

TECHNISCHE UNIVERSITÄT MÜNCHEN



Lehrstuhl für Computergestützte Modellierung und Simulation

Automated BIM-based construction progress monitoring by processing and matching semantic and geometric data

Alexander Braun

Vollständiger Abdruck der von der Ingenieur fakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Uwe Stilla

Prüfer der Dissertation: 1. Prof. Dr.-Ing. André Borrmann
2. HEA Fellow Dr. Frédéric Bosché

Die Dissertation wurde am 01.07.2020 bei der Technischen Universität München eingereicht und durch die Ingenieur fakultät Bau Geo Umwelt am 22.10.2020 angenommen.

Abstract

The digitization of the construction industry has created a multitude of new optimization opportunities. Despite careful planning, construction progress can be delayed due to unforeseeable influences such as delivery difficulties, planning errors, or weather conditions. Besides, building owners demand that deadlines are met, whereby it must be checked that subcontractors meet the contractually agreed delivery and completion dates. All this is ensured by continuous monitoring of construction progress, with quality control, and, in particular, adherence to schedules playing a key role. Nowadays, this process is mostly performed manual and, therefore, very laborious, time-consuming, and error-prone, which can lead to costly planning changes.

In this context, the method of Building Information Modeling (BIM) introduces a digital, three-dimensional design model, which is supposed to contain all information of a building. This data includes, for example, detailed component information, but also construction process data. The model is intended to be used over all phases of usage, including design, planning, construction, and operation, and is already established in many of these areas. However, there are still unused potentials in the area of design and construction. The automated construction progress control is supposed to overcome these issues and allow decision-makers to have detailed and on-time insight into the progress of construction progress and make informed decisions in case of delays or disturbances.

The research approach developed in this dissertation focuses on the use of three-dimensional point clouds, which represent the as-built state of a building. For this purpose, images acquired by, for example, unmanned aerial vehicles (drones), are used to generate point clouds using photogrammetric methods. In the developed approach, these point clouds are matched against the digital building geometry in order to be able to recognize components that have been finished or are under construction. This purely geometric process is further refined by the semantic component information to recognize objects shaded during the recording or which could not be reconstructed.

In order to further increase the accuracy of the detection process, shadowing analyses are carried out based on the camera positions. Besides, the components are projected into the captured images, allowing to conduct an image-based analysis. The Computer Vision methods applied here are used with Machine Learning algorithms to perform automated object detections on an image basis to gain additional knowledge about the presence of individual components from the image data. Based on the temporal data, a final statement can be made about the current construction progress.

The developed methods were verified and further optimized based on several real-world case studies. Accuracies of over 90% could be achieved in the detection of individual components. Besides, the researched method of component projection at image level can be used to generate data for neural networks. These networks require a large amount of training data, which would otherwise have to be manually labeled by qualified personnel. With the methods developed, it has become possible to increase automation in construction progress monitoring significantly.

Zusammenfassung

Die Digitalisierung der Baubranche hat eine Vielzahl neuer Optimierungsmöglichkeiten geschaffen. Trotz sorgfältiger Planung kann sich der Baufortschritt durch unvorhersehbare Einflüsse wie Lieferschwierigkeiten, Planungsfehler oder Witterungseinflüsse verzögern. Zudem verlangen Bauherren die Einhaltung von Terminen, wobei zu prüfen ist, ob die Subunternehmer die vertraglich vereinbarten Liefer- und Fertigstellungstermine einhalten. All dies wird durch eine kontinuierliche Überwachung des Baufortschritts sichergestellt, wobei der Qualitätskontrolle und insbesondere der Termintreue eine Schlüsselrolle zukommt. Dieser Prozess wird heutzutage meist manuell durchgeführt und ist daher sehr mühsam, zeitaufwendig und fehleranfällig, was zu kostspieligen Planungsänderungen führen kann.

Die Methode des Building Information Modeling (BIM) führt in diesem Zusammenhang ein digitales, dreidimensionales Entwurfsmodell ein, das alle Informationen eines Gebäudes enthalten soll. Dazu gehören z.B. detaillierte Bauteilinformationen, aber auch Daten zum Bauprozess. Das Modell soll über alle Phasen der Nutzung, einschließlich Entwurf, Planung, Bau und Betrieb, verwendet werden und ist in vielen dieser Bereiche bereits etabliert. Allerdings gibt es noch ungenutzte Potenziale im Bereich des Entwurfs und der Konstruktion. Die automatisierte Baufortschrittskontrolle soll diese Probleme überwinden und es den Entscheidungsträgern ermöglichen, einen detaillierten und zeitnahen Einblick in den Baufortschritt zu erhalten und im Falle von Verzögerungen oder Störungen fundierte Entscheidungen zu treffen.

Der in dieser Dissertation entwickelte Forschungsansatz konzentriert sich auf die Verwendung von dreidimensionalen Punktwolken, die den Ist-Zustand eines Gebäudes darstellen. Dazu werden Bilder, die z.B. von unbemannten Luftfahrzeugen (Drohnen) aufgenommen wurden, verwendet, um mit photogrammetrischen Methoden Punktwolken zu erzeugen. In dem entwickelten Ansatz werden diese Punktwolken mit der digitalen Gebäudegeometrie abgeglichen, um fertige oder im Bau befindliche Bauteile erkennen zu können. Dieser rein geometrische Prozess wird durch die semantische Komponenteinformation weiter verfeinert, um Objekte zu erkennen, die während der Aufnahme verschattet wurden oder die nicht rekonstruiert werden konnten.

Um die Genauigkeit des Erkennungsprozesses weiter zu erhöhen, werden Verschattungsanalysen auf Basis der Kamerapositionen durchgeführt. Außerdem werden die Komponenten in die aufgenommenen Bilder projiziert, so dass eine bildbasierte Analyse durchgeführt werden kann. Die hier angewandten Computer Vision-Methoden werden mit Algorithmen des maschinellen Lernens eingesetzt, um automatisierte Objektdetektionen auf Bildbasis durchzuführen und aus den Bilddaten zusätzliche Erkenntnisse über das Vorhandensein einzelner Komponenten zu gewinnen. Auf Basis der zeitlichen Daten kann eine abschließende Aussage über den aktuellen Baufortschritt getroffen werden.

Die entwickelten Methoden wurden anhand mehrerer Fallstudien aus der Praxis verifiziert und weiter optimiert. Es konnten Genauigkeiten von über 90% bei der Detektion der einzelnen Komponenten erreicht werden. Zudem kann das erforschte Verfahren der Komponentenprojektion auf Bildebene zur Datengenerierung für neuronale Netze genutzt werden. Diese Netze erfordern eine große Menge an Trainingsdaten, die sonst von qualifiziertem Personal manuell beschriftet werden müssten. Mit den entwickelten Methoden ist es möglich geworden, die Automatisierung in der Baufortschrittskontrolle deutlich zu erhöhen.

Acknowledgements

This PhD thesis has been conducted during my ongoing BIM research at the chair of Computational Modeling and Simulation at Technical University of Munich.

I would like to express my sincere gratitude to Prof. André Borrmann for giving me the opportunity to research freely and open minded but always with valuable guidance to stay focused and on a good path.

Prof. Uwe Stilla supported this PhD not only by the collaboration throughout the joint research project but also with valuable advice as mentor and chair of the final examination. Thank you!

A big thank you also to Dr. Frédéric Bosché who invited me to visit him and his research group at the University of Edinburgh and for the valuable discussions. Thank you Fred for also taking over the task of the second examiner for my thesis!

Next, I want to thank Prof. Ernst Rank for his support even during my early studies and for his motivation to focus myself on construction informatics.

Thanks to all colleagues from our research group for the continuing exchange, fun and support. It would not have been the same without you!

Last but not least I want to thank my friends and especially my family for their support throughout this whole time. Thanks to my parents and also my parents-in-law for supporting me and my wife with any kind of possible help. Thanks to my sister Sarah for her mental support and of course my wife Regina and my kids Annika and Tobias. I love you.

Contributions

This cumulative dissertation is based on four published, peer-reviewed research papers, which are presented in chapters 4 to 7.

Paper I

Braun A, Tuttas S, Borrmann A, Stilla U (2015). *A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds*, Journal of Information Technology in Construction (ITcon) Vol. 20, DOI: <https://www.itcon.org/paper/2015/5>

Contributions:

Alex Braun and Sebastian Tuttas developed the underlying method for progress monitoring. While Sebastian Tuttas focussed on the photogrammetric aspects for the as-built acquisition, Alex Braun developed methods to interpret all geometric and semantic information of the digital model (BIM), as well as data preparation and result visualization. Uwe Stilla and André Borrmann supervised this study and reviewed the manuscript. All authors approved the final version.

Paper II

Tuttas, S., **Braun, A.**, Borrmann, A., Stilla, U., *Acquisition and Consecutive Registration of Photogrammetric Point Clouds for Construction Progress Monitoring Using a 4D BIM*, Journal of Photogrammetry, Remote Sensing and Geoinformation Science 85 (2017) DOI: 10.1007/s41064-016-0002-z

Contributions:

Sebastian Tuttas extended current research on as-built acquisition and as-planned vs. as-built detection with intensive parameter studies. Alex Braun contributed 4D scheduling and geometric information models as well as self-developed software tools for the combination of big data analysis. Uwe Stilla and André Borrmann supervised this study and reviewed the manuscript. All authors approved the final version.

Paper III

Braun, A., Tuttas, S., Stilla, U., Borrmann, A., *Improving progress monitoring by fusing point clouds, semantic data and computer vision*, Automation in Construction 116, 2020, DOI: 10.1016/j.autcon.2020.103210

Contributions:

Alex Braun developed new methods for the as-planned vs. as-built comparison of construction projects based on Computer Vision (CV) and semantic BIM data. Sebastian Tuttas provided the photogrammetric as-built data, necessary for the comparison. Uwe Stilla and André Borrmann supervised this study and reviewed the manuscript. All authors approved the final version.

Paper IV

Braun, A., Borrmann, A.: *Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning*, Automation in Construction 106, 2019, DOI: 10.1016/j.autcon.2019.102879

Contributions:

Alex Braun developed a new method for automated labeling of construction site images that can be used to train Convolutional Neural Networks (CNN). The method has been tested and validated on multiple construction sites. André Borrmann supervised this study and reviewed the manuscript. All authors approved the final version.

Further Related Scientific Contributions

Book chapters

- Braun, A.; Tuttas, S.; Stilla, U.; Borrmann, A.: BIM-Based Progress Monitoring
In: Borrmann, A.; König, M.; Koch, C.; Beetz, J. (Eds): Building Information Modeling, Springer, 2018

Peer reviewed conference papers

- Jahr, Katrin; **Braun, A.**; Borrmann, André: Formwork detection in UAV pictures of construction sites In: Proc. of the 12th European Conference on Product and Process Modelling, Copenhagen, Denmark, 2018
- Jahr, Katrin; **Braun, A.**: Maschinelles Sehen für die automatische Erkennung von Baubehelfen In: Proc. of the 30. Forum Bauinformatik, Weimar, Germany, 2018
- **Braun, A.**; Tuttas, S.; Stilla, U.; Borrmann, A.: Process- and computer vision-based detection of as-built components on construction sites In: Proc. of the 35nd ISARC 2018, Berlin, Germany, 2018
- **Braun, A.**; Vilgertshofer, S.; Daum, S.: Automatisierte Prozessneugenerierung basierend auf Ergebnissen aus der Baufortschrittskontrolle In: Proc. of the 28th Forum Bauinformatik, Hannover, Germany, 2016
- **Braun, A.**; Tuttas, S.; Stilla, U.; Borrmann, A.: Classification of detection states in construction progress monitoring In: 11th European Conference on Product and Process Modelling, Limassol, Cyprus, 2016
- Tuttas, S.; **Braun, A.**; Stilla, U.; Borrmann, A.: Evaluation of acquisition strategies for image-based construction site monitoring In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic, 2016
- **Braun, A.**; Tuttas, S.; Stilla, U.; Borrmann, A.: Incorporating knowledge on construction methods into automated progress monitoring techniques In: 23rd International Workshop of the European Group for Intelligent Computing in Engineering, Krakow, Poland, 2016
- Daum, S.; Vilgertshofer, S.; **Braun, A.**: Service-basierte Analyse von Bauwerksmodellen: der QL4BIM Server In: Proc. des 27. Forum Bauinformatik, Aachen, Germany, 2015
- **Braun, A.**; Tuttas, S.; Borrmann, A.; Stilla, U.: Automated progress monitoring based on photogrammetric point clouds and precedence relationship graphs In: The 32nd International Symposium on Automation and Robotics in Construction and Mining (ISARC 2015), Oulu, Finland, 2015
- Tuttas, S.; **Braun, A.**; Borrmann, A.; Stilla, U.: Validation of BIM components by photogrammetric point clouds for construction site monitoring In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W4, 2015, Munich, Germany, 2015

- **Braun, A.;** Tuttas, S.; Borrmann, A.; Stilla, U.: Towards automated construction progress monitoring using BIM-based point cloud processing In: eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2014, Vienna, Austria, 2014
- Daum, S.; **Braun, A.:** BIM-basierte Bauprozessplanung: Automatisierte Ableitung von Precedence Relationships durch eine räumliche Anfragesprache In: Proc. des 26. Forum Bauinformatik, Darmstadt, Germany, 2014
- Tuttas, S.; **Braun, A.;** Borrmann, A.; Stilla, U.: Comparison of photogrammetric point clouds with BIM building elements for construction progress monitoring In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3, Zurich, Switzerland, 2014
- Tuttas, S.; **Braun, A.;** Borrmann, A.; Stilla, U.: Konzept zur automatischen Baufortschrittskontrolle durch Integration eines Building Information Models und photogrammetrisch erzeugten Punktwolken In: DGPF Tagungsband 23 / 2014, Hamburg, Germany, 2014

Contents

1	Introduction	1
1.1	Building Information Modeling	2
1.2	As-Built acquisition	3
1.3	Research question and challenges	4
1.3.1	Research question.....	5
1.3.2	Challenges	5
1.4	Structure of the thesis	6
2	Overall concept for automated construction progress monitoring	7
2.1	Overview	7
2.1.1	Data sets.....	8
2.1.2	Classification	10
2.1.3	Knowledge-based refinements	11
2.2	Involved challenges.....	12
2.2.1	Geometric comparison	12
2.2.2	Occlusions.....	14
2.2.3	Object detection	15
2.3	Objectives of the thesis	15
3	State of the art.....	17
3.1	Building Information Modeling and Open Standards	18
3.1.1	Industry Foundation Classes.....	18
3.1.2	Geometry	18
3.1.3	Semantic information	20
3.1.4	Process planning and dependencies.....	20
3.2	Manual and contact-based acquisition methods.....	22
3.2.1	Digital construction site diary	23
3.2.2	Contact-based methods	23
3.3	Point cloud-based acquisition methods.....	23
3.3.1	Laser scanning	23
3.3.2	Photo- and Videogrammetry	24
3.3.3	Distance sensors and AR	28

3.4	Scaling and Registration	29
3.5	Computer Vision and Machine learning	30
3.5.1	Ray-casting	30
3.5.2	Machine learning	31
3.6	Scan-to-BIM	33
3.6.1	Geometry fitting	33
3.6.2	Segmentation and object detection on point clouds	33
3.7	Summary	34
4	A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds	35
4.1	Introduction	36
4.2	Related work	37
4.2.1	Monitoring and object verification	37
4.2.2	Process information and dependencies	38
4.3	Concept	39
4.3.1	Generation of as-built data	39
4.3.2	Process information and technological dependencies	41
4.3.3	Comparing as-built and as-planned state	43
4.4	Case Study	46
4.5	Discussion and future work	48
4.6	Acknowledgements	49
5	Acquisition and consecutive registration of photogrammetric point clouds for construction progress monitoring using a 4D BIM	50
5.1	Introduction	51
5.1.1	Motivation	51
5.1.2	As-built as-planned comparison	51
5.1.3	Consecutive registration	53
5.1.4	Structure of the paper	54
5.2	State of the art and contribution	54
5.3	Point-cloud-generation	56
5.4	Co-registration procedure	57
5.4.1	Concept	57
5.4.2	Proposed procedure	60
5.5	Experiments	62
5.5.1	Geometric aspects	62
5.5.2	Temporal aspects	63
5.5.3	Control points network	64
5.6	Results	66
5.6.1	Remarks on results	66
5.6.2	Test Site A	68
5.6.3	Test Site B	72

5.7	Discussion	75
5.8	Future Work	76
6	Improving progress monitoring by fusing point clouds, semantic data and computer vision.....	77
6.1	Introduction.....	78
6.1.1	Automated progress monitoring.....	78
6.1.2	Problem statement	78
6.1.3	Contributions	81
6.2	Related Work	82
6.2.1	Scan vs. BIM	82
6.2.2	Computer vision and deep learning.....	84
6.3	Concept	84
6.3.1	Objective	84
6.3.2	Point of departure	85
6.3.3	Concept overview	85
6.3.4	Sets of elements and detection status	87
6.3.5	Process sequencing and precedence relationships.....	88
6.3.6	Identified tasks during construction	89
6.3.7	Color detection.....	89
6.3.8	Visibility analysis and projection of elements	90
6.3.9	Image-based object detection	92
6.4	Case Study	94
6.4.1	Precedence Relationship Graph	96
6.4.2	Varying dimensions	98
6.4.3	Color detection for formwork and reinforcement.....	99
6.4.4	Visible elements	100
6.4.5	Image-based object detection	100
6.4.6	Results	103
6.5	Discussion and Outlook.....	104
6.5.1	Conclusion	104
6.5.2	Limitations	105
6.5.3	Outlook	105
7	Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning.....	106
7.1	Introduction.....	107
7.2	Related work.....	109
7.2.1	Automated construction monitoring.....	109
7.2.2	Computer Vision.....	110
7.3	Problem statement.....	111
7.4	Automated labeling of images	112
7.4.1	Camera positions	113

7.4.2	4D process data and as-planned vs. as-built comparison	115
7.4.3	Projection	115
7.4.4	Render model based on camera position.....	117
7.4.5	Generating Labels for Machine Learning	119
7.5	Case study	122
7.5.1	BIM element projection onto images	122
7.5.2	Automated labeling and validation	123
7.6	Discussion	127
8	Discussion	129
8.1	Review	129
8.1.1	Acquisition.....	130
8.1.2	Scan-vs-BIM matching	131
8.1.3	Analysis of semantic information for Precedence Relationship Graph generation and color-based detection	131
8.1.4	Computer Vision and Machine learning	132
8.1.5	Automated labeling	134
8.2	Research contributions.....	135
9	Possible Extensions and Outlook	137
9.1	Analysis and Performance Monitoring	137
9.2	Extension to indoor acquisition	138
9.3	Quality assessment and monitoring	138
9.4	Digital construction diary and documentation	138
9.5	Scan-to-BIM.....	139

List of Figures

Figure 1	BIM cycle: Sections, relevant for process monitoring are marked bold.	3
Figure 2	As-built acquisition: sample of a reconstructed point cloud.	4
Figure 3	Overview of the proposed approach.	5
Figure 4	Classification of construction elements based on the as-planned vs. as-built comparison results.	7
Figure 5	Construction sites monitored during the project for automated progress monitoring. Top row shows the models, bottom row one of the observed point clouds.	9
Figure 6	Screenshot of the developed tool for Scan-vs-BIM comparison. A specific observation is selected to visualize the detected construction elements at that time. Details on selected elements are shown in a separate viewer (bottom-right).	10
Figure 7	Classification of construction elements based on the Scan-vs-BIM comparison results. Each category represents a boolean value and every element is classified in exactly one category. The colors are used to mark all elements in the 3D view accordingly.	11
Figure 8	Segment of a point cloud in combination with its corresponding as-designed geometry. 3D view on the left, section cut in the middle and schematic visualization of corresponding thresholds on the right.	13
Figure 9	Scope of the publications presented in the following chapters.	16
Figure 10	Visualization of a triangulated abutment geometry, overlaid with a corresponding as-built point cloud.	20
Figure 11	Connected model with corresponding schedule, visualized as Gantt diagram. .	21

Figure 12 Schema of the SfM method. Feature points are identified on images from varying positions.....	25
Figure 13 Image taken from an UAV during acquisition.	26
Figure 14 Reconstruction process using VisualSfM with visualized camera positions including their viewing directions (camera cone).	26
Figure 15 Proposed methods for image acquisition that were investigated during this research. Left: crane cameras, top right: UAVs, bottom right: manual acquisition.	27
Figure 16 Registration of SfM-based point clouds using geodetic markers.....	30
Figure 17 construction progress monitoring schema	37
Figure 18 Image orientation process using control points and tie points	40
Figure 19 Sample building used to illustrate technological dependencies	42
Figure 20 The technological dependencies for the sample building depicted in Fig. 3 in a precedence relationship graph.....	43
Figure 21 a) Rasterization of a rectangular part of building element, split into two triangular model planes. Each triangle is split into raster cells of size x_r (here in blue), for each raster cell points are extracted using a bounding box of size δd ; b) Weighting function $g(d_i)$ with $d_{min} = 0.5cm$ and $\delta d = 2.5cm$	45
Figure 22 a) Snippet of the point cloud; b) Result for the as-built as-plant comparison, model planes marked blue having the state not built or unknown, model planes marked in red having the state built; The yellow frame shows the area used for evaluation.....	46
Figure 23 Confusion matrix for the evaluation of the area marked with the yellow frame in Fig. 21 b).	47
Figure 24 precedence relationship graph with corresponding BIM	48
Figure 25 3D model of the building (a) and overview image in the construction phase (b).	52
Figure 26 General procedure for construction progress monitoring using photogrammetric point clouds and BIM.	52

Figure 27	Example for results retrieved from an as-built as-planned comparison. (a) Building elements are labeled in green (before schedule), gray (in time) and red (delayed). (b) Detection result for one acquisition date, elements are labeled according to schedule and ground truth.	53
Figure 28	Schematic of the general procedure: I: Generation of a network of photogrammetric markers in the coordinate system of the construction site; II: The initial Bundle Block Adjustment (BBA) incorporates the coordinates of the photogrammetric markers as ground control points and the tie points created in the Structure-from-Motion (SfM) process; III: Consecutive registration of images based on the exterior orientations calculated in the previous step....	58
Figure 29	Types of control points	59
Figure 30	Camera configuration and example images on Test Site A with hand-held camera. The gray block represents a coarse model of the building. Images are acquired on the ground level and are supplemented by images from elevated positions on adjacent buildings and the crane.	63
Figure 31	Camera configuration and example images on Test Site B with UAV acquisition. The gray block represents a coarse model of the building. Images are acquired in nadir view from two different heights and in oblique view.	64
Figure 32	Acquisition dates at a) Test Site A and b) Test Site B.....	64
Figure 33	Surveying Network and 2D plan of Test Site B. Red dots: photogrammetric control points; black triangles: surveying points; blue lines indicate the tachymetric measurements	65
Figure 34	Results of consecutive registration for a) Test Site A and b) Test Site B. The symbols and their colors are explained in Section 5.6.1	67
Figure 35	Point cloud of three selected acquisition dates (right) and the corresponding as-planned model (left) for Test Site A.	68
Figure 36	Point cloud of three selected acquisition dates (right) and the corresponding as-planned model (left) for Test Site B.	72
Figure 37	Process modeling problems depicted by a) as-planned modeling, b) as-built modeling, c) as-built image, d) as-built point cloud on sample construction site.	79

Figure 38 Occluded construction elements in generated point cloud caused by scaffolding, formworks, existing elements and missing information during the reconstruction process	81
Figure 39 Concept for the enhancement of element detection. The highlighted process steps are introduced in this paper.	86
Figure 40 Projection of a selected 3D geometry into the 2D plane of a corresponding image.....	92
Figure 41 Occluded construction elements in generated point cloud caused by scaffolding, formworks, existing elements and missing information during the reconstruction process	93
Figure 42 Sample image of the labeling process. Displayed are the labeled formwork (blue) and column (green) elements. During this research, Labelbox (Labelbox, 2020) is used for labeling.	94
Figure 43 Screenshot of a developed tool for as-planned vs. as-built comparison. A specific observation is selected to visualize the detected construction elements at that time. Details of selected elements are shown in a separate viewer.....	95
Figure 44 Generated precedence relationship graph for Test Site A. Elements marked blue were derived from the PRG in combination with the detected elements marked in green and yellow.....	97
Figure 45 Point cloud of a finished, plain wall and formwork overlaid with the corresponding 3D geometry on Test Site B.....	98
Figure 46 Triangles detected during the time-step shown in Figure 45. a) with 1cm gaps and $\rho > 1000pts/m^2$, b) with 10cm gaps and $\rho > 1000pts/m^2$	99
Figure 47 Distribution of frequency in the HSV color space shows clear deviations between concrete and formwork elements with the Hue value represented by blue bars and Saturation value represented by orange bars.	99
Figure 48 Visibility analysis with rendered geometry of set $E_P(t)$ for several construction sites and observations. All elements are rendered in different colors to distinguish them from each other.	100
Figure 49 Precision-recall Curve for the trained network.....	102

Figure 50 Formwork and column elements detected by a trained CNN using Mask R-CNN on Test Site C.....	102
Figure 51 progressTrack: 4D BIM viewer incorporating detection states, process information and point clouds from observations	108
Figure 52 Occluded construction elements in generated point cloud caused by scaffolding, formworks, existing elements and missing information during the reconstruction process	112
Figure 53 Proposed workflow for the automated labeling toolchain	114
Figure 54 Estimated camera positions during point cloud generation (in this example using VisualSFM (Wu, 2013a))	114
Figure 55 4D building information model including all additional construction materials like scaffolding and formwork	115
Figure 56 Sample of projected, triangulated column geometry into a corresponding picture	118
Figure 57 Using projection methodology for model rendering based on the Painter's Algorithm and 4D semantic information	120
Figure 58 Considered set of elements based on previous results from as-planned vs. as-built comparison. For CV-based methods, visibility plays a crucial role, leading to reduced data sets from construction monitoring.	121
Figure 59 Detected construction elements from one observation. Green elements were successfully detected, yellow elements were not detected but are built.	123
Figure 60 Sample sub-set of auto-labeled columns in one picture from a construction site.....	124
Figure 61 Sample sub-set of auto-labeled columns and walls in another picture from a construction site.	125
Figure 62 Validation of label correctness with cyan poly-lines representing the automatically generated labels and orange poly-lines representing the manually generated validation set.....	126
Figure 63 Automatically generated PRG visualized together with the corresponding model.....	132
Figure 64 Detected construction elements on an image taken during monitoring.	133
Figure 65 Manually labeled scaffolding structures on an image from our data-set.	134

Figure 66 Automatically generated labels for all types of construction elements, present
on the case-study construction sites..... 135

List of Tables

Table 1	Photogrammetric methods suitable for construction monitoring.	28
Table 2	Currently available acquisition methods suitable for construction monitoring.	29
Table 3	Parameters and variables	45
Table 4	3D point accuracy (3D) and depth accuracy D for the dense point clouds of Test Site A.....	69
Table 5	Deviations (variation) between point cloud and as-planned model for Test Site A. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1. No reappearing planes could be reconstructed at the first two acquisitions.	70
Table 6	Deviations (degrees) between point cloud and as-planned model for Test Site A. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1. No reappearing planes could be reconstructed at the first two acquisitions.	71
Table 7	The standard deviations and number of rays for the control points calculated from images acquired on 13 February on Test Site A and the deviations from the geodetic measurement.	71
Table 8	3D point accuracy (3D) and depth accuracy D for the dense point clouds of Test Site B.....	73
Table 9	Deviations [cm] between point cloud and as-planned model for Test Site B. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1.	73
Table 10	Deviations [deg] between point cloud and as-planned model for Test Site B. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1.	74

Table 11	The standard deviations and number of rays for the control points calculated from images acquired on 26 February on Test Site B and the deviations from the geodetic measurement.	74
Table 12	Test sites monitored during this case study.....	95
Table 13	Enhancing results by applying the introduced PRG for Case Study site A.....	97
Table 14	Visible elements based on the introduced algorithm for Test Site A.	101
Table 15	Detected formwork elements during the observations for Test Site A.....	103
Table 16	Resulting element sets for Test Site A.....	103
Table 17	Enhanced results for the detection with the newly introduced methods for Test Site A.....	104
Table 18	Sample labels of two categories for different ML use cases	122

Abbreviations

Numbers

2D - two dimensional.....	18, 24, 133
3D - three dimensional.....	18, 24, 133, 135
4D - 3D and time.....	2, 8, 18, 129, 131

A

AEC - Architecture, Engineering and Construction.....	19
AI - Artificial Intelligence.....	31
AR - Augmented reality.....	28, 30

B

BIM - Building Information Model.....	2, 3, 5, 8, 11, 12, 15, 18, 19, 21, 22, 28, 29, 33, 36, 106–108, 112, 131, 133, 135, 138
---------------------------------------	---

C

CAD - Computer Aided Design.....	18
CNN - Convolutional Neural Network.....	6, 31, 32, 34, 108, 134
CSG - Constructive Solid Geometry.....	33
CV - Computer Vision.....	6, 8, 11, 15, 30, 34, 130, 132, 133, 136, 138, 139

D

DIKW - Data, Information, Knowledge, Wisdom.....	17
--	----

G

GPS - Global Positioning System.....	27, 29, 130
GPU - Graphics Processing Unit.....	30
GUID - Globaly Unique ID.....	21

H

HSV - Hue, saturation, value.....	132
HVAC - Heating, Ventilation, Air Conditioning.....	107

I

ICP - Iterative Closest Point Algorithm.....	29, 30, 37, 131
IFC - Industry Foundation Classes.....	8, 18–21
L	
LCA - Life Cycle Analysis.....	18
LOD - Level of Detail.....	18, 19
M	
MEP - Mechanical, electrical, and plumbing.....	33
ML - Machine Learning.....	6, 11, 15, 31, 34, 134, 136, 139
MVD - Model View Definition.....	19
N	
NDA - Non-disclosure agreement.....	8
Q	
QTO - Quantity take-off.....	137
R	
RANSAC - Random Sample Consensus.....	24
RFID - radio frequency identification.....	23
S	
SfM - Structure from Motion.....	8, 11, 23, 24, 31, 33, 34, 38, 110, 130, 132, 133, 135, 136, 138
SIFT - Scale-invariant feature transform.....	24, 25
U	
UAV - Unmanned Aerial Vehicle.....	8, 17, 24, 25, 27, 28, 111, 130, 131, 135, 138
UWB - ultra-wide band.....	23
W	
WPF - Windows Presentation Foundation.....	107

1. Introduction

The continuing development of information technologies opens up new opportunities for all industry sectors, including the construction industry. Business processes can be streamlined and accelerated and thus made more efficient. The construction industry benefits from these advantages through the implementation of Building Information Modeling (BIM). This digital, model-based method is promoted to make planning more accurate, flexible, and transparent, making construction processes on schedule, faster, and more cost-effective. Errors and conflicts are easier to detect and hence, can be avoided.

In particular, progress monitoring on construction sites is very complex and currently still primarily carried out manually. On large construction sites, many processes take place in parallel. A precisely defined construction sequence is difficult to determine or predict. Construction scheduling is an important task that incorporates many dependencies between different tasks but also relies on the experience of planners that estimate duration based on quantities and available resources. These issues create imponderables in the field of construction logistics and the overall planning of a building. External dependencies like delivery times, availability of workforce, or weather conditions add up to the uncertainties.

These issues make process monitoring an essential part of the organization of a construction site. It is crucial to know the current state of construction of a project to make informed decisions on schedule changes if any issues occur. This is especially so for general planners and main contractors with various subcontractors. General requirements are the creation of performance records for the client and making recourse claims against the subcontractors in case of delays.

Although tasks on most construction sites are already quite structured, it is worth taking a closer look at the construction processes. Currently, progress is usually monitored manually or via mobile devices (smartphones, tablets) directly on the construction site (Saidi et al., 2003). In any case, the progress is manually recorded in construction diaries by the site manager. In addition to the high personnel costs, this also entails various risks concerning inconsistencies and timeliness. An automated procedure would be recommended here.

Construction progress monitoring is particularly important on large construction sites, where storage capacities are rare. Many construction elements and supplementary building materials are delivered "just in time". Besides, prefabrication becomes more prevalent, resulting in new work routines where elements are installed directly and without storage in order to save time and costs. If delays occur on a construction site, it is essential to recognize them as quickly as possible to be able to take countermeasures and, if necessary, postpone deliveries. This problem also applies to small construction sites in inner-city areas with even less available storage areas.

On the one hand, delays are crucial regarding personnel management and construction machinery, as these domains are accountable for significant parts of construction costs. On the other hand, process monitoring is associated with high costs. Due to the complexity of construction projects in general, and a large number of individual components in particular, progress monitoring is challenging to implement and ties up many resources. Therefore, this task benefits from automation in order to be able to react quickly and precisely to changes.

1.1. Building Information Modeling

Automation of construction processes requires a structured and high quality, digital representation of the design model. Building Information Model (BIM) is a methodology based on the use of digital building models over the entire life cycle. Progress monitoring is one of the many use cases and is part of the "execution" phase (Borrmann et al., 2018). A BIM model is intended to describe a building over its entire life cycle (see Figure 1).

This includes planning, design, construction, operation, maintenance, and, where appropriate, decommissioning, or conversion (Eastman, 1999). The model is intended to hold all relevant information for all project participants. In addition to the 3D geometry itself, material information (e.g., steel or concrete) as well as quantities, costs, and process information are also stored. The model, in combination with its corresponding schedule, is usually referred to as 3D and time (4D) model, while the fourth dimension is the time.

Construction companies benefit from these model-based processes since the detailed planning upfront helps to identify planning errors in advance and solve clashes or inconsistencies. Facility Management (FM) also benefits from an exact BIM model during operation, as all data on technical equipment in a building is available for any maintenance requirement.

Recent developments in Germany require all public infrastructure projects to use BIM, starting in 2020 (Borrmann et al., 2018). It is expected that these requirements will subsequently be adopted for any other type of construction. In addition to the apparent advantages in the design phase, this will help ensure that digital models are available for all construction projects in the (nearer) future.

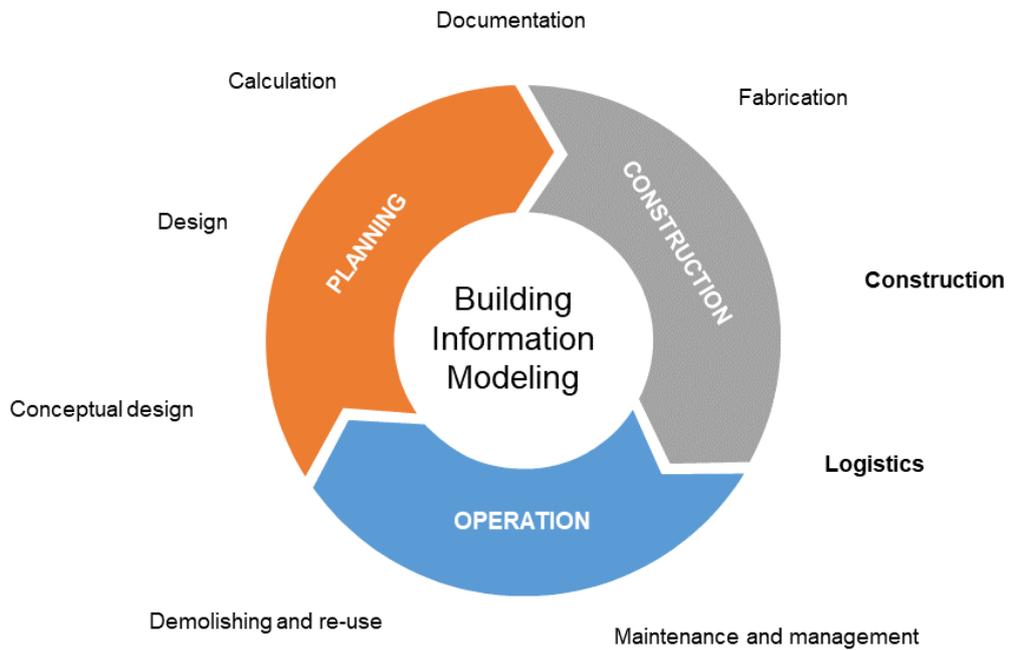


Figure 1 BIM cycle: Sections, relevant for process monitoring are marked bold.

As already mentioned, construction progress monitoring is currently still carried out manually. The phases relevant to progress monitoring are especially the construction phase, including logistics. These are shown in Fig. 1 in the bold printed sections.

Automated progress monitoring is intended to provide decision-makers with information to detect deviations earlier and form better decisions. This process includes the acquisition of the current construction status (as-built), the comparison with the as-planned status, and the detection of deviations in the process schedule (and possibly in the geometry). A BIM model provides the required basis for this since the as-planned state in terms of the 4D model, and geometric representation can be queried here, providing the necessary data to be able to react to any temporal or geometric deviations as resource-saving as possible.

1.2. As-Built acquisition

One significant task in progress monitoring is the actual acquisition of the current status on site. On large construction sites with many elements and parallel running tasks, the site manager can quickly lose track of the current status. Hence, it requires trained personnel with a complete overview of all ongoing construction tasks. In order to automate this process, methods are needed to gather information on each construction element's status and make

it available for a computer-aided management process.

Currently, many different approaches are tested in this regard and are presented in Chapter 3. An often-used approach is the usage of point clouds. A point cloud consists of a set of points in three-dimensional space. Each point consists of at least the three coordinates in x-, y- and z-direction, describing its position. Besides, color values for each point and information about the position of the recording source (normal vector) are usually stored. A sample that has been acquired during one of the case studies is depicted in Fig. 2.

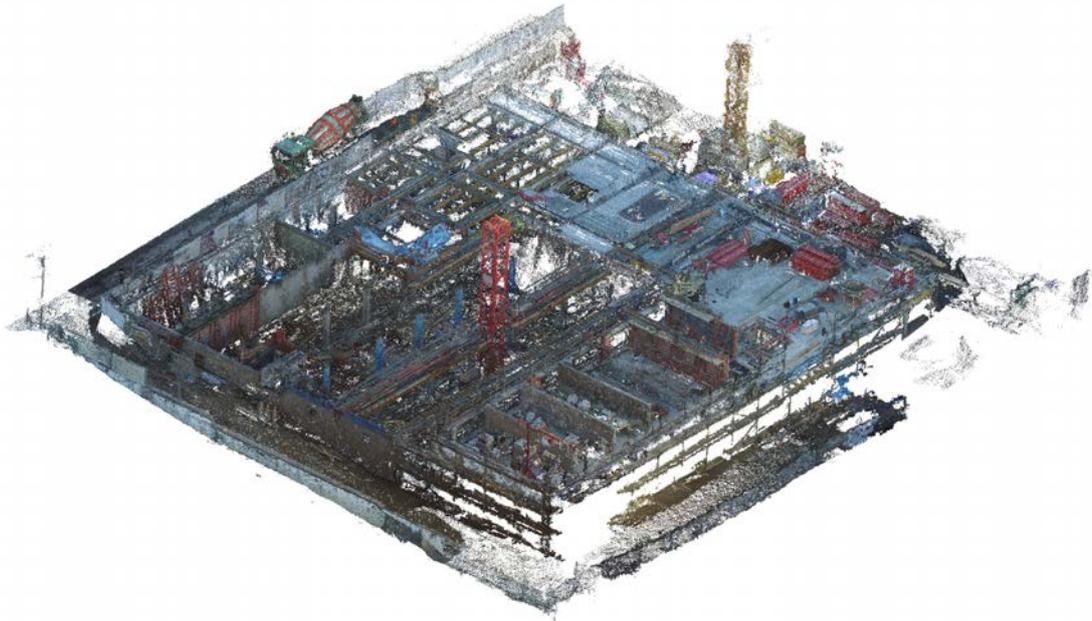


Figure 2 As-built acquisition: sample of a reconstructed point cloud.

State-of-the-art methods to acquire these point clouds are either by laser scanners or photogrammetric methods. Detailed information on these methods is presented in Chapter 3. The described point cloud represents the "as-built" state of the building.

1.3. Research question and challenges

The overall research topic is the automation of construction progress monitoring. The underlying project was funded by the German Research Foundation (DFG) and has been conducted in close collaboration with the Chair of Remote Sensing and Photogrammetry (RSP) under the lead of Prof. Uwe Stilla. The research at RSP has been carried out by Sebastian Tuttas and focused on the photogrammetric acquisition as well as point cloud processing. The results of this part of the project have been presented in Tuttas' PhD thesis (Tuttas, 2017).

1.3.1. Research question

Based on the introduced concept, the following research questions arise:

- How to match BIM model and point clouds?
- How to deal with inaccuracies?
- What are the possibilities to infer information about not directly recognized components?
- How can image information be used to recognize components or component states?

The focus of this thesis lies in the development of a methodology for the matching of photogrammetric point clouds and the BIM models to enable a Scan-vs-BIM comparison of the construction process. In detail, the collected data has to be evaluated, processed, and finally matched efficiently and reliably against an as-designed model. The comparison includes a geometric part and semantic checks that require further information such as color, structural logic, and knowledge on construction methods. Subsequently, the data gathered in images can be analyzed to gain further information on the current construction status.

Finally, the knowledge gained is used to update the schedule and, if necessary, to incorporate new knowledge about the duration of individual sub-processes, thus updating the schedule. This approach is also depicted in Fig. 3.

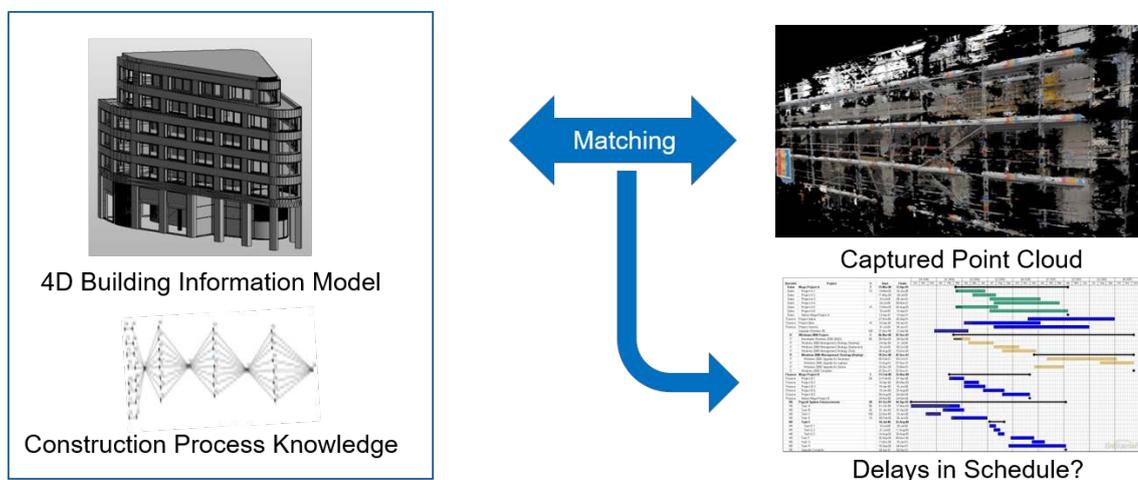


Figure 3 Overview of the proposed approach.

1.3.2. Challenges

The methods of photogrammetry pose particular challenges in recording the current state of construction. After the acquisition with cameras, only the respective field of view can be analyzed. Due to the shading of components or due to other components or construction aids, the relevant elements might be occluded. This problem can partly be solved by taking pictures from different perspectives. However, components inside the building cannot be captured from the outside.

A further challenge when recording the current construction state is the inaccuracy of the point cloud. This is due, for example, to the great distance between the camera and the object to be acquired. After the point clouds have been generated from several composite images, additional inaccuracies can result from too small local distances between the recording points.

In addition to the possible inaccuracies in the point cloud, the alignment with the geometry of the as-designed model is also a reason to introduce further errors. Optimized algorithms and approaches are necessary, especially concerning the calculation time for the alignment of a point cloud with several million points and geometry with several thousand components.

1.4. Structure of the thesis

This cumulative thesis divides into several parts. The following chapter is starting with a general introduction to the developed concept to tackle the identified challenges, followed by the current state of the art in all related research areas. In the following, four journal publications are presented that further detail the individual steps. They are sorted chronologically in terms of the introduced tasks that are required on construction sites.

A general introduction to the proposed approach is presented in Chapter 4. It also focuses on all related problems that occur during monitoring as well as the as-planned vs. as-built comparison.

Chapter 5 details acquisition methods suitable for such mutable environments such as construction sites and presents an in-depth analysis of their advantages and challenges.

In the following Chapter 6, all methods are introduced that help to enhance as-built vs. as-planned comparisons with the help of semantic data from digital construction models and Computer Vision methods is available from the acquisition process itself.

In this scope, a new method is presented in Chapter 7 that uses the gained knowledge from the Computer Vision (CV) process for the labeling of construction elements on images. The gained data set can be used for training a Convolutional Neural Network (CNN) in the scope of Machine Learning (ML).

All introduced topics are evaluated in Chapter 8 that ends with an outlook in Chapter 9 on future research questions in the field of construction progress monitoring.

2. Overall concept for automated construction progress monitoring

The concept for automating construction progress monitoring developed in the course of this thesis is based on matching a photogrammetric point cloud captured from the site in regular intervals against the as-designed 4D BIM that includes the as-planned construction schedule. With this approach, deviations between the as-planned and the as-performed schedule can be detected automatically.

2.1. Overview

Based on the findings from the research community (presented in Chapter 3) and the new findings introduced in the subsequent chapters, the following process is deemed feasible (Figure 4). It is detailed in Chapter 4 and briefly introduced here.

