This is a pre-print for personal use only.
The paper will be published in Springer's Lecture Notes in Geoinformation and Cartography (LNG&C)
series in a book titled “Recent Advances in 3D Geoinformation Science - Proceedings of the 18th 3D GeoInfo Conference”.
This article was selected based on the results of a double-blind review of an extended abstract.

Artificial Intelligence for the automated creation of multi-scale digital twins of the built world – AI4TWINNING

André Borrmann, Manoj Biswanath, Alex Braun, Zhaiyu Chen, Daniel Cremers, Medhini Heeramaglore, Ludwig Hoegner, Mansour Mehranfar, Thomas H. Kolbe, Frank Petzold, Alejandro Rueda, Sergei Solonets, and Xiao Xiang Zhu

Technical University of Munich, Munich, Germany,
andre.borrmann@tum.de,
Home page: https://www.tum.de

Abstract. The AI4TWINNING project aims at the automated generation of a system of inter-related digital twins of the built environment spanning multiple resolution scales providing rich semantics and coherent geometry. To this end, an interdisciplinary group of researchers develops a multi-scale, multi-sensor, multi-method approach combining terrestrial, airborne, and spaceborne acquisition, different sensor types (visible, thermal, LiDAR, Radar) and different processing methods integrating top-down and bottom-up AI approaches. The key concept of the project lies in intelligently fusing the data from different sources by AI-based methods, thus closing information gaps and increasing completeness, accuracy and reliance of the resulting digital twins. To facilitate the process and improve the results, the project makes extensive use of informed machine learning by exploiting explicit knowledge on the design and construction of built facilities. The final goal of the project is not to create a single monolithic digital twin, but instead a system of interlinked twins across different scales, providing the opportunity to seamlessly blend city, district and building models while keeping them up-to-date and consistent. As testbed and demonstration scenario serves a urban zone around the city campus of TUM, for which large data sets from various sensors are available.

Keywords: Digital Twin, Artificial Intelligence, Point Clouds, BIM

1 Overview

Today, technologies are available that allow capturing built environment with a wide range of technologies and from a broad range of platforms like UAVs, planes, satellites, mobile mapping cars, building cranes or robots on construction sites. The resulting point clouds or 3D mesh models provide a high level of geometric detail, and are well suited for realistic high-resolution 3D visualization. However, these urban data cannot be directly used for applications requiring semantically

This is a pre-print for personal use only.
The paper will be published in Springer's Lecture Notes in Geoinformation and Cartography (LNG&C)
series in a book titled “Recent Advances in 3D Geoinformation Science - Proceedings of the 18th 3D GeoInfo Conference”.
This article was selected based on the results of a double-blind review of an extended abstract.

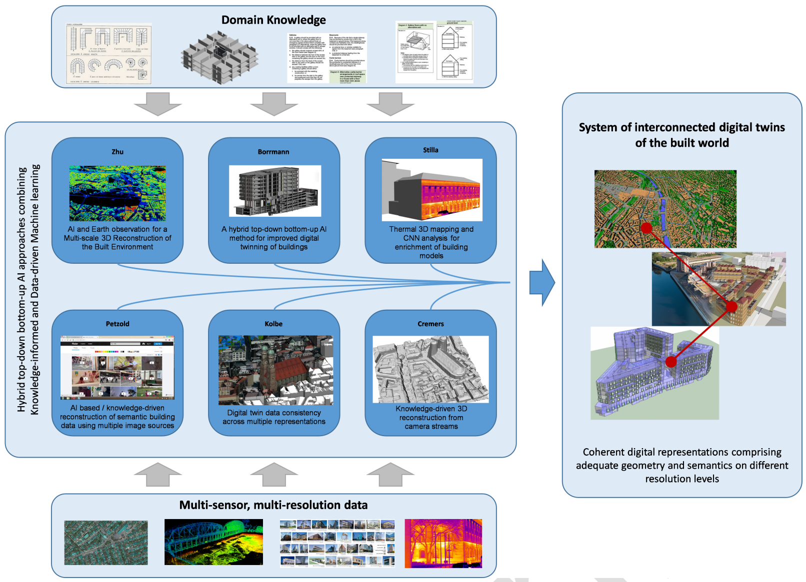


Fig. 1. Overall structure of the AI4TWINNING project including its subprojects and their interaction.

structured models, such as Building Information Modelling (BIM) and Urban Information Modelling (UIM).

The AI4TWINNING project that started in 2022 and runs for four years aims at realizing significant progress in automated creation of digital twins of the built world. The goal is to create digital replicas of buildings and infrastructure assets at varying, yet interconnected levels of detail comprising both geometry and rich semantics.

The project implements a sandwich methodology where bottom-up data-driven methods are combined with top-down knowledge-driven approaches, thus combining the strengths of both approaches in order to achieve a new level of robustness in digital twinning. For the top-down direction, human knowledge on the structure of buildings and the functional relationships of their components are digitized and made available for computational processing. For the bottom-up direction, different sensing technologies on different scales are combined with a multitude of AI-based processing methods.

