IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 17, 2021, accepted January 12, 2022, date of publication January 18, 2022, date of current version January 21, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3144150

Rough Registration of BIM Element Projection for Construction Progress Tracking

JINGGUO XUE[1], XUELIANG HOU[1], AND YING ZENG[2]
[1]School of Economic and Management, North China Electric Power University, Beijing 102206, China
[2]State Grid Mianyang Power Supply Company, Mianyang 621000, China

Corresponding author: Jingguo Xue (xue-jingguo@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71171081, and in part by the Natural Science Foundation of Beijing Municipality under Grant 9162014.

ABSTRACT Monitoring construction progress is essential for project management. A variety of excellent automatic schedule acquisition methods have been proposed, particularly 3D reconstruction. However, both laser-based and image-based 3D reconstruction methods rely on point clouds which have some inherent defects. In this study, a rough registration method was proposed to obtain construction progress. The method does not generate point clouds during the entire process but determines the actual construction progress by registering the target detection results and the projection of building information model elements. The method was tested at the construction sites of commercial and residential buildings. The experimental results indicated that the registration accuracy reached **95.13%**. The average external parameter calibration time and the registration time of each image are **18.57s** and **50.59ms**, respectively. Compared to similar 3D reconstruction methods, the proposed method is realistic, fast, and simple. This provides a promising method for 3D reconstruction from unordered construction-site images.

INDEX TERMS 3D reconstruction, building information modeling, computer vision, construction component detection, rough registration.

I. INTRODUCTION

Owing to the great investment, long duration, and many risk factors, construction projects require lean and efficient management. Traditional information collection methods that rely heavily on manual work can no longer satisfy the current management needs of large-scale projects. First, traditional manual acquisition is time-consuming, as it takes approximately 20–30% of the feeders' daily efforts to update the construction activities, which obviously cannot meet the requirements for fine supervision of construction [1]. Moreover, tedious processes may result in human error and reduce the quality of the data. Delays and errors in information tend to lead to wrong decisions that affect personnel and projects [2]–[3]. Therefore, it is a challenge for the architecture, engineering, and construction (AEC) industry to improve the information feedback speed, quickly and accurately statistics the project data, and keep abreast of the construction site in time [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Jiju Poovvancheri.

To this end, several automatic technologies have been applied to improve the accuracy and feedback speed of information, such as global positioning system (GPS), bar codes, radio frequency identification (RFID), video and audio technology, and computer vision technology [5]–[9]. In particular, the vision-based 3D reconstruction provides a feasible solution for effective monitoring of the construction process. For example, Braun *et al.* [10] created point clouds from the fusion of disparity maps and matched the point clouds with the building information modeling (BIM) models. Subsequently, the density of points within a certain distance from the surface was used to infer the existence of building components. Omar *et al.* [1] imported images into a software (Agisoft PhotoScan Pro) to generate point clouds and inferred the height of cubes by calculating the density of point clouds between the internal and external envelope boxes of BIM elements. Bonczak and Kontokosta [11] developed a digital surface model derived from aerial LiDAR point cloud data to calculate the building massing, height, volume, exposed surface area, and compactness ratios for every building.

A common feature of these studies was that point clouds were created as a medium to retain the 3D information of

the as-built building. However, as a key part, point clouds are complex and messy. First, if images are captured and improperly processed, point clouds are generally noisy and largely affected by unwanted objects [12]–[13]. For various forms of point clouds, it is time consuming to remove all points related to the background and objects of no interest [2]. Additionally, there is no guarantee of completeness of point clouds, and sufficient overlaps among images are necessary to cover all areas of interest [12], [14]. Therefore, it is necessary to find a new way to infer the progress deviation from images without generating point clouds.

Recently, with the breakthrough of deep learning, target detection technology has made unprecedented development, which provides a new direction for image-based construction process monitoring. In the AEC industry, target-detection technology has been applied to the detection and tracking of construction personnel, materials, and equipment. Park and Brilakis [15] focused on the continuous localization of construction workers through the integration of detection and tracking. Zhu *et al.* [16] identified and tracked the workforce and equipment from construction jobsite videos. Doukari and Greenwood [17] automatically extracted useful information from rasterized plans using image processing techniques.

Therefore, a new 3D reconstruction method based on target detection is proposed. The building components are detected from construction site images, and the detection results are mapped to BIM models, which provides an approach for comparing the detection results with the as-planned model to judge the changes in the construction site within a period. In previous studies, a method was proposed to identify partial building components in images [18]. Therefore, the key is to register the detection results and the BIM elements.

Based on this, a method to register the target detection results and BIM elements was proposed. This was achieved by comparing the polygonal envelopes projected from BIM elements with rectangular frames generated by target detection. The remainder of this paper is organized as follows. First, literature on image-based 3D reconstruction and image-based building component recognition is briefly introduced. The next section presents the process of rough registration of BIM element projection for the 3D reconstruction of building structures, and the main steps are discussed in detail. Finally, the method was tested at a construction site of a commercial and residential building, and the experimental results and challenges were discussed.

II. RELATED WORK

A. IMAGE-BASED 3D RECONSTRUCTION

Existing computer vision-based 3D reconstructions can be classified into active and passive sensing. Active sensing produces point clouds by direct scanning with a 3D laser scanner [11], [19]–[20] or 3D ranging cameras [14], [21]–[22]. Passive sensing produces pictures from a camera [3], [23] or video frames from monitors [16], [24]–[25].

3D laser scanning technology, based on the laser ranging principle, can quickly reconstruct the 3D point cloud model of a measured object, which has unique advantages in terms of efficiency and accuracy [26]. However, the high cost, large amount of data collected on-site, long time consumption, and high technical requirements for operation prevent its wide application [19], [27]–[28]. Similarly, a 3D ranging camera obtains a depth image using a small laser scanner mounted on the helmet and converts the depth image into 3D point clouds via software, which has unique advantages in terms of size and cost. However, its measurement range is limited, and it requires a lot of registration work for point clouds in the later stage [22], [29]–[30]. In this research, a passive sensing method called image-based technology is used, and further analysis is as follows [31].

1) GENERAL FRAMEWORK FOR IMAGE-BASED 3D RECONSTRUCTION

The general framework for image-based 3D reconstruction is shown in Figure 1. Point clouds are generated from images/frames, and the point clouds or images/frames are aligned with the BIM. Subsequently, two different inference paths exist: geometry-based and appearance-based. The former infers the BIM element status by calculating the density of the point clouds in the occupied space of the BIM element. Owing to the inevitable occlusion and incomplete data, some researchers combined the logical relationship of the construction process and the physical relationship between the building components in the reasoning process [10]. The latter projects BIM elements registered with 3D point clouds into the image planes and extracts the texture information of different building components for reasoning [32]–[33].

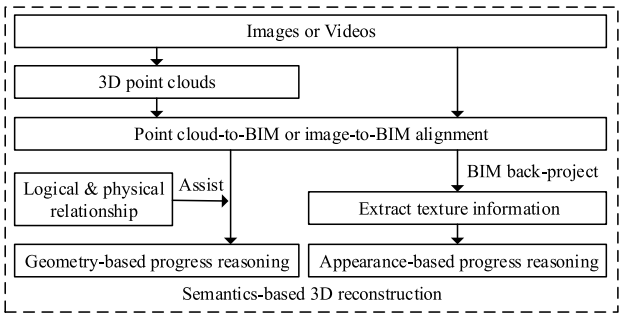


FIGURE 1. General framework for image-based 3D reconstruction.