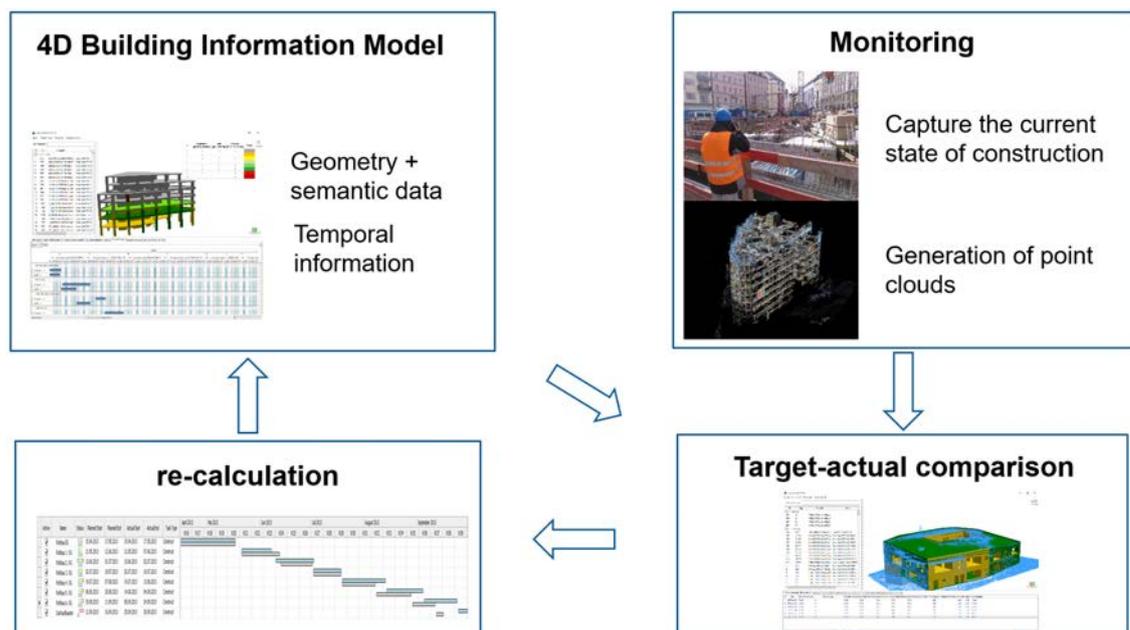


Figure 4 Classification of construction elements based on the as-planned vs. as-built comparison results.

On the one hand, the BIM model is provided as Industry Foundation Classes (IFC), including 4D process data. The acquisition of the as-built state, on the other hand, is achieved with cameras, incorporating Structure from Motion (SfM) methods for the point cloud reconstruction. Chapter 5 investigates different ways of monitoring, including Unmanned Aerial Vehicle (UAV)-based methods. The comparison is based on geometric distance and density checks, as introduced in Section 2.2.1 and refined by additional methods, like CV. Details on these methods are extensively discussed in Chapter 6.

Subsequently, the resulting sets of detected elements can then be compared to the expected elements per schedule.

2.1.1. Data sets

The conducted research relies on available data sets to test and validate the proposed methods. However, there are no data sets publicly available, as many construction companies fear the inadvertent publication of confidential company information. During this research, seven construction sites were monitored with different observation methods to gather research data and validate the introduced concepts (see Fig. 5). The construction sites are all located in Germany and cover a variety of structural engineering buildings as well as infrastructure (one bridge, one wastewater treatment plant). Details on any of these construction sites cannot be made available, as Non-disclosure agreements (NDAs) were signed for all of them.

The gathered data has been used to verify all newly introduced methods and concepts presented in the following chapters. Additionally, a software framework has been developed, that is depicted in Figure 6. To visualize the comparison results and the detected elements, and to verify the algorithms used, all gathered data is stored in a comprehensive database accessible via this software. The tool displays all geometric and semantic building element information as well as scheduling data that has been parsed from IFC data models (see Section 3.1.1). The detected elements are highlighted for easy identification.

Figure 6 depicts the software interface with the example of one of the construction site case studies used in this research. The building mainly consists of in-situ concrete elements that were cast using formwork on site. In the figure, one specific observation is selected, and all detected elements are highlighted. Green coloring represents elements that have been built and are correctly detected and confirmed through the point cloud. All yellow elements are built but were not confirmed through the point cloud.

The overall concept incorporates several data sources: The model of the building, including geometry and component information, is required. Additionally, the point clouds with the respective timestamp of the acquisition are needed. Another data source is the construction schedule, which is linked to the respective components.

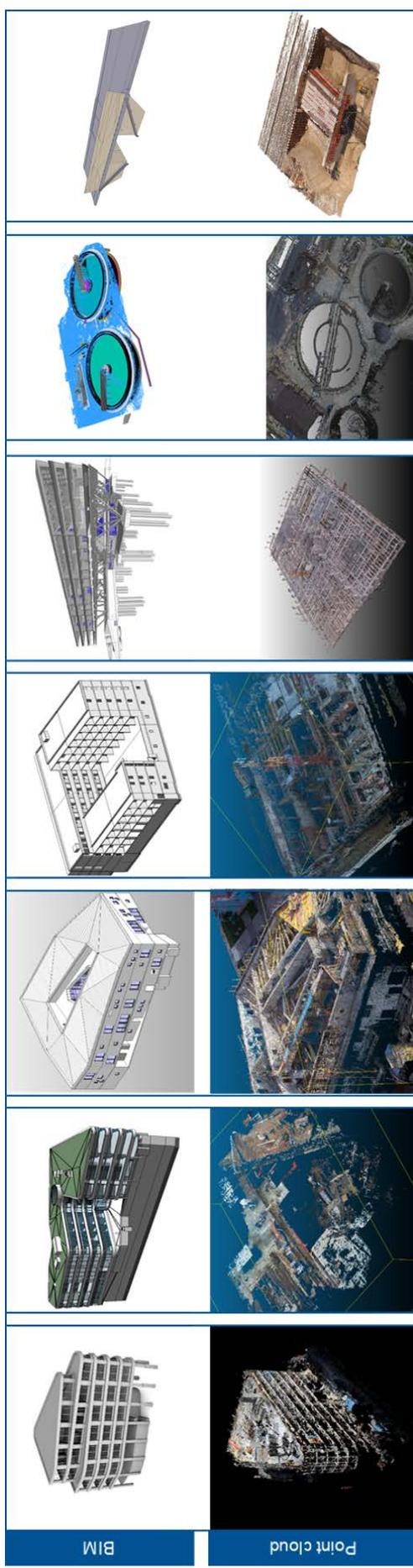


Figure 5 Construction sites monitored during the project for automated progress monitoring. Top row shows the models, bottom row one of the observed point clouds.

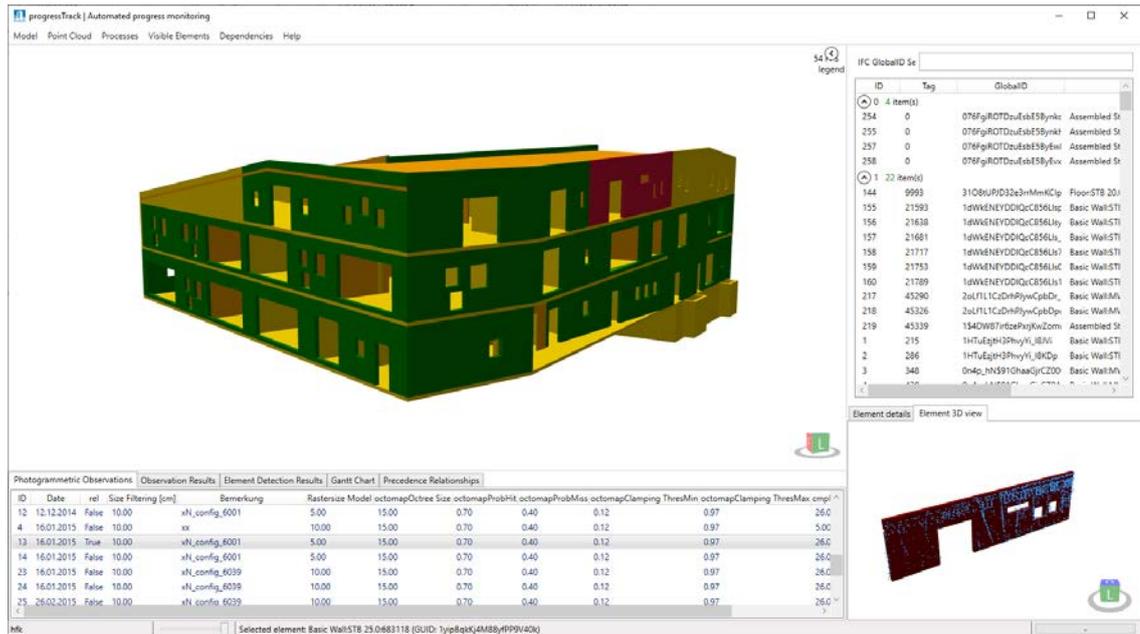


Figure 6 Screenshot of the developed tool for Scan-vs-BIM comparison. A specific observation is selected to visualize the detected construction elements at that time. Details on selected elements are shown in a separate viewer (bottom-right).

2.1.2. Classification

During the as-planned vs. as-built comparison, different detection states occur. As introduced, a classification scheme is required to visualize all possible detection states precisely. In this case, the temporal factor needs to be addressed. The generated point cloud for the Scan-vs-BIM comparison represents the actual situation on-site at a specific time of t . However, this point cloud might not be perfect and have holes or low densities in several spots. These errors are introduced by occlusions or reconstruction errors, and will be discussed in Chapter 3.

For this reason, the detection algorithms might not identify all elements present. In order to correctly handle those elements, each element is considered and categorized independently. This is done for every observation, resulting in a classification at each observation step t .

As shown in Figure 7, each building element is assigned a boolean value for each of the three states:

- **As-planned:** Indicates whether the element should have been built at the considered point in time, according to the as-planned schedule. The main idea behind as-built vs. as-planned comparisons is to detect any deviations on the site compared to this state.
- **As-built:** Indicates whether the element is present or not. This state represents the ground truth and is only available in test scenarios. It is not available for Scan-vs-BIM comparison under real conditions. Corresponding data is gathered manually as the basis for the scientific experiments conducted in this thesis.

#	as-planned [process planned]	built [ground truth]	detected [PC to Geometry]	Color
1	-	-	-	Grey
2	x	-	-	Orange
3	x	x	-	Yellow
4	-	x	-	Light Green
5	-	x	x	Green
6	x	x	x	Dark Green
7	x	-	x	Red
8	-	-	x	Dark Red

Figure 7 Classification of construction elements based on the Scan-vs-BIM comparison results. Each category represents a boolean value and every element is classified in exactly one category. The colors are used to mark all elements in the 3D view accordingly.

- **Detected:** Indicates whether the detection algorithms have detected the element. This state should equal the "as-built" state (ground truth).

Derived from these three states with every two options (true/false), the shown 3 x 8 matrix (2^3) shows all possible combinations.

2.1.3. Knowledge-based refinements

The introduced Scan-vs-BIM approach is made as described in the following Section 2.2.1. However, this comparison does not always provide correct results due to inaccuracies and occlusions in the point cloud.

Accordingly, the concept introduces several new enhancements, like the analysis of the underlying BIM model. In detail, the structure of the model and its technological dependencies are derived. This requires detailed knowledge of the geometric structure of the BIM model. Hence, an approach is presented in Chapter 4, that relies on a query language (Daum et al., 2014) and is also detailed in the state of the art (Section 3.1.3).

Besides, the acquired images provide a high density of additional data that can be analyzed. In this regard, CV- as well as ML methods proved helpful and provided valuable additional input. Especially for the detection of shadings but also for object detection, valuable options are available. In combination with SfM methods, visibility analysis, and object detection, further enhancements the detection results for construction objects can be achieved. These are introduced in Section 3.5 and presented in detail in the Chapters 6 and 7.

2.2. Involved challenges

There are several reasons why some elements may not be detected that have been actually built. The most prominent reason is the occlusions that occur on-site. During construction, large amounts of temporary structures like scaffolds, construction tools, and construction machinery obstruct the view on the element surfaces. Limited acquisition positions further reduce the visible surfaces and hence, the overall quality of the generated point clouds. Additionally, elements inside of the building are also occluded by other building elements for acquisitions outside of the building.

2.2.1. Geometric comparison

After the acquisition, scaling, and registration, an aligned as-built point cloud of an observation time t and the as-designed model with interconnected schedule data is available. The initial step of a Scan-vs-BIM comparison is the plain geometry comparison of the as-built point cloud vs. the as-designed geometry from the BIM. This approach is also known as "Scan vs. BIM". During the construction phase, the actual as-built process can deviate from the original as-designed process.

In the context of this thesis, barycentric coordinates (Coxeter, 1969) are used for this geometric comparison. For each point p in a point cloud, the distance to any triangulated surface (defined by three points $t1, t2, t3$) can be computed as follows:

with

$$q = t1; \vec{u} = t2 - t1; \vec{v} = t3 - t1; \quad (2.1)$$

$$\vec{n} = \vec{u} \times \vec{v} \quad (2.2)$$

$$\vec{n}_0 = \frac{\vec{n}}{\sqrt{\vec{n} \cdot \vec{n}}} \quad (2.3)$$

$$r = \vec{n}_0 \times \vec{q} \quad (2.4)$$

$$dist = | (\vec{n}_0 \cdot p) - r | \quad (2.5)$$

Finally, Equation 2.5 provides the distance of any point to a triangle. This distance is valid for all equations that fulfill these requirements of barycentric coordinates:

$$\vec{w} = p - q \quad (2.6)$$

$$\gamma = \frac{\vec{u} \times \vec{w} \cdot \vec{n}}{n \cdot n} \quad (2.7)$$

$$\beta = \frac{\vec{w} \times \vec{v} \cdot \vec{n}}{n \cdot n} \quad (2.8)$$

$$\alpha = 1.0 - \beta - \gamma; \quad (2.9)$$

The distance holds true for all α , β and γ within:

$$0 \leq \gamma \leq 1; 0 \leq \beta \leq 1; 0 \leq \alpha \leq 1 \quad (2.10)$$

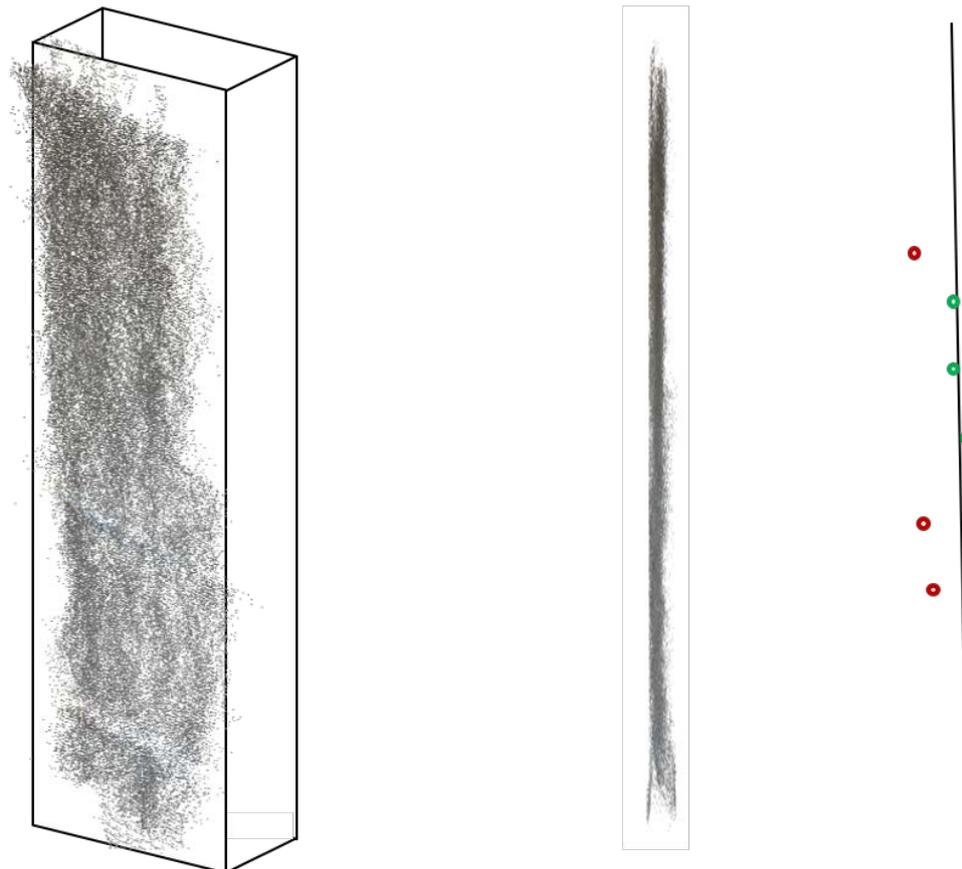


Figure 8 Segment of a point cloud in combination with its corresponding as-designed geometry. 3D view on the left, section cut in the middle and schematic visualization of corresponding thresholds on the right.

Due to reconstruction errors and measurement inaccuracies, point clouds introduce noise. This is shown as exemplary in Figure 8. The left side of the Figure shows a 3D view of the

column, while the right side shows a cross-section to visualize the noise of the point cloud better. Due to these inaccuracies, thresholds need to be introduced to still receive satisfactory results from the comparison. These project-specific thresholds depend on many factors, such as acquisition method, image resolution, among others.

In summary, the geometric comparison provides the purely geometric results for the Scan-vs-BIM approach, while several issues remain. This approach is computationally inefficient if each point is compared against each triangle of all objects of the as-designed geometries. Performance optimization can be achieved by geometric filtering to reduce computation times, as described in Section 3.5.1.

The proposed solution for these challenges is presented in the Chapters 4 and 6.

2.2.2. Occlusions

Another reason for weak detection rates is building elements that are currently under construction. As those elements count towards the overall progress, they must not be missed and play a crucial role in defining the exact state in the current process. Challenging are, in general, all construction methods, whose geometry under construction differs largely from the final element geometry or that needs temporary construction objects. This applies, e.g., for reinforced concrete or multi-layered walls.

On the one hand, formwork, which is used for concrete pouring, may obstruct the view of the element, making it impossible to be detected. On the other hand, the plane surface of formwork for a slab might be detected as the surface of the slab itself and thus would lead to a false positive. Due to these challenges, further enhancements to the comparison and detection algorithms are needed.

Since the digital model contains information on construction methods, the authors propose to use this knowledge for the overall detection process. By deducing the precedence relationships with a query language, assumptions on occluded elements can be made. Construction methods and derivation of expected elements lead to new Scan-vs-BIM comparison capabilities like extended thresholds or computer vision methods to detect objects like formwork on the raw observation images, taken for the point cloud generation.

In previous research projects (Turkan et al., 2013; Golparvar-fard et al., 2009; Bosché and Haas, 2008b), a direct Scan-vs-BIM comparison of point cloud with 3D geometry has been proposed. In order to further increase the accuracy, it makes sense to use the additional information available from the Building Information Model. One aspect is the color of the original materials to distinguish them from construction site equipment and, e.g., formwork.

The available process information is particularly important. It provides additional information about objects that should be available at an observation time t . Besides, this information

shows which components are potentially "under construction". Thus, these components can be tested with increased tolerance against the point cloud to provide a detailed overview of the current state of construction.

A further basis is the "Precedence Relationship Graph", which offers additional possibilities to identify occluded objects despite missing information from the point clouds. In detail, precedence relationships are derived from the structure of the BIM model by a query language, as introduced in Section 3.1.3. This semantic knowledge provides information on elements that are occluded but are required to be built based on precedence requirements.

A concept to solve the mentioned issues is presented in the Chapters 4 and 6.

2.2.3. Object detection

Current methods do not provide sufficiently reliable results for object detection based on point clouds. Due to this reason, a new approach is presented that uses the gained knowledge on camera positions to achieve better results on an image-level. This approach requires several methods, starting with CV as well as ML as introduced in the Sections 3.5. In detail, the known positions of construction elements are combined with the estimated positions of the cameras. This knowledge allows the projection of element bounding boxes into the images at the expected positions. Training of ML networks can benefit from this approach and is presented in detail in Chapter 7.

2.3. Objectives of the thesis

The following accepted and peer-reviewed journal publications propose solutions to the challenges identified in the Section above. As summarized in the Figure 9, each publication focuses on a different objective.

While Chapter 4 presents the overall approach in a detailed manner and highlights the prevailing problems, Chapter 5 presents the current status of as-built acquisition. In addition, methods for Scan-vs-BIM approaches are compared and tested.

The following Chapter 6 presents new approaches to solve many of the stated problems and provides a detailed overview of the key innovations of this thesis. Finally, Chapter 7 presents a new approach to facilitate gisML methods by automating labeling, which would not be possible without the previous work presented in these chapters.

	Concept	Acquisition	Comparison	ML / CV
Objectives	Develop a concept for automated construction progress monitoring	Analyze acquisition methods for SfM-based monitoring	Enhance comparisons with respect to occluded elements and elements under construction	Find an automatable way to label images for Machine Learning
Hypotheses	Combination of photogrammetric methods and BIM can provide suitable data for monitoring	UAVs and crane cameras are suitable for (semi-) automated as-built acquisitions	Semantic data from BIM and Computer Vision allow more precise comparison results	Camera pose estimation in combination with an aligned BIM can automate labeling
CH	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Publication	Braun A, Tutas S, et.al., Journal of Information Technology in Construction, 2015	Tuttas, S., Braun, A., et.al., Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 2017	Braun, A., Tuttas, S., et.al., Automation in Construction, 2020	Braun, A., Borrmann, A., Automation in Construction, 2019

Figure 9 Scope of the publications presented in the following chapters.

3. State of the art

Automated construction monitoring has gained increased research interest during the last decade. Besides the urgent need for accurate monitoring of growing construction sites with faster and faster construction methodologies, recent advancements in technology only make it possible to implement these new methods.

There have been many significant developments in the field of digital construction methods that form the basis for this thesis:

- the introduction and implementation of digital, three-, and four-dimensional planning
- more accurate acquisition techniques and even faster processing times
- introduction of UAV and increasing camera quality combined with more affordable prices
- increasing computational power for methods like Computer Vision or Machine Learning

This chapter will introduce these technologies and discuss the current state of the art in the context of progress monitoring. A detailed review of the respective state of the art is also conducted within the presented papers in the Chapters 4 to 7.

During this research, a lot of different data sources are used in combination with information to gain additional knowledge on construction processes. The Data, Information, Knowledge, Wisdom (DIKW) pyramid has been introduced by Rowley (2007) to define these terms and their relation. Based on this definition, *data* is defined as sensor data without any context. In the context of this research, point clouds can thus be defined as data. *Information* is defined as data, enriched with meaning, in this use case, for example, the definition of a specific object's material. Finally, *knowledge* can be the result of the interpretation of given *data* and *information*. Hence, the knowledge of the current status of the progress state is the declared goal of this thesis.

3.1. Building Information Modeling and Open Standards

Digital planning has been part of the construction industry since the introduction of Computer Aided Design (CAD). While being limited to two dimensional (2D) sections and views in the beginning, visualization of three dimensional (3D) geometries in combination with process data (4D) helped to identify planning errors at early design stages (McKinney et al., 1998). More advanced analyses such as clash detection or quantity take-off are available in combination with semantic data such as construction materials. Physical properties of the construction elements enable more accurate calculations and simulations like Life Cycle Analysis (LCA) (Forth et al., 2019).

This purely digital, model-based planning method is referred to as Building Information Modeling (BIM). In this thesis, the expected state of a construction site according to the construction schedule is referred to as the "as-designed" state, while any representation of the actual state is referred to as the "as-built" state. The BIM model represents the "as-designed" status of a construction project at all times, whereas the "as-built" status needs to be acquired by methods, as introduced in the following section 3.2. It always represents the construction state at one particular time of observation.

3.1.1. Industry Foundation Classes

The manufacturer-neutral IFC format was developed so that high-quality geometric-semantic data can be exchanged between different applications. The IFC are developed and maintained by the international non-profit organization buildingSMART and focus on interoperability between BIM software applications by different software vendors (BuildingSmart, 2014). The data exchange format is standardized in ISO 16739. Since this data format is open source and defined by an open schema (EXPRESS, as defined in ISO 10303), the model data can be read without proprietary software. In this schema, each item is represented by an entity that is set into relation to another entity. The schema is based on objectified relationships. An individual object defines each entity and relation. This allows to derive all properties for any component.

3.1.2. Geometry

The geometric representation of all construction elements forms the basis of this research. A geometric definition is required for all as-designed elements that shall be included in an as-designed vs. as-built comparison.

For a valuable and precise comparison, the model itself needs to fulfill requirements regarding the detailing of all construction elements. In the scope of digital element representation, a schema has been developed to classify the detailing of construction elements: the Level of Detail (LOD). According to this schema, a BIM necessitates at least LOD 300 for accurate construction site monitoring. As stated by The American Institute of Architects (2013), "The Model Element is graphically represented within the model as a specific system, object or

assembly in terms of quantity, size, shape, location, and orientation. Non-graphic information may also be attached to the Model Element."

Since the exact position, shape, and measurements are required, this LOD is sufficiently accurate for the desired purpose. Furthermore, buildingSMART defined so-called Model View Definitions (MVDs) that describe the content of a BIM regarding the included elements and exchange requirements in the Architecture, Engineering and Construction (AEC) industry (BuildingSmart, 2016). To cover different use cases and scopes of usage, predefined views are established. The most common views are the "Coordination View" (CV) or as per the newly defined IFC4 "Reference View" (RV). The mentioned views include all modeled elements with details and constructive parts like reinforcement; however, reduce the available geometric representations for unified interpretability across different software vendors. On the downside, geometric modifications to (parametric) objects are not possible anymore since these representations are not supported. In the context of this thesis, this factor can be neglected in favor of precise geometric representations.

Another important aspect regarding model quality is measurement rules for element boundaries. According to German standards, general construction requires accuracy of around 1cm for 1 meter of element length and up to 3cm for 30 meters of element length (DIN, 2013).

Geometry can be described by explicit and implicit representations (Gheorghiu, 1978; Borrmann et al., 2018). Explicit representation schemes describe volumes indirectly via the surfaces of the objects, while implicit methods use combinations of defined volumetric bodies.

The essential geometric elements (primitives) are points, edges, and surfaces (VEF = vertex edge face). Higher-level explicit geometry models also introduce shells and other elements. Explicit representations usually require more storage capacity. However, they allow access to the surface description without further computation. In the context of geometric comparisons, explicit representations like tessellated surfaces are computationally cheaper. Figure 10 illustrates the triangulated geometry of an abutment during bridge construction.

The IFC schema provides different geometry representations, including Constructive Solid Geometry, Swept Area Representations, and boundary representations. In this particular use case, tessellated formats like *IfcTriangulatedFaceSet* are most suitable. The provision of this type of geometry representation can be enforced by a dedicated Model View Definition (MVD).

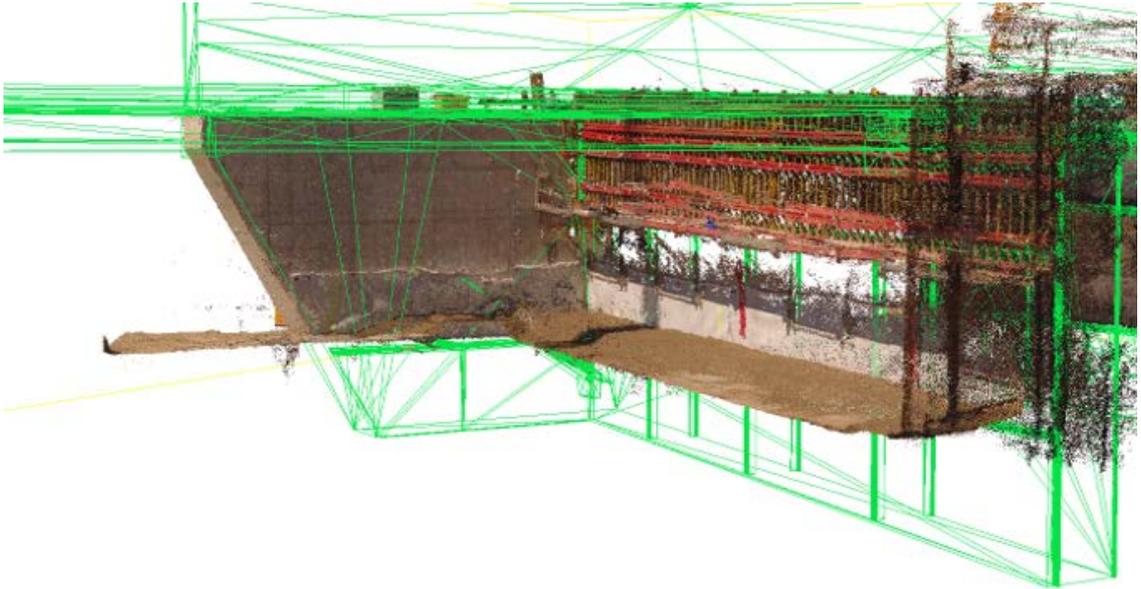


Figure 10 Visualization of a triangulated abutment geometry, overlaid with a corresponding as-built point cloud.

3.1.3. Semantic information

Due to the objectified relationships by the IFC data schema, every property of a construction element is directly linked to it. Thus, semantic information such as material properties is always connected to the entity that represents the corresponding element. In the scope of this research, the following properties were identified as essential to the study:

- type of element (i.e., slab, wall, column)
- material (i.e., concrete, wood)
- layers (for multi-layered elements)
- construction method (i.e., in-situ, prefabricated)
- load-bearing (boolean - yes / no)
- 4D scheduling data (corresponding task, duration, ...)

The listed properties define the general structure of a building as well as the planned schedule. In combination with additional information, the colors of elements can be derived. This information is especially valuable when image-based monitoring is used, where no additional semantic data is available. In this case, information can be derived based on color.

3.1.4. Process planning and dependencies

The enrichment of building models with their corresponding construction schedule is one of the critical requirements for this research. In current construction projects, scheduling is done based on the estimated element volumes, but usually independently from the actual model, based on the scheduler's experiences. Thus, the model and the process schedule are not

interconnected.

This issue can be overcome with the *IfcTask* extension to map processes, introduced with the IFC 4 schema (BuildingSmart, 2014). It enriches the BIM with element-based process information that can be assigned and grouped as required with property sets. In this context, the *IfcProcess* entity represents all individual events. Another possibility is the use of standard tools like MS Project and an additional reference table that holds all IFC Global Unique ID (GUID)s and maps them to the defined process.

Figure 11 visualizes a model, that is linked to its progress schedule, connected via its GUID. Furthermore, dependencies can be made visible. Processes that run independently of each other can be executed simultaneously, i.e., in parallel.

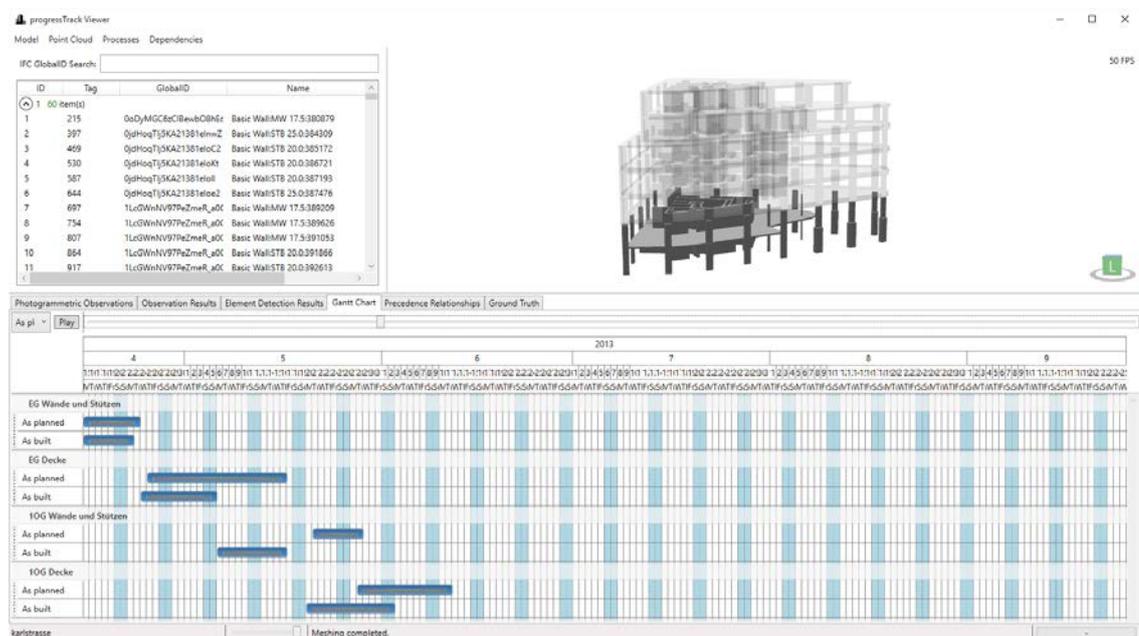


Figure 11 Connected model with corresponding schedule, visualized as Gantt diagram.

However, there are already some approaches to enable automation of scheduling in this area. This task involves many factors that need to be taken into account. The quantities of all elements, as well as personnel effort and resource requirements, add up to the length of an individual task. Additionally, dependencies in relation to other tasks or external influences need to be considered (Hardin and McCool, 2015). By linking process information and the underlying building data model, additional information is available that can be used in connection with progress monitoring. Tauscher (2011) describes a procedure that allows the automation of the generation of construction process planning at least partially. He chooses an object-oriented approach in which each component is categorized according to its properties. Accordingly, each component is assigned to a construction process. Subsequently, characteristic properties of the components are compared with a process database to group them and assign them to comparable processes. Suitable properties for the recognition of similarities are, for example, the component thickness or the building materials used.

With this method, a semi-automated system is developed to support the process planning. In Huhnt (2005), a mathematical formalism based on set theory is used to determine the technical dependencies for automated construction progress planning introduced. In Enge (2010), a Branch-and-Bound algorithm is introduced to allow an optimal decomposition of processes into planning information and process information to determine. Dori (2016) presents an approach in which a methodology for generating the schedule from information about the scope of the tasks, the available resources, and the performance values is described. This approach has not yet been adopted by the industry in broad.

Another vital aspect of Scan-vs-BIM comparisons is the dependencies between the individual construction processes. For example, technological process dependencies prescribe the order in which individual objects are dependent on each other. An object cannot be built before another object on which it is dependent. For example, a slab cannot be built before the load-bearing columns and walls are finished. These dependencies can be represented in so-called precedence relationships (Wu et al., 2010). An established method to represent these dependencies is to use graphs (Szczyzny et al., 2012). These approaches for modeling form an excellent basis for automated construction monitoring, which has not yet been applied in this scope.

Daum and Borrmann (2014) introduced a query language for digital models that allows us to identify geometric orders and topologically query a BIM. With these queries, load-bearing construction elements can be aligned in their vertical order. Thus, the query language enables the automatic generation of these dependency graphs. With these graphs, assumptions on elements are possible, that were not detected, despite being expected to be built. More details are presented in Chapters 4 and 6.

3.2. Manual and contact-based acquisition methods

The acquisition of construction progress on construction sites introduces many challenges. To receive a precise and correct representation of the current construction status, the quickly changing environment needs to be taken into account. Moving parts on-site, like excavators, as well as temporary construction equipment like formwork, hinder to gain an unobstructed view on all construction elements.

In recent studies, Omar and Nehdi (2016), as well as Moselhi et al. (2020), analyzed various acquisition methods and summarized the individual advantages and disadvantages for activity and material tracking on construction sites. Their comprehensive studies categorized these tracking methods in the following sections.

3.2.1. Digital construction site diary

Digital construction site diaries are a first step in digitizing the logging of construction processes. There are several companies on the market that provide services in this area. Usually, a cloud-based service is offered in combination with on-site tablet computers to track the construction process. Despite still being a manual process, the results are stored on the web and are immediately available to all project participants.

The monitoring itself is mostly focused on construction issue tracking. In this regard, recent approaches combine error reports with location-aware tracing on 2D drawings. In conclusion, these methods do not provide automated progress monitoring.

3.2.2. Contact-based methods

These methods use technologies from the field of barcoding, radio frequency identification (RFID), or ultra-wide band (UWB) (Omar and Nehdi, 2016). The main idea of this approach is to tag any delivered item to the construction site and scan it at specific points of interest, for example, the entry of the site or storage places. Using barcodes or QR codes as established standards is a very low-cost option (Shehab and Moselhi, 2005). An effort is required for marking all badges and elements need to be scanned manually.

This approach can also be extended to RFID (i.e., NFC) or also Bluetooth Low Energy (LE). The advantage of these technologies is that there is no direct visual contact with the placed markers needed. They can be scanned in batches upon delivery or while being loaded onto the crane for placement. Thus, this method does not allow continuous tracking and is not suitable for any detailed checks for the exact positioning of the construction elements.

3.3. Point cloud-based acquisition methods

Modern automated acquisition methods of 3D objects are in most cases resulting in 3D point clouds. Commonly used methods to acquire the as-built state of an object are relying this type of representation (Golparvar-fard et al., 2009; Turkan et al., 2011; Bosché and Haas, 2008b; Pučko et al., 2018). Two acquisition methods can be used to generate point clouds. Lasers shoot single light waves at objects, which reflect the light and shoot them back to the receiver at the scanner. This results in the distance information for one point. SfM methods rely on capturing images from different perspectives. They also result in point representations.

3.3.1. Laser scanning

Laser scanners are used in a multitude of applications. A laser scanner emits laser light at high wavelengths that are usually within the range of infrared light (Cracknell and Hayes, 2007). The distance to an object is measured by a sensor that detects the reflected beams of light and records the time taken. Finally, the speed of light is used to calculate the distance between the scanner and the reflecting object. Current models of laser scanners provide options to control the density of the produced point cloud. The higher the density, the longer

the scanning time. The accuracy of current scanner models is documented at around 2mm per 10m distance. These values always depend on scanning times and can be reduced by multiple scanning of every point.

Laser scanners were introduced to construction quality control (Akinici et al., 2006) and monitoring (Bosché and Haas, 2008b) several years ago. While initial steps focused on dimensional controls, these were soon extended to progress estimation (Turkan et al., 2012).

One of the main problems of laser scanners is the high costs involved. Besides, scanning takes long, especially if high-quality point clouds are required. Each measurement requires a new setup of the scanner and later registration of the sub-point clouds to each other. This results in a high effort for post-processing in this approach.

3.3.2. Photo- and Videogrammetry

Photogrammetry describes the process of gaining information on 3D objects based on 2D data such as photographs or videos. The goal is to receive an estimated 3D representation of the object of interest that matches the original's dimensions as closely as possible. Photogrammetry, in particular, has gained more attention with the broader availability of UAV, making this method more flexible in terms of camera positions (Lin et al., 2015).

The main idea is not to use laser scanners but conventional camera equipment on construction sites to capture the current construction state ("as-built"). Since the acquisition from different perspectives is significantly faster than laser scanners, the building can be captured comprehensively with comparatively low effort. A slightly manual effort remains, but different points of observation are acquired very fast (Golparvar-fard et al., 2009).

Reconstruction

The first step in such an SfM process is the detection of image features and the calculation of their descriptors. Usually, Scale-invariant feature transform (SIFT) operators are used for this task (Lowe, 2004). It consists of a detector and a descriptor. The detector is used to identify features that are recognizable in as many scales and from as many directions as possible. The descriptor describes these features as invariant as possible to their properties, i.e., location, scale, and orientation, in the respective image. In Visual-SfM (Wu, 2013b) an implementation of these algorithms on the graphics card (GPU) is used.

Figure 12 illustrates the SfM process, where feature points are identified on images from varying positions during monitoring. The features recognized in the SIFT algorithm are determined using the Random Sample Consensus (RANSAC) algorithm. This estimation algorithm is used to extract incorrect measured values and to identify only those values that are most likely to correspond to the desired solution. Estimation algorithms can also lead to incorrect results, especially if the surveyed building consists of several repetitive and similar geometric objects.

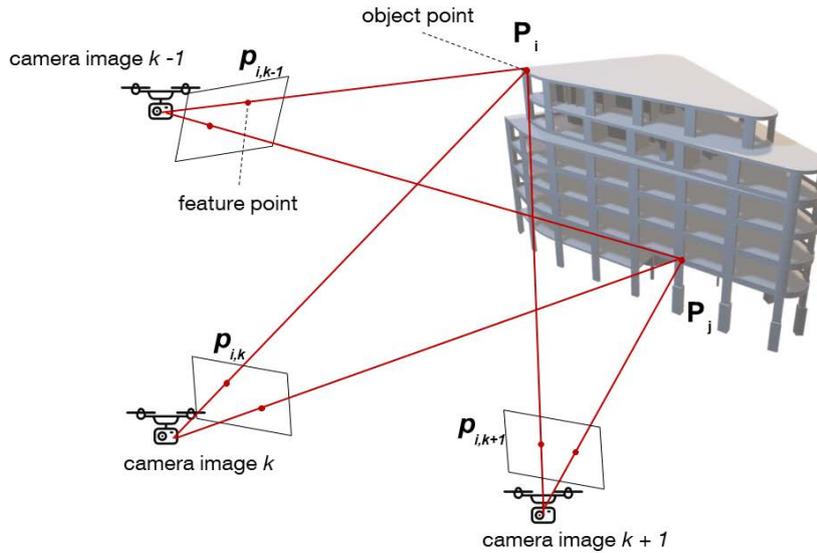


Figure 12 Schema of the SfM method. Feature points are identified on images from varying positions.

Surveying methods

Surveying construction sites poses different challenges on camera positions than, e.g., digital surface models for terrain modeling. Geometries are more complex, and greater detail is required. There are currently several methods available for the acquisition of construction sites. Next to regular manual monitoring with cameras, webcams mounted on cranes or other exposed positions are considered suitable. Another option are UAVs that monitor the construction progress from above. A sample for construction monitoring by UAV is shown in Fig. 13.

Position and distance can be adjusted according to the current progress on-site. For optimal results, each feature must be identified on at least two images (Rothermel et al., 2012). Best results are achieved by flying on a spherical routed flight for optimal coverage (Cesetti et al., 2011). Since the reconstructed points are derived from pixels, a higher image resolution leads to higher densities in the cloud.

During reconstruction, the camera positions are estimated in order to create the 3D point cloud. Figure 14 shows this estimation for a construction site, monitored by two crane cameras. In this figure, the positions of the cameras are visible, but also the camera cone that is visualizing the view direction. Also visible are the SIFT features that are identified on at least three images (Wu, 2013b) and finally result in the point cloud.



Figure 13 Image taken from an UAV during acquisition.



Figure 14 Reconstruction process using VisualSFM with visualized camera positions including their viewing directions (camera cone).

Automation of image acquisition has been investigated. An overview of possible solutions is depicted in Figure 15. These methods were tested and compared in Chapter 5.

The manual approach, shown in the bottom right, allows for high flexibility in viewing angles, positions, and the number of pictures taken. However, positions are limited to areas that are accessible by humans, and the approach is highly manual. In summary, besides thorough monitoring of the as-built status, this method introduces a similar effort compared to entirely manual work. In comparison, crane cameras, as shown in Fig. 15, left, represent a highly automated solution. As validated in Chapter 5, this methodology lacks accuracy in z-direction due to missing information around a physical object.

Another suitable method is the usage of UAVs, as proposed by (Han and Golparvar-Fard, 2015; Lin et al., 2015). They offer great flexibility for varying camera positions outside of the building while also reaching positions inaccessible to humans. Currently, almost all countries regulate the use of UAVs to prevent illegal usage (EASA, 2020). In the context of construction site observations, mostly safety concerns need to be addressed.



Figure 15 Proposed methods for image acquisition that were investigated during this research. Left: crane cameras, top right: UAVs, bottom right: manual acquisition.

Especially the overflight prohibition of areas where people are working and the requirement to always be able to control the UAV pose impact on autonomous usage. On the downside, they are usually not able to fly inside a building due to lacking Global Positioning System (GPS) connectivity. Some recent approaches try to overcome this issue by using different sensor types like distance measurements (Dowling et al., 2018). As a result, they are only suitable for usage outside of a building, monitoring the outer hull.

As described in detail in Chapter 5, UAVs have the highest flexibility for monitoring the outer hull of a building, while crane cameras have the lowest effort for continuous monitoring (see Table 1). Monitoring the inside of a building requires different technologies, i.e., Augmented reality (AR)-based systems, as presented in Section 3.3.3.

Acquisition	Advantages	Limitations
Manual	fast, cheap, focus on points of interest, easy maintenance	completely manual approach
UAV	flexible view on outer hull	manual, due to regulations
Crane camera	no manual effort, high automation	no maintenance possible, limited views from sides, higher errors in z-direction

Table 1 Photogrammetric methods suitable for construction monitoring.

Challenges

In conclusion, photogrammetry enables a quick and cheap solution for as-built acquisition. Besides the manual work needed for the image capturing itself, several issues arise. Reconstruction of low textured or translucent surfaces leads to errors and holes in the point clouds (Hirschmuller, 2005). Reflections also introduce reconstruction errors since the reflected features lead to misaligned points.

3.3.3. Distance sensors and AR

Special devices, such as Microsoft Kinect, combine multiple sensors and can be used for progress monitoring (Pučko et al., 2018). The most recent advancements in the scope of mobile distance measurements combine Time-of-Flight sensors with AR displays.

The primary sensor is a Time of Flight camera with similar functionality in comparison to the laser scanner but works on the whole picture. Therefore it is faster but only provides a low resolution (Hübner et al., 2020). Besides, it covers a limited range of up to 40m, which restricts its application on large scale construction sites.

Kopsida and Brilakis (2020) used this device on construction sites for indoor monitoring. Orientation and registration to a BIM need to be adjusted up front, but the device keeps its orientation during movement.

Table 2 summarizes all evaluated methods with their respective advantages and limitations.

