Process-oriented progress monitoring of cast-in-place shell constructions based on computer vision

Schlenger J., Vilgertshofer S., and Borrmann A.
Technical University of Munich, Germany
jonas.schlenger@tum.de

Abstract. Automated progress monitoring builds an important foundation for objective productivity analysis of construction processes. Digital twins of the construction phase rely on fully automated approaches to acquire near real-time progress information. This is essential for identifying bottlenecks during construction and supporting future project planning. Many existing vision-based methods lack automated image acquisition, fast computation times, or fine-grained progress information. This paper presents a new vision-based construction monitoring approach that reduces the geometric information provided in exchange for a higher time resolution and a higher level of automation. Instead of the detailed geometry, the real-time status of the building elements is provided. It is applied to cast-in-place concrete columns, identifying individual operational steps. The approach is based on projecting building elements from a building model onto images of a fixed on-site camera to then classify them according to the current element status with the help of a CNN. Using image sequences additionally allows accounting for moving objects and other outliers, which makes the approach robust and reliable.

1 Introduction

Construction projects are considered to be successfully finished when they are completed within the planned time and cost and the resulting building complies with specified quality standards (Bannerman, 2008). Continuous progress monitoring is paramount to identify deviations from the plan as early as possible and initiate timely countermeasures.

Traditionally, progress information is collected and documented by hand, which is very time-consuming and error-prone, especially for large construction sites. Recently, many researchers have focused on automating the monitoring process with computer vision-based approaches. As Rehman et al. (2022) pointed out, many of these approaches rely on construction site images used for a photogrammetric reconstruction of the building that is compared against the as-designed building model. However, full automation has yet to be accomplished in data collection and processing. For example, regular site visits are necessary to capture construction site images, e.g., with the help of a drone. This introduces significant manual effort and does not comply with applications that require continuous progress information.

In digital twin construction (Sacks et al., 2020) and productivity analysis, up-to-date progress information is very relevant since many use cases rely on information about the building elements already built. Here, most use cases other than those directly concerned with detailed geometric deviations and defects benefit more from rough status information in high time resolution than from highly detailed geometric information provided in limited time intervals (Mediavilla et al., 2021).

This paper presents a new construction monitoring approach that reduces the geometric information provided in exchange for a higher time resolution and a higher level of automation. Instead of the detailed geometry, the real-time status of the building elements is provided. The developed vision-based approach is validated with a proof-of-concept implementation limited to the progress monitoring of the shell construction of high-rise buildings. Fixed on-site cameras with known positions are used to superimpose expected building elements onto the captured

images and identify regions of interest. These regions are classified by a Convolutional Neural Network (CNN) according to the current status of the building elements into the classes *not started*, *rebar*, *formwork*, and *finished*. The developed approach is applied to cast-in-place concrete columns and slabs in the case study. It is tracked how the status of building elements changes in consecutive images, which allows for using outlier removal in a post-processing step. This makes the outcome of the automated progress monitoring more robust and reliable.

2   Related Work

Automated progress monitoring of construction sites is a very active field of research, partially because the construction industry still has a lot of potential for improvement in productivity and efficiency. In this area, image-based approaches have gained a lot of attention. Compared to other sensors, cameras are relatively affordable, and the image quality has significantly improved over the last few years (Fini et al., 2022). Moreover, drones have facilitated making images from otherwise unreachable points of view. Existing image-based progress monitoring methodologies are discussed, emphasizing their degree of automation, capabilities of near real-time progress updates, and possibilities to detect individual construction steps, to set the present paper in the context of the current state-of-the-art. All publications focus on shell constructions of high-rise buildings.

Rehman et al. (2022) show that many researchers rely on photogrammetric reconstruction through Structure from Motion (SfM) techniques for progress monitoring. A 3D representation of the construction site can be reconstructed by identifying geometric features on several images from different angles. For example, Golparvar-Fard et al. (2015) create point clouds from unstructured images taken by site personnel. These are overlaid with the as-designed BIM model. Based on a voxel grid, locations with expected building elements are checked for the existence of points in the reconstructed point cloud. Depending on the number of points detected, the building elements are either assigned with the status *existing* or *not existing*. With this approach, formwork is falsely classified as a finished building element. In contrast, Braun et al. (2015) match individual points to expected element surfaces to confirm their existence. Here, thresholds are adjusted, and colour information from the images is analysed to distinguish between concrete and formwork surfaces. However, photogrammetric reconstruction is generally a time-intensive process with specific requirements for the image dataset. Its application for real-time monitoring is hindered by long computing times and the manual effort required to acquire the image dataset through drone flights or smartphone images. Furthermore, identifying additional construction steps like the installed rebar is challenging because the point cloud resulting from SfM is sparse and noisy (Reja et al., 2021).

