Point-Sampled Surfaces. *Computer graphics forum* 22 (3): 281–289.
- Perera, S. N., Nalani, H. A., Maas, H.-G., 2012. An automated method for 3d roof outline generation and regularization in airborne laser scanner data. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, 281–286.
- Perissin, D., Ferretti, A., 2007. Urban-Target Recognition by Means of Repeated Spaceborne SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 45 (12): 4043–4058.
- Perko, R., Raggam, H., Deutscher, J., Gutjahr, K., Schardt, M., 2011. Forest Assessment Using High Resolution SAR Data in X-Band. *Remote Sensing* 3 (12): 792–815.
- Persson, A., Holmgren, J., Söderman, U., 2002. Detecting and Measuring Individual Trees Using an Airborne Laser Scanner. *Photogrammetric Engineering & Remote Sensing* 68 (9): 925–932.
- Pingel, T. J., Clarke, K. C., McBride, W. A., 2013. An improved simple morphological filter for the terrain classification of airborne LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 77: 21–30.
- Portales, C., Lerma, J. L., Navarro, S., 2010. Augmented reality and photogrammetry: A synergy to visualize physical and virtual city environments. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (1): 134–142.
- Pouliot, J., Daniel, S., Hubert, F., Zamyadi, A., 2012. *Progress and New Trends in 3D Geoinformation Sciences*. Springer Science & Business Media.
- Poullis, C., You, S., 2009. Automatic reconstruction of cities from remote sensor data. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2775–2782.
- Priestnall, G., Jaafar, J., Duncan, A., 2000. Extracting urban features from LiDAR digital surface models. *Computers, Environment and Urban Systems* 24 (2): 65–78.
- Quartulli, M., Datcu, M., 2004. Stochastic geometrical modeling for built-up area understanding from a single SAR intensity image with meter resolution. *IEEE Transactions on Geoscience and Remote Sensing* 42 (9): 1996–2003.
- Rabbani, T., Heuvel, F. A. v. d., Vosselman, G., 2006. Rabbani et al.-Segmentation of point clouds using smoothness constraints. In: *ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*.
- Rau, J.-Y., Lin, B.-C., 2011. Automatic roof model reconstruction from ALS data and 2d ground plans based on side projection and the TMR algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (6): S13–S27.
- Reale, D., Fornaro, G., Pauciullo, A., 2013. Extension of 4-D SAR Imaging to the Monitoring of Thermally Dilating Scatterers. *IEEE Transactions on Geoscience and Remote Sensing* 51 (12): 5296–5306.
- Reale, D., Fornaro, G., Pauciullo, A., Zhu, X., Bamler, R., 2011. Tomographic Imaging and Monitoring of Buildings With Very High Resolution SAR Data. *IEEE Geoscience and Remote Sensing Letters* 8 (4): 661–665.
- Reigber, A., Jager, M., He, W., Ferro-Famil, L., Hellwich, O., 2007. Detection and classification of urban structures based on high-resolution SAR imagery. In: *Proceedings of Joint Urban Remote Sensing Event (JURSE)*, 1–6.
- Reigber, A., Moreira, A., 2000. First demonstration of airborne SAR tomography using multibaseline L-band data. *IEEE Transactions on Geoscience and Remote Sensing* 38 (5): 2142–2152.
- Reitberger, J., Schnörr, C., Krzystek, P., Stilla, U., 2009. 3d segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (6): 561–574.
- Rocca, F., 2007. Modeling Interferogram Stacks. *IEEE Transactions on Geoscience and Remote Sensing* 45 (10): 3289–3299.
- Rottensteiner, F., 2003. Automatic generation of high-quality building models from lidar data. *Computer Graphics and Applications* 23 (6): 42–50.
- Rottensteiner, F., Briese, C., 2002. A new method for building extraction in urban areas from high-resolution LIDAR data. In: *International Archives of Photogrammetry Remote Sensing and Spatial*