While data sources range from satellite-based earth observation, aerial images, radar, to UAVs terrestrial laser-scanning and thermal cameras, the applied AI-based processing methods will include point cloud segmentation, image-processing, etc. A major challenge lies in fusing the information created through different paths and creating coherent digital twins from it. As modern machine-learning methods are based to a large extend on statistics, a major goal of the project is to represent the reliability of the different components of the resulting twin.

The envisioned digital twins will comprise different resolution levels – twins on the coarsest level will represent an entire city and model its buildings and infrastructure assets in comparatively low geometric-semantic resolution. The digital twins on the finest level, however, will model both the exterior and the interior of building in high geometric-semantic resolution, i.e. including individual building components in what is known as a building information model. As result, we aim at creating a network of interconnected digital twins at different resolution levels. Major emphasis is put on ensuring consistency across the different scales and resolutions.

For test, validation and demonstration purposes, the TUM City Campus and its surrounding districts provide a shared example and testbed across all participating groups, taking advantage of the fact that a multitude of different datasets are already available at the different groups, including optical and Radar satellite images, aerial images, StreetView images, 3D point clouds (from mobile mapping, airborne Laserscans, dense stereo reconstruction), 3D meshes, OpenDRIVE datasets, Open StreetMap data, Semantic 3D City Models, and Building Information Models.

The project is composed of six subprojects that are outlined below.

2 Subprojects

2.1 AI and Earth observation for a Multi-scale 3D Reconstruction of the Built Environment

Recently, big Earth observation data amounts in the order of tens of Petabytes from complementary data sources have become available. For example, Earth observation (EO) satellites reliably provide large-scale geo-information of worldwide cities on a routine basis from space. However, the data availability is limited in resolution and viewing geometry. On the other hand, closer-range Earth observation platforms such as aircraft offer more flexibility in capturing Earth features on demand. This combination enables the creation of a multi-scale representation of the built environment.

This sub-project develops new data science and machine learning approaches tailored to optimal exploitation of big geospatial data as provided by the abovementioned Earth observation platforms, to provide invaluable information for a better understanding of the built world. In particular, we focus on 3D reconstruction of the built environment on an individual building level. From the AI methodology perspective, while the large-scale stream puts the focus on the robustness and transferability of the machine learning models to be developed, for the very high-resolution stream we particularly research the fusion of multisensory data, as well as hybrid models combing model-based signal processing methods and data-driven machine learning ones for an improved information retrieval.

In the realm of multi-scale building reconstruction, we have developed method prototypes expressly for synthetic aperture radar (SAR) [1] and light detection and ranging (LiDAR) [2, 3]. SAR imagery, impervious to weather conditions and accessible constantly, is notably advantageous for disaster response and regions with frequent cloud cover, while LiDAR offers greater flexibility in capturing detailed building geometry. Our SAR-based methodology has been validated on four urban data sets using TerraSAR-X images in both high-resolution spotlight and stripmap modes, exhibiting substantial computational efficiency while preserving exceptional accuracy in height estimation. On the other hand, our LiDAR-based technique perceives polygonal reconstruction as a classification problem, utilizing a graph neural network to categorize polyhedra produced through space partitioning based on their occupancy within the underlying building instance. The subsequent step involves extracting the building surface from these classified polyhedra. Our preliminary results illustrating multi-scale 3D reconstruction of the built environment are depicted in Figure 2. We aim to further refine these methodologies to bolster their capability to handle challenges across diverse reconstruction scenarios. We also envisage the development of hybrid models that can more efficiently utilize the strengths of various data sources to generate large-scale, higher-resolution 3D reconstructions. This amalgamation would foster a more integrated model, equipped to navigate the complexities inherent in building reconstructions under differing conditions and scales. Ultimately, these advancements will serve to bridge the gap between extensive geospatial data and the extraction of meaningful, actionable insights from the built environment.

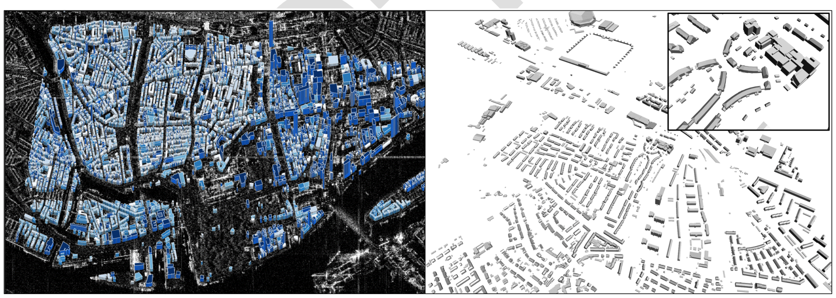


Fig. 2. Multi-scale 3D reconstruction of the built environment. Left: LoD1 building models from TerraSAR-X imagery [1]. Right: LoD2 building models from aerial LiDAR point clouds [2].