Besides photogrammetric reconstruction, other researchers directly analyze individual images or image sequences. A few selected papers are introduced and compared based on the used camera setup, their method to align the as-designed BIM model with reality, and their progress monitoring methodology.

Fini et al. (2022) apply progress monitoring to prefabricated wooden slab panels. Using a fixed camera on a tower crane, they capture images of the building from a top-down angle. By aligning the images with the as-designed 2D floor plans, they can detect the time when individual slab panels are installed and measure productivity but do not consider vertical elements like walls and columns. Also relying on a single camera fixed on the construction site, Wang et al. (2021) monitor the progress of installing precast concrete wall elements. With the help of various neural networks, they perform object detection, instance segmentation, and

multiple object tracking. Their alignment of the BIM model and reality is based on identifying the wall axes on a horizontal plane. As output, their tool detects the time when the wall elements are installed, including the moment when the tower crane moves them. Ibrahim et al. (2009) have yet another approach using a fixed construction site camera. They project the elements from the BIM model onto the camera images. Using the image masks of the building elements as areas of interest, they observe changes in pixel neighbourhoods. Many changes followed by a long changeless period are interpreted as a finished element. No further distinction between individual construction steps is made. Furthermore, changes introduced by moving equipment and construction personnel make the reliability of the results questionable. Also mapping the elements of the BIM model onto individual images, Vincke and Vergauwen (2022) identify the displacement of concrete columns. Opposed to the previous approaches, they manually take images of the construction site from various viewpoints. First, selecting optimal images to detect column displacement, they compare the expected column location with its actual location to identify deviations of 5 mm and above. Focusing on column displacement, they do not propose a methodology to recognise the existence of a particular column but take it as given. Kargul et al. (2015) approach progress monitoring from the side of the construction equipment. Through the internal sensors of a pile boring machine, they evaluate the progress of bored piles. However, such an approach is not feasible for building elements that require primarily manual work for the installation process. All discussed methods are summarized with further details about the needed computation time in Table 1.

Table 1: Comparison of state-of-the-art vision-based progress monitoring.

Reference	Data Acquisition	BIM-reality alignment	Building elements	Operational steps	Computation time
Fini et al. (2022)	Fixed camera	Match 2D plan with top-down image	Wooden slab panels	Existing / not existing	< 30 sec.
Ibrahim et al. (2009)	Fixed camera	Project BIM onto image	Not specified	Existing / not existing	> 20 min.
Vincke et al. (2022)	Images taken manually	Project BIM onto image	Concrete columns	-	> 15 min.
Wang et al. (2022)	Fixed camera	Axis network	Prefabricated walls	Moving / installed	Real-time
Kargul et al. (2015)	Internal sensors of equipment	GPS (presumably)	Bored piles	Not started / drilling / concreting / finished	Not specified

The state-of-the-art shows that none of the existing approaches can detect all relevant operational steps of cast-in-place concrete shell constructions in near real-time. This type of information, however, is valuable for detailed productivity analysis and identification of bottlenecks within construction sequences as part of a digital twin of the construction phase.

3   Methodology

At first, the proposed methodology to fill the identified gap in the state-of-the-art is summarized, showing the complete workflow for the image sequence of a single camera. In the subsequent subchapters, the individual components are explained in more detail. Furthermore,

it describes how the methodology is extended to include rough detection of cast-in-place concrete slabs to cope with multi-storey shell constructions.

3.1 Overview

As the primary data source, a fixed on-site camera captures images in short intervals. It is assumed that the external and internal camera parameters are known. The as-designed BIM model is also required in the form of an IFC file containing a reference to a geodetic reference system. Knowing the location of the camera and the expected location of the building elements from the IFC file, the elements are projected into the camera's 2D image plane. This is done using the approach described in Braun et al. (2020). The image areas that are covered by expected elements from the BIM model correspond to the areas of interest. Image sections are cut out from the original image for every column, reflecting slightly enlarged areas of interest, using rectangular bounding boxes.