- Information Sciences, Vol. 34-3/A, 295–301.
- Sampath, A., Shan, J., 2007. Building boundary tracing and regularization from airborne LiDAR point clouds. *Photogrammetric Engineering & Remote Sensing* 73 (7): 805–812.
- Sampath, A., Shan, J., 2010. Segmentation and Reconstruction of Polyhedral Building Roofs From Aerial Lidar Point Clouds. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3): 1554–1567.
- Sauer, S., Ferro-Famil, L., Reigber, A., Pottier, E., 2009. Polarimetric Dual-Baseline InSAR Building Height Estimation at L-Band. *IEEE Geoscience and Remote Sensing Letters* 6 (3): 408–412.
- Sauer, S., Ferro-Famil, L., Reigber, A., Pottier, E., 2011. Three-Dimensional Imaging and Scattering Mechanism Estimation Over Urban Scenes Using Dual-Baseline Polarimetric InSAR Observations at L-Band. *IEEE Transactions on Geoscience and Remote Sensing* 49 (11): 4616–4629.
- Schimpf, H., Essen, H., Boehmsdorff, S., Brehm, T., 2002. MEMPHIS — A fully polarimetric experimental radar. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1714–1716.
- Schmitt, M., 2014. Reconstruction of Urban Surface Models from Multi-Aspect and Multi-Baseline Interferometric SAR. PhD thesis, Technical University of Munich (TUM), Munich, Germany.
- Schmitt, M., Brück, A., Schönberger, J., Stilla, U., 2013. Potential of airborne single-pass millimeterwave InSAR data for individual tree recognition. In: *DGPF Tagungsband 22 / 2013 – Dreiländertagung DGPF, OVG, SGPF, Vol. 33*, 427–436.
- Schmitt, M., Shahzad, M., Zhu, X. X., 2015. Reconstruction of individual trees from multi-aspect TomoSAR data. *Remote Sensing of Environment* 165: 175–185.
- Schmitt, M., Stilla, U., 2014a. Generating point clouds of forested areas from airborne millimeter wave InSAR data. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1–4.
- Schmitt, M., Stilla, U., 2014b. Maximum-likelihood estimation for multi-aspect multi-baseline SAR interferometry of urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 87: 68–77.
- Schnabel, R., Wahl, R., Klein, R., 2007. Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum* 26 (2): 214–226.
- Scholze, S., Moons, T., Van Gool, L., 2002. A probabilistic approach to building roof reconstruction using semantic labelling. In: *Pattern Recognition. Vol. 2449 of Lecture Notes in Computer Science*. Springer, 257–264.
- Schou, J., Skriver, H., Nielsen, A., Conradsen, K., 2003. CFAR edge detector for polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 41 (1): 20–32.
- Sehrawat, P., Kensek, K., 2014. URBAN ENERGY MODELING: GIS AS AN ALTERNATIVE TO BIM. In: *2014 ASHRAE/IBPSA-USA, Atlanta, USA*.
- Seo, S., Lee, J., Kim, Y., 2014. Extraction of Boundaries of Rooftop Fenced Buildings From Airborne Laser Scanning Data Using Rectangle Models. *IEEE Geoscience and Remote Sensing Letters* 11 (2): 404–408.
- Shahzad, M., Schmidt, M., Zhu, X. X., 2015. Segmentation and crown parameter extraction of Individual trees in an airborne TomoSAR point cloud. In: *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences, Vol. XL-3/W2, Munich, Germany*, 205–209.
- Shahzad, M., Zhu, X. X., 2014. Reconstructing 2-D/3-D Building Shapes from Spaceborne Tomographic Synthetic Aperture Radar Data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XL-3, ISPRS, Zurich, Switzerland*, 313–320.
- Shahzad, M., Zhu, X. X., 2015a. Automatic detection and reconstruction of 2d/3d building shapes from spaceborne TomoSAR point clouds. *IEEE Transactions on Geoscience and Remote Sensing* (accepted).
- Shahzad, M., Zhu, X. X., 2015b. Robust Reconstruction of Building Facades for Large Areas Using Spaceborne TomoSAR Point Clouds. *IEEE Transactions on Geoscience and Remote Sensing* 53 (2): 752–769.
- Shahzad, M., Zhu, X. X., Bamler, R., 2012. Facade structure reconstruction using spaceborne TomoSAR point clouds. In: *IEEE International Geoscience and Remote Sensing Symposium*