Technology	Advantages	Limitations
Contact based	fast, low error rate	manual, no continuous tracking
Laser scanning	highest accuracy	long acquisition times, large files, registration
Photogrammetry	fast acquisition, image data available	long postprocessing, registration
AR / ToF	fast, low processing power, semi-automated registration	very limited range, low resolution

Table 2 Currently available acquisition methods suitable for construction monitoring.

3.4. Scaling and Registration

As-built point clouds - especially from photogrammetry - do not necessarily possess the correct scale and orientation. They require translation as well as transformation to be positioned and oriented at the correct position but also be scaled correctly. Laser-scanned point clouds are scaled directly during monitoring. Images, however, do not possess the possibility to scale correctly. Suitable methods are either the use of GPS positioning data from EXIF information, added by the camera (as done, i.e., by Pix4D).

Another suitable method is the placement of geodetic markers on-site (see Figure 16). These markers are geodetically measured, allowing the point cloud to be transformed into the actual geodetic position. While doing the same for the as-designed BIM model, the registration at the actual position leads to the most accurate results (Tuttas et al., 2017).

If neither geodetic markers nor GPS data is available, the as-built point cloud and the as-designed model need to be aligned by other means.

Registration can be done manually, by point-to-point picking. With this method, corresponding points from two geometric data sets that shall be aligned are selected. Accordingly, the target point cloud can be aligned to the reference point cloud by transformation as well as translation.

Besides this manual approach, the Iterative Closest Point Algorithm (ICP) algorithm provides an automated procedure (Bosché and Haas, 2008b). The ICP algorithm can align multiple point clouds with each other. This is done by minimizing the overall distance between all points. It is an optimization algorithm that searches for the minimum distance and can be used to align an as-built point cloud to an as-designed geometry. This geometry has to be transformed into a point cloud in the first step by randomly placing points on its geometry surfaces.



Figure 16 Registration of SfM-based point clouds using geodetic markers.

However, this approach requires both point clouds to have high similarity. If the as-built point cloud contains many surrounding elements, like scaffolding or formwork, that are not present in the as-designed model, the ICP algorithm will fail to produce an exact solution (Tuttas et al., 2017). This can be partially overcome by using the algorithm in combination with AR methods (Kopsida and Brilakis, 2016).

3.5. Computer Vision and Machine learning

Computer Vision is a research area that focuses on analyzing image and video data with the help of computers (Barlow, 1983). The goal is to identify objects as humans do and semantically enrich the visual data with object information. Advances in this field of research contributed significantly to advances in autonomous driving. Since the basis for this research are images and videos, high computational loads occur in terms of storage as well as data processing. Especially Graphics Processing Unit (GPU) power has risen significantly within the last decade and is one of the critical factors for more advanced CV approaches (NVIDIA, 2018).

3.5.1. Ray-casting

Ray-casting describes a method for determining visible surfaces or points (Bungartz et al., 2002). A virtual beam is shot from a starting point or observation point onto the observed object. On a projected image plane, those elements are then displayed, which have the smallest distance on a virtual line between the observer and the target point. It is a concept within the method of ray-tracing, however, reflections are neglected as the ray is not followed

recursively. This method makes it possible to create a three-dimensional map containing only the visible areas of a room. Thus, weak points can be sorted out from the outset, and unnecessary calculations can be avoided. Octrees allow spatial filtering of objects (Meagher, 1982). With Octomap, a software library was developed which deals with this problem and, based on point clouds and the corresponding camera positions creates individual voxels for defined edge sizes and calculates a probability with which these voxels are visible (Hornung et al., 2013).

SfM methods provide all necessary data for the implementation of a ray-casting supported visibility analysis. This analysis is run on the as-designed model at the time of observation to gain knowledge on the visible elements at this exact time. It is required to gain a precise overview of the as-planned elements and the expected visibility based on the presence of said elements. The point of view, as well as the vector for the viewing direction, is available, allowing to render the view on the model. Detailed explanation on this method is given in Chapter 7, implementing the Painter's algorithm for depth analysis (de Berg, 1993).

3.5.2. Machine learning

Humans perceive their surroundings with a combination of vision, feelings, and acoustics and grasp the context by experience (Gandarias et al., 2019). Computers are currently not capable of imitating this action comprehensively. Artificial Intelligence (AI) aims to overcome this issue by providing functions to computers such as learning or problem solving (Poole et al., 1998). A subset of AI are Artificial neural networks that use neurons that individually decide, whether or not a specific criteria is fulfilled. These criteria are weighted and need to be adjusted during learning, which is a crucial part of these networks. Deep learning (LeCun et al., 2015) and primarily CNNs provide solutions for training computers to learn patterns and apply them to previously unseen data.

They are applicable to any structured input data like text, images, but also point cloud data. Based on seen training sets, a ML network can derive assumptions to unknown data. A particular requirement is a diversified training set. Otherwise, the network can fail to learn on unknown data (also known as overfitting).

Machine learning algorithms have been increasingly used for effective and efficient image analysis and object recognition during the last decades. However, image analysis in the construction sector is a rather new topic.

CNN

A CNN is a particular sub-domain of ML. In Jahr et al. (2018), an overview of the current state of the art is presented. The underlying methods are also referred to as "Deep Learning". CNNs are structured in locally interconnected layers with shared weights (LeCun et al., 2015). Each layer contains multiple neurons. The neurons of the first input layer represent the pixels of the analyzed image; the output layer comprises the predictable object classes. In between

the input and output layer, any number of hidden layers can be arranged. These layers use convolutions as linear operations to reduce the input data for computational effort.

This is particularly important for the training of said networks. During this task, weights are adjusted, and unnecessary information is filtered to focus on the most interesting parts of a data-set. As a result, a trained network can be used on previously unseen data.

In 2012, the CNN "AlexNet" (Krizhevsky et al., 2017) achieved a top-5 error of 15.3% in the ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al., 2015). These results were surprisingly accurate at the time, proving the advantages of using CNN. On this account, the software industry shifted towards using CNN for all machine learning-based image processing tasks (LeCun et al., 2015). There are different tasks to be solved by image processing algorithms. Well known problems include classification, where single-object images are analyzed, object detection, where several objects in one image may be classified and localized within the image, and image segmentation, where each pixel of an image is classified (Buduma and Locascio, 2017).

While AlexNet contains 8 hidden layers, GoogLeNet (Szegedy et al., 2015), and Microsoft ResNet (He et al., 2016) use more than 100 hidden layers. The layers are usually convolution layers (sharpening features), pooling layers (discarding unnecessary information), or fully connected layers (enabling classification) (Buduma and Locascio, 2017; Albelwi and Mahmood, 2017). Many currently used CNNs rely on the COCO Dataset as it provides a vast set of labeled images (Andriluka et al., 2014). Facebook's Mask R-CNN (He et al., 2017) has provided promising results for machine learning in previous applications. To adapt to different problems, such as recognizing formwork elements on images, CNNs must be trained. During training, the connections between specific neurons are increased, while the connections between other neurons are reduced—the weights connection consecutive layers are weighted.

The training is usually carried out using supervised back-propagation, meaning that the network is fed with example input-output pairs (Buduma and Locascio, 2017). The correct solution for each input is called ground truth. A significant amount of training data is required to train a CNN towards reliable predictions, which has to be prepared in a preprocessing step. Weights of previously trained CNNs can be used to accelerate the training processes. The fully connected layers are replaced with layers representing the new data and trained with the new data, to adapt pre-trained CNNs. This process is also known as transfer learning and supports faster learning by adopting information from previously trained networks.

Image datasets

Analyzing images from construction sites, on the other hand, is a rather new topic. Since one of the critical aspects of machine learning is the collection of large data sets, current approaches focus on data gathering. In the scope of automated progress monitoring, Han and Golparvar-Fard (2017a) published an approach for Amazon Turk based labeling. Kropp

et al. (2018) tried to detect in-door construction elements based on similarities, focusing on radiators.

Up to now, the main focus has been on defect detection (for example, cracks) in construction images (Akinci et al., 2006). Crack detection for asphalt roads has also been the subject of research (Nhat-Duc et al., 2018). Chi and Caldas (2011) used initial versions of neural networks to detect construction machinery on images regarding the application of deep learning for construction progress tracking. Kim et al. (2013) analyzed images by filtering them to remove noise and uninteresting elements. They used ML-based techniques for construction progress monitoring to focus the comparison on relevant construction processes. Hamledari et al. (2017) applied CV approaches to indoor appliances like electrical outlets and insulation.

Visibility analysis, in combination with Scan-vs-BIM, can help to tackle this issue. In Chapter 7, a new approach for automated labeling of images is presented that uses the semantic BIM data in combination with the acquired images.

3.6. Scan-to-BIM

In comparison to the extensively described Scan-vs-BIM approach in this chapter, Scan-to-BIM provides a new approach to point-cloud based reconstruction. Scan-to-BIM describes methods that interpret generated point clouds and derive a valid as-built BIM model from them (Xu et al., 2016; Bosché et al., 2015). While Scan-vs-BIM approaches can rely on the as-designed geometry, this information is not present in Scan-to-BIM approaches. This method gained extensive research interest, as the vast amount of currently existing buildings does not have an as-built BIM model. Building owners want to profit from the added value an as-built model provides (Pătrăucean et al., 2015), especially during operation and maintenance. This approach uses the same acquisition methods as Scan-vs-BIM methods and also uses partly similar point cloud processing techniques. In detail, some detection methods on point cloud-level can benefit both progress monitoring as well as reconstruction.

3.6.1. Geometry fitting

Many research groups tried to solve this problem by fitting Constructive Solid Geometry (CSG) geometries into the point clouds. Xu et al. (2015) present an approach that tries to reconstruct scaffolding elements from SfM point clouds. Bosché et al. (2015) focus on cylindrical Mechanical, electrical, and plumbing (MEP) objects, like pipes, and combine this approach with Scan-vs-BIM methods to achieve better accuracy.

3.6.2. Segmentation and object detection on point clouds

Point clouds do not provide any geometric primitives, which makes it hard to segment a point cloud into snippets that represent single objects.

ML methods on point clouds are the logical next step to obtain semantic data from point clouds. With PointNet, a CNN is presented, that can be trained on point clouds and detect pre-trained objects in point clouds (Qi et al., 2017). There are currently no available data-sets available that provide sufficient accuracy to provide consistent detection results for the construction industry. To overcome this issue, artificial data could be used. However, this approach usually has only limited applicability to real-world data (Breu, 2019).

Obrock and Gülch (2020) combine point clouds and images for semantic segmentation. With this combination and a deep-learning approach, they detect edges and surfaces of walls and label the point cloud segments accordingly. Armeni et al. (2019) created an extensive, labeled point cloud data set for indoor office buildings. With the introduced building parser, a large data-set is provided that includes segmented and labeled point cloud data.

3.7. Summary

As presented in this chapter, construction progress monitoring is under substantial research. By automating this process, inaccuracies can be avoided, and working hours could be reduced. However, a multitude of technologies and methods is required for Scan-vs-BIM approaches and their enhancements.

In particular, the combination of geometry and semantic data provides additional value to the Scan-vs-BIM approach. Moreover, CV and ML methods support an SfM-based approach to get more accurate results for progress estimations.

The following Chapters present a more in-depth state of the art review and introduce all methods in detail. Starting with the proposed concept in Chapter 4, image-acquisition is detailed in Chapter 5. The usage of semantic information, as well as CV is introduced in Chapter 6. Finally, a new method for semantic image labeling using SfM is presented in Chapter 7.

4. A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds

Previously published as: Braun A, Tuttas S, Borrmann A, Stilla U (2015). A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds, *Journal of Information Technology in Construction (ITcon)* Vol. 20, pg. 68-79, <https://www.itcon.org/paper/2015/5>

Abstract

On-site progress monitoring is essential for keeping track of the ongoing work on construction sites. Currently, this task is a manual, time-consuming activity. The research presented here describes a concept for an automated comparison of the actual state of construction with the planned state for the early detection of deviations in the construction process. The actual state of the construction site is detected by photogrammetric surveys. From these recordings, dense point clouds are generated by the fusion of disparity maps created with semi-global-matching (SGM). These are matched against the target state provided by a 4D Building Information Model (BIM). For matching the point cloud and the BIM, the distances between individual points of the cloud and a component's surface are aggregated using a regular cell grid. For each cell, the degree of coverage is determined. Based on this, a confidence value is computed which serves as basis for the existence decision concerning the respective component. Additionally, process- and dependency-relations are included to further enhance the detection process. Experimental results from a real-world case study are presented and discussed.

4.1. Introduction

The traditional, manual construction progress assessment with human presence is still dominating. The main reason is the lack of reliable and easy to use software and hardware for the demanding circumstances on construction sites. Automating construction progress monitoring promises to increase the efficiency and precision of this process. It includes the acquisition of the current state of construction, the comparison of the actual with the target state, and the detection of variations in the schedule and/or deviations in the geometry.

A BIM provides a very suitable basis for automated construction progress monitoring. A BIM is a comprehensive digital representation of a building comprising not only the 3D geometry of all its components but also a semantic description of the component types and their relationships (Eastman, 1999). The model is intended to hold all relevant information for all project participants. In addition to the description of the building itself, it also comprises process information, element quantities and costs. A Building Information Model is a rich source of information for performing automated progress monitoring. It describes the as-planned building shape in terms of 3D geometry and combines it with the as-planned construction schedule. The resulting 4D model (Webb et al., 2004) combines all relevant information for the complete construction process.

Accordingly, the planned state at any given point in time can be derived and compared with the actual construction state. Any process deviation can be detected by identifying missing or additional building components. For capturing the actual state of the construction project in an automated manner, different methods can be applied, among them laser scanning and photogrammetric methods. Both methods generate point clouds that hold the coordinates of points on the surface of the building parts but also on all objects, which occlude them.

The main steps of the proposed monitoring approach are depicted in Fig. 17. The minimum information, which has to be provided by the BIM, is a 3D building model and the process information (construction start and end date) for all building elements. From this, the target state at a certain time step t is extracted. Subsequently the target state is compared to the actual state, which is captured by photogrammetric techniques in this study. Finally, the recognized deviations are used to update the schedule of the remaining construction process.

The paper is organized as follows: Section 2 gives an overview on related work in the field. The proposed progress monitoring procedure is explained in detail in Section 3 and first experimental results are presented in Section 4. The paper concludes with a summary and discussion of future work.

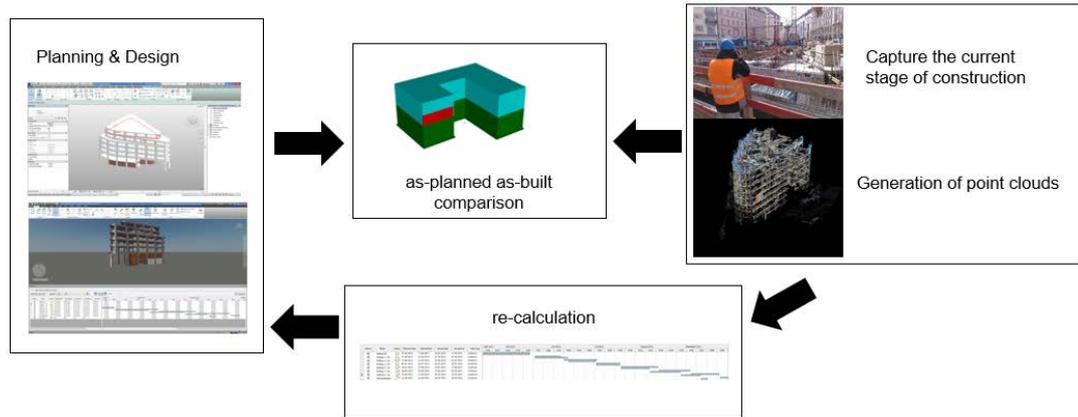


Figure 17 construction progress monitoring schema

4.2. Related work

4.2.1. Monitoring and object verification

Mainly as-built point clouds can be acquired by laser scanning or imaged-based / photogrammetric methods. In Bosché and Haas (2008b) and Bosché et al. (2010) a system for as-built as-planned comparison based on laser scanning data is presented. The generated point clouds are co-registered with the model with an adapted ICP. Within this system, the as-planned model is converted to a point cloud by simulating the points using the known positions of the laser scanner. For verification, they use the percentage of simulated points, which can be verified by the real laser scan. Turkan (2012); Turkan et al. (2013, 2014) use and extend this system for progress tracking using schedule information, for estimating the progress in terms of earned value and for detection of secondary objects.

Kim et al. (2013b) detect specific component types using a supervised classification based on Lalonde features derived from the as-built point cloud. An object is regarded as detected if the type fits to the type in the model. As above, the model also has to be sampled into a point representation here. Zhang and Arditi (2013) introduce a measure for deciding four cases (object not in place, point cloud represents a full object or a partially completed object or a different object) based on the relationship of points within the boundaries of the object and the boundaries of shrunk object. The authors test their approach in a very simplified test environment, which does not include any problems, which occur on data acquired on a real construction site. The usage of cameras as acquisition device comes with the disadvantage of a lower geometric accuracy compared to the laser scanning point clouds.

However, cameras have the advantage that they can be used more flexible and their costs are much lower. This leads to the need for other processing strategies if image data is used. Rankohi and Waugh (2015) give an overview and comparison of image-based approaches for

the monitoring of construction progress. Ibrahim et al. (2009) use a single camera approach and compare images taken over a certain period and rasterize them. The change between two time frames is detected through a spatial-temporal derivative filter. This approach is not directly bound to the geometry of a BIM and therefore cannot identify additional construction elements on site. Kim et al. (2013a) use a fixed camera and image processing techniques for the detection of new construction elements and the update of the construction schedule. Since many fixed cameras would be necessary to cover a whole construction site, more approaches rely on images from hand-held cameras covering the whole construction site as in our and the approaches in the following.

For the scale of the point cloud, stereo-camera systems can be used, as done in Son and Kim (2010); Brilakis et al. (2011); Fathi and Brilakis (2011). Rashidi et al. (2015) propose to use a coloured cube with known size as target, which can be automatically measured to determine the scale. In Golparvar-Fard et al. (2011) image-based approaches are compared with laser-scanning results. The artificial test data is strongly simplified and the real data experiments are limited to a very small part of a construction site. Only relative accuracy measures are given since no scale was introduced to the photogrammetry measurements.

Golparvar-Fard et al. (2011, 2015) use unstructured images of a construction site to create a point cloud. The orientation of the images is performed using a SfM process. Subsequently, dense point clouds are calculated. For the comparison of as planned and as built, the scene is discretized into a voxel grid. The construction progress is determined in a probabilistic approach, in which the parameters for threshold for detection are determined by supervised learning. In this framework, occlusions are taken into account. This approach relies on the discretization of the space by the voxel grid, having a size of a few centimeter.

In contrast to this, we calculate in the approach presented in this paper the deviation of point cloud and building model directly and introduce a scoring function for the verification process. In contrast to most of the discussed publications, we present a test site which presents extra challenges for progress monitoring due to the existence of a large number of disturbing objects, such as scaffolding.

4.2.2. Process information and dependencies

Process planning is often executed independently from conceptual and structural design phases. Current research follows the concept of automation in the area of construction scheduling.

Tauscher (2011) describes a method that allows automating the generation of the scheduling process at least partly. He chooses an object-oriented approach to categorize each component according to its properties. Accordingly, each component is assigned to a process. Subsequently, important properties of components are compared with a process database to group them accordingly and assign the corresponding tasks to each object. Suitable prop-

erties for the detection of similarities are for example the element thickness or the construction material. With this method, a "semi - intelligent" support for process planning is implemented.

Huhnt (2005) introduced a mathematical formalism that is based on the quantity theory for the determination of technological dependencies as a basis for automated construction progress scheduling. Enge (2010) introduced a branch and bound algorithm to determine optimal decomposition of planning and construction processes into design information and process information. These innovative approaches to process modelling form a very good basis for the automated construction monitoring, but have so far not been applied in this context.

4.3. Concept

The developed methodology comprises the following phases:

During the design and planning phase, the building model and the process schedule is modelled and combined in a 4D model. During construction, the site is continuously monitored by capturing images of the as built state. These are processed to create point clouds (Section 4.3.1), which are compared to the as-planned building model (as-built – as-planned comparison), that is described in Section 4.3.3. Process and spatial information can help to further improve the detection algorithms (Section 4.3.2).

4.3.1. Generation of as-built data

The generation of the point cloud consists of four steps: Data acquisition, orientation of the images, image matching and co-registration. Image acquisition: Photogrammetric imaging with a single off-the-shelf camera is chosen as data acquisition since it is inexpensive, easy to use and flexible. When using a camera, some acquisition positions such as on top of a crane can be reached more easily than when using a laser scanner. In addition, a major requirement is that the image acquisition process shall be conducted without any disturbance of the construction process. For performing a suitable acquisition, the construction site should be covered as complete as possible. Orientation: The orientation process is performed using the structure-from-motion system VisualSfM Wu (2013b) for an automatic generation of tie points. By means of the algorithm, also the relative orientations of the cameras are determined.

For the following reasons we also introduce (manually) located control points:

- Having two control points, a distance is introduced and the missing scale is known then.
- With the help of control points, we can combine image groups that could not be orientated relatively to each other by the usage of only the automated measured correspondences.
- Control points are preferably in the same coordinates system as the one that is used for

the construction work itself. If this is ensured, the point cloud is already co-registered to the model (assuming it is having also the same coordinate system).

The joint usage of control points and tie points is depicted in Fig. 18. The red circles are control points on stable points outside the construction site. The yellow lines represent the tie points based on automatically detected features which connect overlapping images of one time step. Depending on the scene content the number of tie points connecting two images is much higher (tens to hundreds) than it is indicated by the four yellow lines in the figure.



Figure 18 Image orientation process using control points and tie points

Finally, a bundle block adjustment is accomplished to determine the exterior orientation of all images and the corresponding standard deviations. Image matching: Using either calibrated parameters or parameters from self-calibration (i.e. determined simultaneously with the orientation parameter), distortion free images are calculated. In this study, a calibration device has been used to calibrate the camera in advance.

As next step, stereo pairs (image pairs which are appropriate for image matching, i.e. they shall be overlapping and shall have approximately an equal viewing direction) have to be determined. This is done based on conditions on the baseline length and the angles between baseline and the camera axes. Every image of each stereo pair is rectified. That means that artificial camera orientations are calculated so that the camera axes of the pair are orientated normal to the base and parallel to each other. The rectified images are resampled from the original images. These images can then be used for dense-matching.

For every pixel, a corresponding pixel in the other image is searched and the disparity is determined. The disparity is the distance of two pixels along an image row. To determine this, semi-global-matching (SGM) has been established in the last years (Hirschmuller, 2005). Different implementations are available, e.g. SGBM in the openCV-library or LibTSGM (Rothermel et al., 2012), which is used here. By means of the disparity (what corresponds to the depth of the point) and the exterior orientation of both images, the 3D point can be triangulated.

To get a more robust estimation of the points, to reduce clutter and to estimate the accuracy of the depth, not simply all 3D-points of all stereo-pairs are combined but overlapping disparity maps are merged and only 3D-points are triangulated which are seen in at least three images. The following procedure follows the approach of Rothermel et al. (2012).

First, an image has to be selected to become a master image. For every pixel of the undistorted master image, the disparities are interpolated from all k disparity maps the master image is involved in. Now for every pixel, n disparity values are available. An interval for the distance D from the camera centre to the 3D-point is determined by adding/subtracting an uncertainty value s from the disparity value. For every pixel, the depth values are clustered into one group if the intervals are overlapping. For calculating the final depth, the cluster having the most entries is chosen. The final value and its accuracy are determined by a least-square adjustment as described by Rothermel et al. (2012). The final 3D-point coordinates (X, Y, Z) are then calculated by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R^t * (n * D) + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (4.1)$$

with rotation matrix R (from object to camera coordinate system), unit vector n from perspective centre to pixel and camera position X_0, Y_0, Z_0 . By applying the law of error propagation, the accuracy of the co-ordinates are calculated, using the standard deviations estimated in the bundle block adjustment (for R and X_0, Y_0, Z_0) and the determination of the depth (D), respectively. As last step, the point clouds of all master images are fused. For every point, the coordinate, the RGB-colour, the accuracy in depth, the accuracy for the co-ordinates and the ID of the reference image are stored. With the latter information, the ray from the camera to the point can be retrieved. This is a valuable information to apply visibility constraints for comparing the as-planned and as-built state.

Co-registration: If the model coordinates as well as the control point coordinates are in a common construction site reference frame, a co-registration is not necessary. Otherwise, corresponding features that can be determined unambiguously in the model and the images have to be measured to calculate the transformation parameters. Of course, only building parts that have been proofed to be built correctly can be used for that. This has to be performed only once in an early time step, since the parameters are constant during the construction process.

4.3.2. Process information and technological dependencies

In principle, a building information model can contain, besides geometry and material information, all corresponding process data for a building. The open source standard file format for storing building information models and related information is called Industry Foundation Classes (IFC). This file format is maintained and developed by the BuildingSMART organisation.

In Version 4 of the IFC data model, the *IfcTask* entity was extended by the subtype *IfcTaskTime* to represent all process information and dependencies for a building element with direct relations to corresponding elements (BuildingSmart, 2014). The *IfcTask* entity holds all task related information like the description, construction status or work method. The *IfcTask* is related to an object using *IfcRelAssignsToProduct* but could also be assigned to another relating process. The complete time-related information is hold in the sub entity *IfcTaskTime*. The information hold here include, next to the process duration, additional process data like *EarlyStart* or *EarlyFinish*. Therefore, this entity gives the possibility to combine geometry and process data with all monitoring related process information in a convenient way in one file.

Technological dependencies

In current industry practice, construction schedules are created manually in a laborious, time-consuming and error-prone process. As introduced by Huhnt and Enge (2007), the process generation can be supported by detecting technological dependencies automatically. These dependencies are the most important conditions in construction planning. In the following, the concept of the technological dependencies is illustrated with the help of a simple two-storey building (Fig. 19).

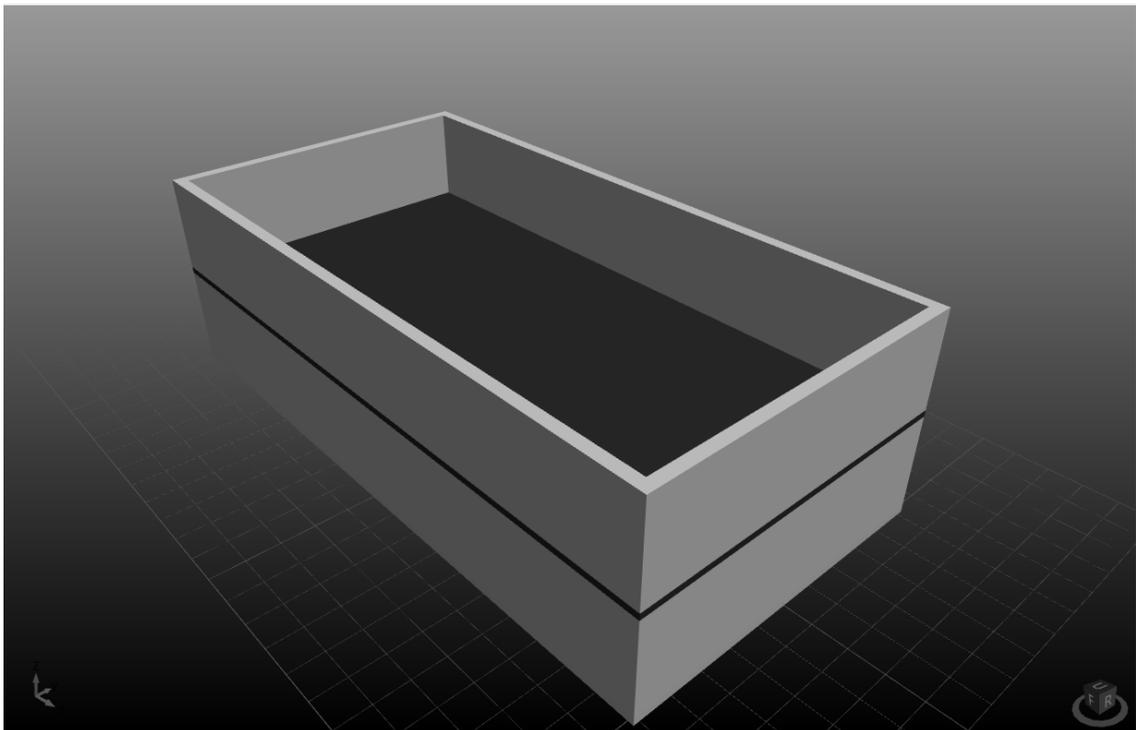


Figure 19 Sample building used to illustrate technological dependencies

The specimen building has four walls and a slab for each floor. One example for deriving dependencies from the model is the following: The walls on the second floor cannot be built before the slab on top of the first floor is finished. The same applies for this slab and the walls beneath it. These dependencies are defined as technological dependencies. Other dependencies that have to be taken into account for scheduling, such as logistical dependencies, are defined by process planners and thus cannot be detected automatically.

A good solution for representing and processing these dependencies are graphs (Enge, 2010). Each node represents a building element while the edges represent the dependencies. The graph is directed since the dependencies apply in one way. By convention, we define the edges as being directed from an object (predecessor) to the depending object (successor). Fig. 20 shows the technological dependencies of the sample building in the corresponding precedence relationship graph.

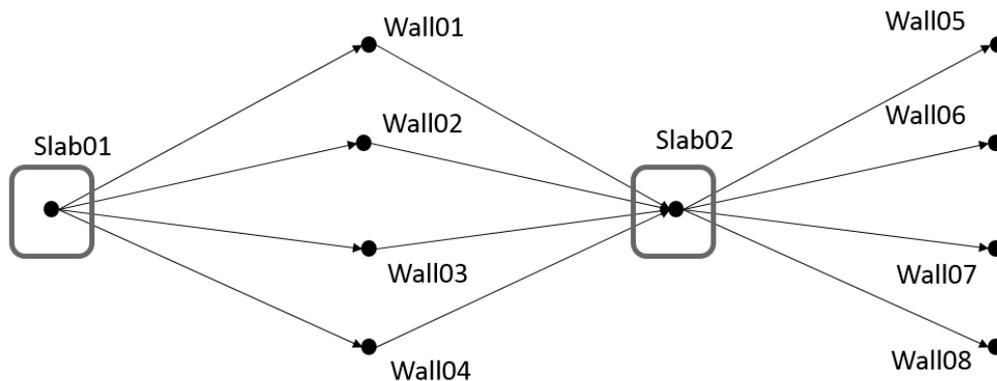


Figure 20 The technological dependencies for the sample building depicted in Fig. 3 in a precedence relationship graph

Checkpoint components

The graph visualizes the dependencies and shows that all following walls are depending on the slab beneath them. In this research, these objects are denoted as checkpoint components. They play a crucial role for helping to identify objects from the point clouds that cannot be confirmed with a sufficient measure by the as built point cloud (see Section 4.3.3).

In graph theory, a node is called articulation point, if removing it would disconnect the graph (Deo, 2004). As defined in this paper, all articulation points represent a checkpoint component. An articulation point is an important feature for supporting object detection, since it depends on all its preceding nodes. In other words, all objects have to be finished before the element linked to the articulation point can be started to be built. As soon as a checkpoint component is detected (represented by an articulation point in the graph), all preceding nodes can be marked as completed. Doing so, also occluded objects can be detected by the proposed method.

4.3.3. Comparing as-built and as-planned state

The "as-planned" vs. "as-built" comparison can be divided into several stages. This includes the direct verification of building components based on the point cloud and the indirect inference of the existence of components by analysing the model and the precedence relationships to make statements about occluded objects.

Matching point cloud and object surfaces

For the verification process, which is based only on geometric condition in a first step, a triangle mesh representation of the model is used. Every triangle is treated individually. It is split into two-dimensional raster cells of size x_r as shown in Fig. 5 a). For each of the raster cells it is decided independently if the as-built points confirm the existence of this part of the triangle surface using the measure M . For the calculation of this measure the points within the distance δd before and behind the surface are extracted from the as-built point cloud. The measure M is based on the orthogonal distance d from a point to the surface of the building part, taking into account the number of points extracted for each raster cell and the accuracy of the points δd , and is calculated as follows:

$$M = \frac{1}{\mu_d} \cdot \sum_i (w_i \cdot g(d_i)) \quad \text{with } g(d_i) = \begin{cases} 1/d_{\min} & \text{if } d_i \leq d_{\min} \\ 1/d_i & \text{if } d_{\min} < d_i \leq \delta d \\ 1/(x - 2 \cdot \delta d) + 1/\delta d & \text{if } \delta d < d_i \leq u \\ -1/d_{\min} & \text{if } u < d_i \end{cases} \quad u = 2\delta d - 1/\left(\frac{1}{d_{\min}} + \frac{1}{\delta d}\right)$$

$$w_i = \frac{w_{d_{\min}} - w_{\delta d}}{d_{\min} - \delta d} \cdot (\sigma_{d_i} - \delta d) + w_{\delta d} \quad \text{with } \sigma_{d_i} = \begin{cases} \sigma_{d_i} = \sigma_{\min} & \text{if } \sigma_{d_i} \leq \sigma_{\min} \\ \sigma_{d_i} = \sigma_i & \text{if } \sigma_{\min} > \sigma_{d_i} > \sigma_{\max} \\ \sigma_{d_i} = \sigma_{\max} & \text{if } \sigma_{d_i} \geq \sigma_{\max} \end{cases}$$

The points are weighted by their inverse distance. This is represented by the weighting function $g(d_i)$ which is shown in Fig. 21 b). The term w_i weights the points by their standard deviation and can take values between 0 and 1. The weight $w_{d_{\min}}$ for standard deviation equal or smaller to $\sigma_{\min} = d_{\min}$ is 1. The weight $w_{\delta d}$ is chosen for a standard deviation having the same value as δd . The parameter δd is the allowed distance tolerance for accepting a point to contribute for the verification of a building part. All σ_{d_i} that are larger than σ_{\max} get the weight 0. The value δd denotes the mean value of the distances of all points to the surface within one raster cell. We compare M against a threshold S to decide if the raster cell is confirmed as existent through the points. S can be calculated based on the chosen values for δd and x_r by defining minimum requirements for a point configuration that is assumed sufficient (see example in Section 4.4). All parameters are shown and explained in Table 3.

Graph-based identification

To further improve the process of comparing actual and target state, checkpoint components and especially articulation points from the precedence relationship graph that represents the technological dependencies help to infer the existence of objects, which cannot be detected by point cloud matching due to occluded objects. Those objects are present on the construction site but are occluded by scaffolding, other temporary work equipment or machines.

Identifying articulation points in a graph can be achieved with the following method:

Loop over all existing nodes in the graph and per-form the following routine:

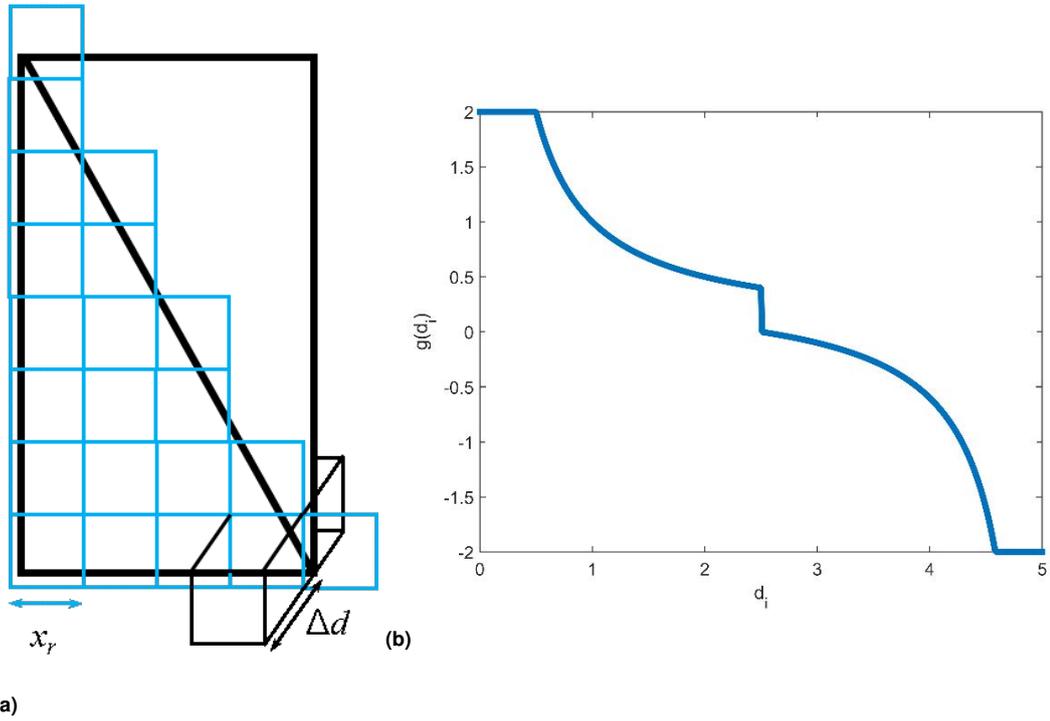


Figure 21 a) Rasterization of a rectangular part of building element, split into two triangular model planes. Each triangle is split into raster cells of size x_r (here in blue), for each raster cell points are extracted using a bounding box of size δd ; b) Weighting function $g(d_i)$ with $d_{min} = 0.5cm$ and $\delta d = 2.5cm$

Parameter	Explanation	Typical Value
d	Absolute value of orthogonal distance between triangle plane and point	0 – 5cm
μ_d	Mean value of all d within one raster cell	0 – 5cm
d_{min}	Smallest value allowed for d for the calculation of M	0.5cm
δd	Maximum distance a point can have to support the decision that a building part exists	2 – 3cm
w_i	Weight for each point based on its standard deviation	0...1
$w_{d_{min}}$	Weight for a point having the standard deviation d_{min} or smaller	1
$w_{\delta d}$	Weight for a point having the standard deviation equal to δd	0.8
σ_{min}	Points with standard deviation σ_{min} and smaller get the highest weight $w_{d_{min}}$	d_{min}
σ_{max}	Calculated for a linear decreasing weight function (standard deviation with the weight 0)	> 5cm

Table 3 Parameters and variables

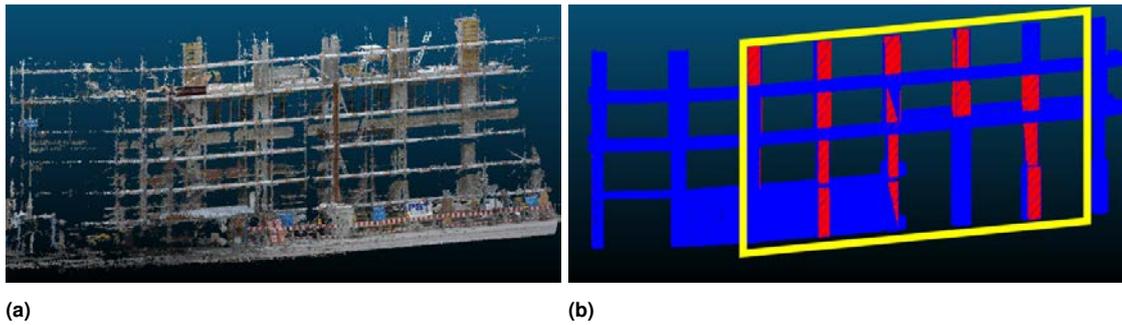


Figure 22 a) Snippet of the point cloud; b) Result for the as-built as-plant comparison, model planes marked blue having the state not built or unknown, model planes marked in red having the state built; The yellow frame shows the area used for evaluation.

- Remove node
- Depth first search (DFS) to check whether the graph is still connected
- Add node

This routine helps to automatically detect checkpoint components.

4.4. Case Study

For a case study, we monitored a recently built 5-storey office building in the inner city of Munich, Germany. In regular time intervals, the building was captured by means of the photogrammetric methods as explained in Section 4.3.1. A snippet of a point cloud created by the procedure is depicted in Fig. 22 a). The accuracy of the points is in the range of one to a few centimetres. Only points that can be seen from three or more images are regarded. For co-registration 11 corresponding points were measured in the images and the model on building parts which were already built.

For the experiment, the model surfaces are split into raster cells with a raster size of $x_r = 10cm$. Points are extracted within the distance $\Delta d = 5cm$. As minimum requirement for S , a point density of 25 points per dm^2 (i.e. in one raster cell) is defined, with all points getting a maximum weight $w_i = 1$ and having the distance to the plane of $\delta d = 2.5cm$. The resulting threshold for S is 5. In Fig. 22 b) all model planes having at least 50 % raster cells with a value M larger S are marked in red are regarded as verified, all other are marked in blue which comprises the states unknown and not built.

Without an exact numerical analysis, the following statements about the quality of the results can be made. All planes that are marked as built (in red) are built on site, except for one, which has a formwork around. However, there are several existing building parts, which are not marked as verified. This has various reasons:

- The acquisition was insufficient, this holds here for the both columns on the right, where

not enough images were taken for 3D reconstruction.

- Occlusions: For the planes which are on the rear part it is obvious that they are not represented in the point cloud, since only images have been taken in front of the building. Another reason are disturbing objects like scaffolding, container, temporary walls or a tram shelter in front of the building.
- Objects, which are not represented in the model: In this building model the insulation in front of the ceiling slabs are represented, but not the concrete slab itself. Since the insulation was not yet installed the areas of the ceiling slabs are not verified, but this is correct in this case.

For the part marked in yellow in Fig. 22 b) an analysis based on the number of triangles which face to the acquisition positions in front of the building is performed. The results are shown in Figure 23. They show what was mentioned above: Nearly all triangles that are classified as built are detected correctly (user's accuracy for "built" is 92.3%), but there is a larger number of triangles that have been classified as not built, even if they already exist.

Triangles		Ground Truth		
		Built	Not built	
Classification	Built	24	2	26
	Not built	20	10	30
		44	12	60,7 % (overall accuracy)

Figure 23 Confusion matrix for the evaluation of the area marked with the yellow frame in Fig. 21 b).

As discussed in Section 4.3.2, additional information can help to identify objects that cannot be detected but must be present due to technological dependencies. Fig. 24 shows the corresponding precedence relationship graph for the monitored building in this case study. It has been generated by means of a spatial query language for Building Information Models (Borrmann, 2007; Daum and Borrmann, 2013). For the generation, the query language was used to select touching elements.

In a further refinement, the graph was produced by ordering the elements in their respective vertical position and by filtering for supporting components. The mentioned articulation points (Section 4.3.2) are clearly visible. In this case, these nodes represent the slabs of each floor, as they are crucial building elements that are necessary for all subsequent parts of the next floor. When a slab is detected and correctly identified, all predecessors in the precedence relationship graph are set to the status "built". Therefor a statement is possible, even for objects that were not identified through visual detection.

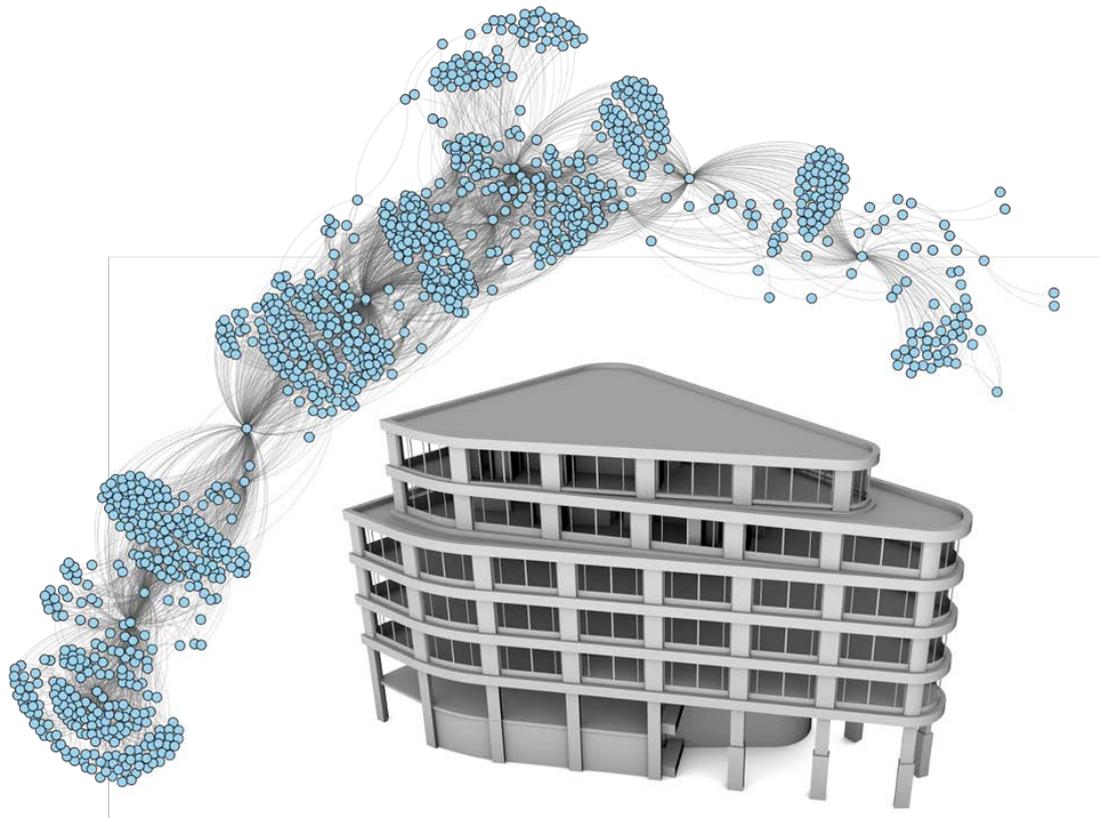


Figure 24 precedence relationship graph with corresponding BIM

4.5. Discussion and future work

This paper presents a concept for photogrammetric production of point clouds for construction progress monitoring and for the procedure for as-planned – as-built comparison based on the geometry only.