In the second step, the image sections are fed to a CNN to be classified according to the current status of the contained cast-in-place concrete column. The CNN was trained using images of concrete columns collected on various construction sites and is expected to be reused for new construction projects without additional training. The previously described steps are executed for all potentially visible columns on the image, repeating the process for all following photos of the same camera. Through a post-processing step, noise and outliers are removed. This is based on observing consecutive image sections and applying knowledge about the expected sequence of construction steps. The complete workflow is visualized in Figure 1. The grey boxes with numbers refer to the corresponding chapters that explain the module in further detail.
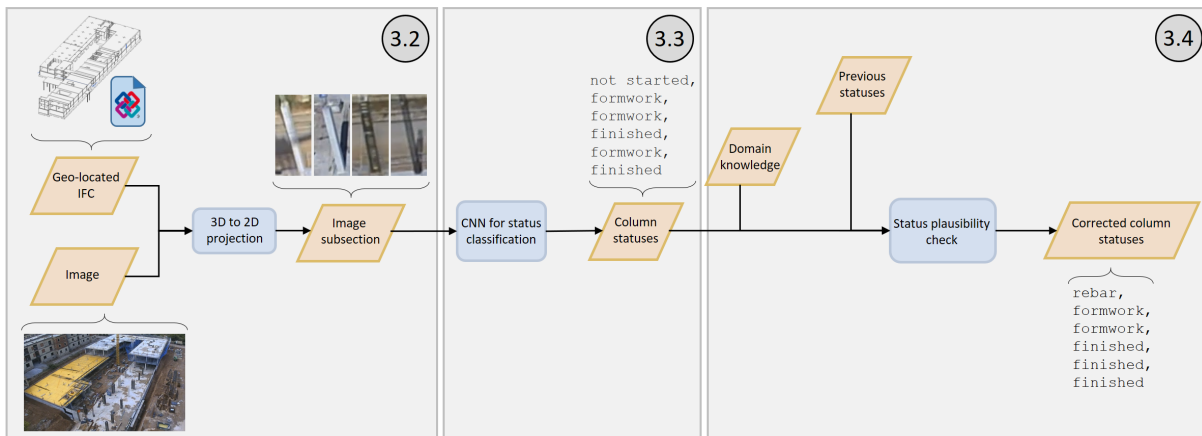

Figure 1: Workflow diagram of the proposed methodology.

Analyzing the status of all concrete columns throughout a construction project complemented with an algorithm for slab detection allows the automatic creation of the as-performed schedule of the shell construction. This comprises information about individual construction steps' start and end dates and can be compared with the as-planned schedule. Since the goal of the proposed approach is to identify deviations from the as-planned schedule, as-planned information was intentionally not integrated into the workflow to ensure that the result will be unbiased.

3.2 Extraction of Elementwise Image Sections

For automated data collection for progress monitoring, a fixed camera is placed on a tower crane, a neighbouring building or a similar location that provides a good overview of a large part of the construction site. With a router and a local server on-site, the images are sent to a remote server to make them accessible over the web. For more details regarding the proposed

data-capturing system, please refer to the hardware setup described in Collins et al. (2022). For the present paper, it is assumed that the external and internal camera parameters are known.

With the known camera position relative to a geodetic reference system and the internal parameters, together with a geolocated IFC file, the three-dimensional building elements from the IFC model are projected onto the 2D plane of the camera. This can be achieved by multiplying the original 3D coordinates with rotation and translation matrices that transform the coordinates into the local coordinate system of the camera, as described by Braun et al. (2020). The IFC model does not include any 4D information. For this reason, all potentially visible columns are projected onto the image independent of their expected construction time. The areas of the image overlaid with element projections are the regions of interest, while one region always corresponds to an individual column. For every region, the minimal axis-aligned bounding box is calculated and slightly enlarged to account for minor movement of the camera. This is also necessary to ensure that the top and bottom parts of the column are entirely included in the bounding box and easily distinguishable from the background. They are relevant features for the following image classification. Based on the bounding boxes, sections from the original images are extracted and treated as small images.

3.3 CNN for Column Status Classification

After extracting the image sections that contain one column each, they need to be classified automatically according to the current status of the column. In the context of cast-in-place concreting, the operational steps comprise placing the rebar, installing the formwork, pouring the concrete and removing the formwork in the given order. Considering the chosen approach, the phases to be identified are *not started*, *rebar*, *formwork*, and *finished*. Some examples of the four stages are shown in Figure 2.