- (IGARSS), Munich, Germany, 467–470.
- Shamaoma, H., Kerle, N., Alkema, D., 2006. Extraction of flood-modelling related base-data from multi-source remote sensing imagery. In: ISPRS mid-term symposium Commission VII, Vol. WG/7, ISPRS.
- Shan, J., Sampath, A., 2005. Urban DEM generation from raw lidar data. *Photogrammetric Engineering & Remote Sensing* 71 (2): 217–226.
- Shapovalov, R., Velizhev, A., Barinova, O., 2010. Non-associative markov networks for 3d point cloud classification. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-3A, Saint-Mandé, France.
- Sheng, Y., Gong, P., Biging, G. S., 2001. Model-based conifer-crown surface reconstruction from high-resolution aerial images. *Photogrammetric Engineering & Remote Sensing* 67 (8): 957–965.
- Shi, J., Malik, J., 2000. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8): 888–905.
- Shiver, B. D., Borders, B. E., 1996. *Sampling Techniques for Forest Resource Inventory*. Hoboken, NJ: Wiley.
- Simonetto, E., Oriot, H., Garello, R., 2005. Rectangular building extraction from stereoscopic airborne Radar images. *IEEE Transactions on Geoscience and Remote Sensing* 43 (10): 2386–2395.
- Sithole, G., 2001. Filtering of laser altimetry data using a slope adaptive filter. In: *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Vol. 34-3/W4, 203–210.
- Sithole, G., Vosselman, G., 2005. Filtering of airborne laser scanner data based on segmented point clouds. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36-part 3/W19.
- Soergel, U. (Hrsg.), 2010. *Radar Remote Sensing of Urban Areas*. Vol. 15 of *Remote Sensing and Digital Image Processing*. Springer Netherlands, Dordrecht.
- Soergel, U., Michaelsen, E., Thiele, A., Cadario, E., Thoennesen, U., 2009. Stereo analysis of high-resolution SAR images for building height estimation in cases of orthogonal aspect directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (5): 490–500.
- Soergel, U., Michaelsen, E., Thiele, A., Thoennesen, U., 2006. Radargrammetric Extraction of Building Features from High Resolution Multi-aspect SAR Data. In: *Proceedings of IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS)*, 3635–3638.
- Soergel, U., Thiele, A., Gross, H., Thoennesen, U., 2007. Extraction of bridge features from high-resolution InSAR data and optical images. In: *Proceedings of Joint Urban Remote Sensing Event (JURSE)*, IEEE, 1–6.
- Soergel, U., Thoennesen, U., Stilla, U., 2003a. Iterative building reconstruction from multiaspect InSAR data. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 34(Part 3/W13), 186–192.
- Soergel, U., Thoennesen, U., Stilla, U., 2003b. Reconstruction of buildings from interferometric SAR data for built-up areas. In: *International archives of photogrammetry and remote sensing*, Vol. XXXIV 3/W8, Munich, Germany, 59–64.
- Sohn, G., Huang, X., Tao, V., 2008. Using a binary space partitioning tree for reconstructing polyhedral building models from airborne lidar data. *Photogrammetric Engineering & Remote Sensing* 74 (11): 1425–1438.
- Sohn, G., Jwa, Y., Kim, H. B., Jung, J., 2012. An implicit regularization for 3d building rooftop modeling using airborne LIDAR Data. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, Melbourne, Australia, 305–310.
- Soille, P., 1999. *Morphological Image Analysis*.
- Sportouche, H., Tupin, F., Denise, L., 2011. Extraction and Three-Dimensional Reconstruction of Isolated Buildings in Urban Scenes From High-Resolution Optical and SAR Spaceborne Images. *IEEE Transactions on Geoscience and Remote Sensing* 49 (10): 3932–3946.
- Stilla, U., Schmitt, M., Maksymiuk, O., Auer, S., 2014. Towards the recognition of individual trees in decimeter-resolution airborne millimeterwave SAR. In: *Proceedings of 8th IAPR Workshop on*