2) DATA COLLECTION AND 3D POINT CLOUDS

The construction-site images include the images captured by the camera [23], [34]–[36] and video frames, which are essentially continuous images [16], [24], [37]–[38]. The image sets collected in different studies are diverse and include time-lapse images or videos from fixed locations, unordered sets of images, and sequenced sets of images [12].

Owing to the difference in image sets, the methods used to generate point clouds also vary. Braun *et al.* [10] acquired pictures from binocular cameras, and then generated 3D point clouds according to the triangulation

principle. The depth information for each point was calculated from the corresponding images. Omar *et al.* [1] obtained images from multiple cameras with fixed positions, shooting directions, and zoom ratios and imported them into the Agisoft PhotoScan Pro software to generate point clouds. Han and Golparvar-Fard [33] obtained point clouds from a video. Based on the SfM technique, they inferred the 3D coordinates of 2D image points from successive frames in a moving image. Golparvar-Fard *et al.* [39]–[41] generated point clouds from unordered construction-site images based on the principles of SfM and scale-invariant feature transform (SIFT).

3) ALIGNMENT

Aligning the as-built and as-planned models is a necessary step in the 3D reconstruction. After alignment, the mapping relationship between the BIM and the actual image was established to facilitate information extraction. There are two alignment methods: point-cloud-to-BIM and image-to-BIM. In some studies, point clouds were generated from images and aligned with BIM. Han and Golparvar-Fard [33] and Golparvar-Fard *et al.* [41] proposed an approach that allows users to select a set of corresponding control points from the as-built point cloud or registered imagery and have those associated with the as-planned model. In other studies, the images were directly aligned using BIM. Kropp *et al.* [42] manually aligned the first frame of a video with the wireframe of a BIM model in an AR manner. The lines in the frames were then analyzed to align the BIM with the subsequent frames of the video. Asadi and Han [43] and Asadi *et al.* [44] identified a possible perspective for localizing a camera by analyzing the missing points and lines in a video. Xue *et al.* [45] transformed image-to-BIM alignment into an optimization problem. They defined the variables, objective functions, and constraints. The optimization process involved continuously adjusting the position and angle of the BIM model through the API.

4) CONSTRUCTION PROGRESS REASONING

There are two reasoning methods: geometry-based and appearance-based.

a: GEOMETRY-BASED REASONING

Braun *et al.* [10] divided the surface of BIM elements by grids, and points within a certain distance from the surface were used to judge whether a component existed. The occluded components were then inferred according to the logical and physical relationships between the components. Omar *et al.* [1] created two envelope boxes inside and outside at a certain distance from the BIM element surface. Then, the redundant point clouds were removed, and only the point clouds between the two envelope boxes were kept. Finally, the height of the cube was determined by calculating the density of the point clouds.

b: APPEARANCE-BASED REASONING

Han and Golparvar-Fard [33] projected registered BIM elements into 2D image planes and determined the component type through texture recognition of the image patches. Compared with the geometry-based method, appearance-based reasoning can recognize different material types and, therefore, detect operation-level progress [2].

These two reasoning methods have their own strengths and weaknesses. Han *et al.* [2] combined two reasoning methods, which made the results more credible.

5) TROUBLE WITH POINT CLOUDS

At present, most 3D reconstruction methods rely on point clouds, but the characteristics of point clouds cause several problems [31]. For example, billions of points need to be processed separately, which consumes computer resources, especially when the background, noise, and objects of no interest need to be deleted [2], [13], [46]. Second, the completeness of the point cloud is difficult to guarantee, and sufficient overlap between images is required to cover all regions of interest [12], [14], [47]. In addition, point clouds have problems such as high noise and difficulty in segmentation and registration [5], [48].

For point cloud defects, some scholars directly perform 3D reconstruction using images without point clouds [32]. Kim *et al.* [36] obtained a black-and-white image containing the target by processing the image and removed the area outside the mask using a mask-based filter. The mask image was compared with the target image to obtain the progress information for a single component. However, the research object was simple because the background of the image was a water surface with little noise. Therefore, it is not applicable for monitoring construction progress. Zhu and Brilakis [49] used machine learning to identify specific areas. These ANNs (artificial neural networks) and support vector machines were trained with segmented image blocks. This method can barely identify the concrete area but cannot distinguish different components; thus, the method fails to achieve image-based 3D reconstruction. Nevertheless, they provide research ideas and methods for developing intelligent algorithms.

Therefore, in this study, a deep learning-based method was used to identify building components from the images. Considering the spatial relationship between the imaging plane and subject, the construction state was inferred by analyzing the location of each component in the images. Point clouds were not created during the entire process.

B. IMAGE-BASED BUILDING COMPONENT RECOGNITION

1) TARGET DETECTION ALGORITHM

Image-based building component recognition mainly relies on the target detection algorithm. The target detection algorithms can be divided into two main categories: traditional target detection algorithm and deep learning-based target detection algorithm. Traditional target detection algorithms

usually include Cascade + HOG/DPM + Haar/SVM and the improvements and optimizations of them. Traditional target detection algorithms mainly include the following steps: image processing, region proposal, feature extraction, training classifier and classification. Since the area selection is not targeted, window redundancy occurs, and the time complexity is high. In recent years, after the breakthrough of neural network algorithm, the target detection algorithm based on deep learning has been favored by scholars. After continuous research, many target detection algorithms are iterated. These algorithms are mainly divided into two categories: region proposal-based and regression-based. The target detection algorithms based on region proposal mainly include algorithms such as R-CNN, SPP-NET, Fast R-CNN, Faster R-CNN, and R-FCN. The accuracy and speed of the CNN series algorithms are constantly improving. But, due to the large number of candidate frames extracted, there are a lot of repeated calculations, which cannot meet the requirements of real-time detection [50]. The target detection algorithms based on regression mainly include YOLO, SSD, YOLO v2, DSSD, and YOLO v3. Compared with the R-CNN series, the YOLO series detects objects with lower accuracy and recall rate, but has stronger real-time performance [51]. And the SSD series algorithms, through experimental tests, show that its performance is better than YOLO and YOLO v2. Although it is not faster than YOLO v3, the detection accuracy is comparable to Faster R-CNN [52]. DSSD is an improved branch of the SSD algorithm, which is mainly to improve the defects of small targets that are not robust enough [53]. In summary, different detection algorithms have their own advantages and disadvantages. For the automatic detection of building components, it is necessary to consider which target detection algorithm is most suitable for the construction field with obvious characteristics and can successfully implement migration learning.

2) TARGET DETECTION IN CONSTRUCTION

At present, in the AEC industry, there are few studies related to ''image-based building component recognition'', but some preliminary explorations have been conducted [3], [17]. For example, some scholars have used target detection technology to identify concrete areas from images. However, because many components are composed of concrete, the detected concrete areas are interconnected, and the building components cannot be subdivided [49], [54]. To overcome this limitation, Zhu and Brilakis [37], [55] proposed an automatic detection method for concrete columns based on visual data. By analyzing the boundary information (such as color and texture) of the concrete columns, the structural columns were separated from the concrete area. However, this method cannot be applied to the detection of building components because of the small difference in the texture and color characteristics of different components in images and the complex spatial relationship between different components.

Based on the bottleneck encountered in this research, through the analysis of various improved target detection algorithms, it was found that the deep supervised object detector (DSOD) model is more suitable for the detection of building components [56]. The use of the DenseNet network significantly reduces the model parameters and guarantees its performance [57]. More importantly, this method overcomes the limitations of traditional detector training based on pretraining and fine-tuning. Without the pretraining model, the highest level of the target detector can be obtained with a limited dataset. Therefore, this study intends to use the DSOD algorithm for automatic detection of building components.