A Convolutional Neural Network was selected to perform this classification task. Neural networks provide very time-efficient automated evaluation methods, which are relevant for the near real-time analysis of construction images. More specifically, CNNs are well-suited for image classification tasks. Several existing CNN models trained on large image datasets can be used as a starting point to be further trained with use case-specific data sets, also referred to as transfer learning. Compared to other types of neural networks, CNNs have a relatively low number of parameters that need to be trained, which allows for achieving good results even with small image data sets (Rehman et al., 2022; Hussain et al. 2018).
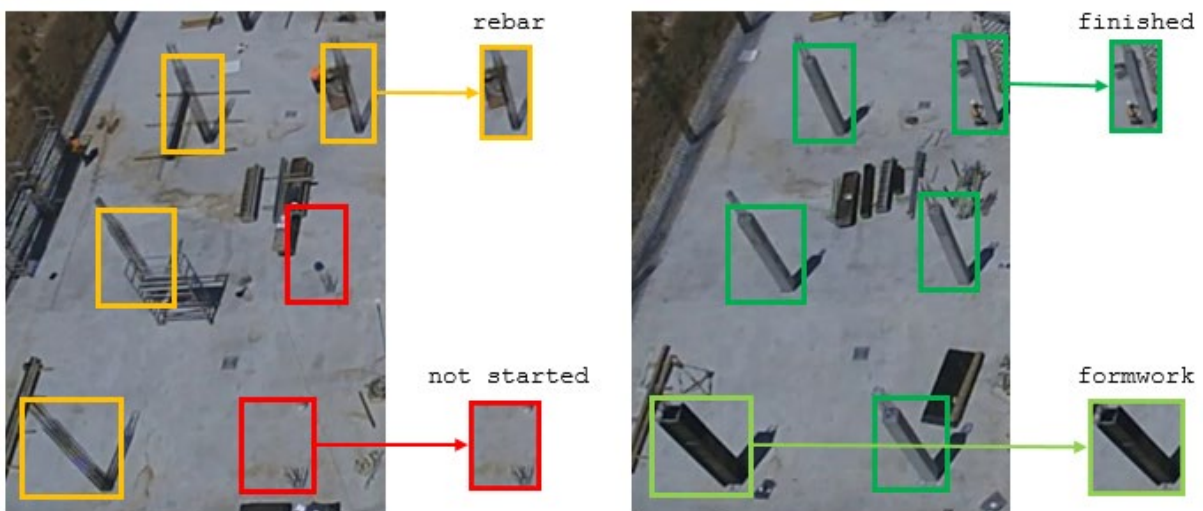

Figure 2: Four construction phases displayed in the example of a group of six concrete columns.