- Pattern Recognition in Remote Sensing (PRRS), 1–4.
- Stilla, U., Soergel, U., Thoennessen, U., 2003. Potential and limits of InSAR data for building reconstruction in built-up areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (1–2): 113–123.
- Straub, B. M., Heipke, C., 2001. Automatic extraction of trees for 3d-City models from images and height data. In: E. Baltsavias, A. Gruen, & L. V. Gool (Eds.), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*. Vol. 3. 267–277.
- Street, J. O., Carroll, R. J., Ruppert, D., 1988. A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician* 42 (2): 152–154.
- Strîmbu, V. F., Strîmbu, B. M., 2015. A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 104: 30–43.
- Sun, S., Salvaggio, C., 2013. Aerial 3d Building Detection and Modeling From Airborne LiDAR Point Clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (3): 1440–1449.
- Susaki, J., 2012. Adaptive Slope Filtering of Airborne LiDAR Data in Urban Areas for Digital Terrain Model (DTM) Generation. *Remote Sensing* 4 (12): 1804–1819.
- Suzuki, S., Abe, K., 1985. Topological structural analysis of digitized binary images by border following.pdf. *Computer Vision, Graphics, and Image Processing* 30 (1): 32–46.
- Tao, J., Palubinskas, G., Reinartz, P., 2011. Automatic interpretation of high resolution SAR images: first results of SAR image simulation for single buildings. In: *Proceedings of ISPRS Hannover Workshop*, 14–17.
- Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., 2007. Hough-Transform and Extended RANSAC Algorithms for Automatic Detection of 3d Building Roof Planes from Lidar Data. In: *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, Vol. XXXVI, 407–412.
- Tebaldini, S., 2009. Algebraic Synthesis of Forest Scenarios From Multibaseline PolInSAR Data. *IEEE Transactions on Geoscience and Remote Sensing* 47 (12): 4132–4142.
- Tebaldini, S., Rocca, F., 2012. Multibaseline Polarimetric SAR Tomography of a Boreal Forest at P- and L-Bands. *IEEE Transactions on Geoscience and Remote Sensing* 50 (1): 232–246.
- Thiele, A., 2013. 3d Building Reconstruction From High Resolution Multi-Aspect Interferometric Synthetic Aperture RADAR Data. PhD thesis, Leibniz University of Hanover, Hannover, Germany.
- Thiele, A., Cadario, E., Schulz, K., Thönnessen, U., Soergel, U., 2007. Building Recognition From Multi-Aspect High-Resolution InSAR Data in Urban Areas. *IEEE Transactions on Geoscience and Remote Sensing* 45 (11): 3583–3593.
- Thiele, A., Hinz, S., Cadario, E., 2010. Combining GIS and InSAR data for 3d building reconstruction. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2418–2421.
- Thiele, A., Wurth, M. M., Even, M., Hinz, S., 2013. Extraction of building shape from TanDEM-X data.pdf. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-1/W1, Hannover, Germany.
- Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2): 411–423.
- Tison, C., Tupin, F., Maitre, H., 2004. Retrieval of building shapes from shadows in high resolution SAR interferometric images. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 3, 1788–1791 vol.3.
- Tison, C., Tupin, F., Maître, H., 2007. A Fusion Scheme for Joint Retrieval of Urban Height Map and Classification From High-Resolution Interferometric SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 45 (2): 496–505.
- Tomaszewski, B., 2014. *Geographic Information Systems (GIS) for Disaster Management*. CRC Press.
- Tomljenovic, I., Höfle, B., Tiede, D., Blaschke, T., 2015. Building Extraction from Airborne Laser Scanning Data: An Analysis of the State of the Art. *Remote Sensing* 7 (4): 3826–3862.
- Tordoff, B., Murray, D., 2005. Guided-MLESAC: faster image transform estimation by using matching priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (10): 1523–1535.

