Automated Construction Progress Monitoring of Partially Completed Building Elements Leveraging Geometry Modeling and Appearance Detection with Deep Learning

Conference Paper · March 2022

DOI: 10.1061/9780784483961.074

Aritra Pal
University of Cambridge

Shang-Hsien (Patrick) Hsieh
National Taiwan University

Automated Construction Progress Monitoring of Partially Completed Building Elements Leveraging Geometry Modeling and Appearance Detection with Deep Learning

Aritra Pal, S.M.ASCE[1]; Jacob J. Lin, A.M.ASCE[2]; and Shang-Hsien Hsieh, M.ASCE[3]

[1]Ph.D. Candidate, Dept. of Civil Engineering, National Taiwan Univ. (corresponding author). ORCID: https://orcid.org/ 0000-0002-1644-7400. Email: apal@caece.net
[2]Assistant Professor, Dept. of Civil Engineering, National Taiwan Univ.
Email: jacoblin@ntu.edu.tw
[3]Professor, Dept. of Civil Engineering, National Taiwan Univ. Email: shhsieh@ntu.edu.tw

ABSTRACT

The exponential growth of on-site visual data and the advent of computer vision techniques have created a unique opportunity to improve automated construction progress monitoring methods. To date, the state-of-the-art vision-based methods are capable of reporting the progress of a building element in terms of binary function. However, for better schedule control and micro-level monitoring, it is necessary to report the partial completion of tasks associated with an element. This research proposes a novel approach for computing and reporting the partial progress of tasks in terms of completion percentage using the on-site visual data, 4D BIM, and deep-learning-based computer vision algorithms. The approach leverages geometry modeling and appearance detection to automatically calculate the percentage completion of tasks associated with each element. The proposed approach is applied to a building construction project, and the preliminary results demonstrate its applicability to generate completion percentage per task in the lookahead schedule for accurate daily progress report generation.

INTRODUCTION

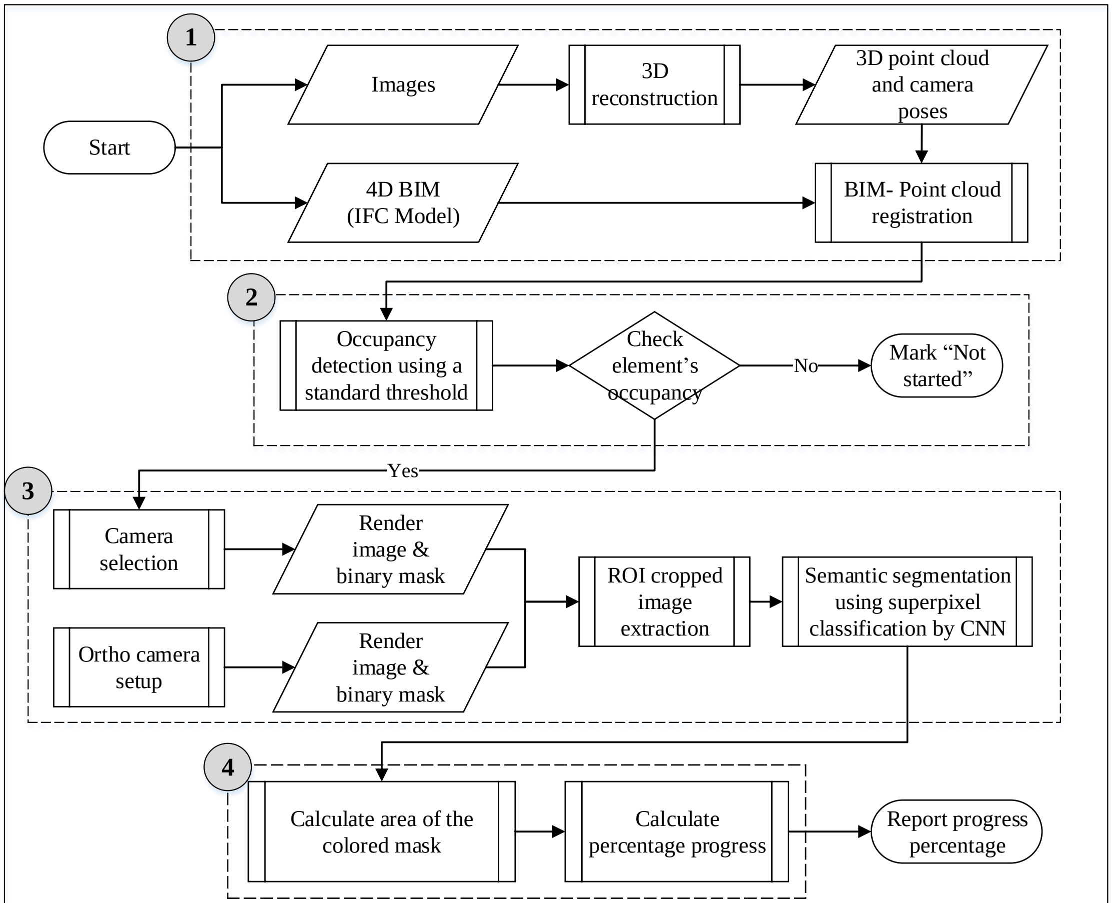
To date, many construction sites rely on conventional daily construction reports for progress monitoring. As these reports are created manually through a labor-intensive process they are often prone to human errors. Also, three-dimensional (3D) progress information cannot be verified instantly through these reports. In the effort of digitizing construction monitoring processes, vision-based methods are found to be very effective. Consequently, researchers started putting effort into developing automated construction progress monitoring methods using vision-based techniques. The comparison of as-planned four-dimensional (4D) Building Information Models (BIM) and as-built 3D point cloud models created using the photogrammetric approach or laser scanning technique was found to be the most studied method for progress monitoring. Currently, such available methods can be broadly classified into two categories, namely occupancy-based methods and appearance-based methods (Yang et al. 2015). These methods output the progress status in binary form, which means they can detect whether there is progress or not. However, this output type limits the usability of such methods for daily progress monitoring and control where partial progress of tasks needs to be reported in terms of completion percentage. Also, as these methods rely on BIM, often it is difficult to estimate the progress of tasks that are usually not modeled in the BIM used in the construction phase. So, to overcome these persisting challenges, this research proposes an improved progress monitoring method that leverages geometry modeling, appearance detection, and deep learning-based semantic segmentation to automatically calculate the percentage completion of partially completed tasks associated with building elements. This method applies to all kinds of tasks whose appearance is detectable, irrespective of whether or not modeled in BIM. The main contribution of this paper is the improved vision-based method that is capable of estimating the progress of partially completed tasks. These estimates are useful for effective daily monitoring and control of construction activities. The rest of the paper highlights the related studies, explains the methodology, describes the proof-of-concept study conducted for evaluating the applicability of the method, and concludes the study.