III. METHODOLOGY

A. OVERVIEW OF THE ROUGH REGISTRATION METHOD

Figure 2 presents the process model of the proposed method. The specific steps are as follows.

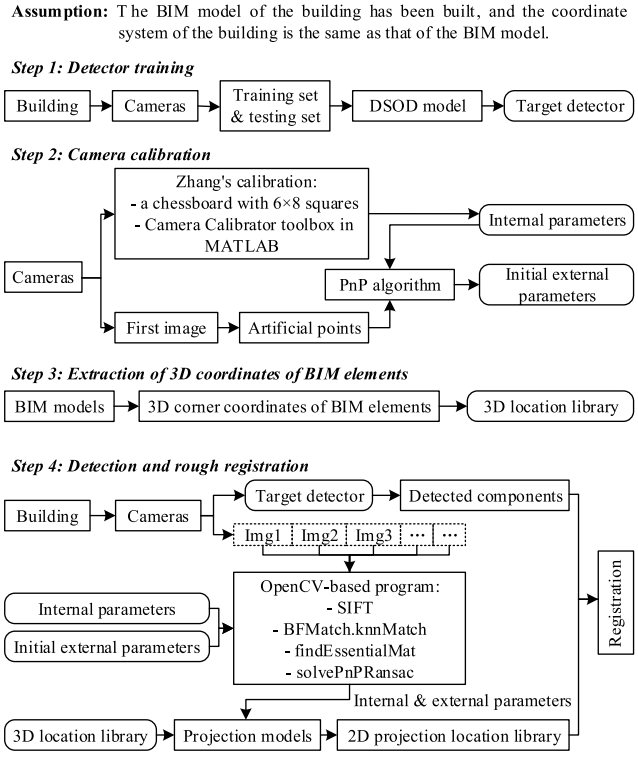


FIGURE 2. Process model of the proposed method.

1) STEP 1: DETECTOR TRAINING

First, a set of construction-site images is collected using multiple cameras. To create a comprehensive dataset, varying degrees of viewpoint, scale, and illumination must be considered. The images are divided into training and testing sets. The training set is annotated using the graphic image annotation tool LabelImg [58], generating XML files in the PASCAL VOC format, and then used to train a DSOD-based detector to automatically recognize the building components from the construction-site images. Using the trained detector/recognizer, building components (including the type and location) are detected from the test set images.

2) STEP 2: CAMERA CALIBRATION

The internal parameters of the cameras are determined using Zhang's calibration method [59]. The initial external parameters are calculated from multiple known points in the testing set image, whereas the others are obtained using an OpenCV-based program. The specific process is described in detail in Section III-C.

3) STEP 3: EXTRACTION OF 3D COORDINATES OF BIM ELEMENT

The 3D coordinates of each corner of the BIM element are extracted manually, the corner coordinates are converted into 2D coordinates on the imaging plane through the projection model, and the BIM element projection location library is established. The specific process is described in detail in Section III-B.

4) STEP 4: DETECTION AND ROUGH REGISTRATION

First, photographs of the construction site were taken. On the one hand, the photos should be input into the target detector to obtain the detection results. On the other hand, these photos should be sent to the OpenCV-based program. By identifying the SIFT feature points, the new images are registered with the original image, and the external parameter matrix is calculated to determine the corresponding projection model of each image (Section III-C). Then, the 3D coordinates obtained in Step 3 are converted into 2D coordinates on the imaging plane through the projection model, and the 2D projection location library of the BIM element is established (Section III-D). Finally, the location rectangle generated by target detection is compared with the item in the 2D projection location library, and the building components identified from the images are associated with BIM elements (Section III-E).

With this method, several manual steps must be clarified. First, the image sets, including the training, validation, and testing sets, are collected manually. Similarly, the chessboard images used to calibrate the internal parameters of the cameras are manually collected. Third, to obtain the initial external parameters for each camera, multiple known points must be manually selected from the initial images, and their 2D and 3D coordinates must be manually provided. Fourth, manual parameter adjustment is necessary during the training process of the DSOD-based detector. Fifth, the 3D coordinates of the BIM element corner are obtained manually in this study, although they could be extracted from the IFC files.

In practical applications, all the manual operations mentioned above are in Step 1-3, which must be performed only once.

B. EXTRACTION OF BIM ELEMENT CORNER COORDINATES

To distinguish each component and extract their 3D coordinates, a set of naming rules is formulated, and the corner coordinates of each component is extracted from the unified coordinate system. The specific steps are as follows:

1) STEP 3-1: NAMING THE GRID AND INTERSECTION

To locate the model, it is assumed that the transverse grid of the model is named with uppercase letters (from bottom to top, grids A, B, and C, respectively), and the longitudinal grid is named with Arabic numerals (from left to right, respectively, grid 1, grid 2…), forming the intersection of the grids (such as A1, A2, D7· · ·).

2) STEP 3-2: DISTINGUISHING THE COMPONENT TYPES BY LETTERS

A representation symbol is set for each type of member; for example, B for beams, F for floors, C for columns, and W for walls.

3) STEP 3-3: NAMING EACH COMPONENT

It is stipulated that the elements, which are in the rectangular grid or on the left and lower edges of the rectangular grid, belong to the lower left intersection. The component is named "building number + floor number + symbol of element type + grid intersection number + supplement number", as shown in Figure 3. For example, in Figure 4, the selected column is named B1F1-C-B011, which is on the first floor and belongs to the B1 grid intersection. The left and lower beams are denoted as B1F1-B-B011 and B1F1-B-B012, respectively.

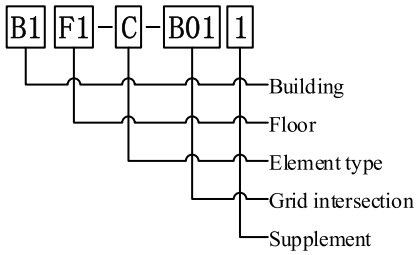


FIGURE 3. Component name number.

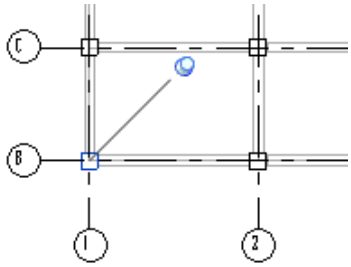


FIGURE 4. F1 plane partial view.

4) STEP 3-4: UNIFYING COORDINATE SYSTEMS

The A1 grid intersection is set as the origin of the coordinate system, the X-coordinate axis coincides with grid A, the

Y-coordinate axis coincides with grid 1, and the Z-coordinate axis is determined by the right-hand rule.

5) STEP 3-5: EXTRACTING THE CORNER COORDINATES OF THE BIM ELEMENT

According to the Cartesian coordinate system, the corner coordinates of each element are calculated, forming a BIM element corner coordinate database.