- Torr, P. H. S., Zisserman, A., 2000. MLESAC: A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding* 78 (1): 138–156.
- Toussaint, G. T., 1983. Solving geometric problems with the rotating calipers. In: *Proceedings of IEEE MELECON*, Vol. 83, Athens, Greece, A10.
- Touzi, R., Lopes, A., Bousquet, P., 1988. Touzi-A statistical and geometrical edge detector for SAR images. *IEEE Transactions on Geoscience and Remote Sensing* 26 (6): 764–773.
- Tóvári, D., Pfeifer, N., 2005. Segmentation based robust interpolation—a new approach to laser data filtering. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 36-3/W19, 79–84.
- Tupin, F., 2003. Extraction of 3d information using overlay detection on SAR images. In: *Proceedings of 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 72–76.
- Tupin, F., Maitre, H., Mangin, J.-F., Nicolas, J.-M., Pechersky, E., 1998. Detection of linear features in SAR images: application to road network extraction. *IEEE Transactions on Geoscience and Remote Sensing* 36 (2): 434–453.
- Tupin, F., Roux, M., 2005. Markov random field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications. *IEEE Transactions on Geoscience and Remote Sensing* 43 (8): 1920–1928.
- Vauhkonen, J., Ene, L., Gupta, S., Heinzl, J., Holmgren, J., Pitkanen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K. M., Lien, V., Packalen, P., Gobakken, T., Koch, B., Naesset, E., Tokola, T., Maltamo, M., 2012. Comparative testing of single-tree detection algorithms under different types of forest. *Forestry* 85 (1): 27–40.
- Verma, V., Kumar, R., Hsu, S., 2006. 3d Building Detection and Modeling from Aerial LIDAR Data. Vol. 2, 2213–2220.
- Vosselman, G., 2000. Slope based filtering of laser altimetry data. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 33-B3/2 PART 3, 935–942.
- Vosselman, G., 2013. Point cloud segmentation for urban scene classification. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-7/W2, 257–262.
- Vosselman, G., Dijkman, S., 2001. 3d building model reconstruction from point clouds and ground plans. In: *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, Vol. 34-3/W4, 37–44.
- Vosselman, G., Maas, H.-G., 2010. *Airborne and Terrestrial Laser Scanning*. Whittles Publishing.
- Waldhauser, C., Hochreiter, R., Otepka, J., Pfeifer, N., Ghuffar, S., Korzeniowska, K., Wagner, G., 2014. Automated classification of airborne laser scanning point clouds. In: *Solving Computationally Expensive Engineering Problems*. Springer, 269–292.
- Wang, C.-K., Tseng, Y.-H., 2010. Dem generation from airborne lidar data by an adaptive dual-directional slope filter. In: Wagner W., Székely, B. (eds.): *ISPRS TC VII Symposium - 100 Years ISPRS*, Vol. XXXVIII, Part 7B, IAPRS, Vienna, Austria, 628–632.
- Wang, D., Terman, D., 1997. Image segmentation based on oscillatory correlation. *Neural Computation* 9 (4): 805–836.
- Wang, J., Shan, J., 2009. Segmentation of LiDAR point clouds for building extraction. In: *Proceedings of American Society for Photogramm. Remote Sens. (ASPRS), Annual Conference*, Baltimore, MD, 9–13.
- Wang, O., Lodha, S. K., Helmbold, D. P., 2006. A bayesian approach to building footprint extraction from aerial lidar data. In: *Proceedings of Third International Symposium on 3D Data Processing, Visualization, and Transmission*, 192–199.
- Wang, Y., Tupin, F., Han, C., Nicolas, J.-M., 2008a. Building detection from high resolution POLSAR data by combining region and edge information. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 4, IV–153.
- Wang, Y., Weinacker, H., Koch, B., 2008b. A Lidar Point Cloud Based Procedure for Vertical Canopy Structure Analysis And 3d Single Tree Modelling in Forest. *Sensors* 8 (6): 3938–3951.
- Wang, Y., Zhu, X., 2015. Automatic Feature-based Geometric Fusion of Multi-view TomoSAR Point