RELATED STUDIES

Progress monitoring is essential for keeping track of project progress and controlling the deviation from the plan promptly. Computer vision techniques promisingly helped in automating the progress monitoring in construction. The early developed vision-based automated construction progress monitoring methods leveraged 4D BIM and 3D reconstructed point clouds for comparing the as-planned and as-built status (Golparvar Fard et al. 2009, 2011). The Structure from Motion (SfM) algorithm was used for reconstructing point clouds from multi-view images and further densification of point clouds was done through the Multi-View Stereo (MVS) approach. Later, Turkan et al. (2012) used point clouds generated from laser scans for automated progress tracking. Kim et al. (2013) also tried a similar approach for progress measurement. However, because those methods relied purely on geometry, inferring progress status information was difficult. To overcome that, Han and Golparvar-Fard (2015) proposed appearance-based methods that could report the progress status through construction material classification. Vision occlusion often leads to the generation of low-quality incomplete point clouds. This affects the performance of the progress monitoring methods severely. To overcome this challenge, construction sequence and precedence relationships checks were incorporated by Han et al. (2015) and Braun et al. (2015) respectively. Later, the vision-based automated progress monitoring methods were extended to infrastructure projects. Vick and Brilakis (2018) studied the road design layer detection problem. The bridge construction scenario was investigated by Puri and Turkan (2020) and the progress was reported in terms of the percentage of completion. However, the method was still limited for reporting the task-wise progress status. Recent studies have focused on various advanced technologies such as deep learning (Braun et al. 2020), virtual reality (Pour Rahimian et al. 2020), and mixed reality (Kopsida and Brilakis 2020) for improving the progress monitoring methods. However, these methods are still insufficient for daily progress monitoring which requires progress status per task in terms of completion percentage. For proactive project control (Lin and Golparvar-Fard 2021) and micro-level monitoring, the proposed improvement of this research is expected to be necessary.

METHODOLOGY

The methodology proposed for automated construction progress monitoring of partially completed building elements is divided into 4 major tasks: 3D reconstruction and BIM-Point cloud registration, occupancy detection, appearance checking, and progress percentage completion estimation. Figure 1 shows the logical flow chart of the methodology.