For the column classification, various neural networks were tested that were pre-trained on large online datasets. These networks served as a baseline and were adjusted by changing the topmost network layers through training with a column-specific dataset. The tested network architectures were VGG16, MobileNetV2, ResNet50V2, InceptionV3, and EfficientNetV2, from which ResNet50V2 performed best. From its 190 network layers, the fifty topmost network layers were retrained using a dataset collected by the authors. It contains roughly 2000 images of columns from various European construction sites. In the dataset, columns with the status *not started* and *finished* are overrepresented. Therefore, weighting of the four classes is applied to give every class the same importance in the training process. Of the complete dataset, 80% of the images of every class were used for the network training, while the remaining 20% were used for testing. Training the model for 50 epochs and restoring the model with the lowest validation loss after 47 epochs results in an overall accuracy of 95.64%. The accuracy and loss functions and the final confusion matrix are displayed in Figure 3.
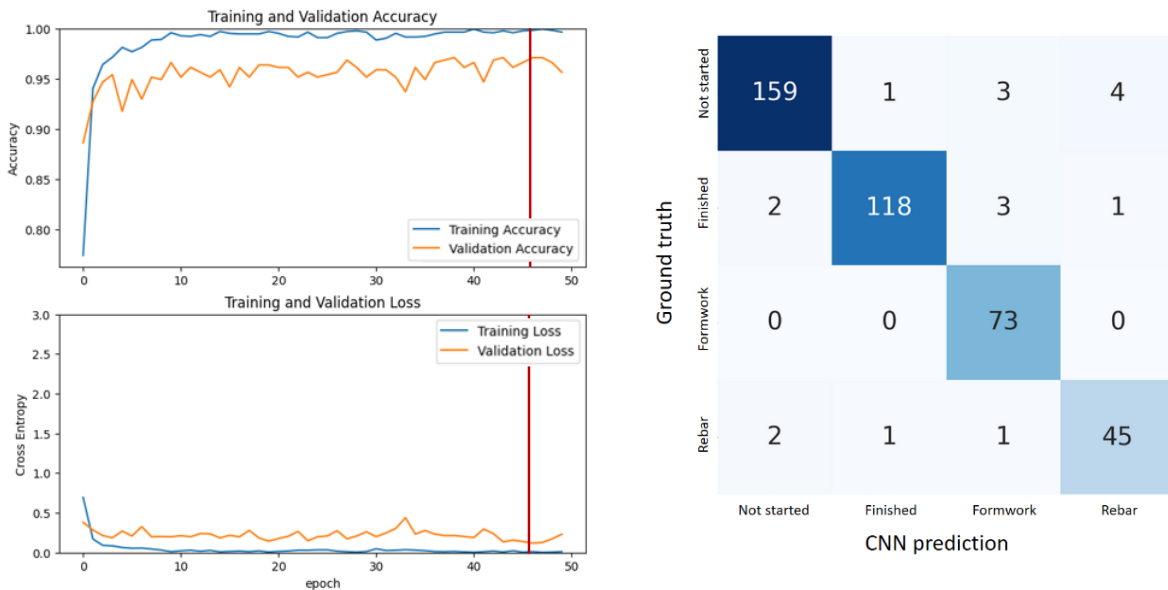

Figure 3: Accuracy and loss function, and confusion matrix of the trained CNN model using the ResNet50V2 network architecture pre-trained on the ImageNet dataset.

With the help of the trained CNN, the extracted image sections of all consecutive images are classified according to their current appearance. The predictions are stored in an object-oriented class structure, grouping columns by the storey and adding pairwise status entries, including the timestamp and status for every column and each image.

3.4 Detection of Phase Start and End Points

Several influencing factors result in erroneous status predictions. Beside the imperfect CNN prediction, the lighting conditions have a considerable impact on the appearance of the columns. Even though images of varying brightness were included in the training dataset, the influence of the lighting conditions cannot be eradicated. Furthermore, moving objects and clutter on the construction site make it more challenging to classify image sections correctly. Especially during ongoing construction works of a specific column, there is a high activity of construction workers and heavy machinery in the surrounding area. These and other reasons make it difficult to correctly identify the exact point in time when a column changes its status. For example, Figure 4a depicts the progress of a single column over time based on analysing several hundred images. The horizontal axis represents the temporal evolvement. It is discretised through the

points in time at which images are captured. The vertical axis describes the columns' progress. It is not to be seen as a specific percentage to which the column is finished but has discrete values for every different type of status. The values increase in the order in which the statuses occur. However, due to the mentioned classification errors, the predicted progress over time includes undesirable irregularities.
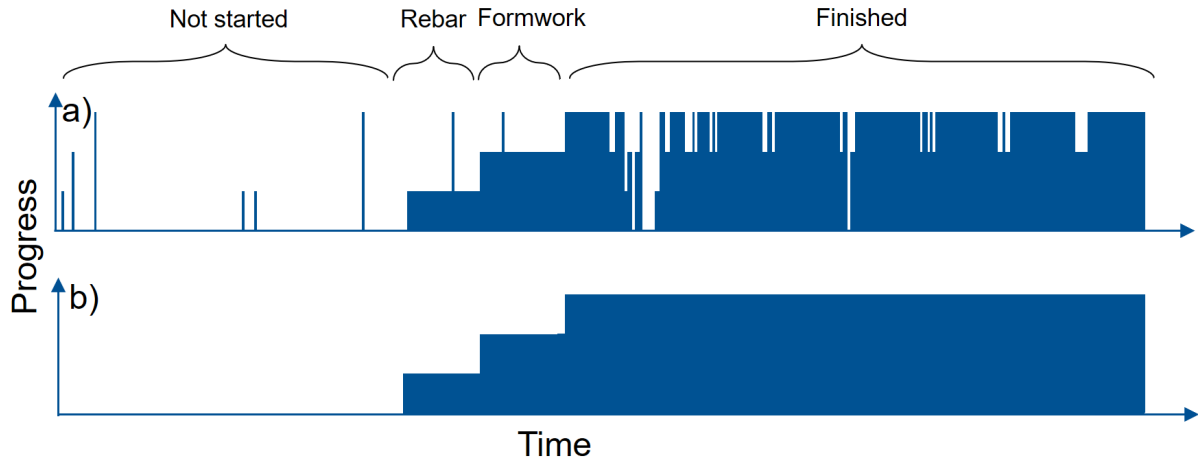