- Clouds in Urban Area. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (3): 953 – 965, 00000.
- Wang, Z., Jiang, L., Lei, L., Yu, W., 2015. Building Height Estimation from High Resolution SAR Imagery via Model-Based Geometrical Structure Prediction. *Progress In Electromagnetics Research M* 41: 11–24.
- Wegner, J. D., Ziehn, J. R., Soergel, U., 2014. Combining High-Resolution Optical and InSAR Features for Height Estimation of Buildings With Flat Roofs. *IEEE Transactions on Geoscience and Remote Sensing* 52 (9): 5840–5854.
- Weidner, U., Förstner, W., 1995. Towards automatic building extraction from high-resolution digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing* 50 (4): 38–49.
- Weinacker, H., Koch, B., Weinacker, R., 2004. A software system for simultaneous 3dreal-Time visualization of DTM, DSM, Laser row data, Multi-spectral data, simple tree and building models. In: *Proceedings of the ISPRS working group on Laser-Scanners for Forest and Landscape Assessment, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36-8/W2, ISPRS, Freiburg, Germany, 90–95.*
- Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015a. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing* 105: 286–304.
- Weinmann, M., Jutzi, B., Mallet, C., 2014. Semantic 3d scene interpretation: A framework combining optimal neighborhood size selection with relevant features. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3, Zurich, Switzerland, 181–188.*
- Weinmann, M., Urban, S., Hinz, S., Jutzi, B., Mallet, C., 2015b. Distinctive 2d and 3d features for automated large-scale scene analysis in urban areas. *Computers & Graphics* 49: 47–57.
- Xiong, B., Jancosek, M., Oude Elberink, S., Vosselman, G., 2015. Flexible building primitives for 3d building modeling. *ISPRS Journal of Photogrammetry and Remote Sensing* 101: 275–290.
- Xiong, B., Oude Elberink, S., Vosselman, G., 2014. A graph edit dictionary for correcting errors in roof topology graphs reconstructed from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 93: 227–242.
- Xu, F., Jin, Y.-Q., 2007. Automatic Reconstruction of Building Objects From Multiaspect Meter-Resolution SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 45 (7): 2336–2353.
- Xu, L., Kong, D., Li, X., 2014. On-the-Fly Extraction of Polyhedral Buildings From Airborne LiDAR Data. *IEEE Geoscience and Remote Sensing Letters* 11 (11): 1946–1950.
- Xu, S., Oude Elberink, S., Vosselman, G., 2012. Entities and features for classification of airborne laser scanning data in urban area. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. I-4.*
- Yamaguchi, Y., Moriyama, T., Ishido, M., Yamada, H., 2005. Four-component scattering model for polarimetric SAR image decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 43 (8): 1699–1706.
- Yan, J., Shan, J., Jiang, W., 2014. A global optimization approach to roof segmentation from airborne lidar point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 94: 183–193.
- Yan, J., Zhang, K., Zhang, C., Chen, S.-C., Narasimhan, G., 2015. Automatic Construction of 3-D Building Model From Airborne LIDAR Data Through 2-D Snake Algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 53 (1): 3–14.
- Yan, L., Zhang, J., Huang, G., Zhao, Z., 2011. Building Footprints Extraction from PolSAR Image Using Multi-Features and Edge Information. In: *Proceedings of 2011 International Symposium on Image and Data Fusion (ISIDF), 1–5.*
- Yao, W., Krull, J., Krzystek, P., Heurich, M., 2014. Sensitivity Analysis of 3d Individual Tree Detection from LiDAR Point Clouds of Temperate Forests. *Forests* 5 (6): 1122–1142.
- Yao, W., Krzystek, P., Heurich, M., 2012. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sensing of Environment* 123: 368–380.
- Yu, X., Hyypä, J., Vastaranta, M., Holopainen, M., Viitala, R., 2011. Predicting individual tree at-

- tributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (1): 28–37.
- Zawawi, A. A., Shiba, M., Jemali, N. J. N., 2015. Accuracy of LiDAR-based tree height estimation and crown recognition in a subtropical evergreen broad-leaved forest in Okinawa, Japan. *Forest Systems* 24 (1): 1–11.
- Zebker, H., Rosen, P., 1994. On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. In: *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Vol. 1, 286–288.
- Zebker, H. A., Goldstein, R. M., 1986. Topographic mapping from interferometric synthetic aperture radar observations. *Journal of Geophysical Research: Solid Earth* 91 (B5): 4993–4999.
- Zhang, J., Lin, X., 2013. Filtering airborne LiDAR data by embedding smoothness-constrained segmentation in progressive TIN densification. *ISPRS Journal of Photogrammetry and Remote Sensing* 81: 44–59.
- Zhang, J., Lin, X., Ning, X., 2013a. SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas. *Remote Sensing* 5 (8): 3749–3775.
- Zhang, J., Sohn, G., Brédifb, M., 2013b. Single Tree Detection from Airborne Laser Scanning Data Using a Marked Point Process Based Method. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W1, ISPRS, Regina, Canada, 41–46.
- Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J., Zhang, C., 2003. A progressive morphological filter for removing nonground measurements from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing* 41 (4): 872–882.
- Zhang, K., Yan, J., Chen, S.-C., 2006. Automatic construction of building footprints from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing* 44 (9): 2523–2533.
- Zhao, L., Zhou, X., Kuang, G., 2013. Building detection from urban SAR image using building characteristics and contextual information. *EURASIP Journal on Advances in Signal Processing* 2013 (1): 1–16.
- Zhao, Z., Saalfeld, A., 1997. Linear-time sleeve-fitting polyline simplification algorithms. In: *Proceedings of AutoCarto*, Vol. 13, 214–223.
- Zheng, Y., Weng, Q., 2015. Model-Driven Reconstruction of 3-D Buildings Using LiDAR Data. *IEEE Geoscience and Remote Sensing Letters* 12 (7): 1541–1545.
- Zhou, Q.-Y., Neumann, U., 2008. Fast and extensible building modeling from airborne LiDAR data. In: *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, 7.
- Zhou, Q.-Y., Neumann, U., 2010. 2.5d Dual Contouring: A Robust Approach to Creating Building Models from Aerial LiDAR Point Clouds. In: *Proceedings of European Conference on Computer Vision (ECCV)*, Springer Berlin Heidelberg, 115–128.
- Zhou, Q.-Y., Neumann, U., 2013. Complete residential urban area reconstruction from dense aerial LiDAR point clouds. *Graphical Models* 75 (3): 118–125.
- Zhu, L., Lehtomäki, M., Hyyppä, J., Puttonen, E., Krooks, A., Hyyppä, H., 2015a. Automated 3d Scene Reconstruction from Open Geospatial Data Sources: Airborne Laser Scanning and a 2d Topographic Database. *Remote Sensing* 7 (6): 6710–6740.
- Zhu, X. X., 2011. Very High Resolution Tomographic SAR Inversion for Urban Infrastructure Monitoring. PhD thesis, Technical University of Munich (TUM), Munich, Germany.
- Zhu, X. X., Bamler, R., 2010a. Tomographic SAR Inversion by L1-Norm Regularization: The Compressive Sensing Approach. *IEEE Transactions on Geoscience and Remote Sensing* 48 (10): 3839–3846.
- Zhu, X. X., Bamler, R., 2010b. Very High Resolution Spaceborne SAR Tomography in Urban Environment. *IEEE Transactions on Geoscience and Remote Sensing* 48 (12): 4296–4308.
- Zhu, X. X., Bamler, R., 2011. Let's Do the Time Warp: Multicomponent Nonlinear Motion Estimation in Differential SAR Tomography. *IEEE Geoscience and Remote Sensing Letters* 8 (4): 735–739.
- Zhu, X. X., Bamler, R., 2012a. Demonstration of Super-Resolution for Tomographic SAR Imaging in Urban Environment. *IEEE Transactions on Geoscience and Remote Sensing* 50 (8): 3150–3157.
- Zhu, X. X., Bamler, R., 2012b. Super-Resolution Power and Robustness of Compressive Sensing for