C. PROJECTION MODEL AND CAMERA CALIBRATION

The purpose of camera calibration is to determine a projection model that refers to the mapping relationship between the points in 3D space and the pixels in the photo. There are four coordinate systems in the projection model: world coordinate system (X_W, Y_W, Z_W) $(\mathrm{X_w}, \mathrm{Y_w}, \mathrm{Z_w})$, camera coordinate system $(\mathrm{X_c}, \mathrm{Y_c}, \mathrm{Z_c})$ (X_C, Y_C, Z_C), imaging plane coordinate system (X, Y), and pixel coordinate system (U, V) (U, V). According to the spatial relationship among these coordinate systems, and considering factors such as the distortion generated by the camera, the projection model is derived as follows:

$$Z_c \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = KT \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{1}$$

where Z_c is the Z coordinate of the point in the camera coordinate system; U and V are the coordinates in the pixel coordinate system; $\alpha_x = f/dx$, $\alpha_y = f/dy$, f is the camera focal length, dx, and dy are the pixel sizes; u_0 and v_0 represent the offset of the camera optical axis in the image coordinate system, in pixels; R and t represent the offset of the camera optical axis in the image coordinate system, in pixels; X_W, Y_W, and Z_W are the coordinates of the point in the world coordinate system; and K and T represent the internal and external parameter matrices, respectively.

The key to camera calibration is to obtain the internal and external parameter matrices, which is a necessary process for 3D reconstruction. Poor camera calibration can cause image defects, such as distortion, which can affect the reliability and accuracy of the information. Therefore, camera calibration is a critical process in computer-vision applications [60]–[61].

To obtain the internal parameter matrix, Zhang's calibration method is used in this study because of its mature application, wide application range, and reliable and accurate calibration results [59]. The internal parameter matrix is calculated with a chessboard using the camera calibration toolbox in MATLAB.

To obtain the external parameter matrix, no more than five images are selected to calculate the initial external parameters of each camera. The contents of these pictures are four side views and a top view of a building. These images can be vacant if the cameras are not used for some views. Subsequently, multiple known points are manually selected from these images, and the initial external parameter matrices are calculated using the efficient perspective-n-point (EPNP) algorithm [62]. Finally, the external parameter matrices of other images are calculated using an OpenCV-based program [63]. The program consists of the following steps: (1) Extract SIFT feature points from images; (2) Match features using KNN (k-nearest neighbor classification) algorithm; (3) Calculate the first external parameter matrix by "findEssentialMat" function; (4) Calculate the rests of the external parameter matrices by "solvePnPRansac" function.

D. ESTABLISHMENT OF BIM ELEMENT PROJECTION LOCATION LIBRARY

After determining the internal and external parameters of the camera projection model, the 3D coordinates of the BIM element corners are substituted into the projection model to calculate the projected pixels, as shown in Figure 5. Each corner of the component corresponds to a pixel point, which is named as 'pixel-corner'. Then, connect the exterior pixel-corners to form a convex polygon, named 'envelope frame', so that all the pixel-corners are wrapped in the envelope frame. Finally, the component location is represented by these pixel corners on the envelope frame, and a 2D projection location library of the BIM element is formed.

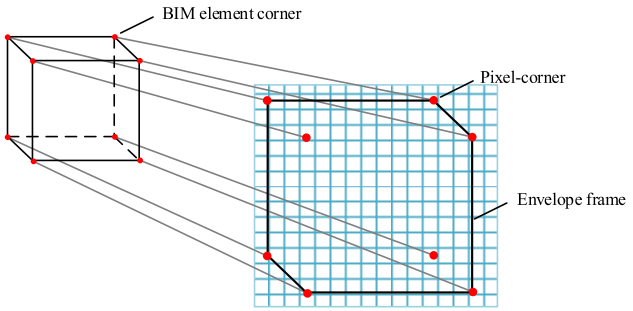


FIGURE 5. Envelope frame.

E. SCREENING AND LOCATION REGISTRATION

The purpose of rough registration is to match the building components identified in the images with items in the BIM element projection library. However, for each registration, most of the information in the BIM element projection library is useless. It is necessary to filter out unrelated envelope frames to improve the registration efficiency and accuracy. Therefore, four constraints are formulated and the following four conditions are screened:

1) COMPONENTS OF UNRELATED FLOORS

For example, the current construction location is on the second floor; therefore, the envelope frames of BIM elements above the second floor are not considered.

2) THE ENVELOPE FRAME WHICH DOES NOT OVERLAP WITH THE LOCATION FRAME OF TARGET DETECTION

The envelope frame is projected from the same direction as the image; therefore, the projection of the component in the image must overlap with the corresponding BIM element projection envelope frame.

3) ENVELOPE FRAMES WITH LARGE SIZE GAP

When the distance between the pixel corners on the envelope frame and the target detection location frame is larger than a certain threshold, the envelope frame should be excluded.

4) ENVELOPE FRAMES OF IRRELEVANT TYPE

The type of image identification result must be consistent with the type of BIM element. For example, if the result of the target detection is a column, the envelope frame of all non-column components will be excluded.

After excluding irrelevant data, the target detection location frames were compared with the remaining envelope frames individually. As shown in Figure 6, the outer rectangular frame is a location frame generated by the target detection, and the inner polygon is an envelope frame generated by the BIM element projection.

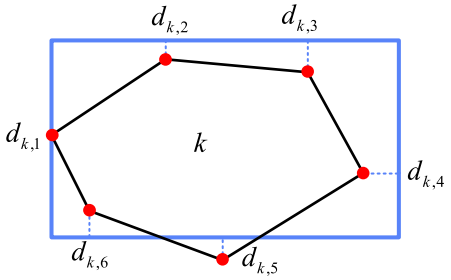


FIGURE 6. Location registration.

For each rectangular frame, the mean-square error between it and the envelope frame k can be calculated as:

$$D_k = \frac{1}{n}\sum_{i=1}^{n} d_{ki}^2, \quad i = 1, 2, \ldots, n \tag{2}$$

where k is the index of the envelope frame, d_{ki} is the minimum distance from the pixel corner i of the envelope frame k to the four edges of the rectangular frame, and n is the number of pixel corners.

D_k is used to represent the positional relationship between the two frames. When D_k is closer to zero, it indicates that the degree of overlap between them is higher, and they represent the same component with a higher probability. Therefore, the BIM element corresponding to the minimum mean-square error D_k is chosen, and the registration result is as follows:

$$K = \{k \mid \min D_k\} \tag{3}$$

Screening and rough registration are performed using MATLAB. See the Appendix for the code.

IV. EXPERIMENT AND DISCUSSION

A. DATA COLLECTION AND EXPERIMENTAL SETUP

At present, research on target detection technology in the AEC industry is still in its infancy, and there are no ready-made, open, and integrated image sets available in the construction industry for detection experiments. Therefore, a new Construction Materials Library (CML) was created, which includes 5200 images (4500/500/200). These images are recorded from the construction sites of commercial and residential buildings. To create a comprehensive dataset, varying degrees of viewpoint, scale, and illumination were considered. Subsequently, the images in the training and validation sets were annotated, including four categories (beam, floor, column, and wall). A detector based on the DSOD algorithm was trained and optimized. Finally, 9872 components were detected in the 200 test set images. After manually removing erroneous data, 8356 components (including 3032 beams, 1498 floors, 1164 columns, and 1863 walls) were retained and used as experimental data in this study.

Four constraints are set to narrow the scope of retrieval: first, the floor information (two floors) is given for the components identified from the top view (*Constraint 1*); second, at least one corner of the envelope frame is inside the rectangular frame (*Constraint 2*); third, the horizontal and vertical dimensions of the envelope frame must be within twice the size of the rectangular frame (*Constraint 3*); and fourth, the component type of the envelope frame must be consistent with the recognition result (*Constraint 4*). Rough registration was performed under four constraints.

B. EXPERIMENTAL RESULTS

Images are selected from the top view and side view to vividly show the experimental results. Figure 7 and 8 show the rough registration results for the top and side views, respectively. The registration results are presented by overlaying the BIM model on the construction image.