1. 3D Reconstruction and BIM-Point Cloud Registration; 2. Occupancy Detection; 3. Appearance Checking 4. Percentage Completion Estimation

Figure 1. Methodology flow chart

3D reconstruction and BIM-Point cloud registration. For 3D reconstruction, a typical SfM-MVS pipeline is used to construct a 3D as-built point cloud model from the aerial images captured on-site. The point cloud model and a production-level BIM model are first registered coarsely by aligning three or more corresponding points that are distinctly visible in both models. Further, the iterative closest point (ICP) algorithm is applied to finely register both of them and a Transformation matrix is generated. The transformation matrix is used to transform the point clouds into the BIM coordinates in the Blender 2.92 environment, a free and open-source 3D creation suite used to visualize the aligned models and for further operations.

Occupancy detection. To ensure that the construction activity has started for a particular element, occupancy detection is checked before progress estimation. A bounding box with a buffer value is created around the element and the number of points from the point cloud present inside that box is counted. The buffer value is estimated in such a way that it can accommodate the registration errors and the space for the activities requiring a larger volume than the element's volume. The case when the number of points exceeds a certain threshold value is considered as the presence of the element in the as-built model, i.e. the production of that element is in progress.

Appearance checking. For enabling the percentage progress measurement, the appearance checking is done in four steps, namely camera selection, ortho-camera set up, region of interest (ROI) cropped image extraction, and semantic segmentation.

Camera selection. Once the occupancy is detected, the BIM element is back-projected to each camera using the transformation matrix, camera intrinsic (optical center, focal length, lens distortion), and extrinsic parameters (rotation, translation, scale). A collection of cameras is created from where the largest face of the element (target face) is visible. From that collection, a camera is selected that covers the maximum area of the target face. A rendered image and a binary mask of the target face are generated from the selected camera view.

Ortho-camera setup. A virtual orthographic camera that looks straight to the target face is set programmatically in the Blender environment. The camera is positioned in the 3D space relating to the center and the face normal of the target face. The orthographic scale and the distance from the face are set in such a way that the entire face is accommodated within the camera field of view (FOV). Finally, another rendered image and binary mask of the target face are generated from the ortho camera view.

ROI cropped image extraction. The target face of the element in the ortho camera view is considered as the region of interest (ROI). The original onsite image that was taken from the selected camera is used for generating the ROI cropped image. First, a homography matrix is calculated between two rendered images generated from the selected camera view and the ortho-camera view. Next, the matrix is applied to the original on-site image for transforming it to the position of the ortho-camera. Finally, the ROI is cropped out using the binary mask of the target face generated from the ortho-camera.

Semantic Segmentation. For semantically segmenting the ROI cropped image, a superpixel classification approach is taken. First, the image is segmented into superpixels of user-defined numbers using the Simple Linear Iterative Clustering (SLIC) algorithm. SLIC clusters the image pixels based on their color similarity and the proximity in the image plane (Radhakrishna et al. 2012). Further, the superpixels are converted to valid image formats using open-source image processing libraries such as OpenCV and Pillow. A convolutional neural network (CNN) is trained through a transfer learning approach for predicting the class of each superpixel. The class represents the construction activities such as concrete, formwork, etc. The classified superpixels with the same class information are masked with a unique color and overlayed on the ROI cropped image. This helps in visualizing the semantic segmentation and percentage progress calculation.

Progress percentage calculation. The pixel area of each colored mask that represents one activity is calculated and considered as the area of work done. Also, the area of the ROI is calculated and considered as the total work to be completed. Finally, the percentage of work progress is calculated using the following formula:

$$Percentage\ of\ progress\ of\ an\ activity = \frac{Area\ of\ the\ mask\ represnting\ the\ activity}{Total\ area\ of\ the\ region\ of\ interest} \tag{1}$$

EXPERIMENT & EVALUATION

For evaluating the methodology proposed in the previous section, a proof-of-concept study was conducted on a building construction project. The as-built point cloud model for a specific scan date was reconstructed from 189 aerial images following a standard SfM-MVS pipeline. The aerial images were captured using a commercial drone and the images without any post-processing were directly used for the point cloud reconstruction. In this experiment, no coordinate correction technique was used for improving the point cloud quality. The point cloud model was finely registered with the production-level 4D BIM in the Blender environment using a transformation matrix. Figure 2 shows two independent models and the finely registered as-planned and as-built model. The starting of construction activities for any given element was checked by comparing the as-planned and as-built models. The occupancy of more than 1000 points inside the elements' bounding box confirmed the presence of the element in the as-built model.