- Spectral Estimation With Application to Spaceborne Tomographic SAR. *IEEE Transactions on Geoscience and Remote Sensing* 50 (1): 247–258.
- Zhu, X. X., Bamler, R., 2014. Superresolving SAR Tomography for Multidimensional Imaging of Urban Areas: Compressive sensing-based TomoSAR inversion. *IEEE Signal Processing Magazine* 31 (4): 51–58.
- Zhu, X. X., Ge, N., Shahzad, M., 2015b. Joint Sparsity in SAR Tomography for Urban Mapping. *IEEE Journal of Selected Topics in Signal Processing* 9 (8): 1498–1509.
- Zhu, X. X., Montazeri, S., Gisinger, C., Hanssen, R., Bamler, R., 2015c. Geodetic SAR Tomography. *IEEE Transactions on Geoscience and Remote Sensing* PP (99): 1–18.
- Zhu, X. X., Shahzad, M., 2014. Facade Reconstruction Using Multiview Spaceborne TomoSAR Point Clouds. *IEEE Transactions on Geoscience and Remote Sensing* 52 (6): 3541–3552.
- Zhu, X. X., Shahzad, M., Bamler, R., 2012. Naples, Italy, 106–113.
- Zhu, X. X., Wang, Y., Gernhardt, S., Bamler, R., 2013. Tomo-GENESIS: DLR's tomographic SAR processing system. In: *Proceedings of Joint Urban Remote Sensing Event (JURSE)*, Sao Paulo, Brazil, 159–162.

Acknowledgments

I am deeply indebted to many individuals for their help and consistent support throughout this work. Now looking back in time, I feel as if I would not have reached this milestone of my life without their encouragement for which I am really thankful to them.

First of all, I am extremely grateful to Prof. Xiaoxiang Zhu for being such a wonderful source of inspiration throughout these four years of work. She has been my supervisor, my friend and my mentor. She provided me the scientific freedom and boosted my confidence by encouraging me to present my work to several international conferences. I have learned quite a lot of things from her which I really appreciate.

Secondly, I would like to express my gratitude to Prof. Richard Bamler for his guidance and interest in my work. I really consider him as a father-like figure and has always been fascinated by his perfectionism. He was the one who actually brought me to the field of SAR remote sensing by hiring me for this openly advertised project. I thank him for many reasons.

I also thank Prof. Bernhard Höfle from Universität Heidelberg for showing interest in my work and agreeing to be the third reviewer. His prompt replies to my emails are highly acknowledged.

I am also grateful to the International Graduate School of Science and Engineering (IGSSE) at TUM for its financial support and Prof. Friedrich Fraundorfer for hosting my 2 month research exchange visit at TU Graz, Austria.

Besides my supervisors, many thanks to all the friends, colleagues and members of SiPEO, LMF-TUM and IMF-DLR for creating an enthusiastic and professional work place. In particular, Dr. Michael Eineder who is the head of SAR signal processing department at DLR where I worked throughout my PhD, Yuanyuan Wang for always being ready to help, Claas Grohnfeldt as inspirational sport lover and friend, Gerald Baier for his humor and in particular the German translation of the abstract of this dissertation, Michael Schmitt for always helping me with the “Diensanreisantrag” thing, Stefan Auer for his initial welcome when I joined LMF-TUM, Stefan Gernhardt especially for the time we spent together while preparing for the DLR Science SLAM in 2012, Michael Baßler for his friendly attitude towards me, and many others (the list goes on and on) ...

Lastly, I would dedicate this work to my family and friends in particular the three most beautiful females in my life: My mother for giving me birth and supporting me in every aspect of my life; my cute little princess Zainab who is my life; and my wife Saadia – a smart and gorgeous companion – for coming into my life. Her unconditional support, quiet patience, unwavering love and consistent support over the last four years was undoubtedly the most vital ingredient that enabled me to finish this work successfully. I love you for all the pain you took for me.