There are 9854 components in the BIM model, so there are 9854 segmented elements (9854 2D BIM elements). A total of 8356 members were identified from the 200 images. Each identified member was registered with filtered 2D BIM elements.

Among the 8356 components in 200 images, the registration rate reached 95.13%, and the average external parameter calibration time and registration time of each image were 18.57 s and 50.59 ms, respectively. To analyze the effects of the four constraints, the registration results under different constraints were counted, as shown in Table 1.

In comparison, it was found that *Constraint 1* has a significant effect on registration accuracy, whereas *Constraints 1, 2, and 4* exert more influence on the average registration time. The floor information (*Constraint 1*) can avoid confusion regarding the same components of the upper and lower floors in the top view. In the actual construction process, the floor shown in the top view does not change in a period; therefore, it is feasible to provide approximate floor information

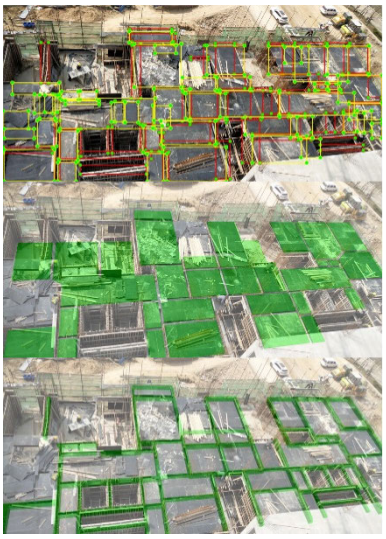


FIGURE 7. Top view (Detection result; Registration results of floors and beams; 1920 × 940 resolution).

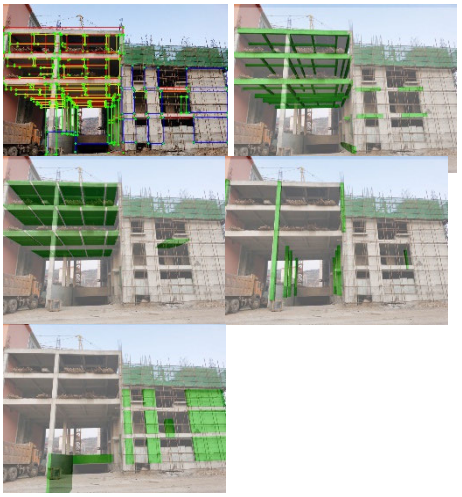


FIGURE 8. Side view (Detection result; Registration results of beams, floors, columns, and walls; 3648 × 2736 resolution).

(for example, with two floors). Excluding non-overlapping components (*Constraint 2*) and known component types (*Constraint 4*) may not improve registration accuracy, but they can significantly shorten registration time. Different allowable size deviations (*Constraint 3*) have no significant effect on the accuracy and speed of registration.

To further analyze the registration results under different shooting angles, the images were divided into multiple blocks based on the building surface. The registration results for each building surface were counted separately, as listed in Table 2.

In the top view, more beams and floors were identified, which were mainly in the formwork and reinforcement states. From Table 2, it can be seen that the smaller the shooting angle, the lower the registration accuracy, because the smaller the angle, the easier it is for the components with equivalent shapes on different floors to overlap. In the side view, more

TABLE 1. Registration results under different constraints.

Constraints		No. images	No. comp.	No. registered comp.	ACC [a]	Time [b] (ms)
Floor (1)	Known	200	8356	8093	96.85%	43.21
	Vague [c]	**200**	**8356**	**7949**	**95.13%**	**50.59**
	Unknown	200	8356	6750	80.78%	263.12
Non-overlapping comp. (2)	**Excluded**	**200**	**8356**	**7949**	**95.13%**	**50.59**
	Unexcluded	200	8356	7910	94.66%	634.54
Allowable size deviation multiple (3)	1.5 times	200	8356	7948	95.12%	50.21
	2 times	**200**	**8356**	**7949**	**95.13%**	**50.59**
	5 times	200	8356	7984	95.55%	51.11
	All	200	8356	7980	95.50%	51.12
Comp. type (4)	**Known**	**200**	**8356**	**7949**	**95.13%**	**50.59**
	Unknown	200	8356	7949	95.13%	95.61

[a] Ratio of the number of registered components (true positive) to the total number of components (true positive + false negative). Because only those correctly detected components were registered, there were no negative samples in the 8356 samples in the experiment. Therefore, ACC = true positive/(true positive + false negative).

[b] Average registration time of 200 images (including screening and location registration).

[c] The floor information (2 floors) was given in this study.

TABLE 2. Registration results under different shooting angles.

View Angle [a]		No. blocks	No. comp.				Registration ACC (%)			
			Beam	Floor	Column	Wall	Beam	Floor	Column	Wall
Top view	0-30°	40	388	313	233	123	95.82	95.91	90.79	94.31
	30-60°	34	347	280	227	103	97.77	96.31	92.75	95.15
	60-90°	28	303	250	213	95	98.77	96.30	96.23	97.89
Side view	0-30°	90	703	120	196	532	93.75	89.90	85.81	95.86
	30-60°	73	632	184	145	503	95.27	94.57	90.20	96.62
	60-90°	76	659	351	150	507	96.11	95.41	89.20	96.45

[a] Angle between the camera's optical axis and the normal of the building surface.

beams and walls were identified, and they were all in a concrete state. Compared with the top view, the accuracy was relatively low because there were no floor restrictions. As the shooting angle increased, the accuracy increased; however, the accuracy improvement was not obvious or even reduced when the shooting angle was too large. Therefore, it is appropriate to have an appropriate angle with the normal of the building surface.

Some examples of the registration errors are shown in Figure 9. From left to right are the actual recognition results, incorrect registration results, expected recognition results, and correct registration results, respectively. The four examples are: a wall covered by debris, a beam covered by the building itself, an internal wall covered by an external column, and a floor template covered by the building itself. The inevitable or accidental occlusion results in the component not pairing with its corresponding BIM elements but is mismatched with BIM elements with similar projected areas.

In Figure 9, Case A is a static occlusion, and Cases B, C, and D are self-occlusions. These occlusions made only a part of the component recognized, and part of the component

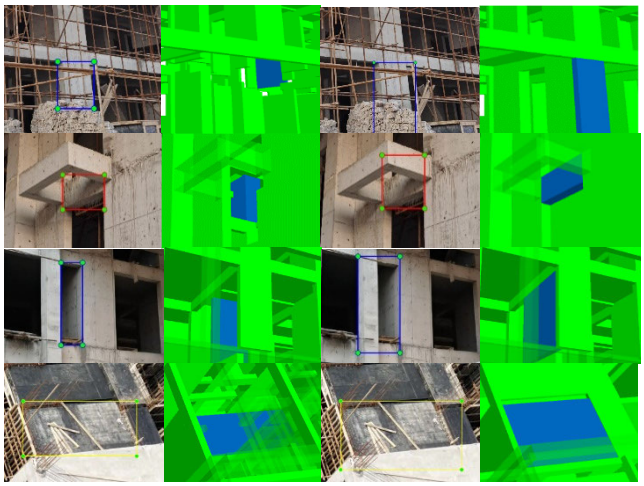


FIGURE 9. Wrong examples (Actual detection results; Wrong registration results; Expected recognition results; Correct registration results).

was registered with other components. Some occlusions can be eliminated by combining the results of multiple-angle images, whereas others cannot. Adding the logical and physical relationships of the components can assist in reasoning and improving the final accuracy. However, some scholars have pointed out that this approach is opportunistic. Only the detected components can be considered to exist. This problem requires further discussion of relevant research.