2.2 Knowledge-driven 3D Reconstruction from Camera Streams

While the reconstruction of the 3D world from moving cameras has advanced enormously over the last few years in terms of precision, density, robustness and speed, to date most power reconstruction algorithms merely generate a sparse point-cloud of the observed world. At the same time there are increasingly powerful model-based representations of the man-made world. Yet, there is a gap between the pointclouds generated by real-time capable visual SLAM algorithms and a fully model-based representation of the world as employed in architecture.

Thus, this subproject aims to develop methods that can create model-based reconstructions of the observed world from an input video. To this end, we combine highly accurate camera-based visual Simultaneous Mapping and Localization (SLAM) methods with the predictive power of deep networks. More specifically, we design and train networks to bridge the gap between purely data-driven point cloud reconstructions and model-based representations of man-made worlds – such as car models or building models. Appropriately designed and trained networks take as input images and/or point clouds and generate as output model-based representations of the observed world.

The subproject also emphasizes the importance of transferability. The goal is to develop a representation that can be effectively transferred from one scene to another, enabling the generalization of learned knowledge across different environments. The project will adopt an approach focused on learning representations of small chunks of 3D data. Instead of attempting to capture the entire scene or object in a single representation, we break down the 3D data into smaller, more manageable parts. By learning representations at a more granular level, the project can achieve transferability across different scenes and objects.

Fig. 3. Reconstruction of a building using model-based approach (on the left). Highlighted structures have close hidden representation (on the right)

2.3 A hybrid top-down bottom-up AI method for improved digital twinning of buildings

With the increasing complexity and dynamic changes in the cities, the need for up-to-date information about buildings and indoor environments for operation and management has risen substantially. Digital twins are beneficial tools for digitally representing an asset's physical and functional properties and a shared data source for building information modeling that provides a reliable basis for decision-making throughout the project life cycle.

Over the past decade, there has been a significant increase in the utiutilization of advanced data acquisition techniques in the built environment, leading to a growing trend in the creation of digital representations of buildings with coherent geometry and rich semantic information. Despite advanced development in engineering and technology, automatic or semi-automatic building digital twin creation using point cloud data and RGB images is still an open topic in the engineering and design society which needs novel approaches.

The subproject focuses on developing knowledge-driven top-down approaches for digital twin generation (Figure 4). While bottom-up approaches typically start from points and successively build up surfaces and volumes, they very often face problems when it comes to creating meaningful building objects and their relations. We therefore apply a top-down approach by fitting a highly parametrized building model to the observed data. Based on a typology of typical office and residential buildings in Germany, we develop a set of highly parametrized and modularized building models that provide sufficient degrees of freedom to allow modeling a wide range of different real-world buildings [4]. These building models will be designed in such a way that they represent human knowledge on building design and engineering.

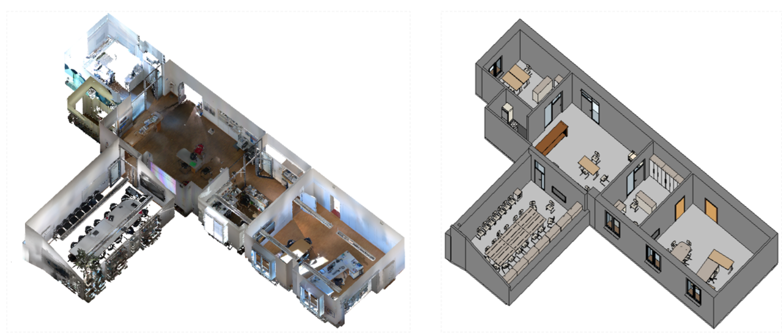


Fig. 4. Building digital twin model creation. Left: raw point cloud of the building's indoor environment. Right: a highly parameterized building digital twin model with rich semantics and coherent geometry.

The process of designing the highly parametrized building models is initially involved by extracting the main structural elements of the building (eg. ceiling, wall and floors) and partitioning the 3D individual spaces whitin indoor point cloud [5]. Major parameters (number of walls, windows, doors and columns) are also identified by a rough pattern-based interpretation of the point clouds and RGB images (bottom-up)(Figure 5). To overcome the data challenges, the domain knowlege in construction and design will be aligned with the capabilities of artifical intellegence methods in scene understanding. The actual fitting of the model to the point cloud is approached as a high-dimensional optimization problem (top-down). For its solution, AI methods based on meta-heuristics are investigated. Thanks to applying the parametric modelling process, the resulting digital twin of individual buildings provide comparatively abstract geometry, but rich semantics and consistent logic (relationships, connection points etc.) which promise significant progress in the field of "Scan-to-BIM".