Figure 2. 4D as-planned BIM (left), 3D as-built point cloud model (middle), and finely registered as-planned and as-built model (right)

For appearance check, first, the detected element's largest face, i.e. the target face, was back-projected to each camera using the transformation matrix and the intrinsic and extrinsic camera parameters. A camera position was selected where the target face occupied the maximum area in the camera's FOV. Next, a rendered image of both the as-planned model and a binary mask of the selected face was generated from the selected camera view. Viewing the element's face orthographically is important for measurement purposes. Perspective projection error occurring from the oblique camera view may excerpt erroneous measurement. For that reason, a virtual orthographic camera was set programmatically in the Blender environment relating to the center and face normal of the target face. Figure 3 shows an example of an ortho camera setup and the view of the point cloud model from that ortho camera. Similar to the selected camera, another rendered image of both the as-planned model and a binary mask of the target face was generated from the ortho camera view. An example of the masks rendered from the selected camera and the ortho camera are shown in Figures 4a and 4b respectively. A homography matrix was calculated between two rendered images. The same matrix was applied to the original onsite image taken from the selected camera to transform it into the orthographic camera position. After that, the binary mask generated from the ortho camera was applied to the transformed image, and the ROI was cropped out. Figures 4c and 4d show an example of the original image taken from the selected camera and the ROI cropped image respectively.

For semantic segmentation and activity-wise progress measurement, the ROI cropped image was segmented into superpixels using the SLIC algorithm. To understand the importance of the number of segments, the experiment was conducted using four different segment values: 25, 50, 100, and 200. The superpixel segmentation on the ROI cropped image can be visualized in Figures 5a snd 5c. The superpixels were further converted to image patches and classified through pre-trained CNN models. Two best performing CNN models in (Pal et al. 2021) study namely VGG19 with batch normalization and ResNet50, pre-trained on Common Object in Context (COCO) dataset and finetuned with Construction Material Library (CML) dataset (Han and Golparvar-Fard 2015) were used in this research. The models were trained and validated with a total of 1,884 images equally distributed among three classes: concrete, formwork, and others. The entire dataset was split into training and validation set with a 4:1 ratio. To enhance the data variability and to avoid overfitting problems, different augmentation techniques, such as resizing, random cropping, padding, flipping, and normalization, were applied during training. The hyperparameters used in the training process are shown in Table 1. The classified superpixels were masked with different colors representing different classes. This approach semantically segmented the appearance image with construction activity information. A visualization of this semantic segmentation can be observed in Figures 5b and 5d. Finally, the percentage progress was calculated using Eq. 1.

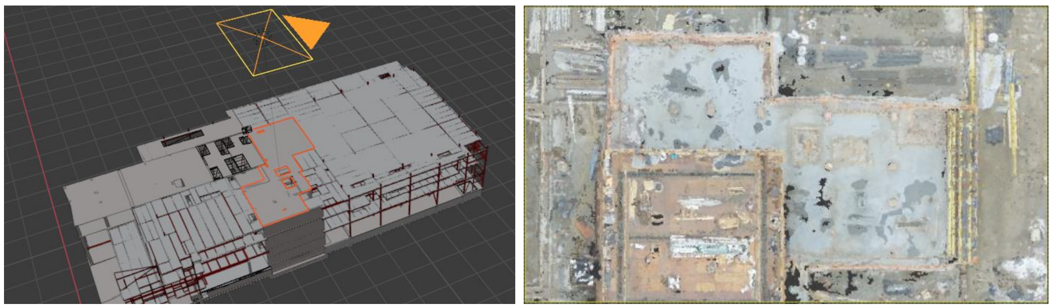


Figure 3. Ortho-camera setup in Blender environment (left), view of the point cloud model from the ortho-camera (right)

(a) (b)

(c) (d)