C. CHALLENGES

Although this paper presented the registration process between building component detection results and BIM elements, several critical challenges remain. Some of the open research problems include the following.

1) OCCLUSION

Occlusion is an inevitable and challenging problem that must be resolved. It was found that most of the components in the side view were blocked by scaffolds and protective nets. Therefore, the acquisition of progress information depends largely on the identification and registration of the components in the top view. However, once the concrete was poured, it was impossible to identify the beams, columns, and walls in the top view. Therefore, further research is required to identify and register these components in indoor images.

2) BIM ACCURACY

Many components were identified in the experiment, but the size and position deviation from the projected envelope were too large to be registered. For example, multiple beams can be identified as a whole, and one beam can be identified as multiple segments. It is difficult to distinguish the specific areas of the different components in an image based on a combination of BIM elements. Therefore, determining the accuracy of the BIM for registering these components is a challenge.

3) DEGREE OF CONSTRAINTS

To avoid errors in rough registration, the types, positions, and sizes of the building components are restricted in Section III-E. Because the same type of components rarely appear together in a local area, it is ideally difficult for this method to misregister. However, there are cases where the recognition result deviates from the actual size, which can easily fail to register if the constraints are too strict. Appropriate reduction of restrictions and manual judgment are optional methods. Further research is required to construct a constraint network. Sometimes, manual processing is necessary because the accuracy of the algorithm is difficult to improve when it reaches a certain level. At this time, manual processing can solve this problem at a low cost.

4) INDOOR SCENE

Indoors, the distance between the camera and the object being photographed is short. A large number of photos must be taken to cover all scenes on a floor. The more the number of photos, the more difficult the camera calibration, and the greater the calibration error. If multiple start scenes are set, registration accuracy can be improved. However, each initial scene requires manual selection of known points, which increases people's workload and does not meet the original intention of automatically collecting progress information. In the future, this method needs to be further improved to realize its application in indoor scenes.

D. COMPARISON WITH SIMILAR METHOD

A similar method uses camera projection models to ''back-project'' BIM onto image planes. In these studies, 4D BIM was projected onto site images, and the image patches corresponding to each BIM element were extracted. Then, the states of the building components were inferred from the change in appearance of the image patches. This approach relies on planned 4D BIM, in which different time periods correspond to different components, and the components of different time periods are generated in turn.

However, the actual construction progress is not similar to that of designed 4D BIM. If the actual progress or sequence is inconsistent with the plan, the projected area is meaningless. In other words, the 4D BIM components generated in the next time period are not necessarily the content to be constructed in practice. Some components are not detected in the back projection, which does not mean that progress has been delayed. It may be that the other components were first constructed. The method proposed in this study detects building components from site images, discretizes each component, and cancels the relationship between them in a time sequence. The detection result is consistent with reality, and the rough registration result between the rectangular frames generated by target detection and the polygonal envelopes projected from BIM elements is also consistent with the actual construction process. Therefore, the 3D reconstruction results are realistic and are not constrained by the planning process.

In addition, the proposed method requires less data and is faster. For example, in the case of Han *et al.*, 69 components generated 6390085 image points, whereas 40 components generated 148622647 laser points. For processing more than a hundred million points, the proposed method requires approximately 30 s, including both filtering and reasoning steps. In contrast, in this study, only 91195 points were processed for the model with 9854 components. The state of the building components was judged using rough registration. For each building component, what needs to be processed is calculating the distances be-tween the corner points of the projected polygon and the rectangular border generated by target detection. In addition, the number of candidate polygons was significantly reduced by the constraints. Therefore, the rough registration process was quite fast (50.59ms/image).

Finally, dimension reduction is used to convert the 3D as-planned model into a 2D polygon. Data comparison must be processed in the location registration. These processes are simple, accurate, and nearly unchallenged.

V. CONCLUSION

To reconstruct a 3D model of a building structure from unordered construction-site images, this study contributes a novel rough registration method. The mapping relationship between the BIM model and construction-site images was established through research on BIM element coordinate extraction, camera calibration, etc.. After detecting the building components from the construction-site images, the target detection results were registered with BIM elements under four constraints. From the experiment, a registration rate of 95.13% and average registration time of 50.59ms were achieved. The effects of various factors, such as floor information, non-overlapping components, allowable size deviation, component types, and shooting angles, were considered. Furthermore, it was found that the floor information had a significant effect on registration accuracy, whereas the non-overlapping components, allowable size deviation, and component types had a significant effect on the average registration time. It was appropriate to have a certain angle with the normal of the photographed building surface.

Compared to similar 3D reconstruction technologies, the proposed method is realistic, fast, and simple. Nevertheless, several critical challenges remain in the research that need to be studied in the future, including the improvement of registration accuracy, recognition and registration of indoor images, and reasonable setting of constraints.

APPENDIX

MATLAB code: https://github.com/yonglong2020/Rough-Registration

REFERENCES

[1] H. Omar, L. Mahdjoubi, and G. Kheder, "Towards an automated photogrammetry-based approach for monitoring and controlling construction site activities," *Comput. Ind.*, vol. 98, pp. 172–182, Jun. 2018, doi: 10.1016/j.compind.2018.03.012.

[2] K. Han, J. Degol, and M. Golparvar-Fard, "Geometry-and appearance-based reasoning of construction progress monitoring," *J. Construct. Eng. Manage.*, vol. 144, no. 2, 2018, Art. no. 04017110, doi: 10.1061/(ASCE)CO.1943-7862.0001428.

[3] F. Pour Rahimian, S. Seyedzadeh, S. Oliver, S. Rodriguez, and N. Dawood, "On-demand monitoring of construction projects through a game-like hybrid application of BIM and machine learning," *Automat. Construct.*, vol. 110, Feb. 2020, Art. no. 103012, doi: 10.1016/j.autcon.2019.103012.

[4] Y. Zhao, X. Deng, and H. Lai, "A deep learning-based method to detect components from scanned structural drawings for reconstructing 3D models," *Appl. Sci.*, vol. 10, no. 6, p. 2066, Mar. 2020, doi: 10.3390/app10062066.

[5] Q. Wang, Y. Tan, and Z. Mei, "Computational methods of acquisition and processing of 3D point cloud data for construction applications," *Arch. Comput. Methods Eng.*, vol. 27, no. 2, pp. 479–499, Apr. 2020, doi: 10.1007/s11831-019-09320-4.

[6] T. Czerniawski and F. Leite, "Automated digital modeling of existing buildings: A review of visual object recognition methods," *Autom. Construct.*, vol. 113, May 2020, Art. no. 103131, doi: 10.1016/j.autcon.2020.103131.

[7] S. A. Salehi and İ. Yitmen, "Modeling and analysis of the impact of BIM-based field data capturing technologies on automated construction progress monitoring," *Int. J. Civil Eng.*, vol. 16, no. 12, pp. 1669–1685, Dec. 2018, doi: 10.1007/s40999-018-0320-1.

[8] J. Liang, J. Zhang, B. Pan, S. Xu, G. Zhao, G. Yu, and X. Zhang, "Visual reconstruction and localization-based robust robotic 6-DoF grasping in the wild," *IEEE Access*, vol. 9, pp. 72451–72464, 2021, doi: 10.1109/ACCESS.2021.3079245.

[9] C. Jiang, Y. He, X. Zheng, and Y. Liu, "OmniTrack: Orientation-aware RFID tracking with centimeter-level accuracy," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 634–646, Feb. 2021, doi: 10.1109/TMC.2019.2949412.