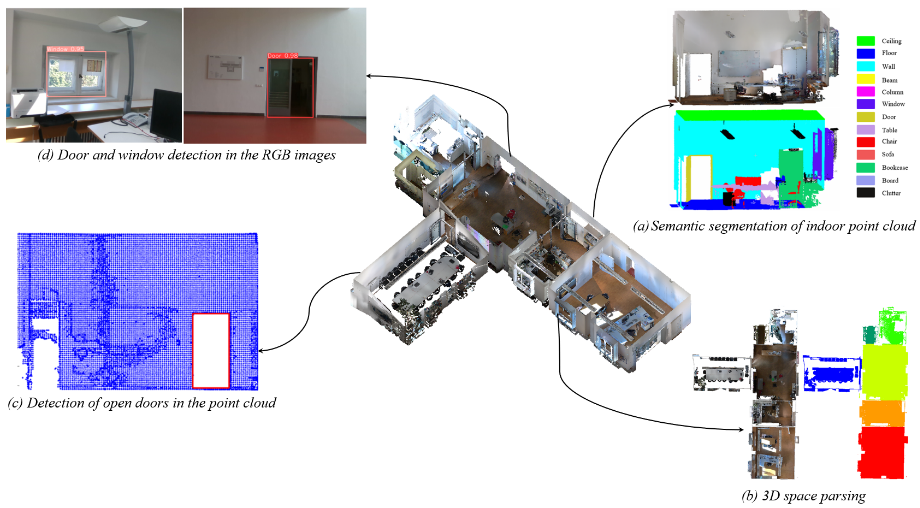


Fig. 5. Extracting semantic information for designing the highly parameterized building's digital twin model.

2.4 AI based / knowledge-driven reconstruction of semantic building data using multiple image data sources

The usage of buildings and their changes over time at the district and city levels are essential parameters for urban planning and strategic urban development [16]. Digital twins of cities are increasingly implemented in municipalities, but currently focus mainly only on geometric representations. Alphanumerical data regarding building usage are partially incorporated from various analog and digital sources of multiple authorities, often in time-consuming manual or semiautomatic procedures, and often they are not up to date. A key challenge is to keep these data up-to-date and consistent, as they are the basis for various applications.

This subproject aims to develop a data- and knowledge-driven methodology using Deep Learning methods combined with a Knowledge Base to reconstruct semantic building data.

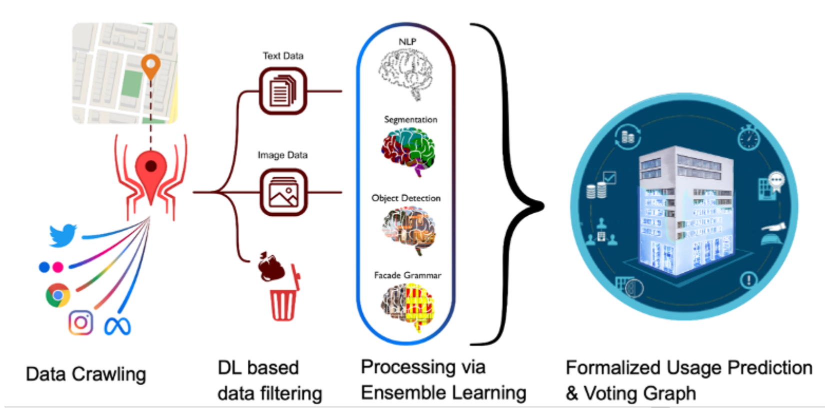


Fig. 6. The proposed pipeline. Data is obtained from openly available sources, filtered, and processed by multiple state of the art deep learning models. A formalized resulting prediction along its voting graph is delivered.

We propose a multi-stage pipeline (Figure 6) to improve the semantic enrichment process of digital twins by predicting the building usage from openly available sources such as Flickr, Twitter, and Google Street View. Since predicting building usage based on facade features is not always unique, additional sources such as comments, ratings, tweets, and hashtags to images are evaluated. It is composed of multiple state-of-the-art fine-tuned deep learning models as well as a novel grammar-based one.

The building architecture is expressed in terms of its underlying compositional logic [23, 20, 18, 17, 24, 26, 25]. This “Façade Grammar” (Figure 7) is utilized to handle hybrid use buildings by analyzing patterns in its structure. To achieve this “Façade Grammar”, context-free grammars (CFGs) are implemented to formally describe this compositional logic. To expand upon existing datasets, a data crawler is implemented to gather additional data from open internet sources. This additional information is filtered, and the façade images and accompanying textual data (i.e., comments, reviews, tweets, and hashtags) are processed using several state-of-the-art deep learning models (segmentation, classification, NLP) via Ensemble Learning. Through this approach, a formalized prediction and its accompanying voting graph is provided to the digital twins group (Subproject Kolbe). Through the use of transfer learning, this pipeline is capable of being refitted for different regional and cultural characteristics. Several existing models have been investigated and evaluated for segmentation, object detection, and classification. In our pipeline, we use and combine DeepLabv3 [21] for segmentation, YOLO [29] for object recognition, and ResNext [28] for classification.