Additionally possibilities to improve these results using additional information provided by the BIM and accompanying process data are discussed. For the determination of the actual state, a dense point cloud is calculated from images of a calibrated camera. To determine the scale, control points are used, which requires manual intervention during orientation. The evaluation measure introduced for component verification detects built parts correctly but misses a larger number of them because of occlusion, noisy points or insufficient input data. Thus there is the need to extend this geometrical analysis by additional information and visibility constraints.

Future research will target at achieving greater automation of image orientation, e.g. by automatically identifiable control points. The as-planned vs. as-built comparison can be improved by additional component attributes provided by the BIM, such as the colour of the components. The automated generation of precedence relationship graphs will be addressed by a spatial query language approach.

The proposed methods and concepts presented in Section 3 introduce new possibilities for an enhancement in progress monitoring. Currently, the effort for photogrammetric techniques and object detection is still very high and needs to be investigated further to improve those methods. However, they can offer a variety of new possibilities for planners and on-site personnel, including:

- Time for photo-documentation can be reduced to a very low level, since the monitoring process is based on images.
- Automated process optimization can be pursued directly from the results of the process monitoring.

4.6. Acknowledgements

The presented research was conducted in the frame of the project „Development of automated methods for progress monitoring based on the integration of point cloud interpretation and 4D building information modeling“ funded by the German Research Foundation (DFG) under grants STI 545/6-1 and BO 3575/4-1.

We would like to thank Leitner GmbH & Co Bauunternehmung KG as well as Kuehn Malvezzi Architects for their support during the case study.

5. Acquisition and consecutive registration of photogrammetric point clouds for construction progress monitoring using a 4D BIM

Previously published as: Tuttas, S., Braun, A., Borrmann, A. et al., Acquisition and Consecutive Registration of Photogrammetric Point Clouds for Construction Progress Monitoring Using a 4D BIM, PFG (2017) 85: 3. <https://doi.org/10.1007/s41064-016-0002-z>

Abstract

In construction progress monitoring, a site must be observed many times throughout the construction process. Aiming at an automatic procedure for progress monitoring using 3D point clouds created from photogrammetric images, these have to be compared to a reference building model created from a Building Information Model (BIM). To accomplish this task, point clouds must be acquired at many acquisition dates and subsequently be co-registered with the building model. This paper proposes a co-registration system for consecutive registration that proceeds in three main steps: (I) Installation of photogrammetric control points and determination of their positions in the construction site coordinate system; (II) creation of an initial image block from these control points, which is co-registered with the building model; (III) registration of images from later acquisition dates based on the initial image block. Therefore, the images at consecutive time steps are registered by a Structure-from-Motion (SfM) process. The approach is evaluated in two scenarios with different acquisition techniques; namely, terrestrial with a hand-held camera and aerial with an Unmanned Aerial Vehicle (UAV). In total 14 acquisitions are investigated. In both scenarios the consecutive registration was successful. The registration accuracy is evaluated based on reconstructed planes which occur in several acquisitions and based on the control points, and was found to be ~ 3 to 5 cm, demonstrating the applicability of the proposed approach.

5.1. Introduction

5.1.1. Motivation

In contrast to industrial production, the processing steps on construction sites are extremely dynamic. Reasons for this are boundary conditions like the weather, which are difficult to predict, the strong dependency between single process steps and the lack of strict process sequences as they exist in assembly line production. Because of the many external dependencies the actual execution of the construction work usually deviates from the planning. Therefore, the monitoring of the construction process is important for recognizing delayed or premature construction steps. The deviations from the planning affect the overall organization, the schedule and the calculated costs and can lead to strong delays and a budget overrun. These effects can be mitigated by early detection of the deviations by a monitoring system. The classical manual recording of the sequential construction steps by a construction diary is labor-intensive, error-prone and allows no transparent evaluation of the productivity. Because of this, methods for the automatic area-wide capturing of the changing 3D structure of a construction site over time are developed in this DFG funded project. The goal is to compare photogrammetric acquisitions of the as-built state automatically with the as-planned state requested to a 4D Building Information Model (BIM). A BIM is a digital representation of a planned or built building. A BIM contains the 3D geometry, process information and schedule, semantic classification of building elements as well as their relations. The BIM stores information for all project partners during the whole life cycle of a building from the early planning and design phase over the construction phase, the operation and maintenance phase to renovation or demolition. In a BIM the object-oriented 3-dimensional geometric model is linked to temporal information like realization periods or predicted completion dates. Additionally, quantities and costs can be modeled. Based on the BIM, several analytic and simulation tools (energy consumption, cost estimation, visualizations, structural analysis, ...) can be applied. The available geometric and semantic data is also used for construction site monitoring and its results are used to keep the BIM up-to-date.

5.1.2. As-built as-planned comparison

For construction progress monitoring the BIM (Figure 25a) has to be compared to the actual state (Figure 25b) at the date of an acquisition t_i . Figure 26 shows the components and general procedure for the monitoring. The as-planned state of the construction is represented by the BIM. The as-built state is created from photogrammetric images, which are processed to create a 3D point cloud. In the following, this paper concentrates on the steps for generating an as-built point cloud to be used for the comparison process. The comparison between as-built and as-planned state is always performed when a new acquisition is available. The planned state is extracted for the respective date from the BIM. For each building element, it is checked if the point cloud proofs its existence. Based on the detected building elements, statements on the overall progress can be derived. The results are used to update the BIM, and particularly to optimize the schedule.

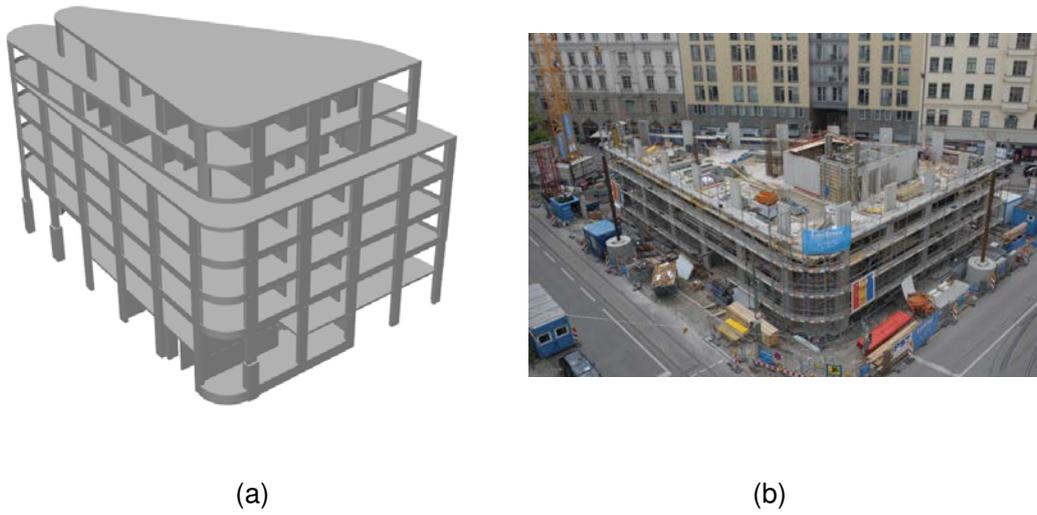


Figure 25 3D model of the building (a) and overview image in the construction phase (b).

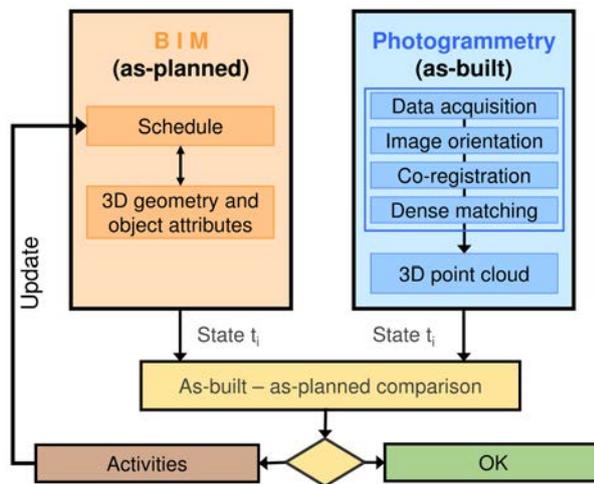


Figure 26 General procedure for construction progress monitoring using photogrammetric point clouds and BIM.

A result of the as-built as-planned comparison is shown in Figure 27. Figure 27a shows the evaluation at the end of construction, elements labeled in green are built before the planned date, gray elements are built in time, while red elements are built with delay. Figure 27b shows the comparison result for one acquisition date. The elements are labeled based on the schedule information and the ground truth. Details on the as-built as-planned comparison are described in (Tuttas et al., 2015), details on the labeling scheme are illustrated in (Braun et al., 2016).

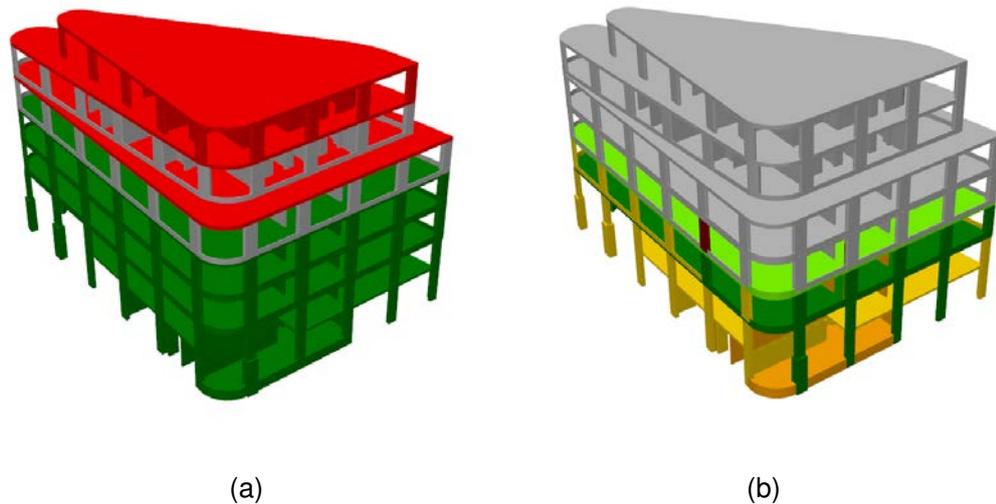


Figure 27 Example for results retrieved from an as-built as-planned comparison. (a) Building elements are labeled in green (before schedule), gray (in time) and red (delayed). (b) Detection result for one acquisition date, elements are labeled according to schedule and ground truth.

5.1.3. Consecutive registration

The abovementioned process must be fed with multiple consecutive acquisitions, and the reconstructed 3D point clouds have to be correctly aligned with high precision to the previous acquisitions and the BIM. In particular, co-registration errors should not lead to indicated building construction differences. In general, three co-registration tasks can be distinguished for consecutive registration as follows:

1. Co-registration of data acquired on the same date, for creating a consistent 3D-point cloud on that date. The relative orientation of the images has to be determined.
2. Co-registration of the complete point clouds (processed by Step (i)) acquired on different acquisition dates.
3. Co-registration of the point clouds and the building model.

The order and implementation of these tasks can be varied. The device positions can be determined from targets or markers with known coordinates (in a local or global coordinate system), which must be identified in the data. Corresponding features (points, lines, planes) are necessary for the co-registration of data acquired at the same time in a relative coordinate

system. A reconstructed scene is linked to the coordinate system of the building model by identifying corresponding points in the model and in the point cloud or images. This is not necessary if the control points are already specified in the model coordinate system. If a navigation unit is available, the data can be directly geo-referenced. Continuous monitoring requires a correct and stable co-registration at each date of the data acquisition. In the following sections we motivate and describe our approach for the co-registration of images on construction sites.

5.1.4. Structure of the paper

The remainder of this paper is structured as follows. In the following Section 5.2 an overview of works on construction site monitoring are given, whereby the focus is on the co-registration methods. Based on the review our contribution is outlined. Section 5.3 shows the procedure for the generation of dense as-built point clouds based on oriented images. Section 5.4 describes our approach for the co-registration. Section 5.5 introduces the two test sites used for evaluation. Section 5.6 presents the results on the test sites considering the success rate and geometric accuracy. The results are discussed in Section 5.7 and future work is proposed in Section 5.8.

5.2. State of the art and contribution

In general, works on construction site monitoring can be distinguished by the sensor technology which is used. Common sensors for construction site monitoring are laser scanners (Bosché et al., 2009, 2010; Bosché, 2012; Golparvar-Fard et al., 2011; Kim et al., 2013a; Maalek et al., 2015; Turkan et al., 2013; Turkan, 2012; Turkan et al., 2014) and cameras (Golparvar-Fard et al., 2011, 2015, 2011; Kim et al., 2013b; Rankohi and Waugh, 2015; Son and Kim, 2010). An essential step in current research on construction monitoring is co-registration between the building model and a point cloud. Registration is most commonly performed by Iterative-Closest-Point (ICP) algorithm (Besl and McKay, 1992) or some variant of it (Pătrăucean et al., 2015).

The co-registration of data from laser scanning is addressed in the following works: (Bosché et al., 2010) used ICP to co-register a laser-scanned point cloud with the mesh of a building model. (Turkan et al., 2013; Turkan, 2012; Turkan et al., 2014) adopted the same approach to progress monitoring. In a later study a modified ICP algorithm based on corresponding planes is presented, but only for coarse registration (Bosché, 2012). It is suggested to fit planes to the point cloud and assign them to corresponding planes in the model. Also (Kim et al., 2013a) use an adapted ICP to register each individual component after a coarse registration step. The ICP approach of (Tang and Rasheed, 2013) rejects points with an overly large data-model difference. Laser scanning point clouds can also be co-registered with a model using targets. (Zhang and Arditi, 2013) adopted this approach but did not explain how the targets and BIM coordinate system were aligned.

The co-registration of image data on construction sites has been addressed in the following works: In Han et al. (2015) and Karsch et al. (2014) corresponding points in a reference camera and the BIM were manually selected for registration. In (Golparvar-Fard et al., 2011) corresponding points are also manually selected in an image and the BIM. They also suggest (but did not show) the alternative use of surveying points. (Kim and Kano, 2008) used fixed cameras whose positions were determined by direct surveying with a total station. An example of direct georeferencing using Unmanned Aerial Vehicle (UAV) and Global Navigation Satellite System (GNSS) data is given in (Zollmann et al., 2014). The accuracy of this approach depends on the GPS conditions. Positional accuracy below one meter is possible only by using Real Time Kinematic (RTK), which requires additional data for correction. (Son and Kim, 2010) use an ICP-based approach to co-register point clouds from a stereo camera system.

We assume that ICP can not be used to perform the co-registration. This has two reasons:

1. At the beginning of the construction site no, or only very few parts of the building are already existing, that means that there is just nothing to co-register with.
2. It can not be assured that the assumption inherent for applying ICP can be ensured on a dynamic scene like a construction site. For ICP it is assumed that the parts which have to be co-registered are the same, but there are two reasons why this cannot be assured: It can be neither guaranteed that the model represents the correct state for the current acquisition (since this is what shall be verified by the monitoring system) nor that consecutive point clouds have included a sufficient number of surfaces which are identical.

In this paper we investigate the consecutive acquisition and co-registration of photogrammetric point clouds for a large number of acquisition dates. Photogrammetric point clouds for construction site monitoring have so far only investigated by the group of Goldpavar-Fard et al. (Bae et al., 2015; Golparvar-Fard et al., 2011, 2015; Ham et al., 2016; Han et al., 2015; Karsch et al., 2014).

The approach for co-registration described in (Han et al., 2015) and (Karsch et al., 2014) is closest to ours. They propose to have a single anchor camera which is aligned manually based on the model. In contrast to this we propose to do a manual measurement of control points on the first acquisition date and align the images of the following acquisition date based on the acquisition of the first one. In (Han et al., 2015; Karsch et al., 2014) the mesh model is used to support the registration process. Our approach is designed to avoid this, because of the same reasons as we want to avoid ICP registration. In (Karsch et al., 2014) the accuracy of the co-registration is evaluated based on the camera positions but not on the deviations between model and point cloud. Also it is not shown how many acquisitions dates could be co-registered and what kind of construction stages were used. In contrast to other

works on construction site monitoring we evaluate the acquisition and co-registration on a large number of acquisitions in different stages of construction. Many works often show only one single acquisition or results only for the ground floor of the buildings. The studies which investigate a larger number of acquisition sets, e.g., (Turkan et al., 2013), perform an ICP at every acquisition date, differ from our tests as they are based on laser scanning data and have a lot of building elements available which are already finished.

5.3. Point-cloud-generation

This section describes the generation of a dense point cloud from oriented images as it is used to create the as-built point clouds in this paper. First, image pairs for dense stereo matching have to be selected. For that the following criteria are considered:

- maximum and minimum baseline
- maximum angle between camera axes
- maximum deviation from a right angle from the angle between baseline and the camera axes (avoids that cameras are located behind each others are used)
- maximum angle between baseline and x- or y- axis of the camera coordinate system (avoids that cameras are used having a strong tilt)

Every image is selected once as master image. For all master images all potential match images are selected based on these criteria. To avoid that a image receives two (or more) match images which are close together (i.e., below the minimum baseline), only this image is selected which has the lowest sum of the angle deviation criteria. For all remaining k match images SGM matching (Hirschmuller, 2008) is performed using LibTSGM (Rothermel et al., 2012) and k disparity maps are calculated. The resulting disparity maps are fused based on a scheme which mainly follows the one presented in (Rothermel et al., 2012). For every pixel of the undistorted master image, the disparities are interpolated from all disparity maps. Now for every pixel, k disparity values are available. An interval for the distance D from the camera center to the 3D-point is determined by adding and subtracting an uncertainty value s from the disparity value. For every pixel, the depth values are clustered into one group if the intervals are overlapping. For calculating the final depth, the cluster having the most entries is chosen. Only clusters are retained which have at least two depth values (i.e., the point will be calculated from at least three image rays). The final value for D and its accuracy σ_D are determined by a least-square adjustment as described by (Rothermel et al., 2012). The focal lengths as well as the image coordinates are fixed here for the estimation of D . The final

3D-point coordinates (X, Y, Z) are then calculated by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{R}_{\text{Cam} \rightarrow \text{Obj}} \cdot (\mathbf{n} \cdot D) + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (5.1)$$

with rotation matrix \mathbf{R} , unit vector \mathbf{n} from perspective center to pixel and camera position X_0, Y_0, Z_0 . By applying the law of error propagation, the accuracy of the coordinates are calculated, using the standard deviations estimated in the bundle block adjustment (for ω, ϕ, κ and X_0, Y_0, Z_0) and in the determination of the depth (D), respectively. As last step, the point clouds of all master images are fused and filtered.

For filtering a voxel grid with cell size v is created. First all points are removed which are the only one in a voxel grid cell. In the remaining cells only the best point is retained. For that the point generated from the most rays is selected. If there are more points having the same amount of rays, the point with the smallest value σ_D is chosen.

5.4. Co-registration procedure

5.4.1. Concept

The proposed registration approach proceeds in three major steps, as depicted in Figure 28:

1. Installation of photogrammetric control points and determination of their coordinates in the construction site coordinate system using the control points of the surveying network.
2. Creation of an initial bundle block incorporating these photogrammetric control points.
3. Acquisition of the image blocks at the required dates and integration to the existing image block of the previous acquisition date.

The tasks performed in each of these steps are outlined below:

(I) The photogrammetric control points must be positioned such that they are visible in the images of the initial block, and can be simultaneously measured by the total station and leveling. Ideally, the control points will be mounted where they are visible during the whole construction process, and are unlikely to be removed or destroyed in the construction activity. If possible, they should be installed outside the active area. Although the markers are required only in the first step of the approach, they provide important backup if the consecutive registration fails at a later acquisition date. The second task is the measurement itself. The photogrammetric

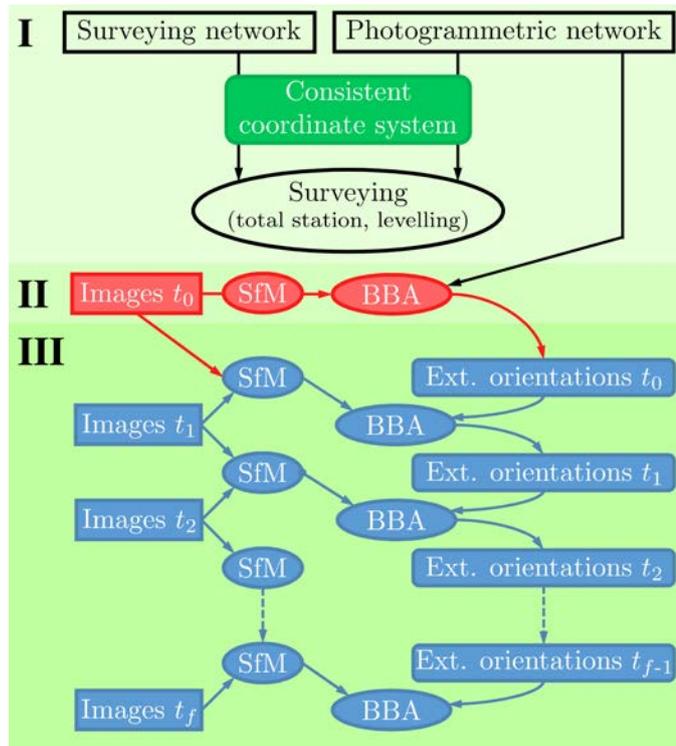


Figure 28 Schematic of the general procedure: I: Generation of a network of photogrammetric markers in the coordinate system of the construction site; II: The initial Bundle Block Adjustment (BBA) incorporates the coordinates of the photogrammetric markers as ground control points and the tie points created in the Structure-from-Motion (SfM) process; III: Consecutive registration of images based on the exterior orientations calculated in the previous step

markers are measured by surveying using the surveying points available at the construction site.

(II) The images for the initial block are captured within this step. For a stable determination of the coordinate system, the photogrammetric control points must be uniformly distributed with respect to the image block. The captured areas should include areas which will be unchanged during the construction process, or at least until the next acquisition. In this step, the image coordinates of the control points are (manually) measured once in all images.

(III) This step captures the images documenting the construction progress, recognizing that a) the changed areas are covered with overlapping images, and b) unchanged areas are also captured.

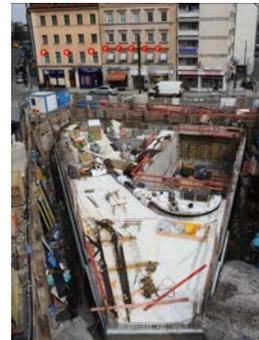
In steps II and III the image orientations incorporating the control point information must be determined. For this purpose a Structure-from-Motion (SfM) process, which recovers the exterior orientation (pose of the camera) and the sparse 3D scene geometry from (unordered) image sequences (Koutsoudis et al., 2015; Snavely et al., 2006; Wu, 2013a), and an additional (BBA) for integrating the control point coordinates is employed. The above tasks are processed on three types of points:

- control points - surveying network
- control points - photogrammetric marker and/or natural points
- tie points - based on point correspondences e.g., from SIFT (scale-invariant feature transform) features (Lowe, 2004)

Figure 29a shows two control points in an UAV image acquired from an approximate height of 25 m. The photogrammetric marker is clearly visible and measurable. The rough position of the surveying point is identifiable, but the actual marker, a chisel mark, cannot be recognized. Figure 29b shows "natural" tie points in an inner city scenario. For this kind of control points, features must be selected which can be clearly and unambiguously identified in the images from different viewing directions. Additionally they must either be measurable reflectorless with a total station or identifiable in a laser scanning point cloud.



(a) Photogrammetric marker (yellow arrow) and surveying point (green arrow)



(b) "Natural" control points, marked with red circles (window corners)

Figure 29 Types of control points

Tie points are generated using a feature detector (and descriptor) such as SIFT. These points are selected by inlier matching of the image pairs during the SfM process.

Since the characteristics of the individual construction site and the acquisition geometry can vary strongly, adaptations to the general approach may be necessary. The experiments in Section 5.5 can be distinguished by their scenario. The first test site (A) is an urban construction site with enclosing streets and buildings which is acquired by a hand held camera. In such terrestrial acquisition the main viewing direction is horizontal. Therefore, the markers must be mounted on vertical walls (such as surrounding building) so that they face the camera. If markers cannot be mounted, natural features such as window corners can be used. It has to be taken care that always unchanged buildings in the surrounding are acquired. The second test site (B) has a free surrounding and is acquired by an UAV. Here, where the viewing direction is towards the ground, the markers can be mounted on the ground, facing the camera. The monitored area should be wider than actually needed and include multiple man-made structures. This recommendation acknowledges that vegetation areas are not suitable for co-

registration of images from different acquisition dates, or even (in many cases) from images acquired at the same time.

5.4.2. Proposed procedure

Above, the requirements for the data acquisition are described. Now the processing steps of the resulting data are discussed. The surveying step (step I in Figure 28) provides the coordinates of the control points. Regardless of the scenario, final registration in the construction site coordinate system is performed by the SfM process and BBA at each acquisition date. Among the many available SfM implementations (Koutsoudis et al., 2015) the well established Visual Structure from Motion (VSfM) tool (Wu, 2013a) is selected here. As features the well known SIFT features are adopted there. The image coordinates of the tie points (obtained from the SIFT features correspondences), 3D-coordinates and image orientations in an arbitrary coordinate system are provided.

For the initial bundle (step II in Figure 28), the images of the first acquisition date are processed by SfM. To approximate the exterior orientation in all images, at least one image with known position and orientation and another images with known position in the construction site coordinate system are required. The initial bundle block adjustment (cf. Equation 5.2) requires the following input parameters (where n , n_P and n_{CP} denote the number of images, tie points and control points, respectively):

x_{ij} : Image coordinates comprising the manually measured coordinates of the control points and the SfM-determined image coordinates of all images in the initial image block.

X_P : Coordinates of all 3D points reconstructed during the SfM processing.

X_{CP} : Coordinates of the photogrammetric control points obtained by surveying.

T_i : Translation of image i , approximated from the SfM results and at least two images registered in the target coordinate system.

R_i : Rotation of image i , approximated from the SfM results and at least two images registered in the target coordinate system.

I_k : Interior orientation of camera k . These parameters can be fixed, if the camera has been calibrated in advance, or estimated during the bundle adjustment.

The initial block adjustment is given by the following, (where estimated parameters are marked with a hat):

$$\min \sum_{i=1}^n \sum_{j=1}^{n_P+n_{CP}} \|\mathbf{x}_{ij} - f(\hat{\mathbf{X}}_{j,P}, \mathbf{X}_{j,CP}, \hat{\mathbf{T}}_{i,t_0}, \hat{\mathbf{R}}_{i,t_0}, \mathbf{I}_k)\|^2 \quad (5.2)$$

The first adjustment determines the positions of the cameras in the target coordinate system. From this result a dense point cloud can be calculated (as described in Section 5.3) which is co-registered with the building model and can directly be used for the as-built - as-planned comparison. It is intended to register images of the next acquisition date in the same coordinate system. However, the control points are not measured manually in the subsequent images. Instead, the procedure searches for correspondences among images of the previous, registered, and new image sets.

To this end, the images acquired at times $t-1$ and t are integrated in a single SfM process. If there are sufficient images with features existing at both acquisition dates, the exterior orientations of all images can be determined in a common coordinate system. As the orientations of the previous acquisition date are known, the exterior orientations of the new images can be calculated. In the final bundle adjustment, the exterior orientations of the previous step are fixed to ensure a consistent coordinate system. They are selected based on their estimated accuracy in the previous bundle adjustment.

The input parameters to the consecutive bundle block adjustments (cf. Equation 5.3) are defined below:

x_{ij} : Image coordinates of the common SfM of two acquisition dates.

X_P : Coordinates of all 3D points reconstructed during the SfM processing.

$T_{i,t-1}$: Fixed translation of image i of the previous acquisition, obtained in the previous bundle block adjustment.

$R_{i,t-1}$: Fixed rotation of image i of the previous acquisition, obtained in the previous bundle block adjustment.

$T_{i,t}$: Translation of image i , approximated from the SfM results and the known orientations of the previous bundle adjustment.

$R_{i,t}$: Rotation of image i , approximated from the SfM results and the known orientations of the previous bundle adjustment.

I_k : Interior orientation of camera k .

Consecutive blocks are then adjusted as:

$$\min \sum_{i=1}^n \sum_{j=1}^{n_P} \|\mathbf{x}_{ij} - f(\hat{\mathbf{X}}_{j,P}, \mathbf{T}_{i,t-1}, \mathbf{R}_{i,t-1}, \hat{\mathbf{T}}_{i,t}, \hat{\mathbf{R}}_{i,t}, \mathbf{I}_k)\|^2 \quad (5.3)$$

In the above procedure, images are added to the SfM system in an unordered manner and correspondences are searched for every image pair. The cameras in the experiments have been calibrated in advance and are fixed for all BBA.

When different parts of the construction site are covered by images with insufficient overlap, the whole process can be separated into multiple blocks that are processed independently. In this case, the process of Figure 28 is performed multiple times between two consecutive time steps. This requires only that all images are oriented with respect to the same initial block. If two consecutive acquisition dates cannot be connected, the images can still be registered by one of the following methods: If (at least three, but better more) control points are visible in the image (sub-block), their image coordinates can be manually measured and the position of the images in the construction site coordinate system can be calculated. Alternatively, these images can be registered using an image block of an earlier date, which may provide more images.

5.5. Experiments

The proposed approach was validated in experiments at two test sites, labeled A and B. Subsections 5.5.1 and 5.5.2 describe the geometric and temporal aspects of the test sites, respectively. The geometric aspects of each experiment are characterized by the selected acquisition device and geometry, a short description of the test site, types of control points and realization of the initial state. Subsection 5.5.3 describes the setup of control points network at Test Site B.

5.5.1. Geometric aspects

Test Site A - Karlstrasse

Acquisition: Terrestrial, hand-held SLR camera

Camera parameters: $c = 24$ mm, sensor size 36 x 24 mm (4256 x 2832 pixel)

Description: Test Site A is an inner city construction site with limited surrounding space. Streets and buildings enclose all sides of the site. The images were taken with a hand-held camera from the sidewalk of the enclosing streets and from elevated positions on the crane and surrounding buildings.

Control points: The control points were extracted from a very dense point cloud acquired by laser scanning during the earlier excavation phase. As control points window corners of the surrounding buildings (cf. Figure 29b) as well as four rotatable laser scanner targets are used. The latter are used for the evaluation shown in Table 7.

Initial state: The initial state comprises data acquired on June 27, 2013. These images best fulfilled the initial-state criterion, as they were obtained from all locations at the test site. All other acquisition dates, both backward and forward in time, were registered from this initial time position.

Typical camera positions and images of this test site are shown in Figure 30.

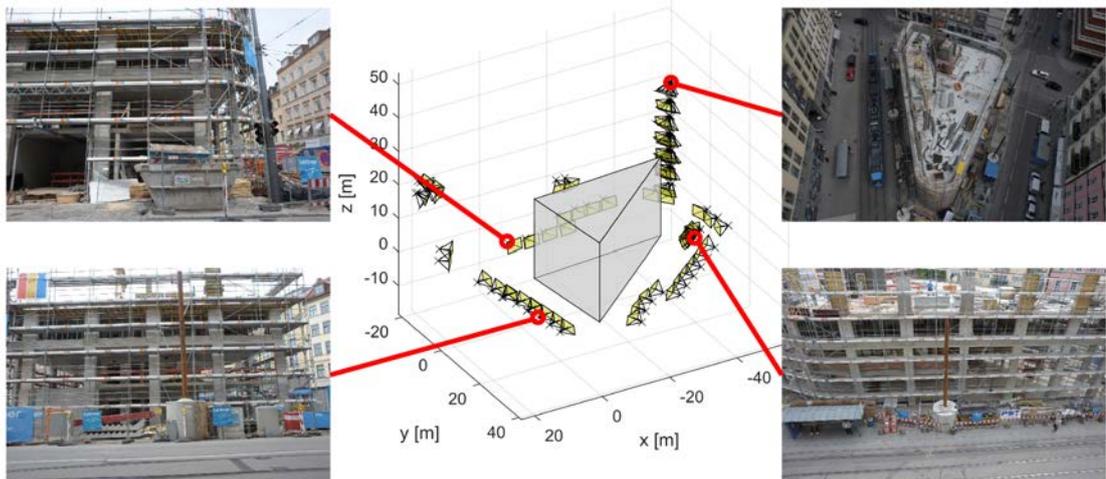


Figure 30 Camera configuration and example images on Test Site A with hand-held camera. The gray block represents a coarse model of the building. Images are acquired on the ground level and are supplemented by images from elevated positions on adjacent buildings and the crane.

Test Site B - Haus für Kinder

Acquisition: UAV, nadir and oblique images (Images were acquired as a photogrammetric block with overlapping stripes at two different heights. Additionally, oblique images were taken)

Camera parameters: $c = 18$ mm, sensor size 23.5×15.6 mm (6000 x 4000 pixel)

Description: Test Site B is a peripheral construction site with neighboring houses at one side and grassland on the other.

Control points: Photogrammetric markers were installed around the construction site and measured by a total station and leveling.

Initial state: The initial state was a combination of the first two acquisitions.

Typical camera positions and images of this test site are shown in Figure 31.

5.5.2. Temporal aspects

Figure 32 presents the timelines of the acquisition dates at both sites. At Test Site A (Figure 32a), images were acquired approximately monthly. At Test Site B (Figure 32b), images

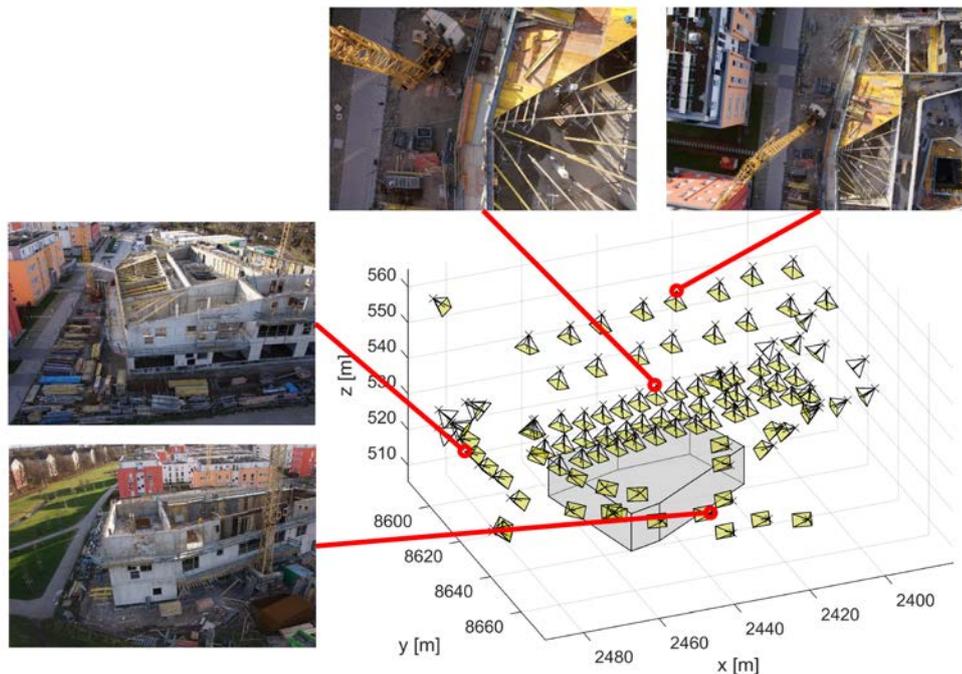


Figure 31 Camera configuration and example images on Test Site B with UAV acquisition. The gray block represents a coarse model of the building. Images are acquired in nadir view from two different heights and in oblique view.

were also acquired approximately monthly but with a longer interval between the last two acquisitions.



Figure 32 Acquisition dates at a) Test Site A and b) Test Site B

5.5.3. Control points network

Figure 33 shows the geodetic network of Test Site B, together with a 2D plan of the building. The figure depicts also the position of the surveying points and photogrammetric markers in the network, along with the point numbers (see Table 11). The network was split into a positional and a height network.

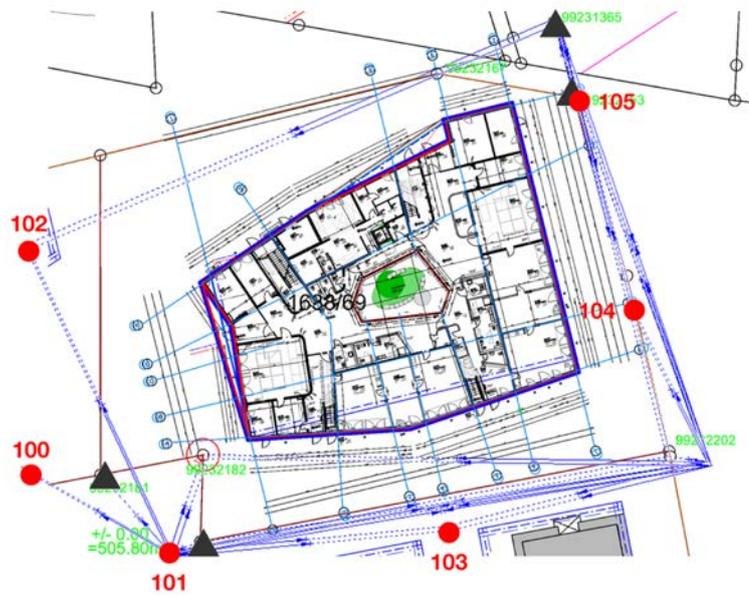


Figure 33 Surveying Network and 2D plan of Test Site B. Red dots: photogrammetric control points; black triangles: surveying points; blue lines indicate the tachymetric measurements

5.6. Results

5.6.1. Remarks on results

The results are visualized and evaluated based on different criteria. For both test sites there are the sections "*Estimated point accuracy*", "*Registration quality*" and "*Control points network*".

The first section gives the estimated accuracies of the dense point clouds estimated as shown in Section 5.3. For all evaluations the point clouds are filtered using a voxel cell size of $v = 1$ cm.

The second section shows the results of the co-registration, which is divided in the parts success and accuracy. First it is shown for which consecutive acquisition dates the registration was successful. For this purpose, the timelines shown in Section 5.5 are complemented with the results of the consecutive registration. It is depicted if the registration was successful (green), not successful (red) or partly successful (yellow). In the partially successful cases, the registration was successful for most of the images, but some images required manual interaction; for example, by measuring the image coordinates of some control points in non-registered images. The initial steps are indicated by blue markers at their respective dates. The dates of correctly registered images are marked in black. Light blue marks indicate that the images were successfully registered at that date, but required some manual intervention. Secondly, the accuracy is evaluated. For that planes of building elements are selected from all acquisition dates which are built according to the ground truth. Furthermore only planes are selected which are at least 3 m^2 large and are covered at least by 80% with points. A plane is fitted to the points extracted around the model plane. The maximum orthogonal distance between this plane and the model plane within the area of the building element surface is calculated. Also the deviation of the plane normal to the normal of the model plane is determined. Finally planes are selected which occur in at least two acquisition dates. Since the model planes can also have a deviation from the true position also the variation of the distance within the different acquisition dates is evaluated as measure for the stability of the co-registration.

The third section evaluates the accuracy based on the control points. For verifying the stability of the photogrammetric control points and the registration solution the image coordinates are measured manually in the images acquired on an acquisition after some steps of the consecutive registration. The 3D coordinates of the control points are then estimated by BBA (Equation 5.4) and compared to the coordinates which were the input to the initial bundle block.

$$\min \sum_{i=1}^n \sum_{j=1}^{n_P+n_{CP}} \|\mathbf{x}_{ij} - f(\hat{\mathbf{X}}_{j,P}, \hat{\mathbf{X}}_{j,CP}, \hat{\mathbf{T}}_{i,t_n}, \hat{\mathbf{R}}_{i,t_n}, \mathbf{I}_k)\|^2 \quad (5.4)$$

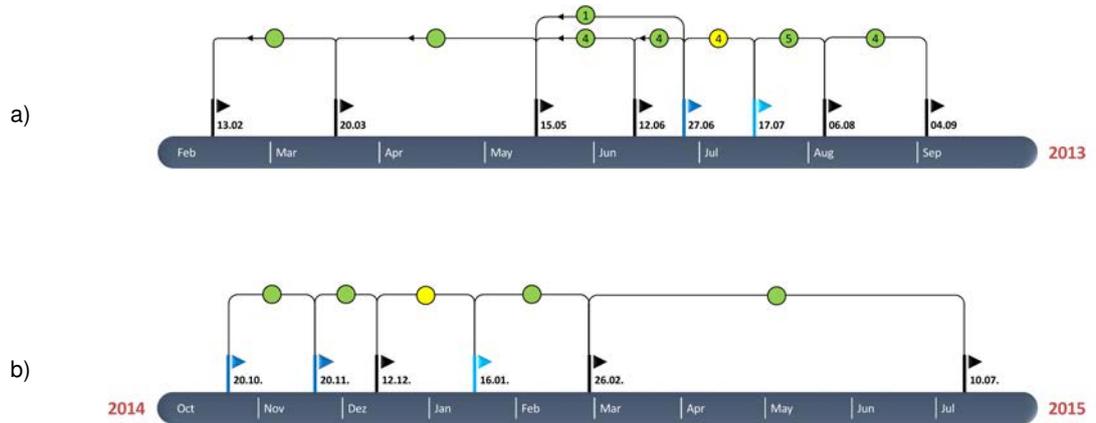


Figure 34 Results of consecutive registration for a) Test Site A and b) Test Site B. The symbols and their colors are explained in Section 5.6.1

5.6.2. Test Site A

Estimated point accuracy

In Figure 35 point clouds and as-planned model for three selected acquisition dates are shown. In Table 4 the estimated accuracy for all points are given. The results are split based on the number of stereo pairs the points are reconstructed from (number of rays = number of pairs + 1). For Test Site A the mean value is around 3 cm for the 3D point accuracy and around 1 cm for the depth accuracy.

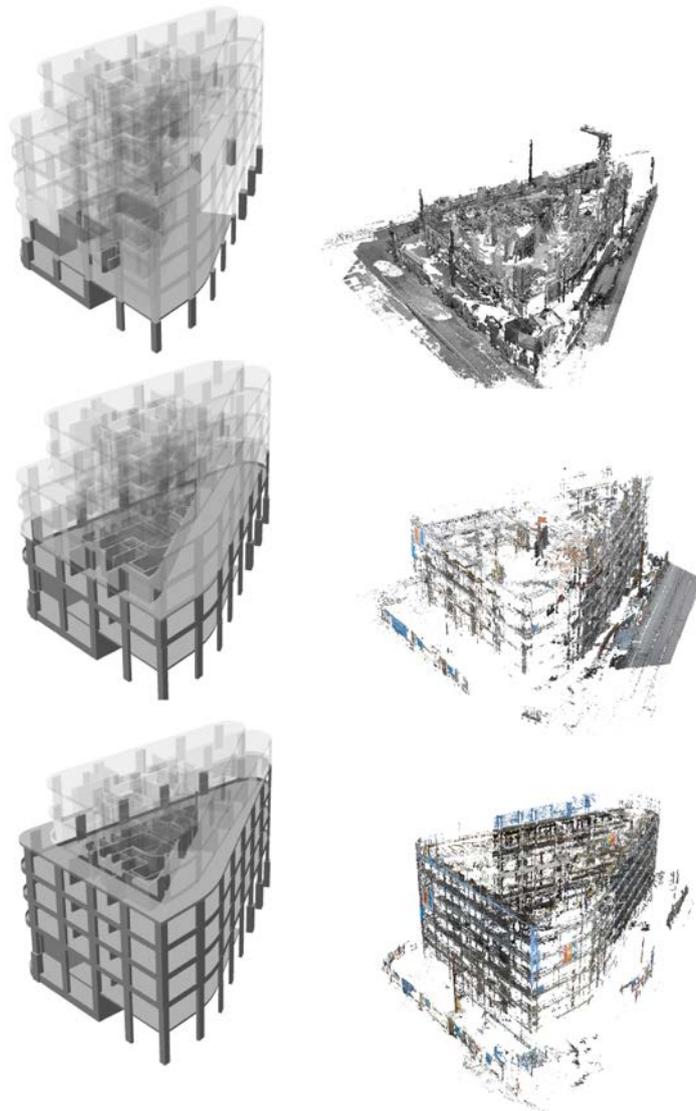


Figure 35 Point cloud of three selected acquisition dates (right) and the corresponding as-planned model (left) for Test Site A.

Registration quality

Figure 30 shows the camera positions at the initial state. The results of the consecutive registration are visualized in Figure 34a. As explained in Section 5.5.1 (*Initial state*), the initial step was not the first step for Test Site A. In this case the experiment was also performed in the backward direction; that is, an earlier acquisition date was registered from the following

Pairs	mean [cm]		median [cm]		std [cm]		number of points
	3D	<i>D</i>	3D	<i>D</i>	3D	<i>D</i>	
all	2.6	1.2	1.4	0.7	2.9	1.5	59 555 741
2	2.9	1.1	1.4	0.6	3.1	1.5	33 638 156
3	2.2	1.4	1.4	0.9	2.4	1.5	21 837 548
4	3.0	1.3	1.7	0.8	3.3	1.3	3 227 209
5	3.5	1.3	2.9	1.1	3.6	1.1	852 557
>5	1.3	0.7	1.3	0.6	0.2	0.1	271

Table 4 3D point accuracy (3D) and depth accuracy *D* for the dense point clouds of Test Site A

acquisition date. In the figure, this situation is marked by a backward-pointing arrow. Also, at this test site, the processing was split into five sub-blocks, indicated by the numbers in the status circles. One of the blocks was skipping one of the acquisition dates. The initial state and the two earliest states are processed as a single image block.