Figure 4: a) Predicted progress of a single column over time, including prediction errors; b) Corrected progress predictions of a single column.

Knowing the order in which the column phases occur helps to reevaluate the prediction results and correct them to a certain degree. To apply the correction, for every stage, the optimal time interval is identified that is as large as possible and contains the largest number of predictions of the current phase of interest while at the same time minimizing the number of predictions of all other stages within the interval. Once they are identified, the transition between them must be assigned to one of the neighbouring intervals. The exact transition point is determined by minimizing the number of predictions that need to be changed due to the correction algorithm. Following this principle, all transition points between the phases are identified. Finally, the status predictions are corrected so all images within the identified intervals are assigned to the same status. Applying the described algorithm to the exemplary column in Figure 3 results in the corrected progress diagram shown in Figure 4b.

3.5 Multi-storey Progress Monitoring

Considering shell constructions with multiple floors, it is necessary to identify when slabs are built to change the column projections from one floor's columns to the next floor's columns. Due to the large extent of the slabs and the broad range of shapes, applying the same approach as for the concrete columns is impossible. The transition from one phase to the next can take multiple hours or even days, making it highly questionable to use discrete progress status values the same way it was done for columns. Nevertheless, a simplified slab detection algorithm was implemented to evaluate the progress of multi-storey buildings. As soon as all columns of a specific floor are detected to be finished, the slab detection algorithm is started. Once the predictions of the majority of the column statuses changed from *finished* to *not started*, it can be assumed that the slab is currently under construction since the formwork of the slab blocks the view of the columns beneath it. In this event, the column projections are updated with the projections of the columns of the following floor, which also updates the bounding boxes for creating image sections.

4   Case Study

4.1 Project Description and Setup

A construction site of a hospital building in Spain was used as a case study to validate the proposed workflow and report on its accuracy. For this, a surveillance camera was installed on the tower crane to capture the building's progress of the shell construction, containing columns and slabs over several months. During working hours, the camera captured one image roughly every one and a half hours. For real-time monitoring, this time interval is rather long. However, the number of images is sufficient for the first validation of the approach. The construction company provided a geo-referenced IFC file used to project expected columns onto the camera images. The automated progress monitoring was tested on two of the building's stories with ten columns each.

4.2 Results and Discussion

During the monitoring period of roughly two months, 5200 image sections were automatically analysed, each containing a single column. This excludes all image sections where columns cannot be built yet because the slab beneath is not finished or the column is not visible anymore since the slab on top is already built. As a basis for comparison, all images were inspected by the human eye to create the ground truth of the progress of the concrete columns over time. Evaluating the classification of the CNN on the image sections without any alteration, the achieved accuracy is 89%. The reasons for a deviation of 6.5% in comparison to the accuracy of the column test dataset are explained later in this section. After applying the outlier removal described in Chapter 3.4, the accuracy was raised to 96.8%.

To visualise the results of the automated progress monitoring software, it contains a UI module that creates a Gantt chart based on the evaluation results using Syncfusion Blazor. A screenshot from the user interface is shown in Figure 6. The chart directly compares the ground truth against the automatically created results. The short bars represent one column each, showing the time of their first detection as rebar until completion. The longer bars belong to the slab construction and the parent processes, like building the complete stories or the complete construction site, respectively. When the ground truth and the automated evaluation coincide, the bar section is coloured in dark blue. The bar is coloured in light blue for the parts where the automatic evaluation results in longer construction times than the ground truth and in grey for the other way round. The detection of the two slabs is currently only implemented in a simplified way, giving a single instance of time rather than the whole construction period. The Gantt chart also includes sub-processes for all individual operational steps of columns, which are not shown in Figure 5 because of the limited space.