CFGs approaches from image analysis [27] and document analysis [19] were explored and applied to a facade context in order to generate grammars. For the analysis of grammars, the framework Agglomerator [22] was used to extract a

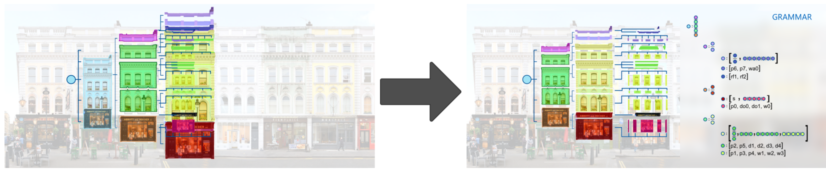


Fig. 7. "Facade Grammar" - parts and components of the facade and their relationships

representation of the part-whole hierarchy from visual features and determine the input distribution according to the conceptual-semantic hierarchical structure between façade parts and components (Figure 8). In contrast to approaches that predict usage based only on image data, hybrid building usage can be predicted by analyzing part-whole trees.

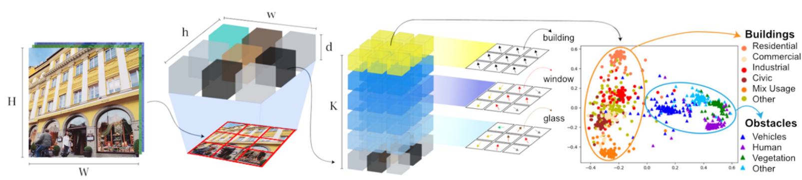


Fig. 8. Analysis of Grammars, Figure adapted from [22]

A further aspect that was investigated was the quality of the freely available valid annotated datasets for training and evaluation. For example, the ETC Dataset, 104 rectified facade images of Haussmann's renovated buildings in Paris, the CMP Facade Database with 602 images, and LabelMeFacade with 945 images were investigated. As a result, it was found that the existing datasets do not have a uniform standard, have different label classes, and are only partially rectified. Freely available valid datasets for training are insufficient, so a hybrid approach was taken using state-of-the-art models and fused datasets. The generation of annotated data using Deep Generative Models was carried out in collaboration with the subproject Prof. Zhu.

The next steps in the research project are developing methods to extract semantic information from texts (tweets, ratings, comments, hashtags) using NLP and the fusion of the results, including explanation as well as the integration of all "components" into the multi-stage pipeline and validation. The process to identify the usage of buildings can be carried out both at specific times and at time intervals to identify changes. For the setup of digital twins information about changes in use in the past will be also integrated.

2.5 Thermal 3D mapping and CNN analysis for enrichment of building models

Thermal imaging cameras record electromagnetic radiation in the infrared range (IR), which is invisible to humans. This makes it possible to determine the characteristics of surfaces or detect objects that remain hidden in the visual range. An interesting field of application is the inspection of buildings in connection with current questions on the efficient use of energy. A common way for inspecting buildings concerning the thermal insulation is to take thermographs of the outer walls by an IR camera and evaluate visually the images in the recorded image geometry. But a direct three-dimensional spatial reference is not established for the measured values.

This deficiency becomes obvious when images of a complex building structure taken from different angles are combined, fused or the measured values of buildings are to be further processed and stored in an object-related manner. Surface-related IR-textures of facades allow a spatial analysis of thermal structures, but only allow very limited statements to be made about the geometry and material of the walls. So, knowledge of the inside and outside temperatures is necessary. The project aims to thermally record both external and internal building surfaces and to spatially locate the measurements with the aid of a building model. Therefore, an extended 3D thermal description of a building model needs to be created, and then subsequent analysis of thermal structures needs to be performed.

As part of creating the 3D thermal description, we introduce a thermal mapping algorithm for generating thermal textures, which is developed and implemented on two facades of a building. In our previous work [14], we extracted the textures from thermal infrared image sequences by matching the images with 3D building models. But there were limited accuracy, blurred edges, and some projection errors because the geometry of building models is often simplified by removing small structures or overhangs. So to overcome those deficiencies, in this paper, we used a 3D thermal point cloud to project onto the facades of a building model. The thermal point cloud is prepared by [15] from mobile laser scanner point cloud extended with thermal intensities from TIR image sequences. The thermal mapping algorithm searches for nearest neighbor points in the point cloud to map the thermal intensities and generate the thermal textures. We use nearest neighbors instead of interpolation because we want to retain the original temperature values, although it results in some blank texels (where no corresponding neighbor points can be found). The method shows benefits such as occlusion detection and removal from textures for better temperature estimation of facade walls. We attempt three different approaches for searching the nearest point: minimize the angle to the normal vector of the texel plane, minimize the perpendicular distance to the normal vector, and minimize only the distance without considering the verticality. The second approach yields the finest quality textures with 0.04m mean perpendicular distance to normal vector and a reasonable computation time. Figure 9 (left) shows preliminary results of thermal textures generated for two facades. In the future, this method will be implemented on the entire university campus buildings by combining both indoor and outdoor mapping. In this way, the building can be observed as a single unit instead of looking at individual surfaces, and we can spatially locate the measurements with a 3D reference to the building and link the inside structures with the outside more efficiently. The results should be consistent considering facades constructed by different materials at different periods of time and the mapping algorithm is expected to have a high detection rate with reasonably fast processing time. Performance evaluation of the methods will be carried out by comparing them with an annotated model.