The accuracy of the co-registration is assessed with the values given in Table 5. The values are created as described in Section 5.6.1. The RMS of the plane fits ranges from 0.5 to 1.5 cm and is 1 cm on average. The mean variation, i.e., the difference in the maximum distance to the reference plane, is 2.9 cm, the mean normal deviation is 0.6° .

Control points network

Four control points are measured in the images acquired on 13 February (i.e., the first acquisition date) and compared to the control point coordinates as they were introduced to the initial BBA on 26 July. The results are shown in Table 7.

15.05.	max deviation [cm]					04.09.	max variation [cm]
	12.06.	27.06.	17.7.	06.08.			
0.7	-	-	1.8	-	-	-	1.0
2.4	-	1.2	-	-	-	-	1.2
-	-	-	1.9	-	-3.5	-	5.4
-	-	-	-1.9	-4.5	-3.8	-	2.6
-	1.2	-	-2.0	-4.2	-	-	5.4
-	-	-	-	2.5	2.8	-	0.3
-	-	0.9	-	3.5	2.0	-	2.6
-	-	0.8	-	3.9	3.0	-	3.1
-	-	2.1	2.6	3.5	2.9	-	1.4
-	-	2.0	2.7	3.2	1.7	-	1.5
-	0.3	-1.3	0.9	-2.1	-3.4	-	4.3
-	-0.7	-1.2	0.9	-1.1	-4.3	-	5.2
-	0.5	-1.1	-1.2	-1.4	-	-	2.0
-	0.8	-0.7	-0.6	-0.7	-5.5	-	6.4
-	-	-	-0.7	-	-3.0	-	2.4
-	-	-	-	-3.6	-1.5	-	2.1
-	-	-1.0	-	-	0.7	-	1.7
-	-	1.1	-	-	2.7	-	1.6
-	-	1.5	2.8	-1.2	-1.5	-	4.4
-	-	-	1.3	-	-1.4	-	2.7
-	-	-	-3.0	-3.2	-	-	0.2
-	-	-	-1.7	2.3	-	-	4.0
-	-	-	-	-5.7	-2.7	-	3.0
-	-	-	-	-5.5	-2.6	-	2.9
-	-	-	-	-0.9	1.4	-	2.3
-	-	-	-	-1.9	2.5	-	4.4

Table 5 Deviations (variation) between point cloud and as-planned model for Test Site A. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1. No reappearing planes could be reconstructed at the first two acquisitions.

deviation [deg]						max deviation
15.05.	12.06.	27.06.	17.7.	06.08.	04.09.	[deg]
0.285	-	-	0.651	-	-	0.651
0.503	-	0.470	-	-	-	0.503
-	-	-	0.658	-	0.810	0.810
-	-	-	0.384	0.559	0.532	0.559
-	0.357	-	0.527	0.878	-	0.878
-	-	-	-	0.469	0.868	0.868
-	-	0.499	-	0.444	0.318	0.499
-	-	0.078	-	0.470	0.267	0.470
-	-	0.507	0.321	0.357	0.621	0.621
-	-	0.564	0.382	0.498	0.670	0.670
-	0.213	0.258	0.241	0.407	0.652	0.652
-	0.425	0.480	0.280	0.344	0.593	0.593
-	0.141	0.431	0.153	0.049	-	0.431
-	0.476	0.069	0.295	0.083	0.247	0.476
-	-	-	0.377	-	0.288	0.377
-	-	-	-	0.830	0.248	0.830
-	-	0.154	-	-	0.236	0.236
-	-	0.308	-	-	0.681	0.681
-	-	0.762	0.124	0.364	0.457	0.762
-	-	-	0.098	-	0.473	0.473
-	-	-	0.890	0.890	-	0.890
-	-	-	0.581	0.628	-	0.628
-	-	-	-	0.629	0.444	0.629
-	-	-	-	0.649	0.667	0.667
-	-	-	-	0.254	0.444	0.444
-	-	-	-	0.701	0.668	0.701

Table 6 Deviations (degrees) between point cloud and as-planned model for Test Site A. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1. No reappearing planes could be reconstructed at the first two acquisitions.

P-Nr.	sx [mm]	sy [mm]	sz [mm]	Rays	dX [mm]	dY [mm]	dZ [mm]	2D [mm]	3D [mm]
1	0.7	1.2	0.6	65	50.0	22.2	-8.8	54.7	55.4
2	1.2	1.0	0.8	62	30.8	21.8	-8.4	37.7	38.6
3	1.1	1.0	0.7	68	55.2	-1.3	-18.6	55.2	58.2
4	0.9	1.0	0.7	55	23.8	-3.2	-17.4	24.0	29.6

Table 7 The standard deviations and number of rays for the control points calculated from images acquired on 13 February on Test Site A and the deviations from the geodetic measurement.

5.6.3. Test Site B

Estimated point accuracy

In Figure 36 point clouds and as-planned model for three selected acquisition dates are shown. In Table 8 the estimated accuracy for all points are given. The results are split based on the number of stereo pairs the points are reconstructed from (number of rays = number of pairs + 1). For Test Site B the mean value is around 1 cm for the 3D point accuracy and around 0.4 cm for the depth accuracy.

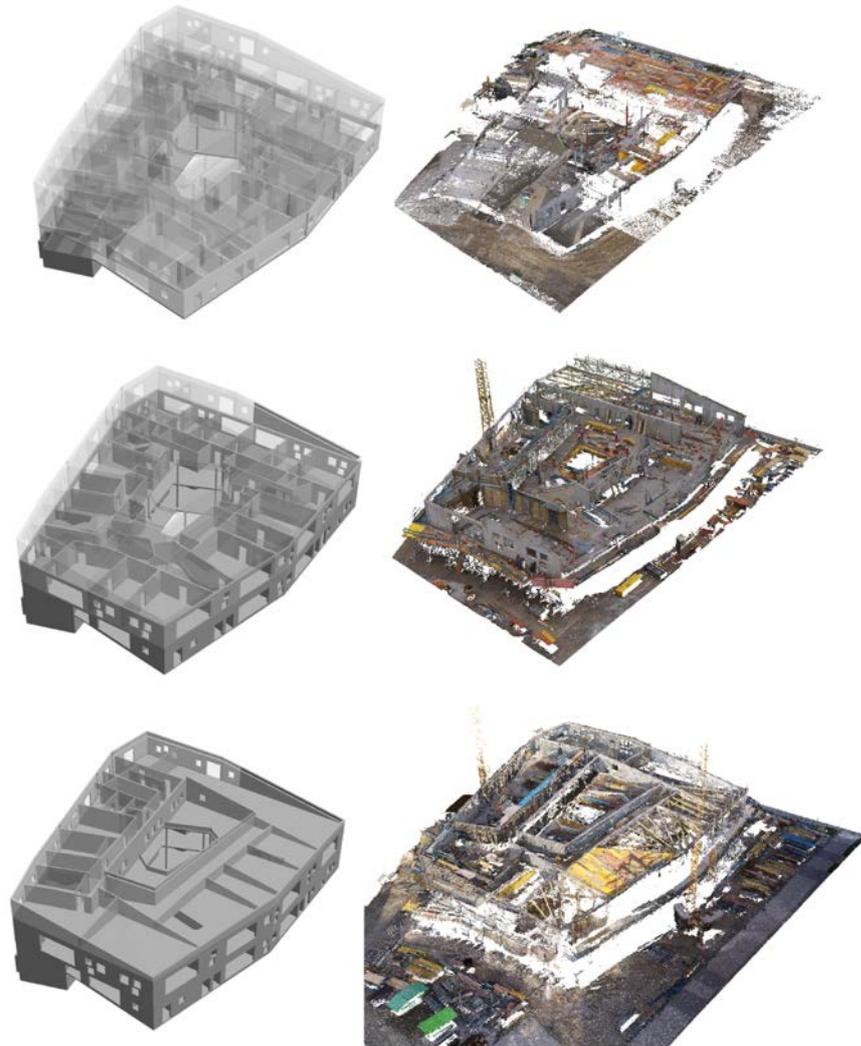


Figure 36 Point cloud of three selected acquisition dates (right) and the corresponding as-planned model (left) for Test Site B.

Registration quality

The camera positions for one of the acquisition dates are shown in Figure 31. At this test site, the consecutive registration was successful at all acquisition dates (Figure 34b). A few images acquired on 16 January required manual intervention for registration.

The accuracy of the co-registration is assessed with the values given in Table 9 and Table 10. The values are created as described in Section 5.6.1. The RMS of the plane fits ranges from

Pairs	mean [cm]		median [cm]		std [cm]		number of points
	3D	<i>D</i>	3D	<i>D</i>	3D	<i>D</i>	
all	1.0	0.4	0.8	0.3	0.7	0.4	242 590 539
2	1.0	0.3	0.8	0.2	0.7	0.4	120 087 989
3	1.0	0.5	0.9	0.3	0.6	0.5	86 094 624
4	1.0	0.4	0.8	0.3	0.6	0.5	26 682 896
5	0.8	0.2	0.6	0.2	0.4	0.2	6 874 435
>5	0.6	0.2	0.6	0.1	0.3	0.1	2 850 595

Table 8 3D point accuracy (3D) and depth accuracy *D* for the dense point clouds of Test Site B

0.6 to 1.4 cm and is 1 cm on average. The mean variation, i.e., the difference in the maximum distance to the reference plane, is 2.2 cm, the mean normal deviation is 0.3°.

	max deviation [cm]				max variation	
	20.10.	20.11.	12.12.	16.01.	26.02.	[cm]
	-	-	-	3.4	4.2	0.8
	-	-	-	2.3	2.7	0.4
	-	3.2	-	2.5	1.9	1.4
	-	-1.6	3.0	2.3	3.3	4.9
	-	-	-	2.1	1.9	0.2
	-	-	-1.5	-2.4	-1.9	0.8
	0.9	-	-	2.6	4.2	3.2
	1.9	-2.0	1.6	2.5	2.7	4.7
	-	-	3.7	3.8	4.3	0.6
	-	-	2.7	1.8	3.7	1.9
	-	-	-	2.1	1.1	0.9
	-	-	-	2.3	1.6	0.7
	-	-	3.1	1.3	1.2	2.0
	-	-	1.2	-1.4	2.3	3.7
	-	-	-	-1.9	1.7	3.6
	-	-3.4	1.8	-1.2	1.0	5.2

Table 9 Deviations [cm] between point cloud and as-planned model for Test Site B. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1.

Control points network

Six control points are measured in the images acquired on 26 February (i.e., the second last acquisition date, since some of the control points have been already destroyed on the last acquisition) and compared to the control point coordinates as they were introduced to the initial BBA. The results are shown in Table 11 which lists standard deviation of the estimated

deviation [deg]					max deviation
20.10.	20.11.	12.12.	16.01.	26.02.	[deg]
-	-	-	0.338	0.213	0.338
-	-	-	0.217	0.397	0.397
-	0.189	-	0.217	0.113	0.217
-	0.142	0.223	0.048	0.142	0.223
-	-	-	0.299	0.352	0.352
-	-	0.014	0.069	0.056	0.069
0.196	-	-	0.221	0.311	0.311
0.150	0.206	0.124	0.195	0.218	0.218
-	-	0.218	0.264	0.210	0.264
-	-	0.044	0.387	0.273	0.387
-	-	-	0.203	0.339	0.339
-	-	-	0.309	0.180	0.309
-	-	0.269	0.151	0.022	0.269
-	-	0.328	0.357	0.235	0.357
-	-	-	0.232	0.305	0.305
-	0.446	0.443	0.176	0.259	0.446

Table 10 Deviations [deg] between point cloud and as-planned model for Test Site B. Planes are used which are visible at least on two acquisitions dates. The calculation of the values is described in detail in Section 5.6.1.

coordinates and the number of rays from which they were calculated. This table shows also the differences from the initial values. The point numbers of the photogrammetric control points are indicated in Figure 33. For most of the points, the horizontal and vertical errors ranges from 1 to 2 cm and up to 3 cm, respectively.

P-Nr.	sx [mm]	sy [mm]	sz [mm]	Rays	dX [mm]	dY [mm]	dZ [mm]	2D [mm]	3D [mm]
100	2.0	2.1	6.2	11	-16.7	-6.3	-31.0	17.9	35.8
101	1.8	2.5	4.7	18	-4.7	-2.7	-12.4	5.5	13.5
102	2.2	3.4	6.4	12	-19.0	2.3	-4.8	19.1	19.7
103	2.2	2.6	4.0	16	-11.5	3.0	-20.8	11.9	23.9
104	1.4	1.4	2.5	22	12.9	13.6	6.8	18.7	19.9
105	3.5	3.8	5.5	12	-30.1	2.7	33.1	30.2	44.8

Table 11 The standard deviations and number of rays for the control points calculated from images acquired on 26 February on Test Site B and the deviations from the geodetic measurement.

5.7. Discussion

The proposed approach for co-registration of point clouds and the building model was successfully applied and validated on real construction sites under different acquisition geometries and scenarios. Better results were achieved at Test Site B which has a slightly higher accuracy and no sub-blocks were constructed (as at Test Site A). Based on the evaluation a registration accuracy of $\sim 3\text{-}5\text{ cm}$ can be expected. The results were determined based on, in total, eleven acquisition dates on two different construction site using 26 (Test Site A) and 16 (Test Site B) planes as reference. The accuracy is sufficient for the verification of the existence of building elements, but not for the detection of small inaccuracies of the execution of construction work.

The accuracy of the dense point is estimated in the range from 1 to 2 cm. The accuracy for the points with few rays (3 to 4) seems to be estimated too positive, possibly because of the small sample of observations for the calculation of σ_D . The deviations of the control points are in the range from 2 to 5 cm and show also larger σ values on Test Site A. In general these deviations can be sourced from the following four errors:

1. Manual measurements of photogrammetric markers, which introduce (sub-pixel) errors in the image coordinates.
2. Uncertainty because of the ray intersection.
3. Additive accumulation of errors during consecutive registration.
4. Instability of ground control points

On Test Site B the larger values in z-direction point to the typical depth error of the ray intersection (case ii). On Test Site A no clear dependencies are identifiable.

The successful registrations at Test Site B can be ascribed to the free area around the construction site, which provides a sufficiently large unchanged area throughout the construction process, and to the UAV acquisition, which provides both nadir and oblique images. The latter allows complete coverage of the construction site and large overlap of images. It was even possible to connect two states having more than four months in between. The limitations of UAVs are related to safety issues.

Namely, the UAV must maintain a safe distance from all towering objects (chiefly the cranes, but also the surrounding buildings). The safety regulations also forbid the UAV from flying over roads, limiting its use in some scenarios, for example at Test Site A. Inaccessibility of observation positions are a main problem, especially for the terrestrial acquisition (Test Site A). This results in the problem of splitting into different sub-blocks at this test site. Although consecutive acquisition dates could be connected, not all of the images acquired on the same date had enough corresponding features. Therefore, the data were processed as four to five indi-

vidual sub-blocks. The advantages and disadvantages of the different acquisition techniques are also discussed in (Tuttas et al., 2016).

5.8. Future Work

In future work the following outstanding issues are addressed: Here, different acquisition techniques are employed for the test sites. It has to be investigated how they can be combined to improve the results in terms of a more stable co-registration and an improved reconstruction of the point cloud. The matching complexity in the two-view matching step of the SfM process could be reduced: To this end, the number of images in the co-matching could be reduced by choosing only the relevant key images from the previous acquisition date, which contain minimal or zero changes. These images could be manually or automatically defined using the orientation and progress information of the BIM. This information would reveal the image areas that are unlikely to change, or that never change because they are outside the active areas. The matching pairs could also be reduced by exploiting the position and orientation information from navigation data, e.g. provided by the navigation unit of the UAV.

The co-registered point clouds are and will be used for reconstruction of scaffolding components (Xu et al., 2015, 2016) and progress monitoring (Braun et al., 2017; Tuttas et al., 2014, 2015).

Acknowledgements

This work is supported by the German Research Foundation (DFG) under grants STI 545/6-1 and BO 3575/4-1. We like to thank Leitner GmbH & Co Bauunternehmung KG and Kuehn Malvezzi Architects (Test Site A) and Baureferat H5, Landeshauptstadt München, Baugesellschaft Mickan mbH & Co KG, h4a Architekten, Wenzel + Wenzel and Stadtvermessungsamt München (Test Site B) for their support during the case studies.

The authors also like to thank Konrad Eder for his support during data acquisition.

6. Improving progress monitoring by fusing point clouds, semantic data and computer vision

Previously published as: Braun, A., Tuttas, S., Stilla, U., Borrmann, A., Improving progress monitoring by fusing point clouds, semantic data and computer vision, Automation in Construction 116, 2020, DOI: 10.1016/j.autcon.2020.103210

Abstract

Automated construction-progress monitoring enables the required transparency for improved process control, and is thus being increasingly adopted by the construction industry. Many recent approaches use Scan-to/vs-BIM methods for capturing the as-built status of large construction sites. However, they often lack accuracy or are incomplete due to occluded elements and reconstruction inaccuracies. To overcome these limitations and exploit the rich project knowledge from the design phase, the authors propose taking advantage of the extensive geometric-semantic information provided by Building Information Models. In particular, valuable knowledge on the construction processes is inferred from BIM objects and their precedence relationships. SfM methods enable 3D building elements to be located and projected into the picture's 2D coordinate system. On this basis, the paper presents a machine-learning-based object-detection approach that supports progress monitoring by verifying element categories compared to the expected data from the digital model. The results show that, depending on the type of construction and the type of occlusions, the detection of built elements can rise by up to 50% compared to an SfM-based, purely geometric as-planned vs. as-built comparison.

6.1. Introduction

6.1.1. Automated progress monitoring

Construction progress monitoring is currently still performed mostly manually, in a laborious and error-prone non-automated process. To prove that all works have been completed as agreed contractually, all completed tasks must be documented and monitored. Complete and detailed monitoring techniques are required for large construction sites where the entire construction area and the number of subcontractors become too large for manual tracking to be efficient. Detecting possible deviations from the schedule provides a benchmark for the performance of the construction site. Regulatory matters add to the requirement of keeping track of the current status on the site. The ongoing establishment of building information modeling (BIM) technologies in the planning of construction projects facilitate the application of digital methods also in the execution phase. In an ideal implementation of BIM, all relevant information on materials, construction methods, and even the process schedule are interlinked. On this basis, it is possible to estimate project costs and project duration more precisely than with conventional methods (Hardin and McCool, 2015).

On top of the digitized construction design process, recent advancements for capturing the as-built geometry by laser scanning (Bosché and Haas, 2008a) or photogrammetry (Golparvarfar et al., 2009) allow using the resulting point cloud data to be compared against the as-planned model. Photogrammetry, in particular, has gained more attention with the broader availability of Unmanned Aerial Vehicles (UAVs), making this method more flexible in terms of camera positions (Lin et al., 2015). The main idea is not to use laser scanners but conventional camera equipment on construction sites to capture the current construction state ("as-built"). Since the acquisition from different perspectives is significantly faster than laser scanners, the building can be captured in a comprehensive manner with comparatively low effort. As soon as a sufficient number of images from different points of view are available, a 3D point cloud can be produced using Structure from Motion (SfM) methods (Wu, 2013a). Finally, the point cloud, representing one particular observation time-point, can be compared against the geometry of the Building Information Model.

6.1.2. Problem statement

Currently, the detection of built elements using SfM methods and other point-cloud-based approaches faces several challenges:

As-planned modeling vs. as-built construction

As introduced in Braun et al. (2016), the as-planned model is represented by a 4D building information model (see Figure 37 a)). All 4D construction processes are linked to their associated elements, allowing for statements regarding the expected construction state at any given observation time. As the relevant model and point cloud are co-registered, an initial detection algorithm can compare the model's geometry with the 3D information from the point cloud.

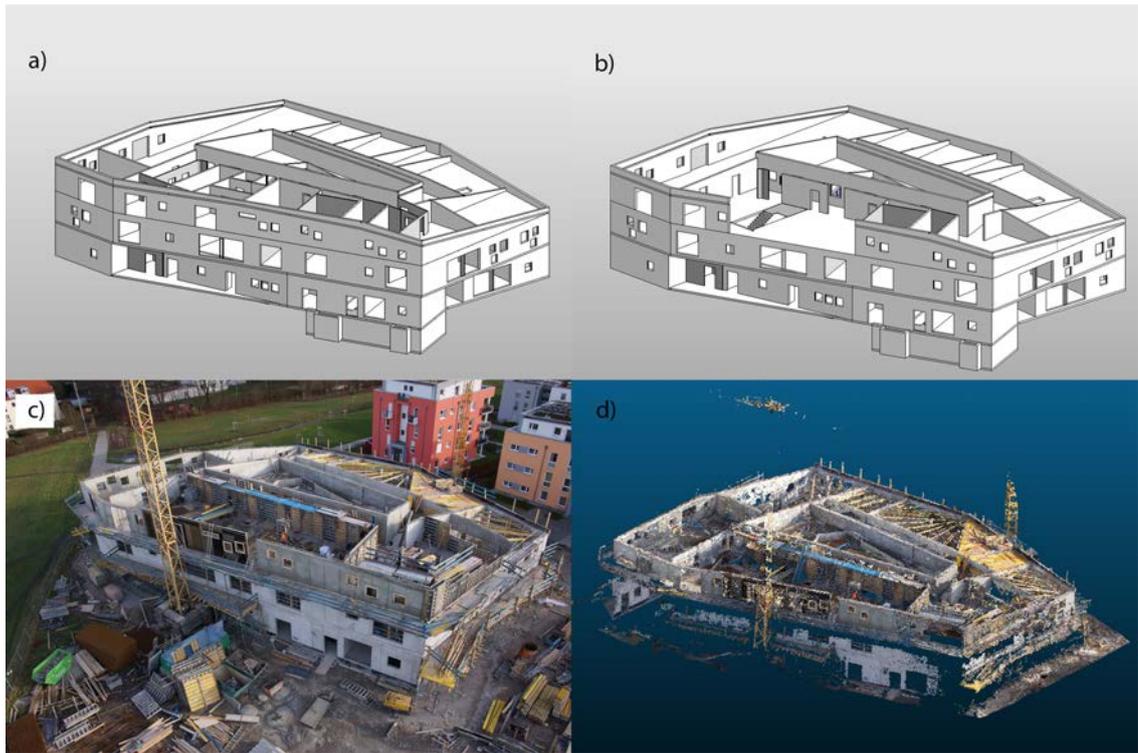


Figure 37 Process modeling problems depicted by a) as-planned modeling, b) as-built modeling, c) as-built image, d) as-built point cloud on sample construction site.

During the construction phase, the actual as-built process can deviate from the original as-planned process. To clarify this deviation, Figure 37 depicts the digital model from one of the test sites, a corresponding UAV-aerial image, and the generated point cloud (a), c) and d)). Accordingly, the as-planned 4D model does not represent the real construction sequencing. This problem is also described in Huhnt et al. (2008); Tulke (2010).

In comparison, Figure 37 b) shows the correct corresponding as-planned model for the given timestamp, with all subsequent elements being removed from the as-scheduled model.

Reconstruction

The monitoring of construction sites by applying photogrammetric methods has become common practice. Currently, several companies (for example, Pix4D or DroneDeploy) provide commercial solutions for end users that permit the generation of 3D meshes and point clouds from UAV or other image-based site observations. All these methods provide proper solutions for finished construction sites or visible elements of interest.

However, UAV-based monitoring of construction sites exhibits several problems. On the one hand, photogrammetric methods are sensitive to low-structured surfaces like monochrome painted walls, or windows. Because of the used method, each element needs to be visible from multiple (at least two) different points of view. Thus, elements inside of a building cannot be reconstructed as they are not visible from a UAV flying outside of the building.

Monitoring inside of a building is currently still the subject of much research (Kropp et al., 2014) and not available via an automated method, as localization in such mutable areas like construction sites is hard to tackle. These problems lead to holes or misaligned points in the final point cloud, which hinders the accurate and precise detection of building elements. On the other hand, laser scanning requires many acquisition points and takes significantly more time and manual effort in acquisition.

Occlusions

Finally, both techniques are challenged by occlusions for regions that are not visible during construction. The as-built 3D point cloud with n points holds all respective coordinates but also color information based on the feature's pixel color value in the initial image. The value n depends on many factors such as

- lighting conditions
- feature detection from different points of view
- surface textures
- amount and resolution of the images taken

A point cloud from one timestamp on one of our test construction sites can be seen in Figure 37 d). Besides scaffolding and formwork, various holes are visible in the point cloud that exist due to insufficient image quality for reconstruction or occlusions. The depicted point cloud matches the expected quality for an as-built acquisition and is incomplete due to changing visibility conditions from working equipment and similar items.

However, it is not sufficient for reliable results in a purely geometric as-planned vs. as-built comparison as significant parts of the actual building are occluded. As seen in figure 38, another problem lies in elements that are occluded by temporary construction elements. In particular, scaffolding and formwork occlude the direct view on walls or slabs, making it harder for algorithms to detect the current state of construction progress.

Currently available methods do not take these problems into account and make only limited use of BIM-related information such as type of construction and the general structure of a building.



Figure 38 Occluded construction elements in generated point cloud caused by scaffolding, formworks, existing elements and missing information during the reconstruction process

6.1.3. Contributions

In this paper, the authors propose a number of inter-related methods to tackle the aforementioned problems. Specifically, this paper presents the following contributions:

- Known technological dependencies of construction sequences are used to enrich the model by precedence relationships, by applying formal graph theory. This allows the inference of the existence of elements, if they have not been directly detected.
- A method is presented that makes use of the knowledge of construction methods and 4D data to adjust the detection thresholds (as-planned vs. as-built deviations allowed) according to their expected construction stage. This permits the detection of elements that are currently under construction and are, for example, covered by formwork.
- We introduce a method based on visibility analysis to identify elements that are detectable from the identified camera positions. Deep learning on projected element positions in the 2D plane of the gathered images for the initial SfM process allows the detection rates of built elements to be further enhanced.

The combined application of these methods helps to significantly improve the accuracy of construction progress monitoring, as documented by the case studies presented in this paper.

The details of the individual methods are described in Section 6.3.

6.2. Related Work

6.2.1. Scan vs. BIM

Progress monitoring has become a heavily researched topic in recent years. Omar and Nehdi (2016) provide an overview of these developments and compare the individual approaches:

The as-built status of a construction site is usually captured by laser scanners or cameras using SfM methods. Laser scanning has the advantage that 3D point measuring is fast and very accurate (within the range of millimeters). However, the equipment is heavy and requires trained personnel.

Additionally, the setup at the point of observation is time-consuming and, depending on the size of the construction site, many observation points are required to scan the whole construction site. Photogrammetric approaches produce less accurate point clouds in comparison to laser scanning and require significant computing power for the reconstruction. However, this method is more flexible and easier in its application, as camera equipment is standard, low-cost, and widely used on UAVs. Other devices, such as Microsoft Kinect, combine multiple sensors and can also be used for progress monitoring (Pučko et al., 2018).

The registration of the acquired point cloud and corresponding as-planned geometry is either performed manually or semi-automatically, e.g. by point-to-point matching through Iterative Closest Point (ICP) algorithms. Here, the algorithm minimizes the distance between the points of the laser scan and the BIM geometry (Bosché, 2012).

Currently, three methods are deemed to be established in the comparison with the as-planned status:

1. comparison of points from the as-planned geometry with as-built point clouds. These methods compare point clouds that are acquired by laser scanners (Bosché, 2010; Turkan et al., 2012) or SfM methods and derived point clouds from as-planned surfaces (Kim et al., 2013). Point proximity metrics mainly do this following a data-alignment process.
2. Feature detection in the acquired images from the as-built state. Using feature detection algorithms to assess the progress of as-planned elements (as the construction site evolves in time) by comparing measurements with dynamic thresholds learned through a Support Vector Machine (SVM) classifier, construction elements are directly identified from the acquired images (Golparvar-Fard et al., 2015).
3. Matching the as-planned geometry surfaces directly with the as-built points. Here, relevant points from the point cloud are directly matched onto triangulated surfaces of the as-planned model after using octree-based checks for occupied regions (Tuttas et al., 2015).

The first approaches involving object detection in laser-scanned point clouds were published by Bosché and Haas (2008a). Turkan et al. (2012) extend this system and use it for progress estimation. Kim et al. (2013b) detect specific component types using a supervised classification based on Lalonde features derived from the as-built point cloud. An object is regarded as detected if the type matches the type present in the model. As above, this method requires that the model is sampled into a point representation. Zhang and Arditi (2013) introduce a measure for deciding four cases (object not in place, point cloud represents a full object or a partially completed object or a different object) based on the relationship of points within the boundaries of the object and the boundaries of the shrunk objects.

However, the authors test their approach in a very simplified artificial environment, which is significantly less challenging than the processing of data acquired on real construction sites. In Mahami et al. (2019), SfM and Multi-View Stereo (MVS) algorithms are coupled with coded targets to improve the photogrammetric process itself. Ibrahim et al. (2009) use a single camera approach and compare images taken during a specified period, and rasterize them. Individual elements are identified for each use case. Most publications focus on identifying one particular type of element like, for example, columns or walls.

Indoor monitoring has been researched by several groups. Asadi et al. (2019) propose a new method to localize and align the camera position and building model in a real-time scenario. Kropp et al. (2018) tried to detect in-door construction elements based on similarities. Turkan et al. (2014) present an approach for detecting elements under construction that uses threshold extensions for those elements. Han and Golparvar-Fard (2015) published another attempt to solve the problem of elements under construction. The focus lies on visibility issues, e.g., assuming that an anchor bolt for a column must be present, despite being invisible, as the column on top of it requires the anchor bolt for structural reasons. Further research has been conducted in regard to multi-layered elements and the introduction of construction sequencing (Han et al., 2015).

Another critical aspect of the as-planned vs. as-built comparison is dependencies. Technological dependencies determine which element is depending on another element, meaning that it cannot be built after the first element is finished. Precedence relationships (Wu et al., 2010) can define these dependencies. Szczesny et al. (2012) discuss a storage solution for these dependencies. The approach with regard to progress monitoring is presented in Braun et al. (2017). Hamledari et al. (2017) introduce an IFC-based schedule updating workflow that relies on detected construction elements.

In their outlook, Turkan et al. (2014) state that further improvements to their work should include color analysis or even image-based methods. Thus, the authors propose incorporating these techniques, as well as the use of semantic data like construction methods, model analysis using technological dependencies, and image-based deep learning, to further enhance the detection of elements in an as-planned vs. as-built comparison.

6.2.2. Computer vision and deep learning

Rising computational power has enabled significant advances in machine learning in recent years. Deep learning (LeCun et al., 2015) and especially Convolutional Neuronal Networks (CNN) provide solutions for training computers to learn patterns and apply them to previously unseen data. In this context, computer vision is a heavily researched topic that has received even more attention through recent advances driven by, for example, the needs of autonomous vehicles.

Image analysis in the construction sector, on the other hand, is a rather new topic. Up to now, the main focus has been on defect detection (for example, cracks) in construction images (Akinci et al., 2006). Crack detection for asphalt roads has also been the subject of research (Nhat-Duc et al., 2018). Since one of the critical aspects of machine learning is the collection of large datasets, current approaches focus on data gathering. In the scope of automated progress monitoring, Han and Golparvar-Fard (2017a) published an approach for labeling image datasets based on the commercial service Amazon Turk. Braun and Borrmann (2019) introduce a method for automated image labeling by fusing semantic and photogrammetric data.

Regarding the application of deep learning for construction progress tracking, Chi and Caldas (2011) used initial versions of neural networks to detect construction machinery on images, and Kim et al. (2013) used ML-based techniques for construction progress monitoring. They analyzed images by filtering them to remove noise and uninteresting elements, so as to focus the comparison on relevant construction processes. Hamledari et al. (2017) applied CV approaches to indoor appliances like electrical outlets and insulation.

These approaches are currently mostly independent from the actual building model, as orientation and scale with respect to the digital twin are neglected or assumed to be given for the application of CV methods. The application of these methods, in combination with SfM-based orientation data, has not been the subject of research to date.

6.3. Concept

6.3.1. Objective

The main goal of this research is to improve the results of element detection from a point-cloud-based as-planned vs. as-built comparison by using additional information provided through the Structure-from-Motion process (images and camera positions), as well as the as-designed building information model (semantic data, geometric representation of elements, and position and dependencies of elements). The following concept presents the proposed solutions to tackle the mentioned challenges with several approaches, such as incorporating additional information on construction methods into the comparison algorithms.

6.3.2. Point of departure

The concept builds upon the body of knowledge of the research community as well as the previous research conducted by the authors. Thus, several steps in the process of automated progress monitoring are assumed to be given. Firstly, image acquisition for the generation of point clouds and camera position estimation is required. The authors provided several studies on image acquisition and proposed a UAV-based method as it is more flexible in comparison to fixed cameras (Tuttas et al., 2017). Secondly, the point cloud and the as-designed building information model must be aligned to one another (also known as registration). According to the well-documented state of the art, this is either performed via geodetic reference points that align the as-planned digital model with the point cloud on the measured geodetic position, via automated ICP methods (as mentioned earlier), or manually via point-to-point picking. The authors provide a detailed description of these approaches in (Braun et al., 2016) and (Braun and Borrmann, 2019). In this paper, we significantly extend the state-of-the-art approach by using computer vision (CV) methods.

6.3.3. Concept overview

The concept presented in this paper relies on the exploitation of as-design building information models to improve the progress-detection process. We assume them to be available as IFC instance models. These models provide a geometric representation of all relevant building components, as well as the related semantic information (such as component type, material or the attribute "load-bearing") as well as 4D process data. The general idea is to enhance the purely geometric as-planned vs. as-built comparison from point-cloud to geometry level, with additional layers of information. Fig. 39 depicts the conceived processing chain. The highlighted process components provide new elements that are introduced in this paper. After defining the different sets of building elements required for the process in Sec. 6.3.4, these new elements are explained in detail in dedicated subsections.

The creation of the precedence relationship graph is discussed in Sec. 6.3.5. The following sections focus on schedule analysis (Sec. 6.3.6), and color detection (Sec. 6.3.7). The latter process step helps to identify whether an element is present or occluded by other structures. Finally, we introduce a method that projects the 3D as-designed geometry provided by the building information model into the 2D plane, so as to apply image analysis techniques for element detection. Sec. 6.3.8 describes the projection process. Subsequently, Sec. 6.3.9 discusses the application of computer vision methods to detect the type of the element that is visible in the projected region of interest.

The combination of these individual processing steps results in a significant improvement in the accuracy of the overall automated progress detection method, as demonstrated through the case studies presented in Section 6.4.6.

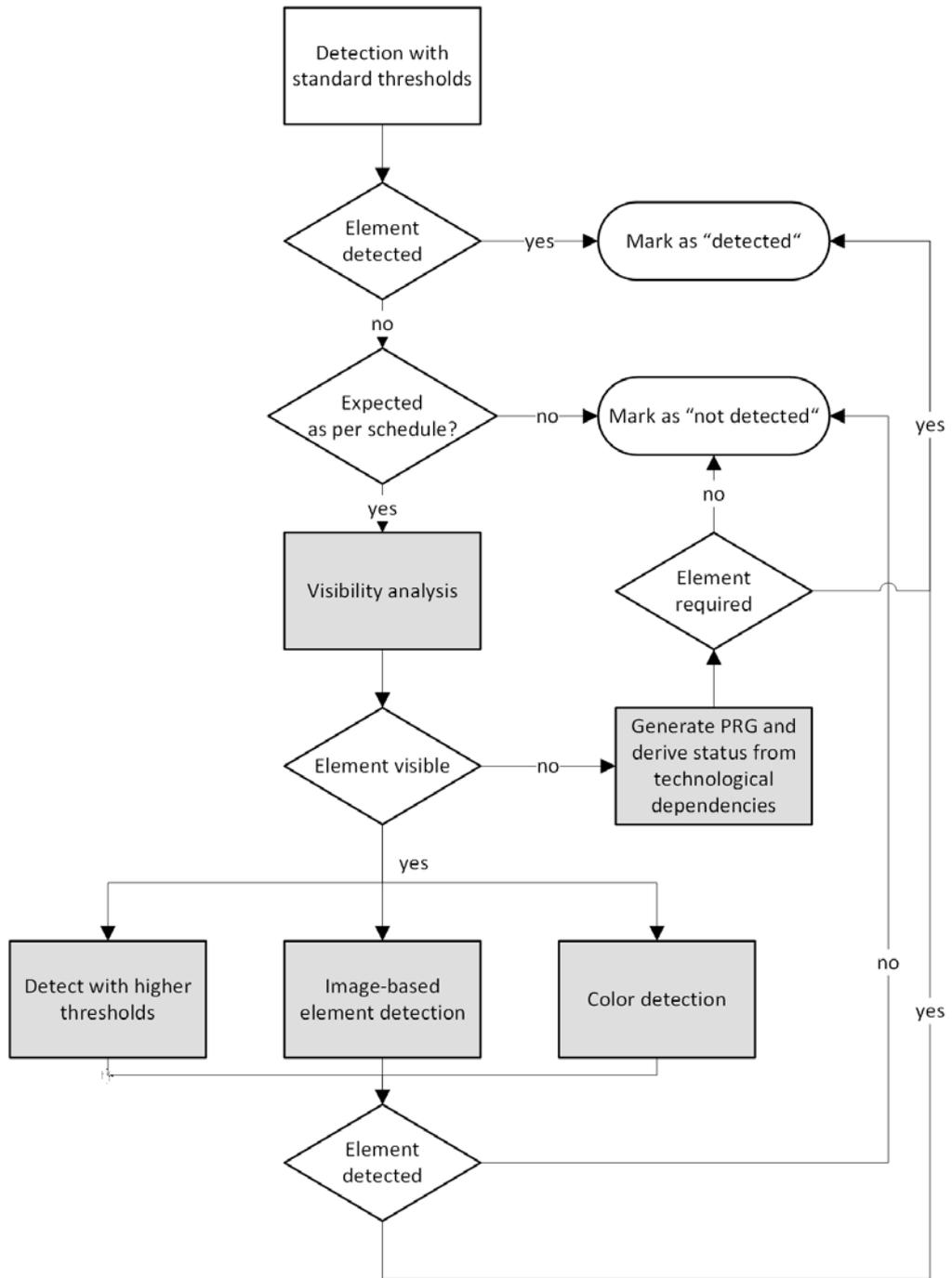


Figure 39 Concept for the enhancement of element detection. The highlighted process steps are introduced in this paper.

6.3.4. Sets of elements and detection status

In the context of the research presented, the following sets of construction elements are defined in regard to as-built vs. as-planned comparison:

- E represents all elements of the current building
- $E_P(t)$ defines the amount of elements that should be present at the time t of observation according to the as-planned schedule
- $E_{GT}(t)$ defines the ground truth as all elements that are built at observation t
- $E_D(t)$ defines all elements that were detected during an observation at timestamp t
- $E_{ND}(t)$ defines all elements that were not detected during an observation at timestamp t
- $E_V(t)$ defines all elements that are visible from the corresponding points of view during observation at timestamp t

t defines the observation timestamp, at which the construction site has been monitored.

The following definitions hold true for all given sets at any timestamp t :

$$E = E_D(t) \cup E_{ND}(t) \quad (6.1)$$

$$E_D(t) \leq E_V(t) \leq E_{GT}(t) \leq E \quad (6.2)$$

According to these definitions, the set of *TruePositives* is defined as

$$E_{TP}(t) = E_D(t) \cap E_{GT}(t) \quad (6.3)$$

while *FalsePositives* are the counterpart:

$$E_{FP}(t) = E_D(t) \setminus E_{GT}(t) \quad (6.4)$$

The goal of this research is to verify as many existing construction elements as possible, so as to minimize the differences between these sets while keeping $E_{FP}(t)$ minimized. Mathematically speaking:

$$E_D(t) \longrightarrow E_{GT}(t) \quad (6.5)$$

It is not possible to define a relation between the planned elements $E_P(t)$ and the ground truth $E_{GT}(t)$ as the progress of the construction site depends on many external factors that cannot be formalized with the given data. The set of $E_P(t)$ can contain more elements than

$E_{GT}(t)$ in the event of a delay on the construction site but also fewer elements in the event of faster construction times.

In addition to the mentioned sets, every construction element is classified individually for each of the following states: built (Ground Truth), detected, planned, encased in formwork, under construction.

These definitions are used in the described concepts.

6.3.5. Process sequencing and precedence relationships

As-built monitoring with SfM methods or laser scanning always captures one particular timestamp.

For automated handling of dependencies, a precedence relationship graph (PRG) is introduced (Braun et al., 2017). The PRG formalizes technological dependencies between construction elements and is defined as a directed, acyclic graph (DAG) with each node representing one construction element (Zobrist and Leonard, 1992). Technological dependencies for load-bearing structures between two elements can be automatically detected when they have a particular spatial constellation that, in combination with the construction method applied, unambiguously defines their sequential order. For example, when conventional in-situ concreting methods are applied, a slab on top of a column can only be built after the column is finished. To generate this graph, the semantic as well as the geometric data from the digital model is used in combination with a knowledge base representing the construction methods. The geometric data is used to identify elements that are touching each other, and for sequencing them in their respective vertical order. Additionally, the semantic data is used to determine the construction method for an individual element, and to filter load-bearing elements. The generation of the initial precedence graph is performed as depicted in Algorithm 1. This method relies on a spatial query language, as introduced in Daum and Borrmann (2014).

Algorithm 1 Pseudo code for the generation of an initial Precedence Relationship Graph

```
1: procedure GENERATEPRECEDENCERELATIONSHIPGRAPH
2:    $E \leftarrow$  set of all construction elements
3:   for all  $E(\text{LoadBearing})$  do
4:     for all  $ET$  do
5:       if  $\text{Above}(E(\text{LoadBearing}), ET)$  then
            $\text{AddDirectedEdge}(E(\text{LoadBearing}), ET);$ 
```

The initial precedence graph is completed manually in order to take project-specific boundary conditions and non-spatial precedence relationships into account.

The PRG is used to identify objects that are possibly under construction at the time of observation.

Using the introduced PRG, it is possible to identify elements that might be under construction and thus are considered for further investigation. The basic flowchart depicted in Figure 39 shows the implemented workflow for enhanced detection.

Based on the construction type and the erection method, different steps follow. As detailed above, walls and other vertically erected elements are considered for an extended threshold in order to identify possible formwork. Additionally, color matching helps to differentiate the material properties.

Moreover, the PRG allows for assumptions with regard to elements that are invisible due to occlusions, and thus not directly detectable. For example, load-bearing columns underneath a detected slab are expected to be built even if it is not possible to verify them via the point cloud.

6.3.6. Identified tasks during construction

Several tasks are required to construct in-situ concrete elements or similar elements. In concrete construction, formwork for in-situ concrete is the most common construction method. Several different methods are depicted in Figure 37 b) and d). All possible elements under construction are considered in order to detect formwork. In general, elements are counted as detected as soon as a certain amount of points per area [Pts/m^2] with a distance of less than 2 cm are matched on the surface of the element (Tuttas et al., 2016). If the expected elements are not detected, the threshold for the maximum distance can be adjusted to take into account the fact that the formwork with a thickness of around 0.20m might be currently in place. If this iteration brings positive results, the element can be marked as "under construction".

6.3.7. Color detection

In general, formwork for walls and columns consists of a wooden, smooth plate on the concrete side, and a steel structure for stability on the backside. This steel structure is often painted red, yellow or orange, and is distinct from the gray concrete. Formwork for slabs usually consists of elevated wooden plates that have the same color range as the steel structure mentioned. This color difference can be measured and may help to further improve the detection quality of formwork. The HSV (Hue-Saturation-Value) color space provided useful data for the color detection (Sural et al., 2003). In contrast to the RGB color space, the HSV color space can describe color as perceived by humans but also saturation and brightness (value). Each value has a range from 0 to 1.

Comparing the color distribution of identified subsections of the point cloud can consequently help to achieve further verification of the existence of an element. The material color, as well as the type of construction, is retrieved from the building information model in order to gather color information. After identifying a gray color distribution for an expected concrete element, this data further confirms the existence of said element. In comparison, a mainly red or orange color distribution leads to the assumption that a formwork element is present, if the

initial element has not been verified but is meant to be constructed with in-situ concrete.

6.3.8. Visibility analysis and projection of elements

Photogrammetry is based on estimating the position of all cameras that are used for the point cloud generation. Since the digital model of the construction sites is aligned to the point cloud during the comparison process, it is possible to project the 3D geometry of all elements into the respective 2D plane of a corresponding image (Braun and Borrmann, 2019). Knowing the expected position of an element in image space enables highly accurate object-detection to be performed, using CV approaches.

More specifically, it is possible to perform a visibility detection by using the camera parameters to compute the projection of the model elements onto image space and of the process information, to define the set of construction elements that are supposed to be built. The building model coordinate system needs to be transformed into the camera coordinate system or vice versa in order to align both models. By applying this method, rendered images from all points of acquisition are generated that allow the determination of which elements are actually visible and can potentially be found in a generated point cloud. The resulting set of visible elements $E_V(t)$ enables greater detection accuracy.

The general approach for this method is explained in Braun and Borrmann (2019), though for a slightly different application scenario. For further clarification, the key steps are explained in this section.