[10] A. Braun, S. Tuttas, A. Borrmann, and U. Stilla, "A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds," *J. Inf. Technol. Construct.*, vol. 20, no. 8, pp. 68–79, 2015. [Online]. Available: http://www.itcon.org/2015/5.

[11] B. Bonczak and C. E. Kontokosta, "Large-scale parameterization of 3D building morphology in complex urban landscapes using aerial LiDAR and city administrative data," *Comput., Environ. Urban Syst.*, vol. 73, pp. 126–142, Jan. 2019, doi: 10.1016/j.compenvurbsys.2018.09.004.

[12] K. K. Han and M. Golparvar-Fard, "Potential of big visual data and building information modeling for construction performance analytics: An exploratory study," *Autom. Construct.*, vol. 73, pp. 184–198, Jan. 2017, doi: 10.1016/j.autcon.2016.11.004.

[13] B. Bortoluzzi, I. Efremov, C. Medina, D. Sobieraj, and J. J. McArthur, "Automating the creation of building information models for existing buildings," *Automat. Construct.*, vol. 105, Sep. 2019, Art. no. 102838, doi: 10.1016/j.autcon.2019.102838.

[14] Y. Li, W. Li, S. Tang, W. Darwish, Y. Hu, and W. Chen, "Automatic indoor as-built building information models generation by using low-cost RGB-D sensors," *Sensors*, vol. 20, no. 1, p. 293, Jan. 2020, doi: 10.3390/s20010293.

[15] M.-W. Park and I. Brilakis, "Continuous localization of construction workers via integration of detection and tracking," *Automat. Construct.*, vol. 72, pp. 129–142, Dec. 2016, doi: 10.1016/j.autcon.2016.08.039.

[16] Z. Zhu, X. Ren, and Z. Chen, "Integrated detection and tracking of workforce and equipment from construction jobsite videos," *Automat. Construct.*, vol. 81, pp. 161–171, Sep. 2017, doi: 10.1016/j.autcon.2017.05.005.

[17] O. Doukari and D. Greenwood, "Automatic generation of building information models from digitized plans," *Automat. Construct.*, vol. 113, May 2020, Art no. 103129, doi: 10.1016/j.autcon.2020.103129.

[18] X. Hou, Y. Zeng, and J. Xue, "Detecting structural components of building engineering based on deep-learning method," *J. Construct. Eng. Manage.*, vol. 146, no. 2, 2020, Art. no. 04019097, doi: 10.1061/(ASCE)CO.1943-7862.0001751.

[19] D. Rebolj, Z. Pučko, N. Č. Babič, M. Bizjak, and D. Mongus, "Point cloud quality requirements for Scan-vs-BIM based automated construction progress monitoring," *Automat. Construct.*, vol. 84, pp. 323–334, Dec. 2017, doi: 10.1016/j.autcon.2017.09.021.

[20] R. Albano, "Investigation on roof segmentation for 3D building reconstruction from aerial LIDAR point clouds," *Appl. Sci.*, vol. 9, no. 21, Nov. 2019, Art. no. 4674, doi: 10.3390/app9214674.

[21] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3D semantic parsing of large-scale indoor spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1534–1543, doi: 10.1109/CVPR.2016.170.

[22] R. Volk, T. H. Luu, J. S. Mueller-Roemer, N. Sevilmis, and F. Schultmann, "Deconstruction project planning of existing buildings based on automated acquisition and reconstruction of building information," *Automat. Construct.*, vol. 91, pp. 226–245, Jul. 2018, doi: 10.1016/j.autcon.2018.03.017.

[23] A. Paixão, R. Resende, and E. Fortunato, "Photogrammetry for digital reconstruction of railway ballast particles—A cost-efficient method," *Construct. Building Mater.*, vol. 191, pp. 963–976, Dec. 2018, doi: 10.1016/j.conbuildmat.2018.10.048.

[24] F. Dai, A. Rashidi, I. Brilakis, and P. Vela, "Comparison of image-based and time-of-flight-based technologies for three-dimensional reconstruction of infrastructure," *J. Construct. Eng. Manage.*, vol. 139, no. 1, pp. 69–79, Jan. 2013, doi: 10.1061/(ASCE)CO.1943-7862.0000565.

[25] D. Acharya, M. Ramezani, K. Khoshelham, and S. Winter, "BIM-Tracker: A model-based visual tracking approach for indoor localisation using a 3D building model," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 157–171, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.014.

[26] C.-C. Chu, N. Nandhakumar, and J. K. Aggarwal, "Image segmentation using laser radar data," *Pattern Recognit.*, vol. 23, no. 6, pp. 569–581, 1990, doi: 10.1016/0031-3203(90)90035-j.

[27] S. Bechtold and B. Höfle, "HELIOS: A multi-purpose LIDAR simulation framework for research, planning and training of laser scanning operations with airborne, ground-based mobile and stationary platforms," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 3, no. 3, pp. 161–168, Jun. 2016, doi: 10.5194/ISPRS-ANNALS-III-3-161-2016.

[28] S. Hong, I. Park, J. Lee, K. Lim, Y. Choi, and H.-G. Sohn, "Utilization of a terrestrial laser scanner for the calibration of mobile mapping systems," *Sensors*, vol. 17, no. 3, pp. 474–497, Feb. 2017, doi: 10.3390/s17030474.

[29] O. Wasenmüller and D. Stricker, "Comparison of Kinect v1 and v2 depth images in terms of accuracy and precision," in *Proc. Asian Conf. Comput. Vis. (ACCV)*. Taiwan, China: Springer, Nov. 2016, pp. 34–45, doi: 10.1007/978-3-319-54427-4_3.

[30] Z. Pučko, N. Šuman, and D. Rebolj, "Automated continuous construction progress monitoring using multiple workplace real time 3D scans," *Adv. Eng. Informat.*, vol. 38, pp. 27–40, Oct. 2018, doi: 10.1016/j.aei.2018.06.001.

[31] J. Xue, X. Hou, and Y. Zeng, "Review of image-based 3D reconstruction of building for automated construction progress monitoring," *Appl. Sci.*, vol. 11, no. 17, p. 7840, Aug. 2021, doi: 10.3390/app11177840.

[32] H. Fathi, F. Dai, and M. Lourakis, "Automated as-built 3D reconstruction of civil infrastructure using computer vision: Achievements, opportunities, and challenges," *Adv. Eng. Informat.*, vol. 29, no. 2, pp. 149–161, Apr. 2015, doi: 10.1016/j.aei.2015.01.012.

[33] K. K. Han and M. Golparvar-Fard, "Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs," *Automat. Construct.*, vol. 53, pp. 44–57, May 2015, doi: 10.1016/j.autcon.2015.02.007.

[34] S. El-Omari and O. Moselhi, "Integrating 3D laser scanning and photogrammetry for progress measurement of construction work," *Automat. Construct.*, vol. 18, no. 1, pp. 1–9, Dec. 2008, doi: 10.1016/j.autcon.2008.05.006.

[35] F. Dai and M. Lu, "Assessing the accuracy of applying photogrammetry to take geometric measurements on building products," *J. Construct. Eng. Manage.*, vol. 136, no. 2, pp. 242–250, Feb. 2010, doi: 10.1061/(ASCE)CO.1943-7862.0000114.