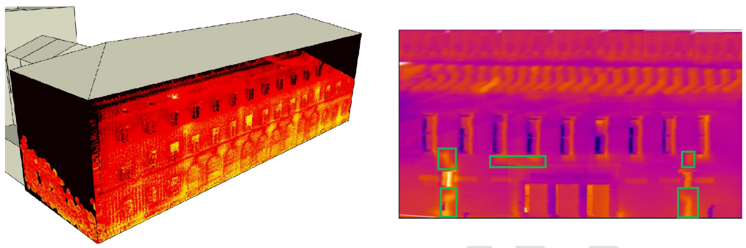


Fig. 9. Thermal mapping and analysis. Left: Thermal textures on a LoD2 building model. Right: Structures such as heating pipes, leakages, etc. are shown in green bounding boxes in a thermal texture.

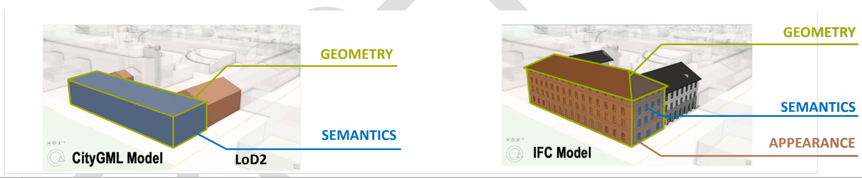
2.6 Digital twin data consistency across multiple representations

With rapid developments on the technological front and improved data acquisition from various sources, different digital representations of urban objects created for varied applications exist next to each other. Though these models might represent the same real-world object, there is no explicit linking between them as they might be of different data types, data formats, geometry types, semantic information, etc. An Urban Digital Twin will not only consist of a single type of representation but must consider these multiple representations simultaneously. To create a 'Digital Twin' of the built environment, all these varied models created for different applications with different data structures, be it 3D point clouds, mesh models, semantic city models or Building Information Models (BIM), will need to be able to interact with each other. This is done by comparing or 'matching' models against each other.Once a match between corresponding models is established, it is possible to express a link between the models and/or the model components. A match also establishes the level of coherence between the models, which can become the foundation of an Urban Digital Twin. This can facilitate information flow between the different representations.

So far, there is no formal and computable framework available that can compare urban models over a broad range of representations on an objective basis.

In the past, this has been addressed only partially by looking, e.g., for change detection between versions of urban objects as discussed in [7] but this is mainly within the same type of representation. As discussed in [8] each of the above-mentioned model types greatly differ from each other in terms of geometry, semantics, topology, appearance, etc. It is important to establish a common representation or a ‘baseline’ for the comparison and evaluation of models against each other. Before a match can be established, the common, comparable modelling aspects between the representations needs to be identified. For example, as seen in Fig.10, if Model A (a CityGML LoD2 model) is compared with Model B (an IFC model), the comparable common aspects are geometry, semantics and possibly appearance. Model A would consist of boundary representations (BRep) while Model B would consist of Constructive Solid Geometry (CSG), both of which would need to be converted to a common geometry type. The semantic structure in both models is also very different and would need to be brought into a comparable framework.

In this subproject, we develop an ontology to express the linkage between urban models across different representations. Evaluation measures are developed based on this ontology, which describes the similarity and/or differences between various models. The match is expressed as a similarity measure for each modelling aspect, viz. geometry, semantics, topology, appearance, and change across time. For the above-mentioned representation types, understanding the heterogeneity of modelling aspects is key to establishing a match between models (Fig.10).



Model A			Model B: 3D Point Clouds	Mesh	CityGML LoD1	CityGML LoD2	CityGML LoD3	CityGML LoD4	IFC
Model A		3D Point Clouds	G S* A*	G A*	G S*	G S* A*	G S* A*	G S* A*	G S* A*
Model A		Mesh	G A*	G A*	G	G A*	G A*	G A*	G A*
Model A	CityGML	LoD1	G S*	G	G S	G S	G S	G S	G S
Model A	CityGML	LoD2	G S* A*	G A*	G S	G S A*	G S A*	G S A*	G S A*
Model A	CityGML	LoD3	G S* A*	G A*	G S	G S A*	G S A*	G S A*	G S A*
Model A	CityGML	LoD4	G S* A*	G A*	G S	G S A*	G S A*	G S A*	G S A*
Model A		IFC	G S* A*	G A*	G S	G S A*	G S A*	G S A*	G S A

G-Geometry | S-Semantics | A-Appearance | *May or may not be present in the model at that Semantic Level

Fig. 10. Common comparable modelling aspects between the different types of representations.