In order to calculate the projection, the intrinsic camera matrix for the distorted images that projects 3D points in the camera coordinate frame to 2D pixel coordinates using the focal lengths (F_x, F_y) and the principal point (x_0, y_0) is required. Additionally, the skew coefficient s_k for the camera is required. It is zero if the image axes are perpendicular. The matrix K can be described as defined in equation 6.6.

$$K = \begin{bmatrix} F_x & s_k & x_0 \\ 0 & F_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6.6)$$

The translation of the camera is defined as:

$$T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (6.7)$$

Additionally, the rotation matrix for each image, as defined in equation 6.8 is needed.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (6.8)$$

Both, translation and rotation can be described in one 3 x 4 matrix:

$$RT = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_1 \\ r_{21} & r_{22} & r_{23} & T_2 \\ r_{31} & r_{32} & r_{33} & T_3 \end{bmatrix} \quad (6.9)$$

Using the model coordinates of all triangulated construction elements, it is possible to calculate the projection of each element into the camera coordinate system and therefore overlay the model projection and the corresponding image taken from the point of observation with equation 6.10.

$$t = K * RT * p; \quad (6.10)$$

The resulting 2D coordinates that are rendered into the image are calculated by using the vector t and calculating the x and y coordinates by

$$x = t[0]/t[2] \quad (6.11)$$

and

$$y = t[1]/t[2] \quad (6.12)$$

With this projection, the model can be rendered from the camera's perspective for all images acquired during observation. After including the 4D temporal information from the as-planned model, this information can be fused, to render the model with the expected set of elements $E_P(t)$ from all estimated camera positions. The term "rendering" here refers to the creation of the 2D projection of the model according to the rendering pipeline established in computer graphics (Foley and Foley, 1990), but without applying advanced features such as reflections, light sources or shading. These rendered images are analyzed for all visible elements $E_V(t)$ by applying the Painter's algorithm (Elvins, 2005). With knowledge of this set of elements, the set $E_D(t)$ can be checked for false positives, but also measured for accuracy regarding its true positive rate. This is done by excluding elements from set $E_P(t)$ or $E_{GT}(t)$ that are invisible from any camera position during acquisition.

6.3.9. Image-based object detection

To further enhance the detection of construction elements, we propose making use of the images taken in the course of the initial acquisition for the photogrammetric point cloud generation. By applying the previously described projection technique, all construction elements can be localized on any image taken during the acquisition. A sample is shown in Fig. 40: A column of interest is selected in the 3D view (marked red); detailed information about this element is shown in the lower right. Accordingly, a corresponding image that validated the existence of the selected element - and additionally the 3D to 2D projection described in Section 6.3.8 - is used to display the expected position of the element in this image.

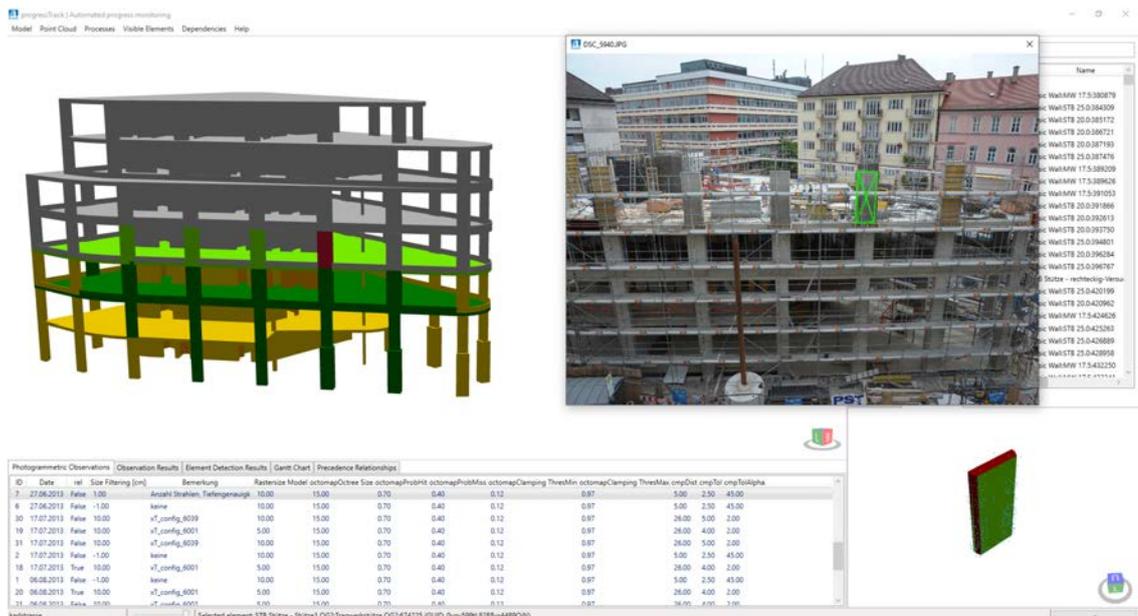


Figure 40 Projection of a selected 3D geometry into the 2D plane of a corresponding image

As machine-learning methods have made significant advancements in recent years, tasks like

image classification or even region detection on images are now being used in various scenarios. For the task of progress monitoring, the authors propose the use of a Convolutional Neural Network (CNN) trained on construction elements and thus able to detect the type and instances of construction elements on the given images. In the case that an element is not detected and validated by the point cloud, the implemented workflow is followed as described in Fig. 41.

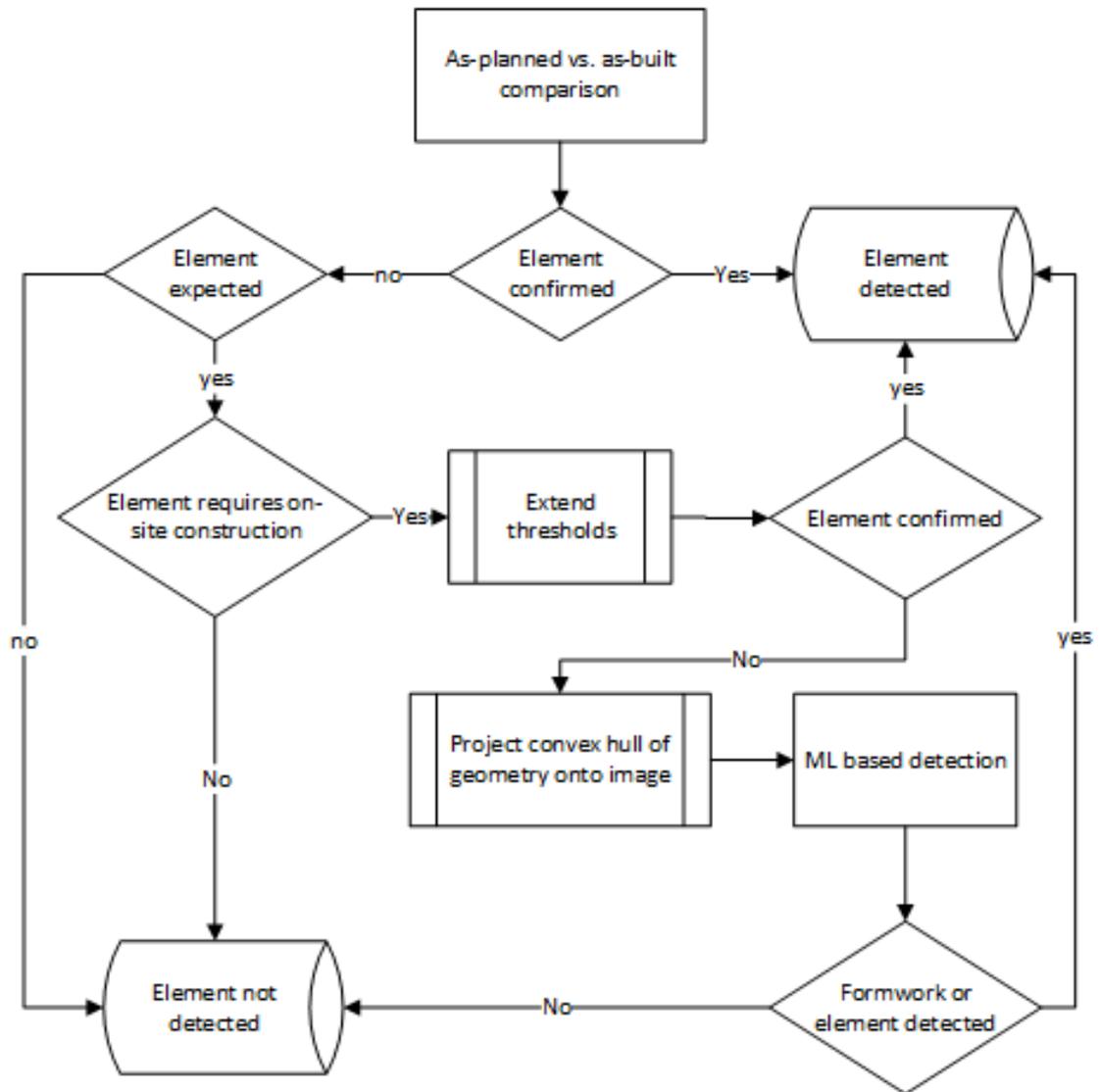


Figure 41 Occluded construction elements in generated point cloud caused by scaffolding, formworks, existing elements and missing information during the reconstruction process

If the element is expected according to the up-to-date schedule and requires in-situ work, in a first step, the thresholds are increased as defined in Sec. 6.3.6. If this helps to validate the elements' existence, it is added to the set of detected elements $E_D(t)$. If not, the 2D projection, as mentioned in Sec. 6.3.8, is used to identify the region of interest in a suitable image.

Subsequently, the trained CNN (He et al., 2016) classifies the region according to the prede-

fined states and thus contributes to a refined state detection. If, e.g., formwork is detected here, the element can be marked as "under construction".

In order to use a CNN for object-based region detection, the training of said network is required. For this purpose, 5,000 images were labeled with the categories *formwork*, *scaffolding*, *columns*, and *walls*. This resulted in 9,700 labeled formwork elements. The labeling procedure is depicted in Fig. 42. The data is converted into the COCO data format (Andriluka et al., 2014) and prepared for training by augmenting the images to enlarge the training set even further.



Figure 42 Sample image of the labeling process. Displayed are the labeled formwork (blue) and column (green) elements. During this research, Labelbox (Labelbox, 2020) is used for labeling.

To sum up, all introduced methods make the overall process much more robust compared to a purely geometry-based approach, and lead to a higher detection accuracy.

6.4. Case Study

Several construction sites were monitored with different observation methods to validate the introduced concepts. The construction sites are all German-based and cover a number of structural engineering buildings as well as infrastructure (one bridge, one wastewater treatment plant). The main construction method is in-situ concreting, this being the most common construction technique in Germany. Listed in Table 12 are the three construction sites that are used as case studies in this section.

Site	Elements	Observations	Pictures taken	Duration
Test Site A	671	6	1,805	5 months
Test Site B	943	9	2,350	10 months
Test Site C	2,229	23	3,144	5 months

Table 12 Test sites monitored during this case study

In this context, the authors published several papers presenting their approach and developed a software framework, which was introduced in Braun et al. (2017) and shown in Figure 43. To visualize the comparison results and the detected elements, and to verify the used algorithms, all gathered data is stored in a database that is accessible via this software. The tool displays all geometric and semantic building element information as well as scheduling data that has been parsed from IFC instance models. The detected elements are highlighted for easy identification. Figure 43 shows the software interface with the example of one of the construction-site case studies used in this research. The building mainly consists of in-situ concrete elements that were cast using formwork on site. In the figure, one individual capturing event is selected, and all detected elements are highlighted. Green coloring represents elements that have been built and are correctly detected and confirmed through the point cloud. All yellow elements are built but were not confirmed through the point cloud.

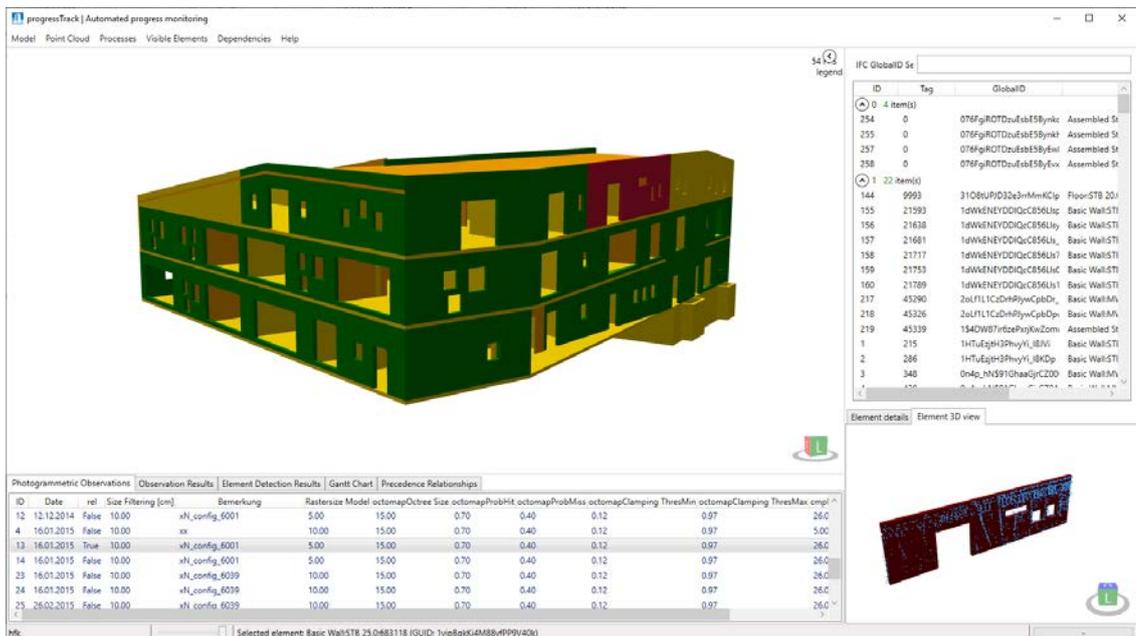


Figure 43 Screenshot of a developed tool for as-planned vs. as-built comparison. A specific observation is selected to visualize the detected construction elements at that time. Details of selected elements are shown in a separate viewer.

There are several reasons why some of those elements may not be detected. The most prominent reason is the occlusions that occur on site. During construction, large amounts of temporary structures like scaffolds, construction tools, and construction machinery obstruct the view of the element surfaces. Limited acquisition positions further reduce the visible

surfaces and hence the overall quality of the generated point clouds. Additionally, elements inside of the building are also occluded by other building elements for acquisitions outside of the building.

Another reason for weak detection rates is building elements that are currently under construction. As those elements count towards the overall progress, they must not be missed, and play a crucial role in defining the exact state in the current process. In general challenges exist for all construction methods, whose geometry under construction differs largely from the final element geometry which requires the use of temporary construction objects. This applies, e.g., to reinforced concrete and multi-layered walls. On the one hand, formwork which is used for concrete pouring, may obstruct the view of the element, making it impossible to be detected. On the other hand, the plane surface of formwork for a slab might be detected as the surface of the slab itself and thus would lead to a false positive. Due to these challenges, further enhancements to the comparison and detection algorithms are needed. Since the digital model contains information on construction methods, the authors propose using this knowledge in the overall detection process. By deducing the precedence relationships with a query language, assumptions regarding occluded elements can be made. Construction methods and derivation of expected elements lead to new as-planned vs. as-built comparison capabilities, such as extended thresholds and computer vision methods to detect objects like formwork on the raw observation images, taken for the point cloud generation.

6.4.1. Precedence Relationship Graph

The PRG for all construction sites is generated by using a query language for Building Information Models (QL4BIM, Daum and Borrmann (2014)). With the algorithm introduced in Sec. 6.3.5, any building information model that has sufficient semantic information can be analyzed, and technological dependencies are formalized by the introduced graph. Fig. 44 shows the PRG for one of the mentioned case studies. Each node represents one construction element; the directed edges show the corresponding dependency.

Based on the detected elements (marked in green and yellow), all dependent elements can be identified via this graph. Specifically, this graph allows one to make assumptions regarding the construction elements that were either invisible during observation, or were not detected due to occlusions or other issues (as mentioned before). The elements marked in blue in Fig. 44 are identified as depending elements with this method.

Table 13 shows detailed enhancements for the introduced PRG. In particular, a significant amount of construction elements were identified as depending upon the detected elements. In this respect, these elements are logically required to be built despite the fact that they were not confirmed visually by the point cloud.

This information helps to obtain additional information for the as-planned vs. as-built comparison: if a slab is built, all load-bearing elements underneath it must have been built, even

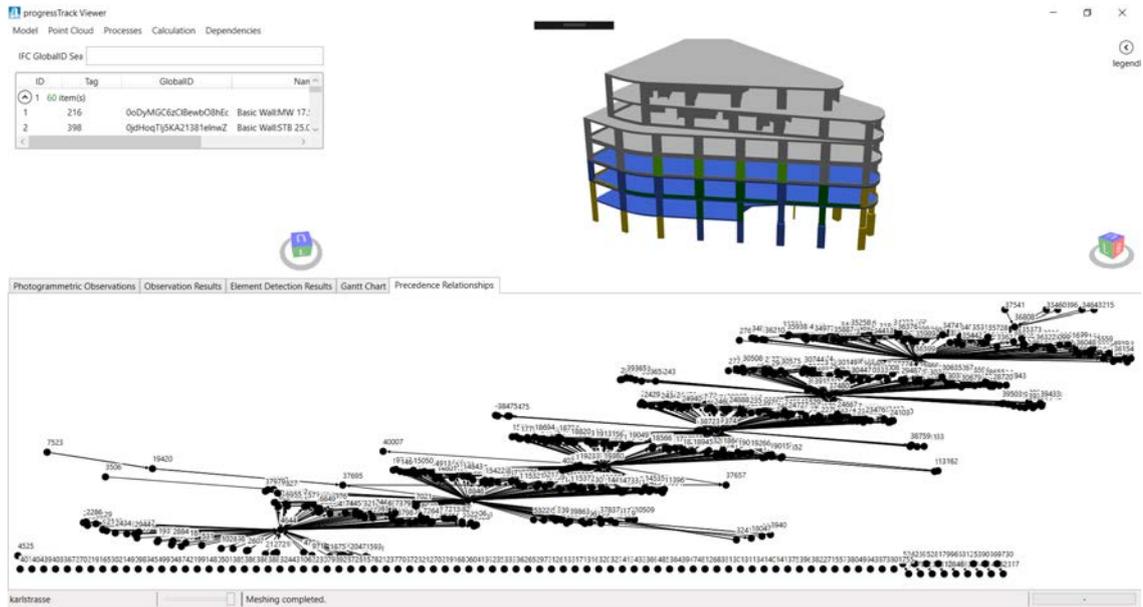


Figure 44 Generated precedence relationship graph for Test Site A. Elements marked blue were derived from the PRG in combination with the detected elements marked in green and yellow.

Date	$E_{GT}(t)$	$E_D(t)$	$\delta E_{PRG}(t)$
15.05.	89	37	20
12.06.	152	32	57
27.06.	184	59	54
17.07.	233	53	85
06.08.	277	95	102
04.09.	342	98	159

Table 13 Enhancing results by applying the introduced PRG for Case Study site A

though they cannot be verified by any visual method.

6.4.2. Varying dimensions

Figure 45 depicts a part of a snippet of a point cloud, generated at one individual time-step during observation. It is overlaid with the corresponding 3D geometry and visualized in green, to symbolize the as-planned as well as the as-built status. Based on this example, the general workflow for elements under construction is shown. As depicted, the front wall is already finished, and the concrete surface is visible. The wall in the second row is currently under construction, and the formwork is present and registered in the point cloud.

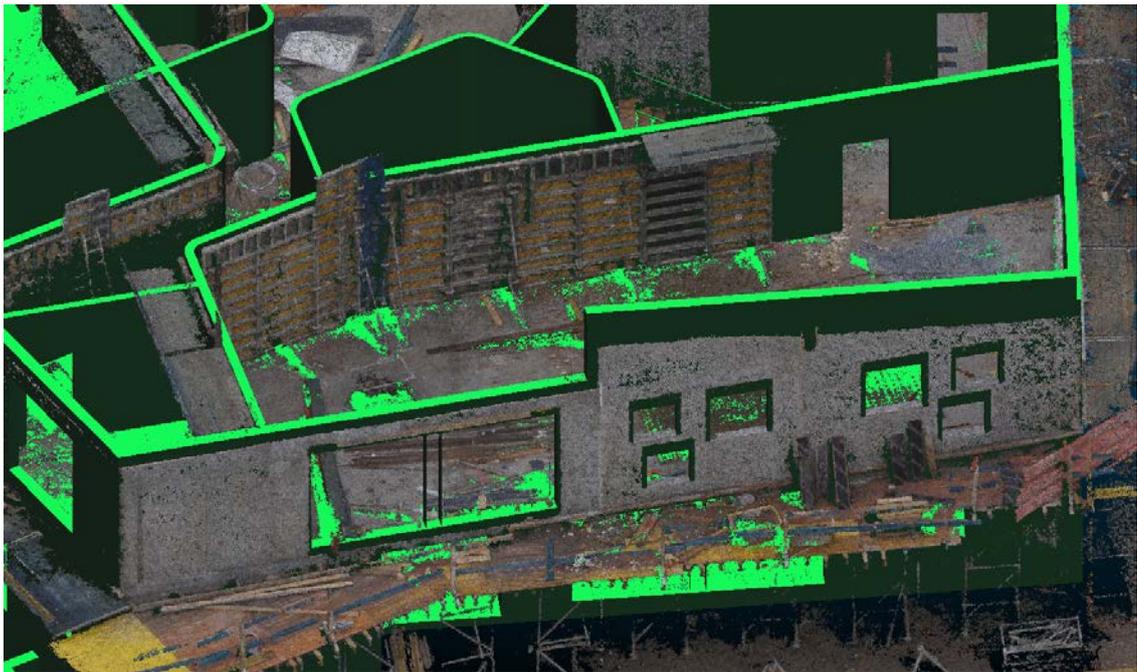


Figure 45 Point cloud of a finished, plain wall and formwork overlaid with the corresponding 3D geometry on Test Site B

During detection, it is expected that the first row of walls will be detected. Due to the threshold of max. 1 cm, the second row should not be detected due to the formwork. Figure 46 a) shows the expected result, with an additional set threshold of $1000 \text{ points}/\text{m}^2$ (in green). Triangles marked in yellow have matching points but do not qualify for the set thresholds, while elements marked red have no qualifying points at all. The walls in the second row are expected to be in progress. As presented in the concept in Section 6.3, the detection is therefore carried out with a larger threshold. Based on this result, the accepted point-to-surface distance is increased to 10 cm, which leads to the results depicted in Figure 46 b).

The increased threshold leads to the expected higher point density on the wall under construction, as the formwork is considered, too. According to the introduced workflow, the wall is now marked as "under construction", leading to a further detailed automated progress monitoring.

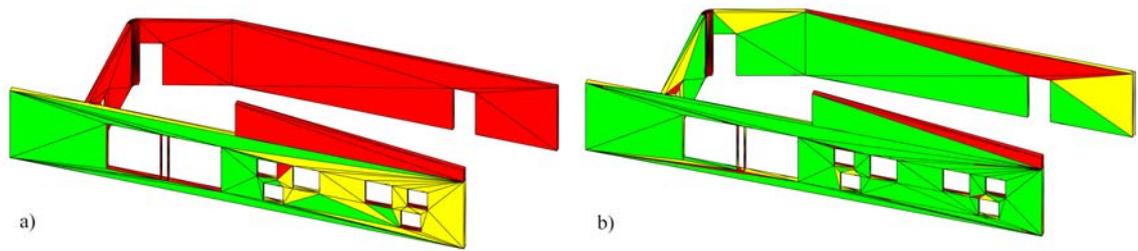
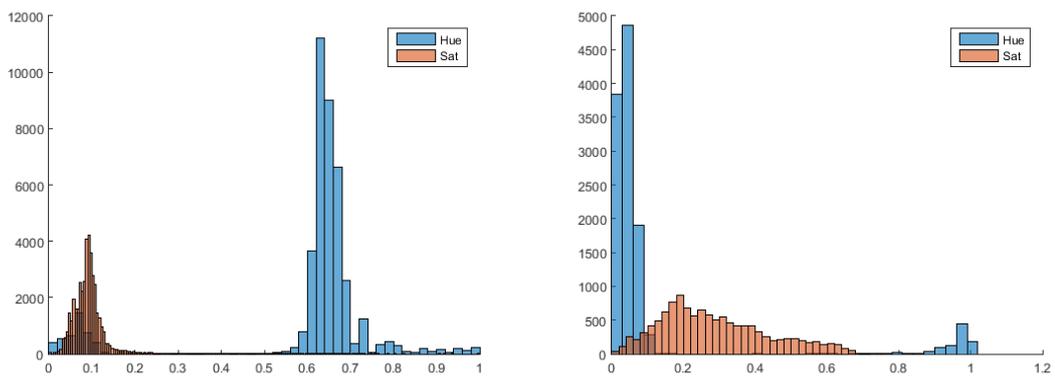


Figure 46 Triangles detected during the time-step shown in Figure 45. a) with 1cm gaps and $\rho > 1000\text{pts}/\text{m}^2$, b) with 10cm gaps and $\rho > 1000\text{pts}/\text{m}^2$



(a) HSV color distribution for concrete

(b) HSV color distribution for formwork

Figure 47 Distribution of frequency in the HSV color space shows clear deviations between concrete and formwork elements with the Hue value represented by blue bars and Saturation value represented by orange bars.

6.4.3. Color detection for formwork and reinforcement

As detailed in Section 6.3, taking colors into account can improve the detection of formwork or reinforcements due to their significantly varying colors, in comparison to the grey colors of the concrete. The color values of the different elements were compared to prove this statement. Figure 47 shows the calculated mean values for different elements under different lighting conditions.

In calculating the mean HSV values, all points relevant to an element are considered, along with the relevant color information. The results show that the brightness (value) varies largely, which is due to the lighting conditions itself. Therefore, this value has no further significance for this study. However, the hue values for formwork fall into the correct range for warm, red colors, whereas the concrete walls are based on "colder" colors. Additionally, the saturation differs by at least a factor 2.3. This color distribution analysis at a point-cloud level allows automated color interpretation to be carried out, and helps to identify differences between expected and actual color ranges based on material properties. The described process is used during the whole comparison to obtain a higher accuracy of information.

6.4.4. Visible elements

The visibility analysis is tested on several construction sites. Fig. 48 shows four samples from different observation times and construction sites. Each element has a unique color for identification purposes.

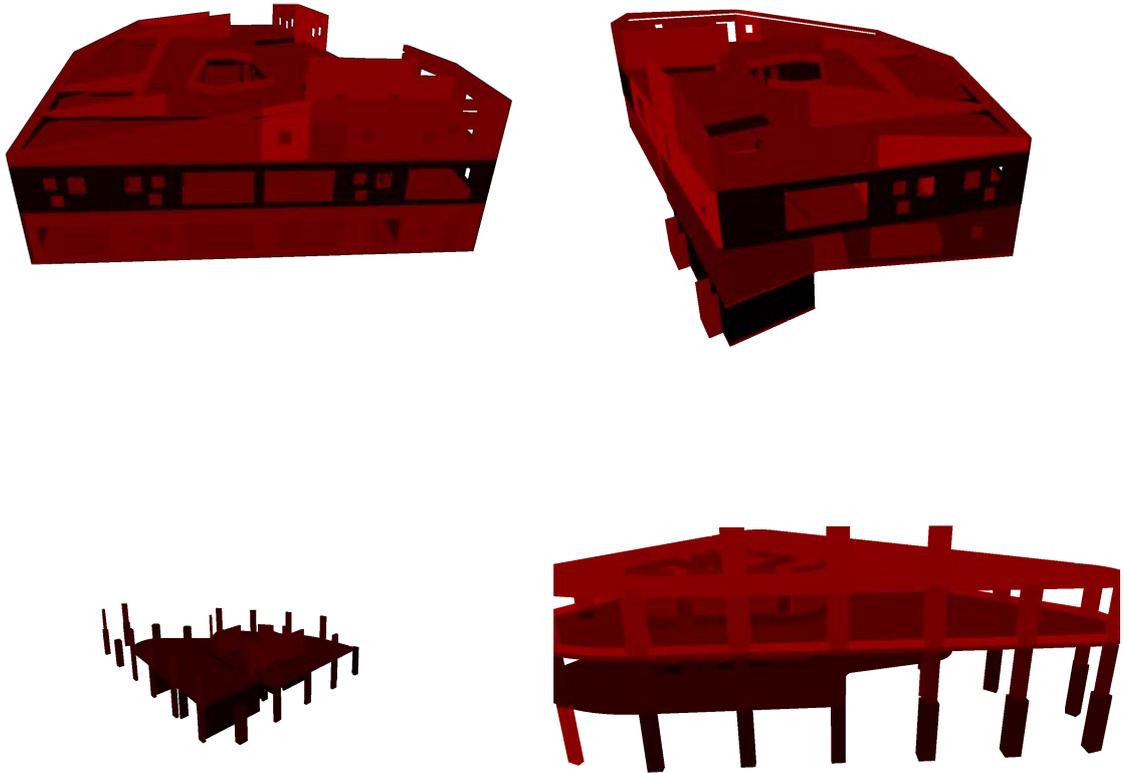


Figure 48 Visibility analysis with rendered geometry of set $E_P(t)$ for several construction sites and observations. All elements are rendered in different colors to distinguish them from each other.

Based on these results, all visible elements are identified and added to a corresponding set $E_V(t)$. This additional step does not detect any additional elements during the as-planned vs. as-built process, however it helps to set the detection results in a more accurate context. In detail, false positives can be reduced by removing invisible elements. Additionally, the thresholds used for the comparison process can be validated in a more precise manner, as the invisible elements are not added to the set of not detected elements.

Table 14 shows this data for one of our case studies during the whole observation period.

6.4.5. Image-based object detection

For the image-based object detection described in Sec. 6.3.9 we trained a Mask R-CNN-based (He et al., 2016) neural network using a training set consisting of over 5,000 images from five different construction sites and 40 observations with 9,700 labeled formwork elements and around 5,000 labeled column elements. Depicted in Figure 50, the results for formwork and column elements are shown in an image that was not part of the training set.

Date	$E_{GT}(t)$	$E_D(t)$	$E_V(t)$	$\%_{Vis}$
15.05.	89	37	73	82.0 %
12.06.	152	32	122	80.3 %
27.06.	184	59	155	84.2 %
17.07.	233	53	214	91.8 %
06.08.	277	95	275	99.3 %
04.09.	342	98	325	95.0 %

Table 14 Visible elements based on the introduced algorithm for Test Site A.

A common method to quantify the estimated result is the mean average precision that calculates as

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (6.13)$$

In combination with the recall

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (6.14)$$

the harmonized F_1 score can be calculated as:

$$F_1score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6.15)$$

An ideal network with perfect precision and recall values would achieve a F_1 score of 1. The trained network has a mean average precision (mAP) of 90.7% with an IoU (Intersection over Union) of 0.5 over all categories. With $TP = 11731$, $FP = 1099$ and $FN = 928$, the precision is at 0.914, the recall at 0.927, resulting in an $F_1Score = 0.920$ proving the suitability of the implemented methods. Fig. 49 shows the corresponding precision-recall curve for the trained network.

It has been tested against previously unknown images from the internet and other construction sites.

The results of this image-based region detection are subsequently used for the as-planned vs. as-built comparison. As introduced in Figure 41, construction elements that have not been

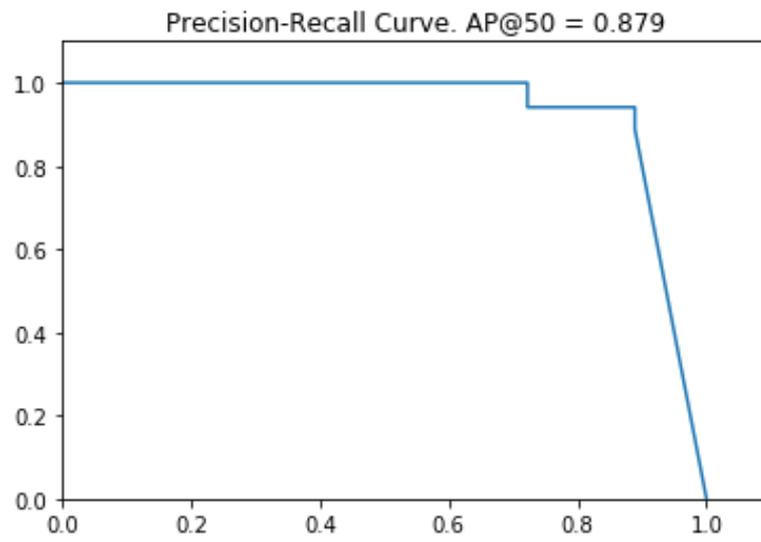


Figure 49 Precision-recall Curve for the trained network

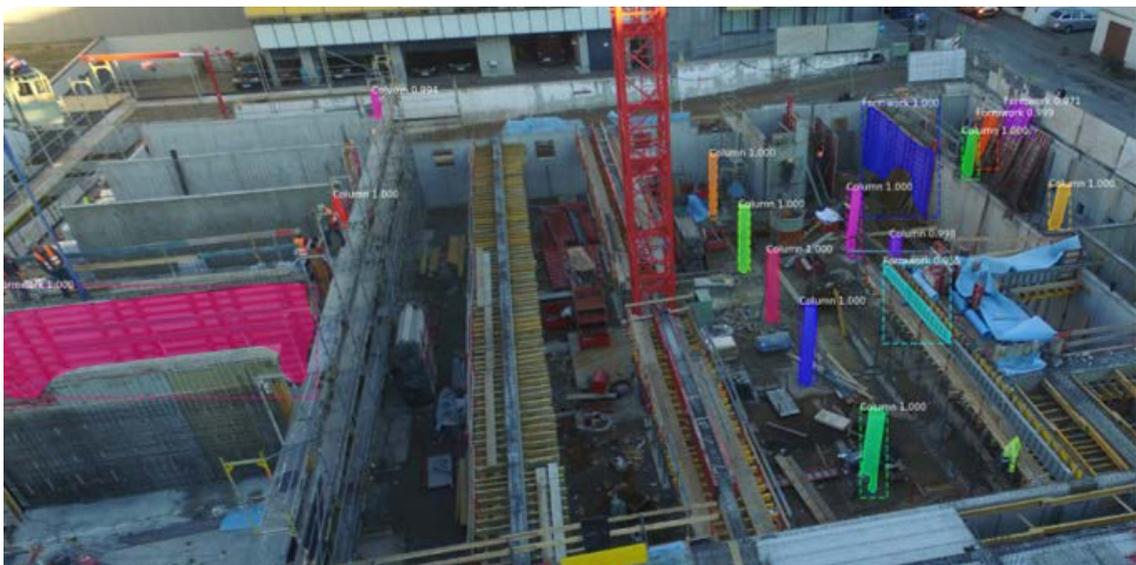


Figure 50 Formwork and column elements detected by a trained CNN using Mask R-CNN on Test Site C

verified by the point cloud, are run through an additional workflow, in order to check for formwork elements. If the CNN verifies the existence of a formwork element, the corresponding concrete structure is labeled as "under construction", making the process estimation more accurate. After testing this approach on a real-world construction site, this additional step proved to be suitable for construction sites that use in-situ concreting as a manufacturing method. Table 15 shows the amount of detected formwork elements with the help of the trained network.

Date	15.05.	12.06.	27.06.	17.07.	06.08.	04.09.
Detected	9	11	8	2	0	8

Table 15 Detected formwork elements during the observations for Test Site A.

6.4.6. Results

After the evaluation of all steps, the methods are incorporated into the presented software framework. Table 16 shows the results for one of our case studies during the complete construction process. During the initial, point-cloud-based comparison, the following data was gathered:

Date	$E_P(t)$	$E_{GT}(t)$	$E_D(t)$	$E_{FP}(t)$	$A_D(t)[m^2]$	$A_{GT}(t)[m^2]$	$\%_A$
15.05.	60	89	37	2	1162.76	1916.95	60.66 %
12.06.	133	152	32	11	1326.95	3557.74	37.3 %
27.06.	240	184	59	0	2244.51	4808.6	46.68 %
17.07.	348	233	53	5	4147.65	6261.07	66.25 %
06.08.	456	277	95	1	4480.78	6773.9	66.15 %
04.09.	569	342	98	1	4763.63	9197.7	51.79 %

Table 16 Resulting element sets for Test Site A

According to this data, the detection rates differ over a range of 37 % to 66 % correctly detected elements, based on the area surfaces. As mentioned above, these results largely depend on the point-cloud density and reconstruction quality from the SfM process. For any construction planner, these results would be insufficient as a comprehensive progress-monitoring tool.

After applying the newly introduced methods to this initial as-planned vs. as-built comparison, these additional results were gathered as shown in Table 17 with detected, cast elements defined as $E_{FW}(t)$ and elements inferred by the PRG, in addition to the previously detected elements, as $\delta E_{PRG}(t)$.

This table summarizes the results of the previous sections.

Date	$E_V(t)$	$E_{FW}(t)$	$\delta E_{PRG}(t)$	$E_{D_{new}}(t)$	$A_D(t)[m^2]$	$A_V(t)[m^2]$	$\%_A$
15.05.	73	9	20	66	1509.42	1681.04	83.8 %
12.06.	122	11	57	100	2792.95	3284.76	85.0 %
27.06.	155	8	54	121	3975.51	4579.60	86.8 %
17.07.	214	2	85	140	4975.65	6059.13	82.1 %
06.08.	275	0	102	197	5780.78	6644.94	87.0 %
04.09.	325	8	159	265	7675.58	9021.86	85.1 %

Table 17 Enhanced results for the detection with the newly introduced methods for Test Site A.

As shown, the number of detected true positives is raised significantly by applying the introduced steps. The newly detected rates all lie in the range between 80% to 90% of the actually built elements. An improvement of more than 100% in detected elements in comparison to the pure point-cloud vs. geometry-based detection methods was achieved. To draw conclusions from the results, there is still potential for further improvements. However, the introduced methods were tested on real-world construction sites over the complete construction cycle, and not only on a limited test area which usually constitutes a more controlled environment. Real-world data from construction sites always introduces many occlusions, and non-modeled elements that make it nearly impossible to detect all elements on a construction site.

6.5. Discussion and Outlook

6.5.1. Conclusion

Detailed progress monitoring is of utmost importance for efficient construction site management as it allows delays to be identified early, and for respective counter-measures to be taken. Matching the as-designed 4D building information model to point clouds provides a suitable basis for automating this process. The general approach of Scan-vs-BIM has been proposed and investigated by a number of researchers in recent years.

In this paper, a number of methods are introduced that further improve the accuracy of the detection process of the as-planned vs. as-built comparisons. The common approach lies in fusing information generated by different techniques and from different sources, namely the images, the point cloud and the building information model. The formal description of the technological dependencies in the construction process in the form of a precedence relationship graph allows the inference of status information on components that are not directly detectable. Image-based color detection and a higher threshold for elements with possible formwork in place enable the correct identification of elements that are under construction at the time of capturing the site.

As a core contribution, the paper presents how CNN-based object-detection methods are applied to the captured images to correctly detect elements that tend to be otherwise falsely classified. Significant synergies are created by training the network with images that are automatically labeled, by applying Scan-vs-BIM techniques. The use of image-based object detection extends the reliability of the status-detection process significantly, due to the larger density of pixel-based information, in comparison with a pure point-cloud-based approach.

6.5.2. Limitations

It is crucial to note that the image data can only be used thanks to the photogrammetric process and the underlying camera pose estimation. Laser scanners usually do not provide this data and are therefore not suitable for this approach. Another limitation is the requirement for a well-aligned BIM. In our approach, this is achieved by markers on site. However, a minor manual step is required in order to find the exact orientation and scaling. Only after combining this data with the aligned building information model is it possible to gather additional information from the images in relation to the building model.

The described ML approach is limited to the provided training data. This data currently only includes construction sites in Germany, which might make the network biased and unsuitable for different regions that use different construction methods. The observed construction sites so far mainly used in-situ concreting and a small number of prefabricated elements.

6.5.3. Outlook

All introduced methods enhance the automated construction progress monitoring workflow. However, it is still the case that not all elements can be detected. Better acquisition methods will play an essential role in solving these issues. Several research groups have proposed different acquisition methods to detect indoor elements, too. A combination of all these methods could help to improve element detection even further.

More comprehensive data sets for image-based ML are required to cover different construction methods and materials from other regions.

7. Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning

Previously published as: Braun, A., Borrmann, A.: Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning, *Automation in Construction* 106, pp. 1-13, 2019, DOI: 10.1016/j.autcon.2019.102879

Abstract

Image-based object detection provides a valuable basis for site information retrieval and construction progress monitoring. Machine learning approaches, such as neural networks, are able to provide reliable detection rates. However, labeling of training data is a tedious and time-consuming process, as it must be performed manually for a substantial number of images. The paper presents a novel method for automatically labeling construction images based on the combination of 4D Building Information Models and an inverse photogrammetry approach. For the reconstruction of point clouds, which are often used for progress monitoring, a large number of pictures are taken from the site. By aligning the Building Information Model and the resulting point cloud, it is possible to project any building element of the BIM model into the acquired pictures. This allows for automated labeling as the semantic information of the element type is provided by the BIM model and can be associated with the respective regions. The labeled data can subsequently be used to train an image-based neural network. Since the exact regions for all elements are defined, labels can be generated for basic tasks like classification as well as more complex tasks like semantic segmentation. To prove the feasibility of the developed methods, the labeling procedure is applied to several real-world construction sites, providing over 30,000 automatically labeled elements. The correctness of the assigned labels has been validated by pixel based area comparison against manual labels.

7.1. Introduction

Large construction projects require a variety of different manufacturing companies of several trades on site (for example masonry, concrete and metal works, Heating, Ventilation, Air Conditioning (HVAC), ...). An important goal for the main contractor is to keep track of accomplished tasks by subcontractors to maintain the general schedule. Additionally, the documentation of correctly executed tasks plays a crucial role for all involved parties. In construction, process supervision and monitoring is still a mostly analog and manual task. To prove that the work has been completed as defined per contract, all performed tasks have to be monitored and documented. The demand for a complete and detailed monitoring technique rises for large construction sites where the complete construction area becomes too large to monitor by hand, and the number of subcontractors rises. Main contractors that control their subcontractors' work need to keep an overview of the current construction state. Regulatory issues add up on the requirement to keep track of the current status on site.

The ongoing digitization and the establishment of building information modeling (BIM) technologies in the planning of construction projects help to establish new methods for process optimization. In an ideal implementation of the BIM concept, all semantic data on materials, construction methods, and even the process schedule are connected. On this basis, it is possible to make much more precise estimations about the project costs and its duration. Most importantly, possible deviations from the schedule can be detected early, and the resources can be adapted accordingly.

This technological advancement allows new methods in construction monitoring. In Braun et al. (2017), the authors propose a method for automated progress monitoring using photogrammetric point clouds and 4D Building Information Models. The central concept is to use standard camera equipment on construction sites to capture the current construction state by taking pictures of the complete facility under construction at regular intervals. As soon as a sufficient number of images from different points of view are available, a 3D point cloud can be reconstructed with the help of photogrammetric methods. This point cloud represents one particular time-stamp of the construction progress (as-built) and is subsequently matched against the geometry of the BIM (as-planned) on a per-element basis.

Figure 51 shows the C#-based Windows Presentation Foundation (WPF) software tool, developed in the scope of this research. The tool visualizes a building information model and all corresponding semantic data. Additionally, the observation results can be selected and are supported by the possible overlay of the corresponding point clouds.

The presented approach can be varied in terms of acquisition method (laser scanning, manual acquisition, ...) and matching methods (as discussed in Section 2 - Related work). However, none of the methods is capable of providing absolute reliability due to occlusions or other boundary conditions. To further improve the reliability of the methods mentioned above,

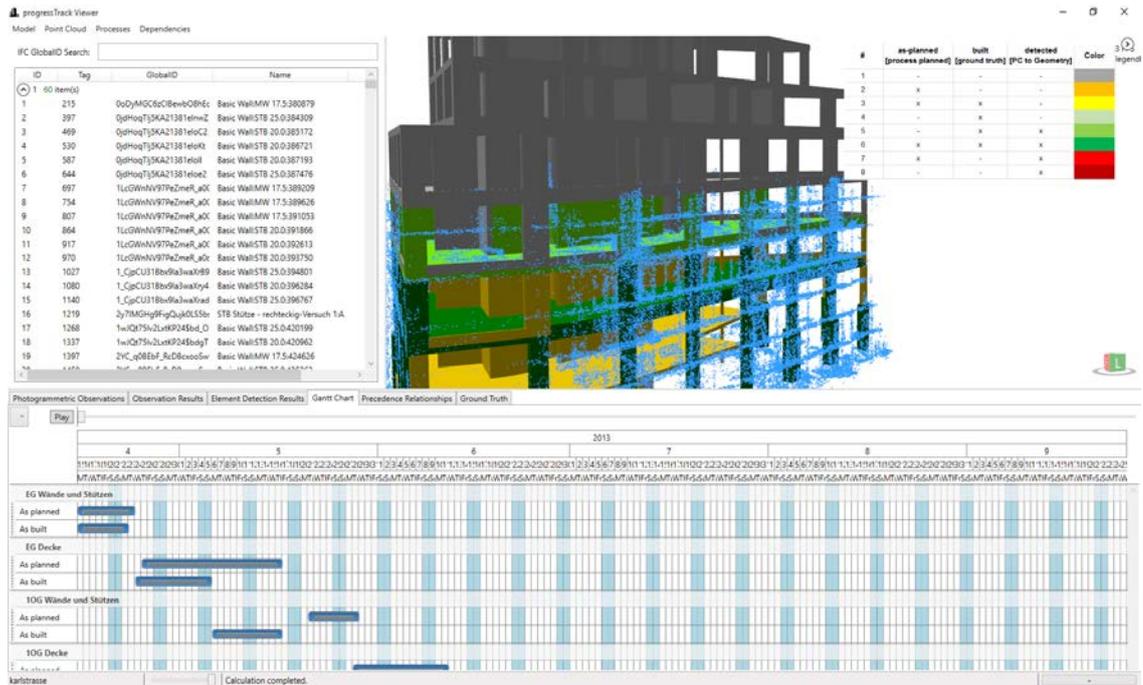


Figure 51 progressTrack: 4D BIM viewer incorporating detection states, process information and point clouds from observations

image-based machine learning techniques offer a promising approach. These techniques allow to analyze pictures based on their contents and even mark and classify specific regions of pictures. This new information can further improve the geometric as-planned vs. as-built comparison based on point clouds by increasing the reliability of made assumptions while comparing semantic data from the BIM with classified categories on similar pictures.