[36] C. Kim, B. Kim, and H. Kim, "4D CAD model updating using image processing-based construction progress monitoring," *Automat. Construct.*, vol. 35, pp. 44–52, Nov. 2013, doi: 10.1016/j.autcon.2013.03.005.

[37] Z. Zhu and I. Brilakis, "Concrete column recognition in images and videos," *J. Comput. Civil Eng.*, vol. 24, no. 6, pp. 478–487, 2010, doi: 10.1061/(ASCE)CP.1943-5487.0000053.

[38] I. Brilakis, H. Fathi, and A. Rashidi, "Progressive 3D reconstruction of infrastructure with videogrammetry," *Automat. Construct.*, vol. 20, no. 7, pp. 884–895, 2011, doi: 10.1016/j.autcon.2011.03.005.

[39] M. Golparvar-Fard, F. Penña-Mora, and S. Savarese, "D^{4}AR—A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication," *J. Inf. Technol. Construct.*, vol. 14, no. 13, pp. 129–153, 2009. [Online]. Available: http://www.itcon.org/2009/13.

[40] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, and F. Peña-Mora, "Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques," *Automat. Construct.*, vol. 20, no. 8, pp. 1143–1155, Dec. 2011, doi: 10.1016/j.autcon.2011.04.016.

[41] M. Golparvar-Fard, F. Peña-Mora, and S. Savarese, "Automated progress monitoring using unordered daily construction photographs and IFC-based building information models," *J. Comput. Civil Eng.*, vol. 29, no. 1, 2015, Art. no. 04014025, doi: 10.1061/(ASCE)CP.1943-5487.0000205.

[42] C. Kropp, C. Koch, and M. König, "Interior construction state recognition with 4D BIM registered image sequences," *Automat. Construct.*, vol. 86, pp. 11–32, Feb. 2018, doi: 10.1016/j.autcon.2017.10.027.

[43] K. Asadi and K. Han, "Real-time image-to-BIM registration using perspective alignment for automated construction monitoring," in *Proc. Construct. Res. Congr.*, New Orleans, LA, USA, Apr. 2018, pp. 388–397.

[44] K. Asadi, H. Ramshankar, M. Noghabaei, and K. Han, "Real-time image localization and registration with BIM using perspective alignment for indoor monitoring of construction," *J. Comput. Civil Eng.*, vol. 33, no. 5, 2019, Art. no. 04019031, doi: 10.1061/(ASCE)CP.1943-5487.0000847.

[45] F. Xue, W. Lu, and K. Chen, "Automatic generation of semantically rich as-built building information models using 2D images: A derivative-free optimization approach," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 11, pp. 926–942, Nov. 2018, doi: 10.1111/mice.12378.

[46] J. Chen, Z. Kira, and Y. K. Cho, "Deep learning approach to point cloud scene understanding for automated scan to 3D reconstruction," *J. Comput. Civil Eng.*, vol. 33, no. 4, Jul. 2019, Art. no. 04019027, doi: 10.1061/(ASCE)CP.1943-5487.0000842.

[47] S. Yan, Y. Peng, G. Wang, S. Lai, and M. Zhang, "Weakly supported plane surface reconstruction via plane segmentation guided point cloud enhancement," *IEEE Access*, vol. 8, pp. 60491–60504, 2019, doi: 10.1109/ACCESS.2019.2946456.

[48] L. Bai, Y. Zhao, M. Elhousni, and X. Huang, "DepthNet: Real-time LIDAR point cloud depth completion for autonomous vehicles," *IEEE Access*, vol. 8, pp. 227825–227833, 2020, doi: 10.1109/ACCESS.2020.3045681.

[49] Z. Zhu and I. Brilakis, "Parameter optimization for automated concrete detection in image data," *Automat. Construct.*, vol. 19, no. 7, pp. 944–953, Nov. 2010, doi: 10.1016/j.autcon.2010.06.008.

[50] K. Zhang and X. Li, "Infrared small dim target detection based on region proposal," *Optik*, vol. 182, pp. 961–973, Apr. 2019, doi: 10.1016/j.ijleo.2019.02.008.

[51] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[52] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[53] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*.

[54] H. Son, C. Kim, and C. Kim, "Automated color model–based concrete detection in construction-site images by using machine learning algorithms," *J. Comput. Civil Eng.*, vol. 26, no. 3, pp. 421–433, 2012, doi: 10.1061/(ASCE)CP.1943-5487.0000141.

[55] Z. Zhu and I. Brilakis, "Comparison of optical sensor-based spatial data collection techniques for civil infrastructure modeling," *J. Comput. Civil Eng.*, vol. 23, no. 3, pp. 170–177, May/Jun. 2009, doi: 10.1061/(ASCE)0887-3801(2009)23:3(170).

[56] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "DSOD: Learning deeply supervised object detectors from scratch," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 1919–1927.

[57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[58] GitHub. *LabelImg: A Graphical Image Annotation Tool*. Accessed: Jan. 15, 2020. [Online]. Available: https://github.com/tzutalin/labelImg

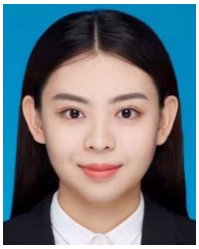
[59] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. 7th IEEE Int. Conf. Comput. Vis. (ICCV)*, Kerkyra, Greece, Sep. 1999, pp. 666–673, doi: 10.1109/ICCV.1999.791289.

[60] A. Fetić, D. Jurić, and D. Osmanković, "The procedure of a camera calibration using camera calibration toolbox for MATLAB," in *Proc. 35th Int. Conv. MIPRO*, Opatija, Croatia, May 2012, pp. 1752–1757. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6240932.
[61] G. Percoco, M. G. Guerra, A. J. S. Salmeron, and L. M. Galantucci, "Experimental investigation on camera calibration for 3D photogrammetric scanning of micro-features for micrometric resolution," *Int. J. Adv. Manuf. Technol.*, vol. 91, nos. 9–12, pp. 2935–2947, Aug. 2017, doi: 10.1007/s00170-016-9949-6.
[62] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EP*n*P: An accurate $O(n)$ solution to the P*n*P problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, Feb. 2009, doi: 10.1007/s11263-008-0152-6.
[63] GitHub. *Sfm-Python*. Accessed: Jan. 15, 2020. [Online]. Available: https://github.com/adnappp/Sfm-python

JINGGUO XUE was born in Jinzhong, Shanxi, China, in 1994. He received the B.S. degree in engineering management from North China Electric Power University, Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree in management science and engineering. His current research interests include image-based 3D reconstruction and computer vision.

XUELIANG HOU was born in Taiyuan, Shanxi, China, in 1966. He received the Ph.D. degree from the Xi'an University of Architecture and Technology, in 2006.

He is currently a Professor at the School of Economics and Management, North China Electric Power University, the Director of the Institute of Engineering Technology and Management, and a Ph.D. Supervisor. His current research interests include construction technology and project management. He is a member of the Engineering Project Management Steering Committee of the Chinese Architectural Society. He is the winner of the New Century Excellent Talents of the Ministry of Education. He is the Director of the Construction Economy Branch of the Chinese Architectural Society. He is a Peer Review Expert of the National Natural Science Foundation, a Review Expert of the Beijing Natural Science Foundation, an Expert Database Member of the Ministry of Construction, and an Expert Group Member of the Shanxi Provincial Government.



YING ZENG received the B.S. degree in information and computing science and the M.S. degree in management science and engineering from North China Electric Power University, Beijing, China, in 2016 and 2019, respectively. She is currently working with State Grid Mianyang Power Supply Company, Mianyang, Sichuan, China.