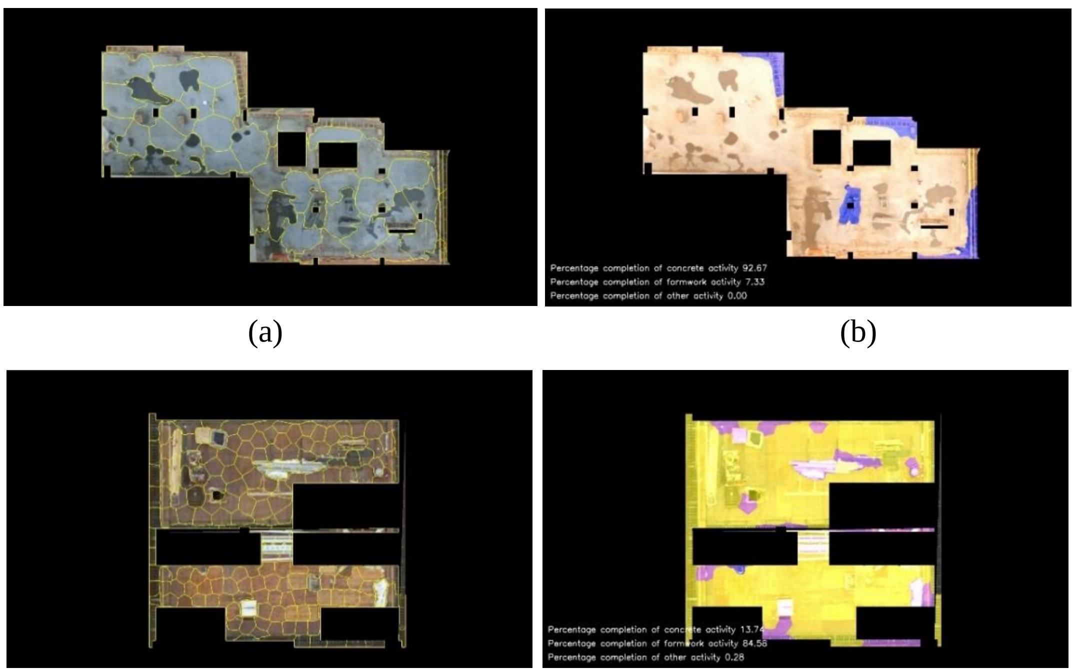
Figure 4. Binary mask rendered from the selected camera (a) and ortho-camera (b); (c) original image taken from the selected camera (c); ROI cropped image (d)

Five randomly selected elements were checked for two types of construction activities such as concrete and formwork. Three of the selected elements were slabs and the other two were walls.

Table 2 shows a comparison between the actual and the predicted progress of each element and Table 3 shows the activity-wise mean absolute error (MAE) of the predicted progress. It is observed that although both the CNN models have comparable performance, ResNet50 performed slightly better with an overall MAE of less than 11.5%. It is also noticed from the MAE values that the prediction performance for concrete activity is better than that for formwork. It may be because of the more complex and varied appearance of formwork than that of concrete.

Table 1. Hyperparameter information

Hyperparameters	Settings
Input image size	224 x 224 x 3
Number of epochs	500
Batch size	50
Learning rate	0.0001
Optimizer	Adam



(a) (b)

(c) (d)

Figure 5. (a) and (c) ROI cropped images with 50 and 200 superpixels respectively; (b) and (d) Images with predicted semantic segmentation masks (Pal et al. 2021).

The number of superpixel segments divided in the test image also played an important role in the prediction performance. It is observed in Table 4 that the concrete prediction performance was much better in the case of the lower segment numbers whereas it was reversed in the case of formwork activity. This indicates that the larger image patches with a similar appearance are easily predicted as concrete. On the contrary, for detecting the formwork activity, the models require a smaller image patch with a more uniform appearance.

Table 2. Actual progress vs predicted progress

Element Id #	Element Type	Activity Name	Actual Progress	Predicted Progress	
				VGG-19_BN	ResNet50
589546	Roof Slab	Concrete	100.00%	86.16%	**94.24%**
1287837	Floor Slab	Formwork	90.00%	61.18%	**74.07%**
5880772	Wall	Concrete	100.00%	**95.38%**	93.17%
5883142	Wall	Formwork	45.00%	**48.29%**	63.32%
630490	Roof Slab	Formwork	100.00%	82.65%	**84.58%**

Table 3. Mean absolute error (MAE) of predicted progress

Prediction Error	VGG-19_BN	ResNet50
Concrete (MAE)	9.23%	6.30%
Formwork (MAE)	16.49%	16.56%
Overall (MAE)	**12.86%**	**11.43%**

Table 4. Effect of number of superpixel segments

Element Id #	Seg#	Activity	Actual Progress	Predicted Progress	
				VGG-19_BN	ResNet50
5880772	**25**	Concrete	100%	**95.38%**	**93.17%**
	50			76.43%	70.56%
	100			55.11%	47.32%
	200			51.78%	51.16%
630490	25	Formwork	100%	39.72%	26.50%
	50			54.99%	62.93%
	100			66.51%	82.46%
	200			**82.65%**	**84.58%**