Analysing the results, it becomes clear that the overall methodology is feasible. However, it is also evident that there is still room for improvement. While the complete time of the construction of the columns was roughly detected, some transitions from one status to the next are detected more reliably than others. The transitions from the status *not started* to *rebar* sometimes turned out to be hard to detect since the CNN can easily mistake surrounding clutter with linear shape with the narrow rebar. This type of error can be observed in Figure 6 on all columns that start with a light blue section on the left. Similarly, the status *formwork* and *finished* can look very alike. In bad lighting conditions, the material structure of the formwork is hardly visible and mistaken for a finished column. For this reason, some columns are predicted to have been finished earlier than they were (see columns in Figure 6 with a grey section on the right end of the bar). Overall, two primary sources of erroneous

progress detection were identified. Firstly, some images were taken during challenging lighting conditions, even including some images taken late in the evening and therefore being almost entirely black. Second, the ongoing construction activities using movable scaffolding and depositing construction materials and formwork close to the location of expected columns also led to misclassification. Combining these two factors, it was sometimes challenging to correctly classify the image sections, even with the human eye.
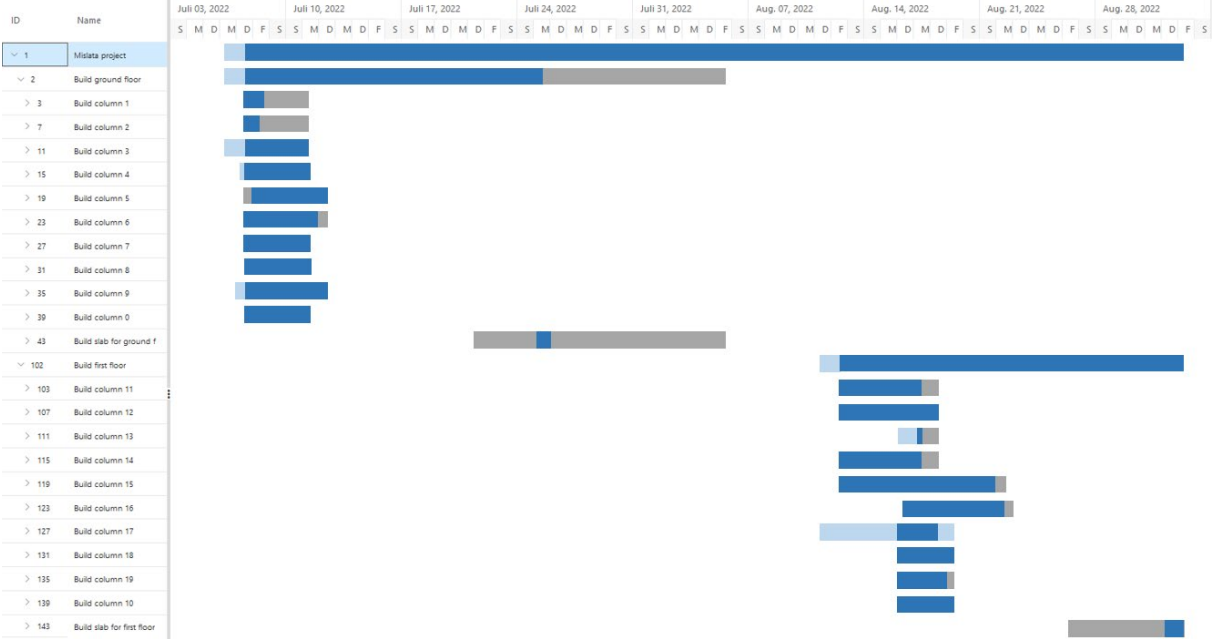

Figure 5: Ground truth schedule and automatically generate schedule in comparison.

Finally, it should be critically mentioned that the proposed approach only identifies the needed time, starting with placing the rebar up until the completion of the column. It does not enable us to identify the time construction workers spent actively working on the column. In that case, the focus needs to be shifted to detecting the workers and identifying their interaction time with a particular column. This requires a different methodology. Nevertheless, the presented approach can provide times of status changes that already allow to significantly narrow down the time spans when the workers' activities should be carefully assessed.