The first step in this process is to convert all models to a common representation type so that the integrity of the modelling aspects from the original representation is preserved. For example, any 'common representation' for semantic city models and BIM would require the transfer of not just geometry but also the semantic information as well as the organisational hierarchy and logical relationships between components. Though both the models, i.e., semantic city model (based on CityGML) and BIM, have semantics, the organisational structure of the information is very different in each model. The organisation of the information also needs to be translated into a cohesive and comparable structure. On the other hand, for 3D point clouds, the common representation would need to incorporate geometry, unstructured semantics, and possible appearance data. All the varying modelling aspects (as shown in Fig. 10) need to be translated without any loss, to the common representation for comparison. Therefore the common representation needs to have these characteristics - a) unambiguous location, b) geometric primitive that is compatible with points, polygons and solids, c) ability to hold semantic information long with the organisation of information, d) ability to support appearance information, etc. Voxels are an ideal primitive that support most of the above mentioned criteria as seen in [11].

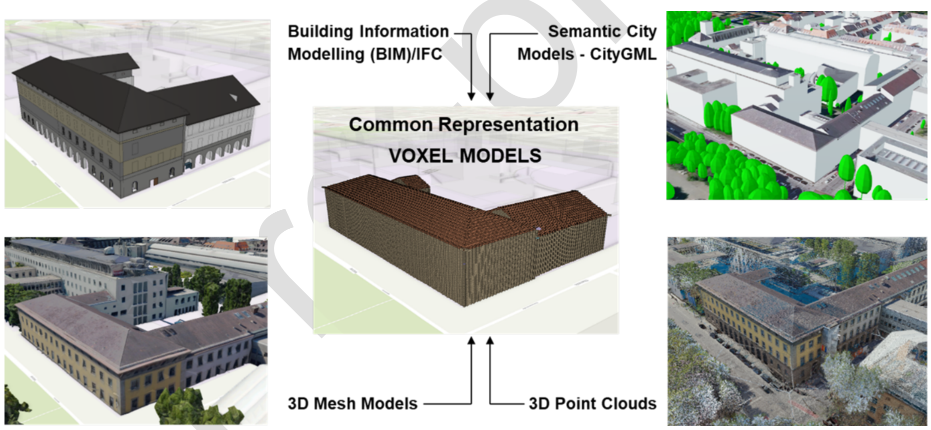


Fig. 11. RichVoxel models [6] as a common representation to match the different 3D city/building models

Voxels in the field of 3D city modelling and urban research are not a new concept. Voxels have been used in various BIM/GIS applications such as path planning as seen in [9], as volumetric modelling primitives [10], for space planning and semantic segmentation of 3D point clouds, [12], etc. Further analysis using voxels is also popular in many applications such as thermal analysis of buildings where voxelised IFC models have been used which is seen in [13] or in [12] where voxels have been used to semantically segment 3D point clouds. But, the concept of a voxel as it exists is not enough to deal with the semantic organisation of city models especially those that follow the international standards of CityGML and IFC. For this, [6] have developed the concept of a RichVoxel or a 'Semantically Enriched Voxel' which can map the information from the modelling aspects of all other representation types without loss of information.

For this joint case study in Munich, we employ the ontology to link the urban models given as separate datasets using a common representation in the form of 'Semantically Enriched Voxels' or 'RichVoxel' (Fig.11) models. Once voxelised, the models are initially compared against the level of semantic information (explained in [6]) and a common baseline is established. The models are also compared against their spatial and semantic extents followed by a component to component match than an individual voxel match. The methods developed here perform qualitative and quantitative determination of the similarity measures for concrete datasets against modelling aspect as parameters which is explained in subsequent publications.

Acknowledgments

The AI4TWINNING project is funded by the TUM Georg Nemetschek Institute for Artificial Intelligence for the Built World, which is thankfully acknowledged.

References

1. Sun, Y., Mou, L., Wang, Y., Montazeri, S. and Zhu, X.X., 2022. Large-scale building height retrieval from single SAR imagery based on bounding box regression networks. ISPRS Journal of Photogrammetry and Remote Sensing, 184, pp.79-95.
2. Chen, Z., Shi, Y., Nan, L., Xiong, Z. and Zhu, X.X., 2023. PolyGNN: Polyhedron-based Graph Neural Network for 3D Building Reconstruction from Point Clouds. arXiv preprint arXiv:2307.08636.
3. Chen, Z., Ledoux, H., Khademi, S. and Nan, L., 2022. Reconstructing compact building models from point clouds using deep implicit fields. ISPRS Journal of Photogrammetry and Remote Sensing, 194, pp.58-73.
4. Mehranfar, M., Braun, A, Borrmann, A., 2023. Automatic creation of digital building twins with rich semantics from dense RGB point clouds through semantic segmentation and model fitting. In EG-ICE 2023; 30th International Conference on Intelligent Computing in Engineering.
5. Mehranfar, M., Braun, A, Borrmann, A., 2022. A hybrid top-down, bottom-up approach for 3D space parsing using dense RGB point clouds. Proc. of European Conference on Product and Process Modeling 2022.
6. Heeramaglore, M., Kolbe, T. (2022). Semantically enriched voxels as a common representation for comparison and evaluation of 3D Building Models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. X-4/W2-2022. 89-96. 10.5194/isprs-annals-X-4-W2-2022-89-2022.
7. Nguyen, S., Yao, Z., Kolbe, T. (2017). Spatio-Semantic Comparison of Large 3D City Models in CityGML Using a Graph Database. Proceedings of the 12th International 3D GeoInfo Conference 2017 (ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences), ISPRS, 2017, 99-106