CONCLUSION

This research proposes an improved vision-based construction progress monitoring method that is capable of computing and reporting the percentage progress of tasks associated with a building element. The proposed method leverages on-site visual data, 4D BIM, and a deep learning-based computer vision algorithm. In this study, SLIC-based superpixel segmentation and CNN-based image classification are used for semantically segmenting the orthographically projected appearance image of the element. A proof-of-concept study is conducted on a building construction project and the progress of tasks, such as concrete and formwork, are estimated. It is observed that the proposed method can predict the progress percentage of tasks with less than 12% MAE. However, the prediction accuracy for concrete is found to be better than that for formwork. Also, the number of superpixel segments plays an important role in prediction accuracy. Despite interesting outcomes, this research has a few limitations. As the appearance detection method is applied on the largest face of the elements, this method will not be applicable in case the target face is heavily occluded. Also, as of now, only two construction tasks are tested for progress estimation. Future studies will focus on resolving current limitations and will try to test this method for more construction activities. Also, state-of-the-art CNN-based semantic segmentation methods will be investigated for improving the prediction accuracy.

ACKNOWLEDGEMENT

The author would like to thank the Ministry of Science and Technology, Taiwan for supporting this research through research grant Nos. MOST 109-2621-M-002-012 and MOST 109-2622-E-002-027.

REFERENCES

Braun, A., Tuttas, S., Borrmann, A., and Stilla, U. (2015). "Automated progress monitoring based on photogrammetric point clouds and precedence relationship graphs." *32nd International Symposium on Automation and Robotics in Construction and Mining: Connected to the Future, Proceedings*.

Braun, A., Tuttas, S., Borrmann, A., and Stilla, U. (2020). "Improving progress monitoring by fusing point clouds, semantic data and computer vision." *Automation in Construction*, Elsevier, 116(August 2019), 103210.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2011). "Integrated sequential as-built and as-planned representation with D 4AR 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., Feniosky, P. M., and Savarese, S. (2009). "D4AR-A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication." *Electronic Journal of Information Technology in Construction*, 14(June), 129–153.

Han, K., and Golparvar-Fard, M. (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., Cline, D., and Golparvar-Fard, M. (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*, Elsevier Ltd, 29(4), 889–901.

Kim, C., Kim, C., and Son, H. (2013). "Automated construction progress measurement using a 4D building information model and 3D data." *Automation in Construction*, Elsevier, 31, 75–82.

Kopsida, M., and Brilakis, I. (2020). "Real-Time Volume-to-Plane Comparison for Mixed Reality–Based Progress Monitoring." *Journal of Computing in Civil Engineering*, 34(4), 04020016.

Lin, J. J., and Golparvar-Fard, M. (2021). "Visual and Virtual Production Management System for Proactive Project Controls." *Journal of Construction Engineering and Management*, 147(7), 04021058.

Pal, A., Lin, J. J., and Hsieh, S.-H. (2021). "Semantic Segmentation of Superpixels for Vision-based Automated Construction Progress Reporting." *Proceedigs of the 25th Symposium on Construction Engineering and Management*, Taipei, Taiwan.

Pour Rahimian, F., Seyedzadeh, S., Oliver, S., Rodriguez, S., and Dawood, N. (2020). "On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning." *Automation in Construction*, Elsevier, 110(August 2019), 103012.
Puri, N., and Turkan, Y. (2020). "Bridge construction progress monitoring using lidar and 4D design models." *Automation in Construction*, Elsevier, 109(August 2019), 102961.
Radhakrishna, A., Appu, S., Kevin, S., Aurelien, L., Pascal, F., and Sabine, S. (2012). "SLIC Superpixels Compared to State-of-the-art Superpixel Methods." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2274–2282.
Turkan, Y., Bosche, F., Haas, C. T., and Haas, R. (2012). "Automated progress tracking using 4D schedule and 3D sensing technologies." *Automation in Construction*, Elsevier B.V., 22, 414–421.
Vick, S., and Brilakis, I. (2018). "Road Design Layer Detection in Point Cloud Data for Construction Progress Monitoring." *Journal of Computing in Civil Engineering*, 32(5), 04018029.
Yang, J., Park, M. W., Vela, P. A., and Golparvar-Fard, M. (2015). "Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future." *Advanced Engineering Informatics*, Elsevier Ltd, 29(2), 211–224.