5   Future Work

This paper presented a new methodology using fixed construction site cameras to automatically detect the progress of cast-in-place shell constructions, including the individual construction steps, in near real-time. Opposed to many existing approaches, the focus lies on capturing rough status information in high time resolution instead of a geometrically accurate as-built model, which cannot be created multiple times a day for several reasons. Aggregating the status changes of all elements over time allows us to automatically create the as-performed construction schedule of the shell construction that is of immense value for, e.g., digital twin applications of the construction phase for progress and productivity analysis.

The presented implementation focused on concrete columns as primary elements. The workflow should be extended to concrete walls in the future since they also form an essential part of the shell construction. Additionally, the accuracy and completeness of the progress monitoring system could be improved, e.g., by analyzing images from several cameras simultaneously and fusing the contained progress information.

6 Acknowledgement

References

Bannerman, P.L. (2009). Defining Project Success: A Multilevel Framework. In: Project Management Institute Research Conference, 2008, Warsaw, Poland.

Braun, A., Tuttas, S., Borrman, A. and Stilla, U. (2015) A concept for automated construction progress monitoring using BIM-base geometric constraints and photogrammetric point clouds, Journal of Information Technology in Construction. Vol. 20, pp. 68-79, https://www.itcon.org/paper/2015/5.

Braun, A., Tuttas, S., Stilla, U. and Borrmann, A. (2020) Improving progress monitoring by fusing point clouds, semantic data and computer vision, Automation in Construction. 116, doi: 10.1016/j.autcon2020.103210.

Collins, F., Pfitzner, F. and Schlenger, J. (2022) Scalable construction monitoring for an as-performed progress documentation across time. In: 33. Forum Bauinformatik, Munich, Germany.

Fini, A.A.F, Maghrebi, M., Forsythe, P.J. and Waller, T.S. (2022) Using existing site surveillance cameras to automatically measure the installation speed in prefabricated timber construction. Engineering, Construction and Architectural Management. Vol. 29, no. 2, pp. 573–600, doi: 10.1108/ECAM-04-2020-0281.

Golparvar-Fard, M., Peña-Mora, F. and Savarese, S. (2015) Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models, Journal of Computing in Civil Engineering. Vol. 29(1), doi: 10.1061/(ASCE)CP.1943-5487.0000205.

Hussain, M., Bird, J.J. and Faria, D.R. (2018) A Study on CNN Transfer Learning for Image Classification. In: Advances in Computational Intelligence Systems, pp. 191-202, https://doi.org/10.1007/978-3-319-97982-3_16.

Ibahim, Y.M., Lukins, T.C., Zhang, X., Trucco, E. and Kaka, A.P. (2009) Towards automated progress assessment of workpackage components in construction projects using computer vision, Advanced Engineering Informatics, Vol. 23, pp. 93-103, doi:10.1016/j.aei.2008.07.002.

Kargul, A., Günthner, W.A., Bügler, M. and Borrmann, A. (2015) Web based field data analysis and data-driven simulation application for construction performance prediction, Journal of Information Technology in Construction, Vol. 20, pp. 479-494, http://www.itcon.org/2015/28.

Mediavilla, A., San Mateos, R. and Torres, J. (2021) D1.2: Definition of the digital workflows for the construction process. BIM2TWIN: Optimal Construction Management & Production Control.

Reja, V.K., Varghese, K. and Ha, Q.P. (2021) Computer vision-based construction progress monitoring, Automation in Construction, Vol. 138, https://doi.org/10.1016/j.autcon.2022.104245.

Rehman, M.S.U., Shafiq, M.T. and Ullah, F. (2022) Automated Computer Vision-Based Construction Progress Monitoring: A Systematic Review, Buildings. 12, doi: 10.3390/buildings12071037.

Sacks, R., Brilakis, I., Pikas, E., Xie, H., and Girolami, M. (2020) Construction with digital twin information systems, Data-Centric Engineering, Vol. 1, E14, doi:10.1017/dce.2020.16.

Vincke, S. and Vergauwen, M. (2022) Vision based metric for quality control by comparing built reality to BIM, Automation in Construction. Vol. 144, https://doi.org/10.1016/j.autcon.2022.104581.

Wang, Z., Zhang, Q., Yang, B., Wu, T., Lei, K., Zhang, B. and Fang, T. (2021) Vision-based Framework for Automatic Progress Monitoring of Precast Walls by Using Surveillance Videos during the Construction Phase, Journal of Computing in Civil Engineering. Vol. 35, no. 1, doi: 10.1061/(ASCE)CP.1943-5487.0000933.