Recently, Convolutional Neural Networks (CNN) were introduced in this context (Girshick et al., 2014; He et al., 2016). These networks require large training sets to learn similarities of provided data-sets to make assumptions on unknown data. Applications of CNNs range from face-detection in security-related applications to autonomous driving (Redmon et al., 2016). With respect to automated construction monitoring, these methods can help to detect construction elements on pictures and provide an alternative method for detection in case of low point cloud densities and to improve the overall accuracy of detection (Dimitrov and Golparvar-Fard, 2014; Brilakis et al., 2005). However, data pre-processing and labeling of test-sets for the training of said algorithms is a laborious and time-consuming task since common CNNs require large amounts of labeled data (Lin et al., 2015).

This paper presents a method to automate the process of construction-site image labeling. The proposed method makes use of available information on image localization from the photogrammetric process as well as information on the presence of individual construction elements from the as-planned vs. as-built comparison by the process described above. The resulting availability of training data provides the basis for applying the trained CNN for image-based object detection on any construction site, in particular, those where a 4D-BIM does not

exist or only a limited number of images are taken, and the generation of a point cloud is not possible. However, this paper does not report on these next stages but focuses on the provision of correctly labeled images as an essential first step.

7.2. Related work

7.2.1. Automated construction monitoring

Several methods for BIM-based progress monitoring have been developed in recent years (Omar and Nehdi, 2016). Basic methods make use of minor technical advancements like introducing email and tablet computers into the manual monitoring process. These methods still require manual work, but already contribute to the shift towards digitization. More advanced methods try to track individual building components through radio-frequency identification (RFID) tags or similar methods (for example QR codes).

Current state-of-the-art procedures apply vision-based methods for more reliable element identification. These methods either make direct use of photographs or videos taken on site as input for image recognition techniques or apply laser scanners or photogrammetric methods to create point clouds that hold point-based 3D information and additionally color information.

Bosché and Haas (2008a); Bosché (2012) present a system for as-planned vs. as-built comparisons based on laser-scanning data. The generated point clouds are co-registered with the model using an adapted Iterative-Closest-Point-Algorithm (ICP). Within this system, the as-planned model is converted into a point cloud by simulating the points using the known positions of the laser scanner. For verification, they use the percentage of simulated points, which can be verified by the real laser scan. Turkan et al. (2012) use and extend this system for progress tracking using schedule information for estimating the progress in terms of earned value and for detecting secondary objects.

Kim et al. (2013b) detect specific component types using a supervised classification based on Lalonde features derived from the as-built point cloud. An object is regarded as detected if the type matches the type present in the model. As above, this method requires that the model is sampled into a point representation. Zhang and Arditi (2013) introduce a measure for deciding four cases (object not in place, point cloud represents a full object or a partially completed object or a different object) based on the relationship of points within the boundaries of the object and the boundaries of the shrunk objects. The authors test their approach in a very simplified artificial environment, which is significantly less challenging than the processing of data acquired on real construction sites.

In comparison with laser scanning, photogrammetric methods are less accurate. However, standard cameras have the advantage that they can be used more flexibly, and their costs are much lower. This leads to the need for other processing strategies when image data is

used. Omar and Nehdi (2016) give an overview and comparison of image-based approaches for monitoring construction progress. Ibrahim et al. (2009) use a single camera approach and compare images taken during a specified period and rasterize them. The change between two time-frames is detected using a spatial-temporal derivative filter. This approach is not directly bound to the geometry of a BIM and therefore cannot identify additional construction elements on site. Kim et al. (2013a) use a fixed camera and image processing techniques for the detection of new construction elements and the update of the construction schedule. Since many fixed cameras would be necessary to cover a whole construction site, more approaches rely on images from hand-held cameras covering the whole construction site.

For finding the correct scale of the point cloud, stereo-camera systems can be used, as done in (Son and Kim, 2010; Brilakis et al., 2011). Rashidi et al. (2015) propose using a colored cube of known size as a target, which can be automatically measured to determine the scale. Additionally, image-based approaches can be compared with laser-scanning results (Golparvar-Fard et al., 2011). The artificial test data is strongly simplified, and the real data experiments are limited to a small part of a construction site. Only relative accuracy measures are given since no scale was introduced to the photogrammetry measurements. Golparvar-Fard et al. (2011, 2015) use unstructured images of a construction site to create a point cloud. The orientation of the images is computed using a SfM process.

Subsequently, dense point clouds are calculated. For the comparison of as-planned and as-built geometry, the scene is discretized into a voxel grid. The construction progress is determined in a probabilistic approach, in which the threshold parameters for detection are determined by supervised learning. This framework makes it possible to take occlusions into account. This approach relies on the discretization of space as a voxel grid to the size of a few centimeters. In contrast, the approach presented here is based on calculating the deviation between a point cloud and the building model directly and introduces a scoring function for the verification process.

The mentioned approaches provide valuable enhancements for automated construction progress monitoring. However, so far, not all potential benefits from using semantic BIM data are unlocked to their full extent. Also, current research does not present solutions for occluded elements as well as temporary construction elements like scaffolds. These elements cover large parts of construction sites and thus cannot be neglected. The presented approach tries to solve this issue by analyzing the images taken during the SFM process.

7.2.2. Computer Vision

Computer Vision is a heavily researched topic, that got even more attention through recent advances in autonomous driving and machine learning related topics. Image analysis for construction sites, on the other hand, is a rather new topic. Since one of the key aspects of machine learning is the collection of large data-sets, current approaches focus on data gathering. In the scope of automated progress monitoring, Han and Golparvar-Fard (2017a)

published an approach for labeling based on the commercial service Amazon Turk.

Chi and Caldas (2011) used first versions of neural networks to detect construction machinery on images, Kropp et al. (2018) tried to detect in-door construction elements based on similarities, focusing on radiators. Kim et al. (2013) used ML-based techniques for construction progress monitoring. They analyzed images by filtering them to remove noise and uninteresting elements to focus the comparison on relevant construction processes. Other publications mainly focus on defect detection (like for example cracks) in construction images (Akinci et al., 2006).

Current research mainly uses manual labels for computer vision. Additionally, no construction data set is currently covering the whole amount of construction elements. An automated labeling approach could better this lack of data to further improve machine learning methods in this scope of application.

7.3. Problem statement

Monitoring of construction sites by applying photogrammetric methods has become a common practice. Currently, several companies (for example Pix4D, DroneDeploy) provide commercial solutions for end users that allows to generate 3D meshes and point clouds from UAV-based site observations. All these methods give reasonable solutions for finished construction sites or visible elements of interest.

However, there are still many unsolved problems in monitoring construction sites. Photogrammetric methods are sensitive to low structured surfaces or windows. Because of the used method, each element needs to be visible from multiple (at least two) different points of view. Thus, elements inside of a building cannot be reconstructed as they are not visible from a UAV flying outside of the building. Monitoring inside a building is currently still under heavy research (Kropp et al., 2014) and not available in an automated manner as orientation and observation in such mutable areas like construction sites is hard to tackle.

These problems lead to holes or misaligned points in the final point cloud, that hinder accurate and precise detection of building elements. On the other hand, laser scanning requires many acquisition points and takes significantly more time and manual effort for acquisition. Finally, both techniques remain with occlusions for regions that are not visible during construction.

As can be seen in Figure 52, another problem is elements that are occluded by temporary construction elements. Especially scaffolds and formwork elements occlude the view on walls or slabs, making it harder for algorithms to detect the current state of construction progress.



Figure 52 Occluded construction elements in generated point cloud caused by scaffolding, formworks, existing elements and missing information during the reconstruction process

This paper proposes a method that is meant to overcome some of the limitations of the available methods. It contributes to the final goal of exploiting images as an information source for construction state detection, either as additional information in case one of the methods mentioned above is applied, or even as sole and primary information if a 4D BIM does not exist or an insufficient number of images is available for photogrammetric detection. To achieve this, the authors propose to apply CNNs for automated object detection.

However, a huge set of correctly labeled images is required for training the CNN and achieve high precision and low recall. So far, the labeling process had to be performed manually in a laborious and error-prone process. This is why the authors propose to automate this process by making use of the methods they originally developed for construction progress monitoring. In particular, we use image localization from the photogrammetric process as well as information on the presence of individual construction elements from the as-planned vs. as-built comparison. This results in the availability of the required high quality, high volume training data.

7.4. Automated labeling of images

An essential part of progress monitoring is the detection of an element's status, i.e. to decide whether an element is still under construction (e.g., surrounded by formwork) or finished. Pure point-cloud-to-model matching methods are facing difficulties in this regard as temporary and auxiliary constructions (such as formwork) usually are not included in the BIM model. As proposed in Braun et al. (2018), computer vision based methods can help here and significantly improve the reliability of as-planned vs. as-built comparison. The basic idea is to use visual information to decide upon an element's visibility status.

The authors propose the use of machine learning (ML) methods for image-based detection of a construction element's status. However, ML techniques require a large set of labeled images for training. As currently large labeled sets of construction site images or not available, the labeling has to be performed manually in a tedious and time-consuming process. Generating these labels automatically can drastically reduce preparation efforts for training and improving such networks.

The proposed concept of automatic labeling is based on fusing information available from the photogrammetric process (images and relative position of the camera) and the information available from the 4D BIM (object type, object position). Since the BIM and the resulting point cloud are aligned, each BIM element can be projected onto the image initially taken for the photogrammetric process. This allows to precisely identify the region covered by a building element on a picture.

However, there is a significant problem remaining: Information on the actual presence of the element cannot be reliably taken from the 4D as-planned BIM, as execution time very often deviates from the original schedule (which is the underlying rationale for applying progress monitoring). At this point, we benefit from the original point-cloud vs. BIM matching process outlined in Section 1: It provides reliable information about the actual presence of an element in reality and thus also on the captured images.

Consequently, the proposed method for automated labeling of construction elements uses the data of previously monitored construction sites together with the results from the as-planned vs. as-built comparison to generate valid data sets for the training of neural networks.

The proposed workflow is also depicted in Figure 53.

As soon as the training is successfully completed, these networks can be used on any construction site for an image based detection of elements.

The following subsections describe the process and mathematical background for the projection of construction elements into pictures and the labeling procedure using these results.

7.4.1. Camera positions

In the proposed method, the point cloud is produced using photogrammetric methods. In this process, pictures are taken, for example by UAVs (Unmanned aerial vehicles) from different points of view. These pictures can then be used to generate a 3D point cloud if all elements are visible from a sufficient amount of viewpoints. During the reconstruction process, the camera positions around the construction site are estimated. This is illustrated in Fig. 54. This estimation is refined during the dense reconstruction and can get more accurate by using geodetic reference points on site.

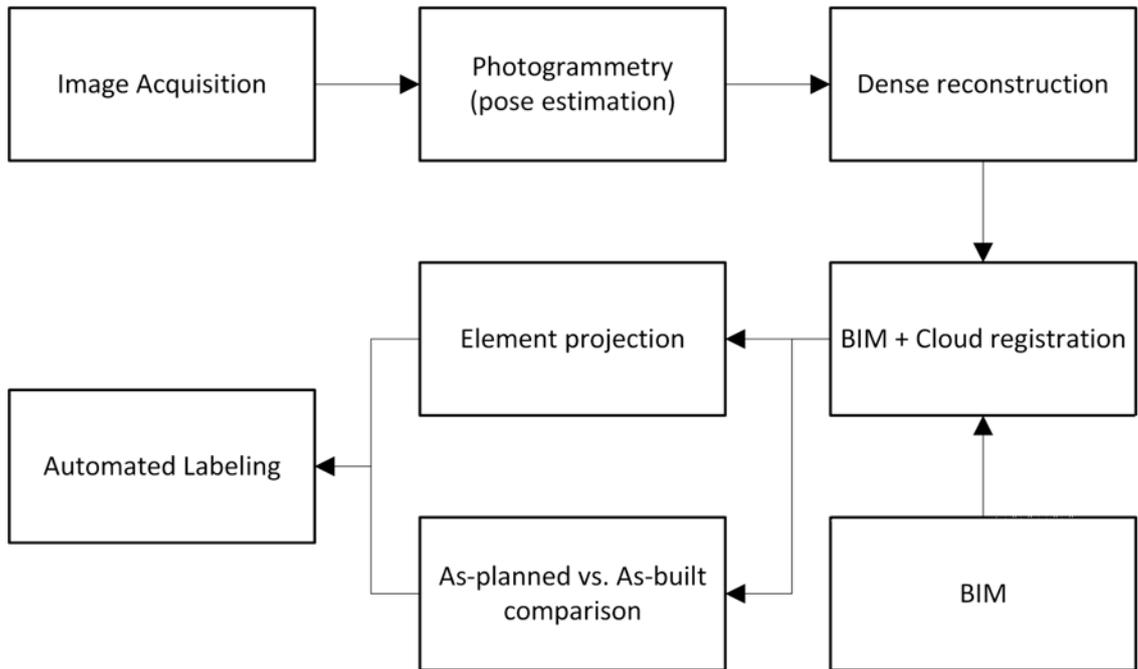


Figure 53 Proposed workflow for the automated labeling toolchain

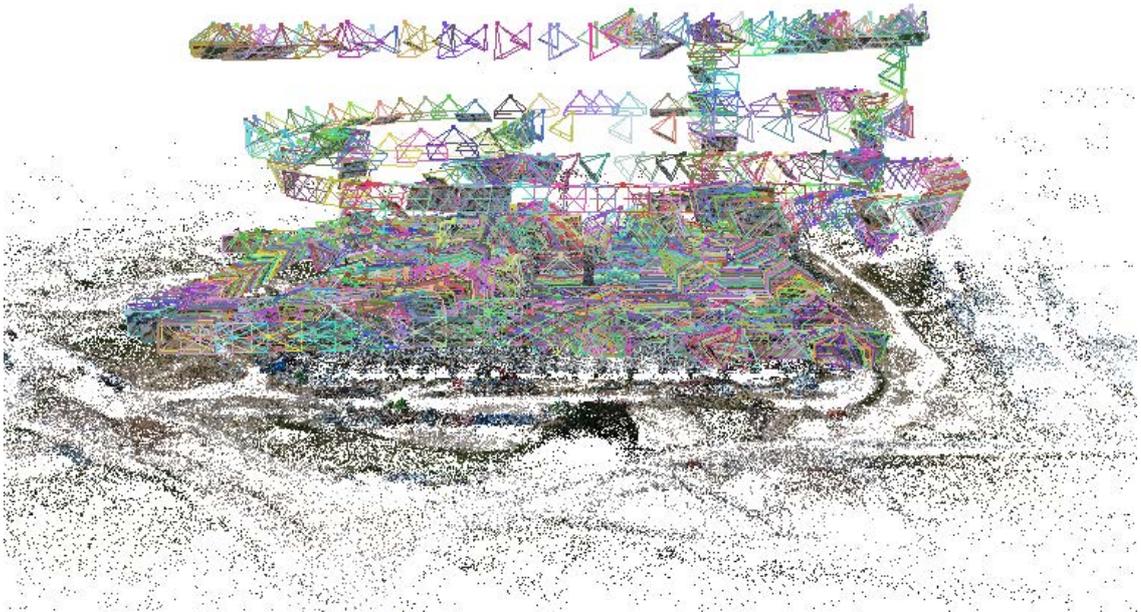


Figure 54 Estimated camera positions during point cloud generation (in this example using VisualSFM (Wu, 2013a))

7.4.2. 4D process data and as-planned vs. as-built comparison

Building information modeling can be used to combine the geometry of construction elements with semantic data such as material information but also process schedules. In the scope of this research, the corresponding process schedule is connected to all elements, resulting in a fine-grained 4D-BIM model. This allows identifying all elements that are expected to be built at each observation time.

As depicted in Fig. 55, the software tool used in this research is capable of integrating the building information model with process data and construction elements such as scaffolding and formwork.

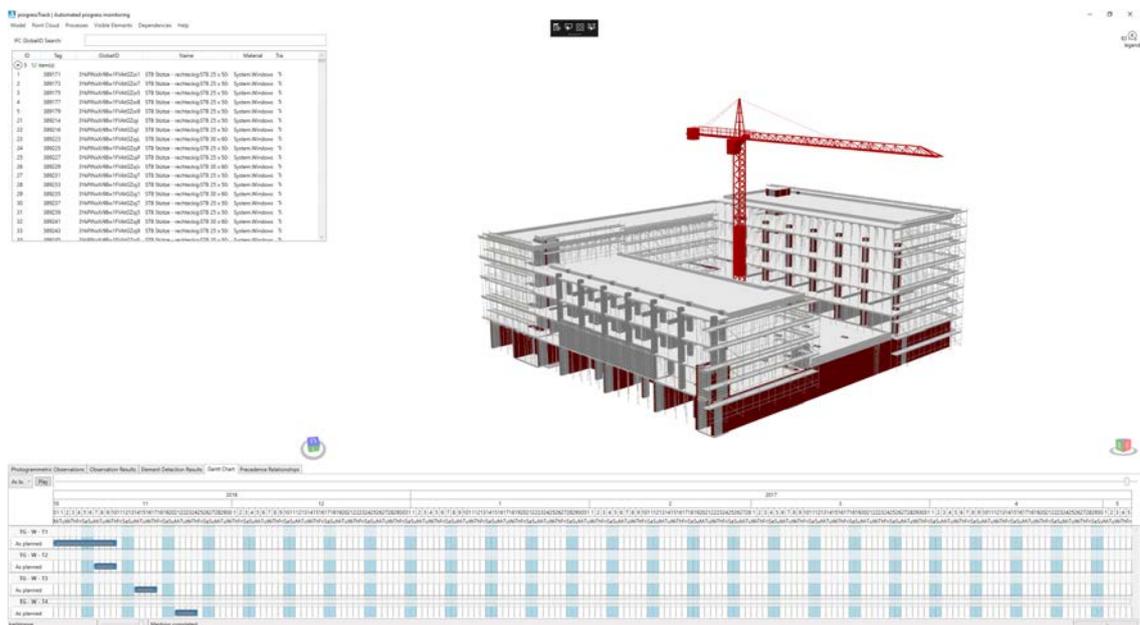


Figure 55 4D building information model including all additional construction materials like scaffolding and formwork

This data is required to define the sets of elements that are used for the labeling method described in this paper. Since the process schedule may change during construction, it is crucial to update the schedule permanently based on the gathered observation data. Since the as-planned vs. as-built comparison has already been conducted for the construction sites in this research, the results are available for all construction elements. This information is crucial since the labeling of elements that were not built yet would lead to incorrect labels.

7.4.3. Projection

Based on the gathered information, it is possible to do a visibility detection by using the camera positions as the point of view, and the process information to define the set of construction elements, that are meant to be built. To achieve this, the building model coordinate system needs to be transformed into the camera coordinate system or vice versa. Several parameters are needed for this transformation.

On the one hand, the intrinsic camera matrix for the distorted images that projects 3D points in the camera coordinate frame to 2D pixel coordinates using the focal lengths (F_x, F_y) and the principal point (x_0, y_0) is required. Additionally, the skew coefficient s_k for the camera is required. This scalar parameter defines the relation between x and y axis. It is zero if the image axes are perpendicular. The matrix K can be described as defined in equation 7.1.

$$K = \begin{bmatrix} F_x & s_k & x_0 \\ 0 & F_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7.1)$$

The translation of the camera is defined as:

$$T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} \quad (7.2)$$

Additionally, the rotation matrix for each image as defined in equation 7.3 is needed.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (7.3)$$

Both, translation and rotation can be described in one 3 x 4 matrix:

$$RT = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_1 \\ r_{21} & r_{22} & r_{23} & T_2 \\ r_{31} & r_{32} & r_{33} & T_3 \end{bmatrix} \quad (7.4)$$

Using the model coordinates of all triangulated construction elements, it is possible to calculate the projection of each element into the camera coordinate system and therefore overlay the model projection and the corresponding picture taken from the point of observation with equation 7.5.

$$t = K * RT * p; \quad (7.5)$$

The resulting 2D coordinates that are rendered into the picture are calculated by using the vector t and getting the x and y coordinates by calculating

$$x = t[0]/t[2] \quad (7.6)$$

and

$$y = t[1]/t[2] \quad (7.7)$$

This is done for each point belonging to the triangulated geometry representation of all construction elements.

As visible in Fig. 56 for an analytical column, the projection works as expected and helps to identify the respective construction element in the recorded picture. The mentioned calculations need to include an optional transformation and rotation if the model is geo-referenced and thus the two coordinate systems differ broadly.

7.4.4. Render model based on camera position

The algorithm introduced in section 7.4.3 enables the element-wise rendering of all construction elements in the respective coordinate system. To get a rendered image of all visible construction elements, the following steps are carried out:

Algorithm 2 Pseudo code for rendering an image of all visible elements

```

1: procedure RENDERVISIBLEELEMENTS
2:    $O \leftarrow$  set of all observations of the construction site
3:    $I \leftarrow$  set of all images of current observation
4:    $E \leftarrow$  set of all construction elements
5:    $C \leftarrow$  set of all coordinates of the triangulated surfaces
6:    $d \leftarrow$  distance of element to corresponding camera position
7:   for all  $O$  do
8:     for all  $I$  do
9:       for all  $E$  do
10:        for all  $C$  do
11:          if isvisible( $c$ ) then  $P(x,y,d,color) = \text{projection}(c)$ ;
12:   for all Pixels do
         $p_{min} = \min(P(d))$ ;
         $p(x,y) = p(color)$ ;

```

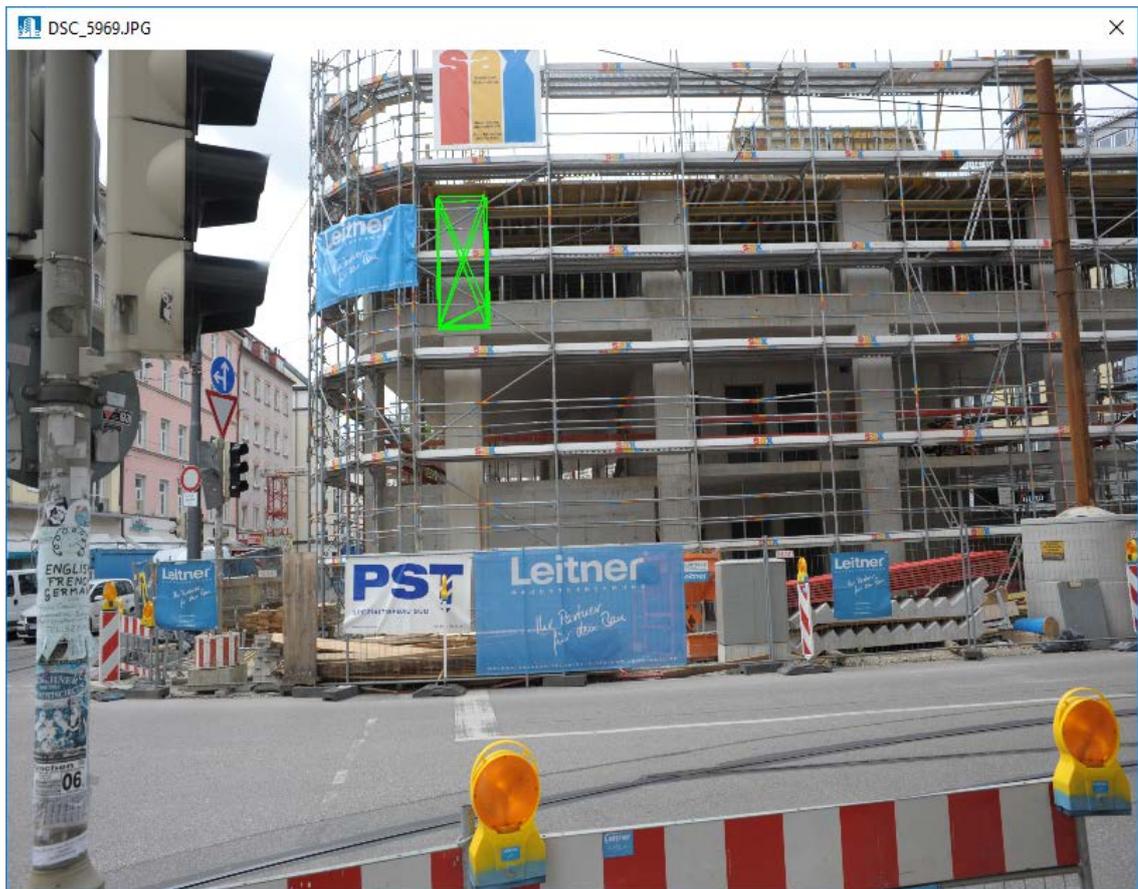


Figure 56 Sample of projected, triangulated column geometry into a corresponding picture

While all geometric information is available, three problems need to be solved for an accurate rendering of all construction elements:

1. For triangulated elements, only the boundaries are known. However, the whole surface needs to be rendered correctly.
2. The rendered surface needs to be connected to the corresponding element since this information is crucial for a proper visibility analysis
3. Elements may blend over from the viewpoint in some circumstances. This needs to be addressed to get a correct rendering.

The first issue is solved by applying necessary inside/outside tests for points inside a bounding box around each triangle. This is combined with min/max tests to verify that all points are inside the given coordinate system of the current picture. The second issue is addressed by assigning an individual color in the RGB color range to every construction element. This allows identifying each element after the rendering is finished.

The third issue is solved by applying the Painter's algorithm (Elvins, 2005) to each pixel in the given picture. In the given challenge, the distance to the point of view is stored for the current construction element and the color information is replaced in case an element has a smaller distance to the point of view and is also visible in the same pixel of the picture.

The applied algorithms result in a rendering as seen in Fig. 57.

After applying this technique to all observations and all camera positions, a distinct list of all visible construction elements can be generated by iterating over all pixels of each rendered image. The color of each pixel is assigned to a construction element, and since the painters' algorithm is applied, only the element is visible, that has the lowest distance to the point of observation. Therefore, all visible, non-occluded elements can be determined with this method.

7.4.5. Generating Labels for Machine Learning

Since machine learning tasks require large training sets for the learning procedure, the labeling and pre-processing of suitable data play a crucial role.

Labeling for ML depends on the desired output of the ML system. A basic ML system for classification is only capable of making general statements on the content of an image and hence only requires a set of images containing the classification category as training input. On the other hand, a system for semantic segmentation can predict the exact location and also the amount of (multiple) elements in one image. Labeling for this class of systems requires detailed convex hull polygons around all instances of elements. Additionally, the category for each label needs to be defined.

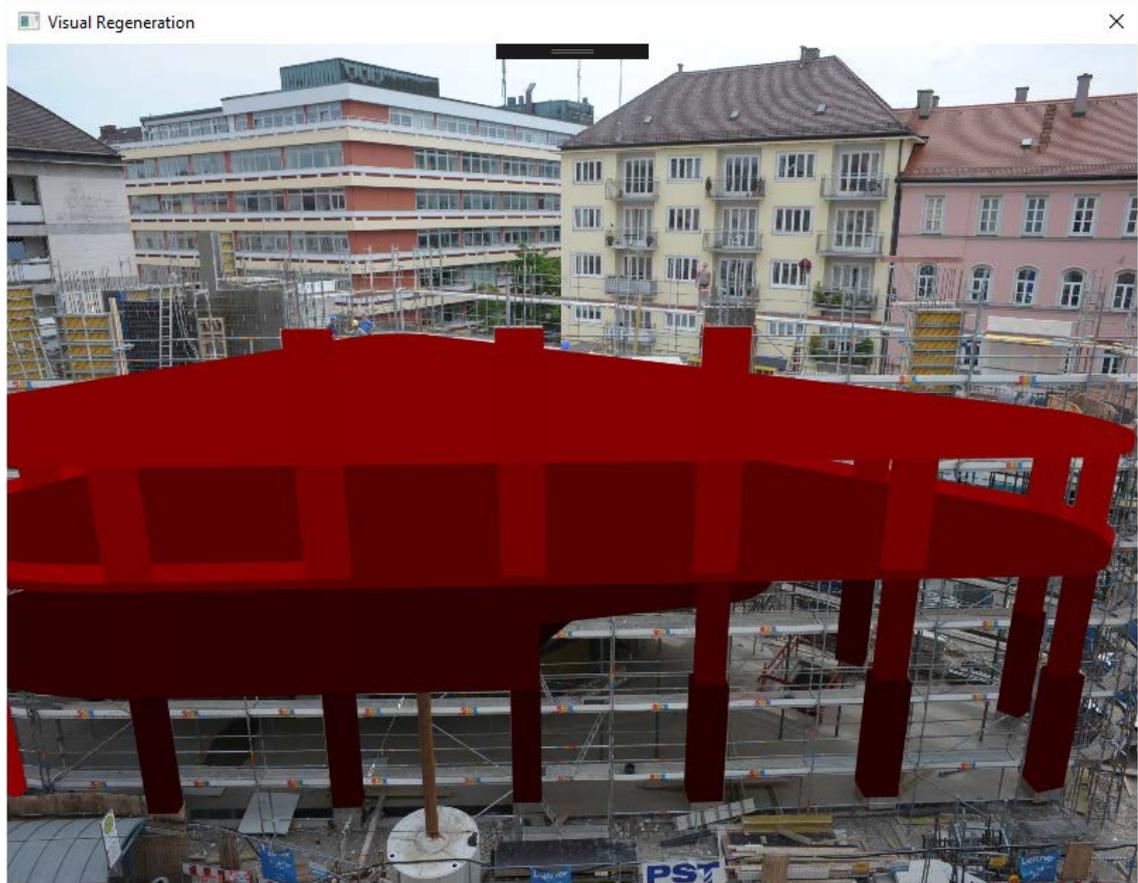


Figure 57 Using projection methodology for model rendering based on the Painter's Algorithm and 4D semantic information

The automated labeling process presented here builds on the previously presented projection algorithm and is capable of generating labels for all sorts of ML systems starting from necessary bounding boxes up to detailed convex hulls around individual element instances. Image-based labeling is realized by defining a polygon line around each object and associating a corresponding category with this label. The polygon label can be generated by the above-mentioned projection and fits precisely around the shape of each construction element. The defining element category can be extracted from the semantic information provided by the building information model. Since geometry and semantic data are connected, any additional information can be added to the generated labels.

Besides using the mathematical algorithms for projection, also the results of the visibility analysis are essential. As discussed before and depicted in Figure 58, labeling cannot only rely on all available elements. A prominent but noteworthy factor is the actual presence of the labeled element. The element must have been built to generate an image valid for training or testing. By extracting this information from the as-planned vs. as-built comparison, the set of available elements is reduced to the set of detected elements. In the next step, the set needs to be further reduced to the set of visible elements for each picture.

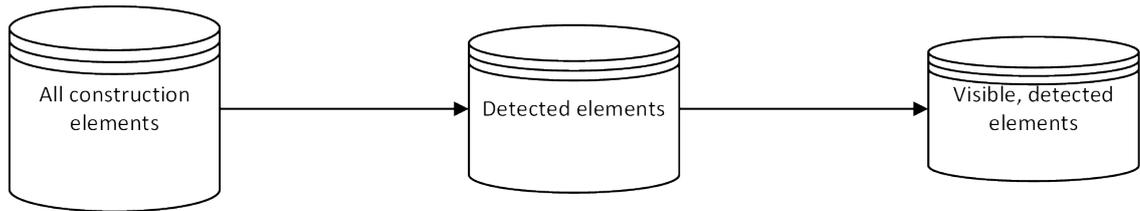


Figure 58 Considered set of elements based on previous results from as-planned vs. as-built comparison. For CV-based methods, visibility plays a crucial role, leading to reduced data sets from construction monitoring.

To sum up the labeling process, the following method is proposed:

Algorithm 3 Pseudo code for labeling all visible elements in an image

```

1: procedure LABELVISIBLEELEMENTS
2:    $I \leftarrow$  set of all images
3:    $List < element, List < P >> LabelList \leftarrow$  set of all labels
4:   for all  $I$  do
5:      $E \leftarrow$  set of all construction elements, visible in current picture
6:     for all  $E$  do
7:        $List < P > ConvexHull = GetConvexHull();$ 
        $LabelList.Add(E.elementtype, ConvexHull);$ 
  
```

The proposed method works for all kinds of label requirements. To illustrate this, Table 18 shows sample labels for a Classification Network (which usually requires image snippets with bounding boxes) and semantic segmentation (which usually requires polygon lines and the corresponding images). The sample shown here uses the well known COCO format. For better understanding, a graphic labeled image for semantic segmentation is added, too.

Category	Classification	semantic seg. [JSON coords]	seg. image
column		[[3778, 1230, 3810, 1230, 3834, 1231, 3837, 1230, 3840, 1230, 3854, 983, 3852, 984, 3848, 984, 3791, 985]]	
formwork		[[2662, 1662, 2666, 1682, 2703, 1682, 2704, 1323, 2702, 1319, 2699, 1314, 2663, 1314]]	

Table 18 Sample labels of two categories for different ML use cases

Current best practice in machine learning proposes to split the labeled data-set into a set of training data for the actual training process, a set of validation that does not contain any data from the training set to validate the current training rates. Finally, a set for testing that is not used for training or validation at all is used for checking the overall performance of the neural network without further changing the learning parameters.

Hence, the labeled images are split randomly into the mentioned categories to fulfill this requirement.

7.5. Case study

To prove the introduced methods, the following case studies were conducted:

7.5.1. BIM element projection onto images

The developed methodology has been applied to several construction sites.

As depicted in Fig. 52, most observations lack details at some point and have mostly occluded areas due to the observation methods. In very disadvantageous observations, the detection rate can drop down to 50% of the overall built construction elements. In this case,

the detection rate d describes the percentage of elements that the proposed method marked as detected over the ground truth of all elements that were built. The latter set of elements has been acquired manually in order to verify used algorithms.

With the help of the presented methods, these rates can be explained since most of the undetected elements were not visible from the observation points. To quantify the efficiency of an algorithm for as-planned vs. as-built detection, it is essential to have a valid ground truth to allow an unbiased evaluation of the used methods. This approach helps to quantify the used methods correctly.

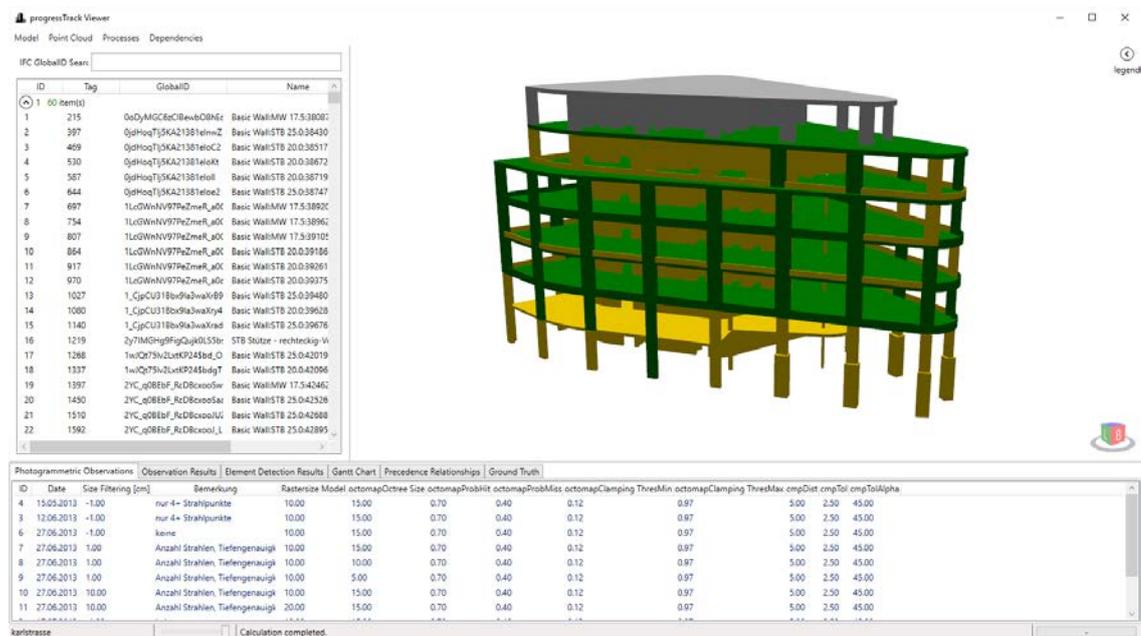


Figure 59 Detected construction elements from one observation. Green elements were successfully detected, yellow elements were not detected but are built.

This concept is illustrated in Fig. 59. The green elements were detected correctly. The yellow elements, however, are built but were not detected. This is because the inner walls were not visible from a sufficient number of viewpoints. Thus there were not enough points in the corresponding point cloud that allowed to validate the existence of the elements. However, these elements were identified as not visible using the method introduced earlier in this paper.

7.5.2. Automated labeling and validation

After successfully testing the projection, the actual labeling is performed being the key contribution of this paper.

Many currently used CNNs rely on the COCO Data-set (Andriluka et al., 2014). Facebook's Mask R-CNN (He et al., 2017) has provided promising results for machine learning in previous applications. The network itself also relies on the COCO data format as a basis. Thus, the authors chose this schema as a basis for the generation of the labels. This schema requires a defined structure for all labels, including information about each image (id, width,

height, license, date captured), all annotations (id, corresponding image, label category, label polygon, bounding box, ...) as well as the defined categories (in this case for example walls or columns, all represented by individual IDs).

The construction projects on which the developed methods have been applied involve mainly the production of concrete elements. The following construction elements and temporary elements were modeled in the corresponding BIM:

- columns
- walls
- formworks
- slabs
- roofs
- stairs



Figure 60 Sample sub-set of auto-labeled columns in one picture from a construction site.

The proposed methods were tested on observation data from multiple construction sites, resulting in 32,787 labeled construction elements on 1,300 images. The machine used for this test is a Windows 10 system equipped with an Intel Xeon E5-1630 CPU @ 3.70GHz, 16 GB DDR4-RAM, AMD Fire Pro W4100 2GB RAM, and a 10Gbit network connection.

The entire automated labeling process took around 20 minutes, outperforming manual labeling significantly. During this time, all images (close to 9 GB) were downloaded from a NAS (Network attached storage) and randomly added to the training, validation, and test data

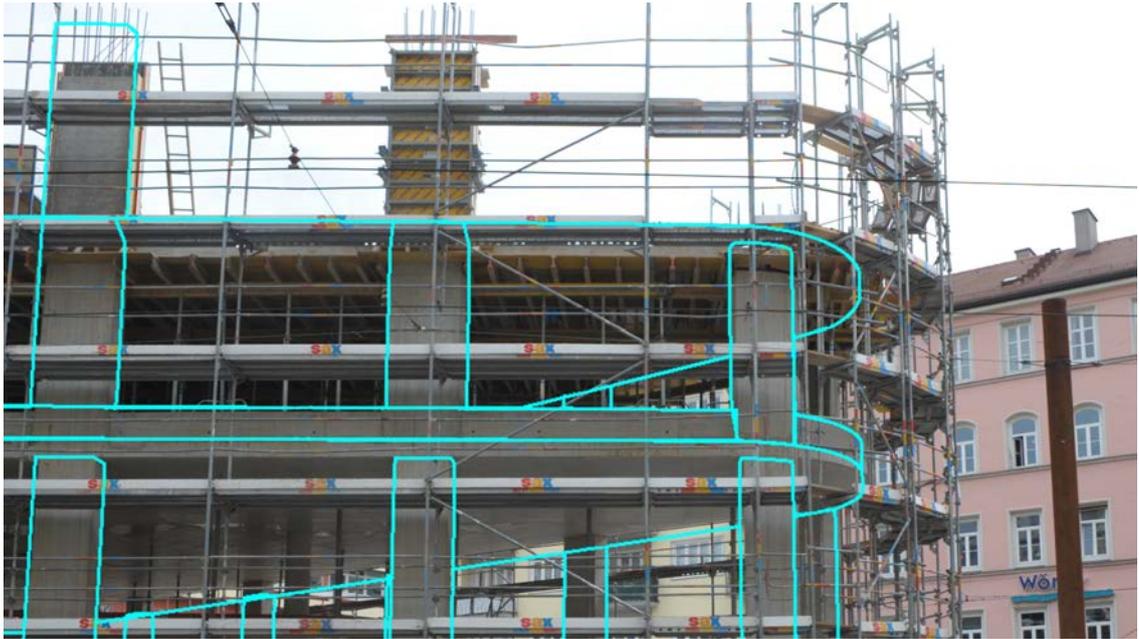


Figure 61 Sample sub-set of auto-labeled columns and walls in another picture from a construction site.

sets. Additionally, the corresponding label files in JSON format were generated. A sample visualization of the generated labels for one picture is depicted in Figure 60 and 61, showing exported columns with their respective convex hull label around them.

The method was validated through human evaluation on all labels for the tested construction sites. The label projection worked without failure for all built construction elements in terms of generating a valid convex hull as the existing elements have been verified against a manually created ground truth. Since no issues were found in a set of over 32,000 snippets, the projection can be regarded as working correctly. However, as depicted in Fig. 62, the automatically generated labels (cyan poly-lines) have a slight deviation from the actual construction elements.

This deviation can have multiple reasons:

- errors in pose estimation during Structure-from-Motion
- large scale deviations when using real-world coordinates
- construction inaccuracies
- modeling inaccuracies

Since all elements were validated in the as-planned vs. as-built comparison, allowing for only very minor construction inaccuracies, construction inaccuracies can be disregarded in this research. Otherwise, the element would not have been classified as "built" and would not have been labeled at all.

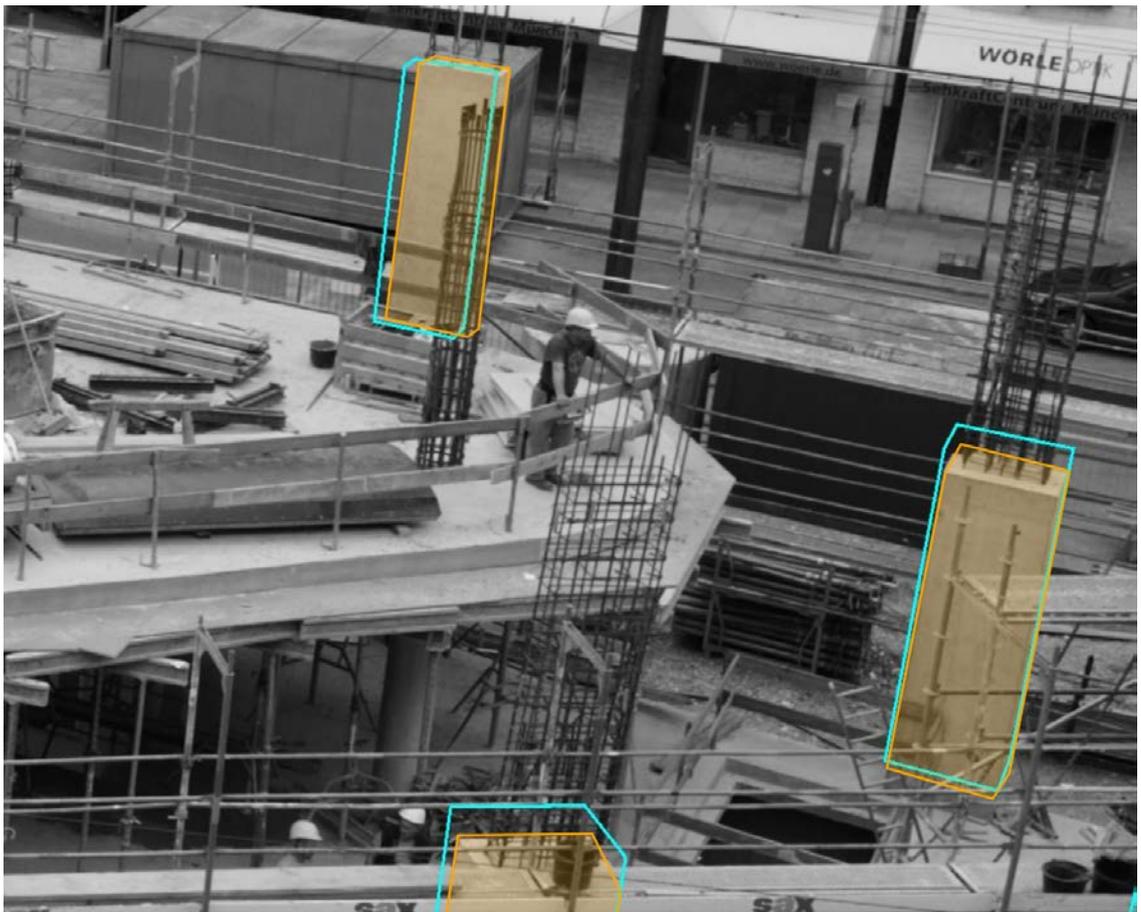


Figure 62 Validation of label correctness with cyan poly-lines representing the automatically generated labels and orange poly-lines representing the manually generated validation set.

Thus, the deviations, in this case, are minor and arise from an aggregation of the mentioned reasons. To quantify the introduced error, a set of 1.000 elements have been labeled manually and tested against the automatically created labels. The labels were then compared pixel-wise. The overall accuracy of the automated system was measured by calculating the overlapping area I of the resulting labels of both labeling methods over the manually labeled area:

$$p_o = I/A_{manuallabel} \quad (7.8)$$

with

$$I = A_{autolabel} \cap A_{manuallabel} \quad (7.9)$$

The resulting accuracy p_o had an average of 91.7% overlap over all checked labels, constantly lying within the bounds of 85% and 97%. The overlap rates give promising results and make the labels usable for machine learning tasks. Rates could be further improved by taking more pictures for the Structure-from-Motion process and enhancing the resulting camera pose estimation.