8. Stadler, A., Kolbe.T.H., (2007) Spatio-semantic coherence in the integration of 3D city models. Proceedings of the 5th International ISPRS Symposium on Spatial Data Quality ISSDQ 2007 in Enschede, The Netherlands, 13-15 June 2007 (ISPRS Archives), ISPRS
9. Wang, Q., Zuo, W., Guo, Z., Li, Q., Mei, T., Qiao, S. (2020). BIM Voxelization Method Supporting Cell-Based Creation of a Path-Planning Environment. Journal of Construction Engineering and Management, 146, 04020080.
10. Li, W., Zlatanova, S., Gorte, B. (2020). Voxel data management and analysis in PostgreSQL/PostGIS under different data layouts. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, VI-3/W1-2020, 35-42.
11. Nourian, P., Gonçalves, R., Zlatanova, S., Ohori, K. A., Vu Vo, A. (2016). Voxelization algorithms for geospatial applications: Computational methods for voxelating spatial datasets of 3D city models containing 3D surface, curve and point data models. MethodsX, 3, 69-86.
12. Poux, F., Billen, R. (2019). Voxel-based 3D Point Cloud Semantic Segmentation: Unsupervised Geometric and Relationship Featuring vs Deep Learning Methods. ISPRS International Journal of Geo-Information, 8(5).
13. Krijnen, T., El-Diraby, T., Konomi, T., Attalla, A. (2021). Thermal analysis of IFC building models using voxelized geometries. In Proceedings of the 38th International Conference of CIB W78 (pp. 437-446). (CIB W78 conference series). https://itc.scix.net/paper/w78-2021-paper-044
14. Hoegner, L. and Stilla, U., 2018. Mobile thermal mapping for matching of infrared images with 3D building models and 3D point clouds. Quantitative Infrared thermography journal, 15(2), pp.252-270.
15. Zhu, J., Xu, Y., Ye, Z., Hoegner, L. and Stilla, U., 2021. Fusion of urban 3D point clouds with thermal attributes using MLS data and TIR image sequences. Infrared Physics , Technology, 113, p.103622.
16. BBSR, 2017. Nutzungsmischung und die Bewältigung von Nutzungskonflikten in Innenstädten, Stadt- und Ortsteilzentren – Chancen und Hemmnisse . (https://www.bbsr.bund.de/BBSR/DE/veroeffentlichungen/bbsr-online/2017/bbsr-online-23-2017-dl.pdf), Accessed: 2023-26-07
17. Berger, R., Berger, E. Bauwerke betrachten, erfassen, beurteilen: Wege zum Verständnis klassicher und moderner Architektur. (Augustus,1999)
18. Bleek, J. Fassade und Ornament. (Brill Fink,2022,7)
19. Chua, F., Duffy, N. DeepCPCFG: Deep Learning and Context Free Grammars for End-to-End Information Extraction. *Document Analysis And Recognition – ICDAR 2021*. pp. 838-853 (2021)
20. Cremers, J. Die gestalterische Wirkung von Öffnungen in der Fassade. *Atlas Gebäudeöffnungen*. pp. 24-31 (2015,12)
21. Chen, L., Papandreou, G., Schroff, F. , Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. (2017,6)
22. Garau, N., Bisagno, N., Sambugaro, Z. , Conci, N. Interpretable part-whole hierarchies and conceptual-semantic relationships in neural networks. *Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition*. pp. 13689-13698 (2022)
23. Herzog, T., Krippner, R. , Lang, W. Fassaden Atlas. (DETAIL,2016,12)
24. Kemp, W. Architektur analysieren: Katalog Haus der Kunst München. (Schirmer Mosel,2006)
25. Mitchell, W. The logic of architecture: Design, computation, and cognition. (MIT Press,1998)

26. Robert, H. Wege zur Kunst: Begriffe und Methoden für den Umgang mit Architektur. (Schroedel,2009)
27. Socher, R., Lin, C., Ng, A. , Manning, C. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. *Proceedings Of The 28th International Conference On International Conference On Machine Learning*. pp. 129-136 (2011)
28. Xie, S., Girshick, R., Dollar, P., Tu, Z. , He, K. Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*. (2017,7)
29. Redmon, J., Divvala, S., Girshick, R. , Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*. (2016,6)