7.6. Discussion

For improving the reliability of construction progress monitoring, this paper introduces a novel concept for automating the labeling process of construction site images. It is based on fusing information available from the photogrammetric process (images and relative position of the camera) and the information available from the 4D BIM (object type, object position). Since the BIM and the resulting point cloud are aligned, a digital element can be projected onto the image, initially taken for the photogrammetric process. Also, matching the point cloud and the BIM allows to make sure that only images are considered where the elements under consideration exist in reality.

From the projected BIM elements, it is possible to automatically connect the covered image segments with the semantic information provided by the Building Information Model. Since the introduced as-planned vs. as-built comparison also offers valuable information on the presence of all elements, the labels can be further refined regarding possible occlusions. As a valid label should only be applied to an at least partially visible element, the gathered knowledge from the previously applied as-planned vs. as-built comparison makes this automated approach even more accurate. Since the comparisons' resulting elements are built at the correct positions, the labels are also correct. On the downside, only elements that were built as-planned can be labeled.

The sample-based validation showed over 91% pixel-wise accuracy of the automated procedure when tested against manual labeling procedures. A previously tested, manual labeling approach took over 100 working hours to accurately label only one category of elements. Labeling and generating the corresponding images folders for this case study took around 20 minutes, including downloading of 9 GB of pictures from a remote NAS folder which takes over 90% of the time used. Additionally, several studies show that manual labeling is also introducing a range of errors due to missed elements or inaccurately labeled elements. As the correct identification of construction elements also requires technical personnel (Han and Golparvar-Fard, 2017b), labeling is hugely cost intensive and danger of bore-out to this group of workers due to the repetitive work.

The construction sites used for this process are located in Germany and apply in-situ concrete pouring as the primary construction methodology. Consequently, the resulting labels and especially the trained network, will only be able to detect construction elements from this domain of manufacturing. However, the presented approach can be easily extended by also including construction sites from other countries or other construction techniques.

Future steps of this research will focus on creating a CNN for detecting the most important construction elements on construction sites. The final objective is to enable a utterly image-based construction monitoring process in the future.

Acknowledgments

We thank the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities (BAdW) for the support and provision of computing infrastructure essential to this publication. We thank the German Research Foundation (DFG) for funding the initial phases of the project.

Additionally, we would like to thank all construction companies, providing data and/or support to the research conducted in this field, including: Leitner GmbH & Co Bauunternehmung KG, Kuehn Malvezzi Architects, Staatliches Bauamt Muenchen, Baugesellschaft Brunner+Co, BKL (Baukranlogistik GmbH), Geiger Gruppe, Baureferat H5, Landeshauptstadt Muenchen, Baugesellschaft Mickan mbH & Co KG, h4a Architekten, Wenzel + Wenzel and Stadtvermessungsamt Muenchen.

8. Discussion

The presented methods in the Chapters 4 to 7 introduce enhancements for an automated way of construction progress tracking. A specific focus lies on the combination of point cloud-based acquisition methods on the one hand and geometric and semantic data from digital planning processes on the other hand. This data is further enhanced by CV and ML methods.

This chapter will discuss the main findings of the conducted investigations and the main characteristics of the introduced methods, while Chapter 9 provides an outlook and introduces possible future extensions in the scope of this research area.

Digital 4D models for construction planning enable planners to combine schedules with the 3D geometry. With this digitized approach, new monitoring techniques become available. Point cloud-based representations that are gathered from laser scanners or photogrammetric methods provide digital representations of the as-built status on construction sites. Combining these two digital data sets makes it possible to automate the labor-intensive task of progress tracking in large parts of the construction site.

8.1. Review

Automated progress monitoring can provide several advantages to this currently manual process. Besides the reduction of manual and error-prone work, it allows for informed decision making based on real-time information on the current progress on-site. The manual assessment of construction progress can be reduced significantly by acquiring the as-built state with automated methods. By transforming this work into a digital process, all acquired data and information can be distributed to construction managers, independently of their current working location. Thus, the whole management can take place independently, and any parties involved in the construction process can immediately check the current status of construction.

The introduced methods in this thesis extend the state of the art in progress monitoring by several new approaches, like semantic analyses and CV. The results show that depending on the type of construction and the type of occlusions, the detection of built elements can rise by up to 50% compared to an SfM-based, purely geometric as-planned vs. as-built comparison (see Chapter 6). As introduced in Chapter 5, acquisition can be automated by using image-based acquisition methods like crane cameras or UAVs.

8.1.1. Acquisition

As elaborated in Chapter 5, the acquisition of construction sites faces many challenges. Currently, laser scanning or photogrammetric methods are used to generate point clouds representing the as-built status of a construction site.

Image acquisition can be automated by either using UAVs or webcam-like cameras mounted on cranes. In this regard, photogrammetric methods proved useful, since they provide faster acquisition times in comparison to laser scanners, also they are cheaper and easier to use. Compared to laser scanners, photogrammetric methods are significantly faster during acquisition, and easier to implement (Omar and Nehdi, 2016). However, they can only provide an accuracy of +/- 5cm, while laser scanners can achieve better accuracy in the region of sub-centimeters. As elaborated in Chapter 6, this accuracy is sufficient for progress monitoring.

As introduced in Chapter 5, crane cameras provide continuous monitoring without any additional monitoring efforts. However, reconstruction accuracy in the z-direction is limited due to insufficient coverage from side-wards facing camera positions. This problem can be solved by placing additional cameras around the construction site. However, these positions need to be carefully selected to ensure a continuous, unobstructed view on the construction site.

The UAV-based acquisition allows to focus monitoring on regions of interest during flight time and additionally provides the possibility to ensure mostly unobstructed views on these regions due to optimal camera positions from above. This method proves very suitable for most situations. However, indoor monitoring is currently not possible with this method. Besides, current regulations in most countries require a pilot to actively overlook the entire flight time, resulting in manual work during acquisition.

As investigated by other research groups, indoor monitoring can also provide valuable results on the construction progress (Kropp et al., 2018). These methods require significant efforts on positioning since GPS signals cannot be received, introducing new vectors of errors with high possibilities of error propagation, the further monitoring goes without any checkpoints.

In conclusion, there is no monitoring solution available that offers 100 % coverage without any manual work during acquisition. Nevertheless, the introduced methods provide significant benefits for capturing the as-built status and facilitate a comprehensive data gathering while minimizing manual labor. Despite the still manual task of flying a drone over the construction site, this method provides valuable data and documentation of construction progress.

In the near future, more automated UAV flights could even further automate the progress. Besides legal reasons, localization and distance sensors become more accurate, providing new possibilities to plan automated flight routes over and even through construction sites. In combination with crane cameras, head-mounted cameras, and additional sensors, these methods will provide additional benefits to the automation of construction monitoring (See Chapter 9).

8.1.2. Scan-vs-BIM matching

The comparison of the as-planned geometry with the captured as-built point cloud data requires registration and geometric alignment of said data. Based on the research results in Chapter 5, it is proposed to position the BIM at the actual geodetic measured reference points manually. Markers on-site can geodetically reference the as-built point cloud to provide reliable accuracy compared to automated methods like ICP. The Scan-vs-BIM comparison is proposed to be performed on a geometric basis in the first place. This method applies to any point-cloud based approach since the general mathematical comparison approaches apply to each of these methods. This approach provides good results for quick feedback on the current construction state (see Section 9.1). It is crucial to note that this part of the comparison solely concentrates on geometric matching. Especially deviations from the initially planned positions cannot be identified with this approach. However, process-based (4D) deviations can be identified if elements are not occluded. Processes like paint jobs that do not introduce any geometric changes can not be assessed with this method.

Due to the nature of this comparison, occlusions on construction sites or inaccurate reconstructions during acquisition lead to inaccurate results. Experiments with real-world cases studies (Section 6.4.6 and 9.1) have shown that true positive detections of construction elements can decrease to around 50% of the overall available elements (ground truth). Based on the results presented in Chapters 5 and 6, it is elaborated that additional measures are required for more precise results. They are discussed in the following Sections.

8.1.3. Analysis of semantic information for Precedence Relationship Graph generation and color-based detection

To further increase robustness, semantic data, including material and process information, is extracted from the digital building model. This data is, among others, used to derive technological dependencies for the building based on its load-bearing structure. It includes color information but also attributes like "load-bearing" or the type of construction. Chapter 6 introduces a workflow that automatically derives precedence relationships from digital models.

The dependencies are stored in a graph to be easily accessible for digital analysis. These relationships are used for inferring the existence of elements that are not visible in the point cloud due to occlusions. As visualized in Figure 63, this graph contains a vertex for every element, making the overall graph quite large, containing several thousands of nodes. Query-wise, however, it is computationally cheap to request depending elements based on their dependencies in a graph (Zhao and Han, 2010).

The PRG allows to identify up to 50% more elements in comparison to a purely geometric approach. It works particularly well for load-bearing elements. Other elements might not possess enough dependency relationships, hindering the inference of the as-built status.

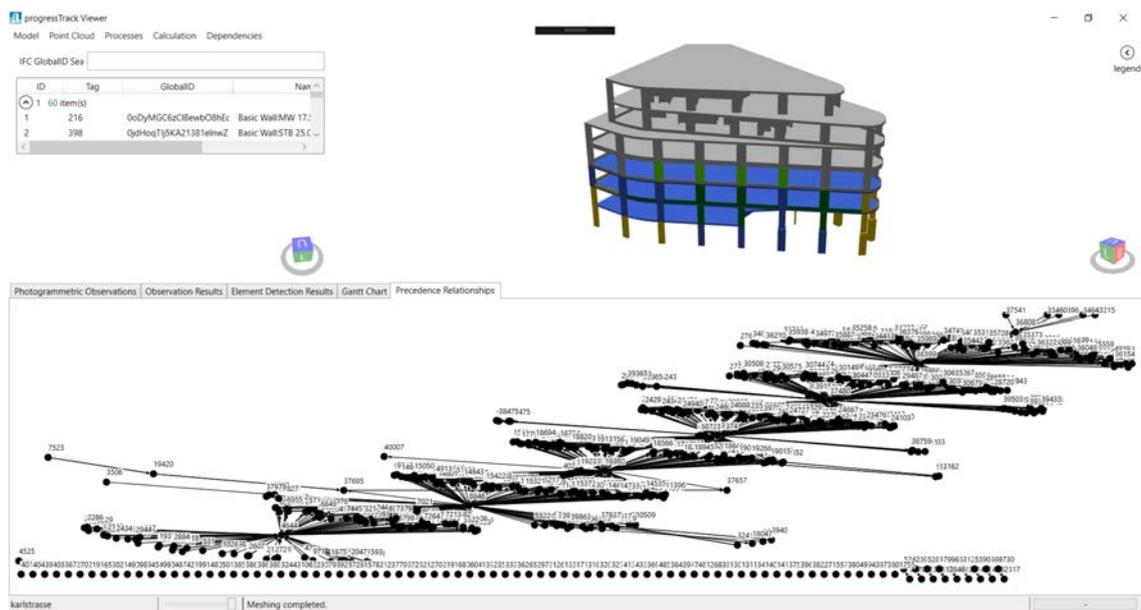


Figure 63 Automatically generated PRG visualized together with the corresponding model.

Besides, material-derived color analyses are introduced. Especially for elements under construction, this approach provides data to distinguish between actual construction elements and temporary elements, as these elements usually require temporary items such as formwork (see Section 6.3.7). The Hue, saturation, value (HSV) color space describes the color as perceived by humans and provides the basis for this method. Construction elements can be assessed regarding their colors. The grey color of concrete can be distinguished clearly from other colors like red or yellow formwork elements. These items' colors differ largely from the element's color. The validation of color values provides further certainty to assess the status of a construction element.

8.1.4. Computer Vision and Machine learning

As introduced in Chapters 6 and 7, computer vision can help to gather further information about the construction progress based on a computational analysis of images. SfM methods are based on estimating the camera positions for all images used for reconstruction. The resulting camera position, including the view vector, is used for all methods applied in the scope of CV.

Firstly, CV methods are applied to identify visible elements during acquisition based on the camera position during acquisition. This step is introduced to distinguish between the set of visible and invisible construction elements. Based on these results, invisible elements can be removed from any additional detection steps, resulting in better computing performance during the as-planned vs. as-built comparison (see Section 6.4.4). Secondly, these elements can be further investigated with the semantic methods, as summarized in Section 6.3.7. This includes, e.g., checks whether an element is encased in formwork or whether it is detectable via the color information from its material.

With the knowledge on the position of all elements on every acquired image, an image-based evaluation follows. While most methods that make use of point clouds can benefit from the 3D spatial information, methods that only rely on images are lacking position and geometry information from the 3D space. However, images have a higher density, resulting in more potential data sources. The introduced projection approach makes use of both data sources, combining the benefits of both approaches. On the one hand, the estimated camera positions from the SfM process are needed to position an image concerning the model. On the other hand, the aligned BIM model is needed to transform the 3D geometry of every element into the 2D plane of an image.



Figure 64 Detected construction elements on an image taken during monitoring.

Image analysis and ML supported object detection on ordinary pictures is well-established, and so, a useful addition to progress monitoring is presented. After training a Mask R-CNN-based network, construction elements can be detected on images at a pixel-wise level (see Fig. 64). Introduced in Section 6.4.5, the network has a mean average precision (mAP) of over 90 %.

8.1.5. Automated labeling

While the developed method is in principle performing well, there is currently no comprehensive data set for construction elements available. These data sets are required to build reliable ML systems that can be trained to detect construction elements. The training process of said networks is computationally expensive and requires large amounts of images. Currently, labeling is done by trained personnel, which is expensive and has a high danger of "bore-out". Manual labels of scaffolding, as shown in Figure 65, impressively visualize the high effort required and the complexity of such tasks that demand a high amount of concentration.



Figure 65 Manually labeled scaffolding structures on an image from our data-set.

To overcome the necessity of laborious manual labeling as introduced in Chapter 7, the projection-algorithm mentioned above can also be adapted to image labeling for ML. This approach is solely possible once the Scan-vs-BIM comparison results are available, and a statement can be made on the presence of an element in an image. Consequently, this method can only be applied in the aftermath of a Scan-vs-BIM comparison, not providing any benefit to this particular construction site. The fusion of all these results from the various introduced methods now provides this by-product that can be used to train a dedicated CNN for construction elements.

Depicted in Fig. 66, a small subset of the generated elements is shown. The quality of the resulting labels depends on the estimated camera positions during the SfM process. In the conducted research in Chapter 7, the resulting accuracy p_o had an average of 91.7% overlap over all checked labels, always lying within the bounds of 85% and 97%. The overlap rates give promising results and make the labels usable for machine learning tasks. Rates could be further improved by taking more pictures for the Structure-from-Motion process and enhancing the resulting camera pose estimation.

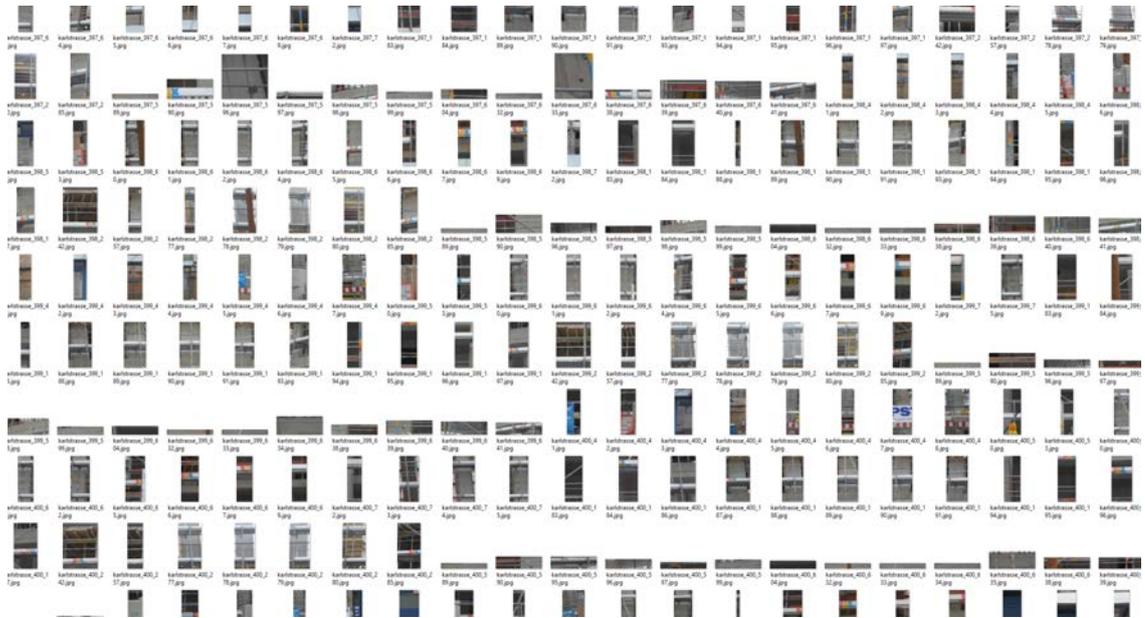


Figure 66 Automatically generated labels for all types of construction elements, present on the case-study construction sites.

8.2. Research contributions

This work was able to advance automation in the area of construction progress monitoring significantly and thus made a notable contribution to increasing transparency as a basis for Informed Decision Making. The acquisition of construction sites has been assessed, and a UAV-based monitoring approach is proposed (Chapter 5) as also introduced by other research groups (Golparvar-fard et al., 2009; Ham et al., 2016). All introduced methods built upon the current state of the art in construction progress monitoring.

The image-based acquisition uses SfM methods to reconstruct 3D point clouds to represent the as-built status of the construction site. In the context of shading and indoor acquisitions, there are still challenges to be met. In detail, the presented approaches make extensive use of semantic information provided by the BIM model. For this purpose, several new methods are introduced.

The structure of the model is analyzed with a query language to gain knowledge on the technological dependencies of the building (Chapter 4). These are stored in a precedence relationship graph that can be used to query depending elements for each detected construction

element. In combination with the introduced visibility analysis (Chapter 6), this information can be used to identify occlusions based on observation points according to the SfM process. The color values derived from the images are used to compare the expected values based on material information (Chapter 6).

ML based object detection on image basis has been introduced in Section 6.4.5. The presented combination of SfM and CV introduced a new way to project elements directly into the image plane. With this knowledge, image-based comparisons become possible that can be directly linked to the corresponding elements.

Finally, a novel approach in the scope of ML image labeling is introduced. It makes extensive use of the Scan-vs-BIM comparison and its result in combination with SfM. This method provides an automated way to label previously detected construction elements on image-basis.

9. Possible Extensions and Outlook

Computational power, as well as automation on construction sites, will rise in the coming years. In combination with more accurate acquisition methods, progress monitoring will become more precise.

This thesis introduced new methods that partially automated this currently tedious work. While the proposed workflows work well for the monitoring of structural works, several steps can be improved. The presented results also show possible extensions for other requirements besides progress monitoring.

9.1. Analysis and Performance Monitoring

This thesis laid a strong focus on the comparison of the as-planned objects that are expected as per the process schedule with the as-built point cloud. The results are not used any further since this has not been part of the research question addressed within this project. Nevertheless, the set of detected elements and, subsequently, the identified stage in the process plan provides data for further analysis.

Especially performance analysis in combination with Quantity take-off (QTO) can be performed, including the determination of performance values such as "used concrete per day" or the number of built walls per section. These results can help construction managers to, i.e., identify staff shortages. Another possible application scenario can be the management of just-in-time deliveries that require adjustments according to the detection results.

9.2. Extension to indoor acquisition

Currently, the introduced approaches only work reliably for the monitoring of structural works. Several approaches show the feasibility of indoor monitoring by overcoming localization issues with markers or, i.e., path-finding algorithms. Also, current research introduces combinations of laser scanning and cameras that are mobile and thus become suitable for construction sites.

The combination of these methods can provide even better results and overcome some of the issues, a solely UAV-based approach needs to tackle.

9.3. Quality assessment and monitoring

Quality assessment has multiple aspects that need to be considered. One crucial aspect of quality control is to validate whether all construction elements are built up as designed. This includes the check for correct placement as well as geometric correctness of the as-built element.

Another part is material check for cracks or similar visual defects (Zhou and Song, 2020). These defects usually occur if too little reinforcement has been placed or sheer forces are too high. Identifying and documenting these issues in an automated manner would provide a great benefit to construction managers. First steps are sketched in Braun et al. (2019), proposing the introduced data from SfM methods.

Another valuable extension to the introduced methods is the monitoring of built structures. Defects due to corrosion or constant overloads lead to cracks in concrete. A comprehensive overview of methods to monitor these defects with CV is provided by Koch et al. (2015).

9.4. Digital construction diary and documentation

Mentioned in Section 3.2.1, digital construction diaries are already implemented on several construction sites. However, these diaries lack a direct connection to the digital model, especially the individual construction elements. While applying the proposed methods, a logical next step can be implementing an interlinked construction diary that incorporates the connection between the acquired images and the BIM model.

9.5. Scan-to-BIM

Current research focuses on Scan-to-BIM approaches, since acquisition methods advance in terms of accuracy and automation. Research in this area dramatically benefits from advancements in Machine Learning. In this regard, the automated labeling approach presented in Chapter 7 can be extended to point clouds to provide additional value to the training of a point cloud-based neural network.

Current research in construction ML focuses on training directly on the point cloud (Charles et al., 2017). Since these methods are still under development, the previously introduced methods regarding CV methods and visibility analysis can provide valuable information to improve future research approaches in this field. Especially the automated labeling approach presented in Chapter 7 can provide significant benefits in this regard as it provides quick labeling results without manual work.

Bibliography

- Akinci, B., F. Boukamp, C. Gordon, D. Huber, C. Lyons, and K. Park (2006). A formalism for utilization of sensor systems and integrated project models for active construction quality control. *Automation in Construction* 15(2), 124–138.
- Albelwi, S. and A. Mahmood (2017). A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy* 19(6), 242.
- Andriluka, M., L. Pishchulin, P. Gehler, and B. Schiele (2014). 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3686–3693. IEEE.
- Armeni, I., J. Z.-y. He, J. Gwak, A. R. Zamir, M. Fischer, J. Malik, and S. Savarese (2019). 3D Scene Graph: A Structure for Unified Semantics, 3D Space, and Camera. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5664–5673.
- Asadi, K., H. Ramshankar, M. Noghabaei, and K. Han (2019). Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction. *Journal of Computing in Civil Engineering* 33(5), 04019031.
- Bae, H., M. Golparvar-Fard, and J. White (2015). Image-Based Localization and Content Authoring in Structure-from-Motion Point Cloud Models for Real-Time Field Reporting Applications. *Journal of Computing in Civil Engineering* 29(4), 1–13.
- Barlow, H. (1983). Vision: A computational investigation into the human representation and processing of visual information. *Journal of Mathematical Psychology* 27(1), 107–110.
- Besl, P. and N. D. McKay (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), 239–256.
- Borrmann, A. (2007). *Unterstützung verteilt-kooperativer Bauplanung durch Integration interaktiver Simulationen und räumlicher Datenbanken*. Ph. D. thesis, Technical University of Munich.
- Borrmann, A., M. König, C. Koch, and J. Beetz (2018). *Building Information Modeling: Technology Foundations and Industry Practice*. Cham: Springer International Publishing.

- Bosché, F. (2010). Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Advanced Engineering Informatics* 24(1), 107–118.
- Bosché, F. (2012). Plane-based registration of construction laser scans with 3D/4D building models. *Advanced Engineering Informatics* 26(1), 90–102.
- Bosché, F., M. Ahmed, Y. Turkan, C. T. Haas, and R. Haas (2015). The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Automation in Construction* 49, 201–213.
- Bosché, F. and C. Haas (2008a). Automated retrieval of 3D CAD model objects in construction range images. *Automation in Construction* 17(4), 499–512.
- Bosché, F. and C. T. Haas (2008b). Automated retrieval of project three-dimensional CAD objects in range point clouds to support automated dimensional QA/QC. *Information Technologies in Construction* 13, 71–85.
- Bosché, F., C. T. Haas, and B. Akinci (2009). Automated Recognition of 3D CAD Objects in Site Laser Scans for Project 3D Status Visualization and Performance Control. *Journal of Computing in Civil Engineering* 23(6), 311–318.
- Bosché, F., Y. Turkan, C. T. Haas, and R. Haas (2010). Fusing 4D Modelling and Laser Scanning for Construction Schedule Control. In *26th ARCOM Annual Conference and Annual General Meeting*.
- Braun, A. and A. Borrmann (2019). Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning. *Automation in Construction* 106, 102879.
- Braun, A., A. Borrmann, S. Tuttas, and U. Stilla (2016). Classification of detection states in construction progress monitoring. In *eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 11th European Conference on Product and Process Modelling, ECPPM 2016*.
- Braun, A., F. Bosché, and A. Borrmann (2019). Towards automated quality assessment of construction elements. In J. O'Donnell, A. Chassiakos, D. Rovas, and D. Hall (Eds.), *Proceedings of the 2019 European Conference on Computing in Construction*, pp. 482–483.
- Braun, A., S. Tuttas, A. Borrmann, and U. Stilla (2017). Automated Progress Monitoring Based on Photogrammetric Point Clouds and Precedence Relationship Graphs. In *Proceedings of the 32nd International Symposium on Automation and Robotics in Construction and Mining (ISARC 2015)*, pp. 274–280.
- Braun, A., S. Tuttas, U. Stilla, and A. Borrmann (2016). Incorporating knowledge on construction methods into automated progress monitoring techniques. In *23rd International Workshop of the European Group for Intelligent Computing in Engineering*, pp. 1–11.

- Braun, A., S. Tuttas, U. Stilla, and A. Borrmann (2018). Process- and Computer Vision-based Detection of As-Built Components on Construction Sites. In *2018 Proceedings of the 35th ISARC, Berlin, Germany*, pp. 7.
- Breu, T. (2019). *Point Cloud Classification as a Support for BIM-based Progress Tracking*. Ph. D. thesis, Technical University of Munich.
- Brilakis, I., H. Fathi, and A. Rashidi (2011). Progressive 3D reconstruction of infrastructure with videogrammetry. *Automation in Construction* 20(7), 884–895.
- Brilakis, I., L. Soibelman, and Y. Shinagawa (2005). Material-Based Construction Site Image Retrieval. *Journal of Computing in Civil Engineering* 19(4), 341–355.
- Buduma, N. and N. Locascio (2017). *Fundamentals of deep learning : designing next-generation machine intelligence algorithms*. O'Reilly Media.
- BuildingSmart (2014). IfcProcess Entity. <https://standards.buildingsmart.org/IFC/RELEASE/IFC2x2/FINAL/HTML/ifckernel/lexical/ifcprocess.html>, Last access: 2020-06-10.
- BuildingSmart (2016). Model View Definitions. <https://technical.buildingsmart.org/standards/ifc/mvd/>, Last accessed: 2020-06-10.
- Bungartz, H.-J., M. Griebel, and C. Zenger (2002). *Einführung in die Computergraphik: Grundlagen, Geometrische Modellierung, Algorithmen (Mathematische Grundlagen der Informatik) (German Edition)*. Vieweg+Teubner Verlag.
- Cesetti, A., E. Frontoni, A. Mancini, A. Ascani, P. Zingaretti, and S. Longhi (2011). A visual global positioning system for unmanned aerial vehicles used in photogrammetric applications. *Journal of Intelligent and Robotic Systems: Theory and Applications* 61(1-4), 157–168.
- Charles, R. Q., H. Su, M. Kaichun, and L. J. Guibas (2017). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77–85. IEEE.
- Chi, S. and C. H. Caldas (2011). Automated Object Identification Using Optical Video Cameras on Construction Sites. *Computer-Aided Civil and Infrastructure Engineering* 26(5), 368–380.
- Coxeter, H. S. M. H. S. M. (1969). *Introduction to geometry*. Wiley.
- Cracknell, A. P. and L. Hayes (2007). *Introduction to remote sensing (Second edi ed.)*. London: CRC Press.
- Daum, S. and A. Borrmann (2013). Definition and Implementation of Temporal Operators for a 4D Query Language. In *Proc. of the ASCE International Workshop on Computing in Civil Engineering*.

- Daum, S. and A. Borrmann (2014). Processing of Topological BIM Queries using Boundary Representation Based Methods. *Advanced Engineering Informatics* 28(4), 272–286.
- Daum, S., A. Borrmann, C. Langenhan, and F. Petzold (2014). Automated generation of building fingerprints using a spatio-semantic query language for building information models. In *European Conference on Product & Process Modelling*.
- de Berg, M. (1993). Depth orders in the plane. In *Ray Shooting, Depth Orders and Hidden Surface Removal*, pp. 115–133. Springer, Berlin, Heidelberg.
- Deo, N. (2004). *Graph theory with applications to engineering and computer science*. Prentice-Hall, Inc.
- Dimitrov, A. and M. Golparvar-Fard (2014). Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. *Advanced Engineering Informatics* 28(1), 37–49.
- DIN (2013). DIN 18202 - Tolerances in building construction - Buildings. Deutsche Industrienorm, Baurechtliche Blätter.
- Dori, G. (2016). *Simulation-based methods for float time determination and schedule optimization for construction projects*. Ph. D. thesis, Technical University of Munich.
- Dowling, L., T. Poblete, I. Hook, H. Tang, Y. Tan, W. Glenn, and R. R. Unnithan (2018). Accurate indoor mapping using an autonomous unmanned aerial vehicle (UAV). *CoRR abs/1808.0*, 1808.01940.
- EASA (2020). Civil drones (Unmanned aircraft) regulations. <https://www.easa.europa.eu/easa-and-you/civil-drones-rpas>, Last accessed: 2020-06-10.
- Eastman, C. (1999). *Building product models: computer environments, supporting design and construction*. CRC Press.
- Elvins, T. T. (2005). A survey of algorithms for volume visualization. *ACM SIGGRAPH Computer Graphics* 26(3), 194–201.
- Enge, F. (2010). *Muster in Prozessen der Bauablaufplanung*. Ph. D. thesis, TU Berlin.
- Fathi, H. and I. Brilakis (2011). Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features. *Advanced Engineering Informatics* 25(4), 760–770.
- Foley, J. D. and J. D. Foley (1990). *Computer graphics : principles and practice*. Addison-Wesley.
- Forth, K., A. Braun, and A. Borrmann (2019). BIM-integrated LCA - model analysis and implementation for practice. In *IOP Conference Series: Earth and Environmental Science*, Volume 323, pp. 012100.

- Gandarias, J. M., A. J. Garcia-Cerezo, and J. M. Gomez-De-Gabriel (2019). CNN-Based Methods for Object Recognition with High-Resolution Tactile Sensors. *IEEE Sensors Journal* 19(16), 6872–6882.
- Gheorghiu, A. (1978). *Geometry of structural forms*. Applied Science Publishers Ltd.
- Girshick, R., J. Donahue, T. Darrell, and J. Malik (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587. IEEE.
- Golparvar-fard, M., F. Pena-Mora, and S. Savarese (2009). D4AR - a 4 dimensional augmented reality model for automation construction progress monitoring data collection, processing and communication. *Journal of Information Technology in Construction* 14(June), 129–153.
- Golparvar-Fard, M., F. Peña-Mora, and S. Savarese (2011). Integrated Sequential As-Built and As-Planned Representation with D4AR Tools in Support of Decision-Making Tasks in the AEC/FM Industry. *Journal of Construction Engineering and Management* 137(12), 1099–1116.
- Golparvar-Fard, M., F. Pena-Mora, and S. Savarese (2011). Monitoring changes of 3D building elements from unordered photo collections. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 249–256. IEEE.
- Golparvar-Fard, M., F. Peña-Mora, and S. Savarese (2015). Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *Journal of Computing in Civil Engineering* 29(1), 04014025.
- Ham, Y., K. K. Han, J. J. Lin, and M. Golparvar-Fard (2016). Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. *Visualization in Engineering* 4(1), 1.
- Hamledari, H., B. McCabe, and S. Davari (2017). Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction* 74, 78–94.
- Hamledari, H., B. McCabe, S. Davari, and A. Shahi (2017). Automated Schedule and Progress Updating of IFC-Based 4D BIMs. *Journal of Computing in Civil Engineering* 31(4), 04017012.
- Han, K. and M. Golparvar-Fard (2017a). Crowdsourcing BIM-guided collection of construction material library from site photologs. *Visualization in Engineering* 5(1), 14.
- Han, K. K., D. Cline, and M. Golparvar-Fard (2015). Formalized knowledge of construction sequencing for visual monitoring of work-in-progress via incomplete point clouds and low-LoD 4D BIMs. *Advanced Engineering Informatics* 29(4), 889–901.

- Han, K. K. and M. Golparvar-Fard (2015). Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs. *Automation in Construction* 53, 44–57.
- Han, K. K. and M. Golparvar-Fard (2017b). Potential of big visual data and building information modeling for construction performance analytics: An exploratory study. *Automation in Construction* 73, 184–198.
- Hardin, B. and D. McCool (2015). *BIM and construction management : proven tools, methods, and workflows*. Wiley.
- He, K., G. Gkioxari, P. Dollar, and R. Girshick (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision 2017-October*, 2980–2988.
- He, K., X. Zhang, S. Ren, and J. Sun (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. IEEE.
- Hirschmuller, H. (2005). Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Volume 2, pp. 807–814. IEEE.
- Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), 328–341.
- Hornung, A., K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard (2013). OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots* 34(3), 189–206.
- Hübner, P., K. Clintworth, Q. Liu, M. Weinmann, and S. Wursthorn (2020). Evaluation of HoloLens Tracking and Depth Sensing for Indoor Mapping Applications. *Sensors* 20(4), 1021.
- Huhnt, W. (2005). Generating sequences of construction tasks. In *Conference on Information Technology in Construction*, pp. 1–6.
- Huhnt, W. and F. Enge (2007). Consistent information management for structuring construction activities. In *Bringing ITC Knowledge to Work, Proceedings of 24th CIB-W78 Conference*, pp. 607–614.
- Huhnt, W., S. Richter, and F. Enge (2008). Modification Management for Construction Processes. *Tsinghua Science and Technology* 13(SUPPL. 1), 185–191.
- Ibrahim, Y., T. Lukins, X. Zhang, E. Trucco, and a. Kaka (2009). Towards automated progress assessment of workpackage components in construction projects using computer vision. *Advanced Engineering Informatics* 23(1), 93–103.

- Jahr, K., A. Braun, and A. Borrmann (2018). Formwork detection in uav pictures of construction sites. In *Proc. of the 12th European Conference on Product and Process Modelling*, Copenhagen, Denmark.
- Karsch, K., M. Golparvar-Fard, and D. Forsyth (2014). ConstructAide. *ACM Transactions on Graphics* 33(6), 1–11.
- Kim, C., B. Kim, and H. Kim (2013). 4D CAD model updating using image processing-based construction progress monitoring. *Automation in Construction* 35(0), 44–52.
- Kim, C. C., H. Son, and C. C. Kim (2013a). Automated construction progress measurement using a 4D building information model and 3D data. *Automation in Construction* 31, 75–82.
- Kim, C. C., H. Son, and C. C. Kim (2013b). Fully automated registration of 3D data to a 3D CAD model for project progress monitoring. *Automation in Construction* 35, 587–594.
- Kim, H. and N. Kano (2008). Comparison of construction photograph and VR image in construction progress. *Automation in Construction* 17(2), 137–143.
- Koch, C., K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics* 29(2), 196–210.
- Kopsida, M. and I. Brilakis (2016). Markerless BIM Registration for Mobile Augmented Reality Based Inspection. In *Proceedings of the International Conference on Smart Infrastructure and Construction*, pp. 1–7.
- Kopsida, M. and I. Brilakis (2020). Real-Time Volume-to-Plane Comparison for Mixed Reality-Based Progress Monitoring. *Journal of Computing in Civil Engineering* 34(4), 04020016.
- Koutsoudis, A., F. Arnaoutoglou, and G. Pavlidis (2015). Structure from Motion – Multiple View Stereovision (SFM / MVS). In *DIGARCH 2015 Notes*, pp. 1–20.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60(6), 84–90.
- Kropp, C., C. Koch, and M. König (2014). Drywall State Detection in Image Data for Automatic Indoor Progress Monitoring. In *Computing in Civil and Building Engineering (2014)*, Reston, VA, pp. 347–354. American Society of Civil Engineers.
- Kropp, C., C. Koch, and M. König (2018). Interior construction state recognition with 4D BIM registered image sequences. *Automation in Construction* 86, 11–32.
- Labelbox (2020). training data workbench. www.labelbox.com, last accessed: 2020-06-10.
- LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep learning. *Nature* 521(7553), 436–444.

- Lin, J. J., K. K. Han, and M. Golparvar-Fard (2015). A Framework for Model-Driven Acquisition and Analytics of Visual Data Using UAVs for Automated Construction Progress Monitoring. In *Computing in Civil Engineering 2015*, Reston, VA, pp. 156–164. American Society of Civil Engineers.
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), 91–110.
- Maalek, R., D. Lichti, and J. Ruwanpura (2015). Development of an automated 3D/4D as-built model generation system for construction progress monitoring and quality control. In *11th Construction Specialty Conference*, pp. 312–322.
- Mahami, H., F. Nasirzadeh, A. Hosseininaveh Ahmadabadian, and S. Nahavandi (2019). Automated Progress Controlling and Monitoring Using Daily Site Images and Building Information Modelling. *Buildings* 9(3), 70.
- McKinney, K., J. Kunz, and M. Fischer (1998). Visualization of construction planning information. In *Proceedings of the 3rd international conference on Intelligent user interfaces - IUI '98*, New York, New York, USA, pp. 135–138. ACM Press.
- Meagher, D. (1982). Geometric modeling using octree encoding. *Computer Graphics and Image Processing* 19(2), 129–147.
- Moselhi, O., H. Bardareh, and Z. Zhu (2020). Automated Data Acquisition in Construction with Remote Sensing Technologies. *Applied Sciences* 10(8), 2846.
- Nhat-Duc, H., Q.-L. Nguyen, and V.-D. Tran (2018). Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. *Automation in Construction* 94(10), 203–213.
- NVIDIA (2018). Nvidia Digits - Deep Learning Digits Documentation. <https://developer.nvidia.com/digits>, last accessed: 2020-06-20.
- Obrock, L. and E. Gülch (2020). Automatisierte semantische Modellierung von Innenräumen aus Bildern und abgeleiteten Punktwolken basierend auf Deep Learning Methoden. In *40. Wissenschaftlich-Technische Jahrestagung der DGPF in Stuttgart – Publikationen der DGPF*, pp. 358–372.
- Omar, T. and M. L. Nehdi (2016). Data acquisition technologies for construction progress tracking. *Automation in Construction* 70, 143–155.
- Pătrăucean, V., I. Armeni, M. Nahangi, J. Yeung, I. Brilakis, and C. Haas (2015). State of research in automatic as-built modelling. *Advanced Engineering Informatics* 29(2), 162–171.
- Poole, D. L., A. Mackworth, and R. G. Goebel (1998). Computational Intelligence and Knowledge. *Computational Intelligence: A Logical Approach* 1(1), 1–22.

- Pučko, Z., N. Šuman, and D. Rebolj (2018). Automated continuous construction progress monitoring using multiple workplace real time 3D scans. *Advanced Engineering Informatics* 38(4), 27–40.
- Qi, C. R., H. Su, K. Mo, and L. J. Guibas (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Volume 2017-January, pp. 77–85. Institute of Electrical and Electronics Engineers Inc.
- Rankohi, S. and L. Waugh (2015). Image-based modeling approaches for projects status comparison. *Annual Conference of the Canadian Society for Civil Engineering 2014: Sustainable Municipalities 1*(Rankohi 2013), 455–464.
- Rashidi, A., I. Brilakis, and P. Vela (2015). Generating Absolute-Scale Point Cloud Data of Built Infrastructure Scenes Using a Monocular Camera Setting. *Journal of Computing in Civil Engineering* 29(6), 04014089.
- Redmon, J., S. Divvala, R. Girshick, and A. Farhadi (2016). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788. IEEE.
- Rothermel, M., K. Wenzel, D. Fritsch, and N. Haala (2012). SURE - Photogrammetric surface reconstruction from imagery. *LC3D Workshop 8*, 1–9.
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science* 33(2), 163–180.
- Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115(3), 211–252.
- Saidi, K. S., A. M. Lytle, and W. C. Stone (2003). Report of the NIST Workshop on Data Exchange Standards at the Construction Job Site. In *Citeseer*.
- Shehab, T. and O. Moselhi (2005). An Automated Barcode System for Tracking and Control of Engineering Deliverables. In *Construction Research Congress 2005*, Reston, VA, pp. 1–9. American Society of Civil Engineers.
- Snavely, N., S. M. Seitz, and R. Szeliski (2006). Photo tourism. *ACM Transactions on Graphics* 25(3), 835.
- Son, H. and C. Kim (2010). 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring. *Automation in Construction* 19(7), 844–854.
- Sural, S., Gang Qian, and S. Pramanik (2003). Segmentation and histogram generation using the HSV color space for image retrieval. In *Proceedings. International Conference on Image Processing*, Volume 2, pp. II–589–II–592. IEEE.

- Szczesny, K., M. Hamm, and M. König (2012). Adjusted recombination operator for simulation-based construction schedule optimization. In *Proceedings of the 2012 Winter Simulation Conference*, pp. 1219–1229.
- Szegedy, C., Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9. IEEE.
- Tang, P. and S. H. Rasheed (2013). Simulation for characterizing a progressive registration algorithm aligning as-built 3D point clouds against as-designed models. In *2013 Winter Simulations Conference (WSC)*, pp. 3169–3180. IEEE.
- Tauscher, E. (2011). *Vom Bauwerksinformationsmodell zur Terminplanung*. Ph. D. thesis, Bauhaus-Universität Weimar.
- The American Institute of Architects (2013). AIA Document G202. <https://www.aiacontracts.org/contract-documents/19016-project-bim-protocol>, last accessed: 2020-06-20.
- Tulke, J. (2010). *Kollaborative Terminplanung auf Basis von Bauwerksinformationsmodellen*. Ph. D. thesis, Bauhaus Universität Weimar.
- Turkan, Y. (2012). *Automated Construction Progress Tracking using 3D Sensing Technologies*. Ph. D. thesis, University of Waterloo.
- Turkan, Y., F. Bosché, C. T. Haas, and R. Haas (2011). Automated progress tracking of erection of concrete structures. In *Annual Conference of the Canadian Society for Civil Engineering*, pp. 2746–2756.
- Turkan, Y., F. Bosché, C. T. Haas, and R. Haas (2012). Automated progress tracking using 4D schedule and 3D sensing technologies. *Automation in Construction* 22, 414–421.
- Turkan, Y., F. Bosché, C. T. Haas, and R. Haas (2013). Toward Automated Earned Value Tracking Using 3D Imaging Tools. *Journal of Construction Engineering and Management* 139(4), 423–433.
- Turkan, Y., F. Bosché, C. T. Haas, and R. Haas (2014). Tracking of secondary and temporary objects in structural concrete work. *Construction Innovation* 14(2), 145–167.
- Tuttas, S., A. Braun, A. Borrmann, and U. Stilla (2014). Comparison of photogrammetric point clouds with BIM building elements for construction progress monitoring. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Volume 40, pp. 341–345.
- Tuttas, S., A. Braun, A. Borrmann, and U. Stilla (2015). Validation of Bim Components By Photogrammetric Point Clouds for Construction Site Monitoring. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4(March)*, 231–237.

- Tuttas, S., A. Braun, A. Borrmann, and U. Stilla (2016). Evaluation of acquisition strategies for image-based construction site monitoring. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Volume 41, ISPRS Congress, Prague, Czech Republic, pp. 733–740.
- Tuttas, S., A. Braun, A. Borrmann, and U. Stilla (2017). Acquisition and Consecutive Registration of Photogrammetric Point Clouds for Construction Progress Monitoring Using a 4D BIM. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 85(1), 3–15.
- Tuttas, S. A. (2017). *Erfassung von Bauteilen durch photogrammetrische Punktwolken und Abgleich eines 4D-Bauwerkmodells zur Baufortschrittskontrolle*. Ph. D. thesis, Technical University of Munich.
- Webb, R., J. Smallwood, and T. C. Haupt (2004). The potential of 4D CAD as a tool for construction management. *Journal of Construction Research* 05(01), 43–60.
- Wu, C. (2013a). Towards linear-time incremental structure from motion. In *Proceedings - 2013 International Conference on 3D Vision, 3DV 2013*, pp. 127–134. IEEE.
- Wu, C. (2013b). Towards linear-time incremental structure from motion. In *Proceedings - 2013 International Conference on 3D Vision, 3DV 2013*, pp. 127–134. IEEE.
- Wu, I. C., A. Borrmann, U. Beißert, M. König, and E. Rank (2010). Bridge construction schedule generation with pattern-based construction methods and constraint-based simulation. *Advanced Engineering Informatics* 24, 379–388.
- Xu, Y., J. He, S. Tuttas, and U. Stilla (2015). Reconstruction of scaffolding components from photogrammetric point clouds of a construction site. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W5*, 401–408.
- Xu, Y., S. Tuttas, L. Heogner, and U. Stilla (2016). Classification of photogrammetric point clouds of scaffolds for construction site monitoring using subspace clustering and PCA. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B3*, 725–732.
- Zhang, C. and D. Arditi (2013). Automated progress control using laser scanning technology. *Automation in Construction* 36, 108–116.
- Zhao, P. and J. Han (2010, 9). On graph query optimization in large networks. *Proceedings of the VLDB Endowment* 3(1), 340–351.
- Zhou, S. and W. Song (2020). Robust Image-Based Surface Crack Detection Using Range Data. *Journal of Computing in Civil Engineering* 34(2), 04019054.
- Zobrist, G. W. and J. V. Leonard (1992). *Progress in simulation*. Intellect Books.
- Zollmann, S., C. Hoppe, S. Kluckner, C. Poglitsch, H. Bischof, and G. Reitmayr (2014). Augmented Reality for Construction Site Monitoring and Documentation. *Proceedings of the IEEE* 102(2), 137–154.