Kalpa Publications in Computing
Volume 22, 2025, Pages 344–355
Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

Enhanced Curtain Wall Construction Progress Monitoring via Tower Crane Perspective and Integrated 3D Reconstruction

Tao Zhong [1*], Yujie Lu[1†], Shuo Wang [1‡], and Xuanjun Li [1§]
[1] College of Civil Engineering, Tongji University, Shanghai, China
2410045@tongji.edu.cn, lu6@tongji.edu.cn,
wangshuo2207@tongji.edu.cn, 2430928@tongji.edu.cn

Abstract

Urban development has driven the widespread proliferation of high-rise buildings, making curtain wall construction a critical aspect of progress management. However, the extensive surface areas of curtain walls and the severe perspective distortions present significant challenges to efficient construction progress monitoring. To address these challenges, a novel method is proposed for tracking curtain wall installation progress in high-rise buildings through integrated 3D reconstruction. Cameras are strategically mounted at the ends of adjacent tower crane booms, with multiple ground control points (GCPs) deployed on-site as hardware anchors. The process begins with capturing multi-view images using the crane-mounted cameras. The COLMAP 3D reconstruction pipeline is enhanced by incorporating GCPs to ensure accurate 3D reconstruction within a real-world coordinate system. The building structure is subsequently extracted from the site model, and rectified facade images are generated using projection techniques. The curtain wall installation progress is assessed at both the floor and overall building levels using the YOLOv8 image segmentation model. The proposed method was validated through a case study on a super high-rise construction site. This approach achieved decimeter-level accuracy in 3D reconstruction and a precision rate of 95.9% in curtain wall progress identification, meeting project requirements. These findings establish a robust framework for managing large-scale outdoor construction progress, particularly for high-rise curtain walls. Additionally, the site modeling methodology enables more refined and timely monitoring practices, offering significant potential for the development of digital twin models.

[*] Conceptualization, Investigation, Writing - Original Draft
[†] Supervision, Funding acquisition
[‡] Writing - Review & Editing
[§] Investigation, Formal Analysis

1 Introduction

As urbanization accelerates and urban land becomes increasingly scarce, high-rise buildings have become a crucial architectural solution, widely adopted for offices, commercial spaces, and other purposes (Quesada-Olmo, Jimenez-Martinez et al. 2018). Effective schedule management in high-rise construction plays a pivotal role in controlling project costs (Baduge, Thilakarathna et al. 2022). Meanwhile, curtain wall projects, known for their intensive workloads, high installation risks, and low levels of automation, are critical elements that directly impact key project milestones (Kim, Kim et al. 2021). Therefore, monitoring and managing the progress of curtain wall installation is essential throughout the construction process of high-rise buildings.

The objectives of monitoring facade installation progress include assessing the progress of each facade and identifying uninstalled areas (Eom and Kang 2022). Manual inspections face difficulties in accurately quantifying progress and are labor-intensive. Recent studies have explored the use of algorithms such as computer vision (Lu, Wei et al. 2023) and 3D reconstruction (Lei, Zhou et al. 2019) to automate the identification of construction elements for progress calculation. Computer vision techniques analyze two-dimensional visual features of construction elements to determine progress based on the proportion of object pixels within the image. In contrast, 3D reconstruction methods create 3D point clouds from images or laser scans, followed by semantic annotation through BIM-based comparison to calculate construction volume. However, these approaches pose significant challenges. Visual recognition methods that rely solely on images, captured manually or via drones, are often limited by shape distortions and the lack of accurate size information, making it difficult to obtain precise progress data. Furthermore, due to the reflective and transparent properties of glass, the resulting point clouds of curtain walls are typically too sparse to form a complete 3D representation, leading to the loss of key visual features (Lin and Golparvar-Fard 2020).

Additionally, existing methods typically rely on either manual inspections or drone flights to monitor construction progress and quality. However, manual inspections are inefficient for large construction sites, while the use of drones is constrained by the need for professional operators, favorable weather conditions, and compliance with licensing requirements. As a result, these approaches fail to support efficient daily site monitoring. In contrast, crane-mounted cameras provide an effective alternative by autonomously capturing construction data during routine operations without requiring manual intervention or permits (Masood, Aikala et al. 2020). Positioned on the crane's boom, these cameras capture multi-view imagery of the construction site, which is subsequently utilized for 3D reconstruction of site point clouds. As autonomous hardware systems, they operate continuously without human oversight, enabling real-time monitoring of curtain wall progress and overall site activity.

This paper proposes a method for monitoring curtain wall installation progress in high-rise buildings based on tower crane perspectives combined with 3D reconstruction (see **Figure 1**). Our approach uses tower crane cameras as the primary hardware, leveraging 3D reconstruction for accurate dimensional analysis and computer vision techniques for visual progress tracking.

The structure of this paper is organized as follows: The first section presents the importance of monitoring curtain wall progress in high-rise buildings and the limitations of current research. The second section outlines the research methods, including the integration of ground control points (GCPs) for accurate 3D reconstruction, extraction of rectified facade images from the site model, and curtain wall identification for progress analysis. In the third section, we present the experimental setup along with the results of site modeling and progress quantification. The fourth section discusses the benefits and practicality of the proposed method. Finally, the contributions and limitations of this work are summarized in the fifth section.

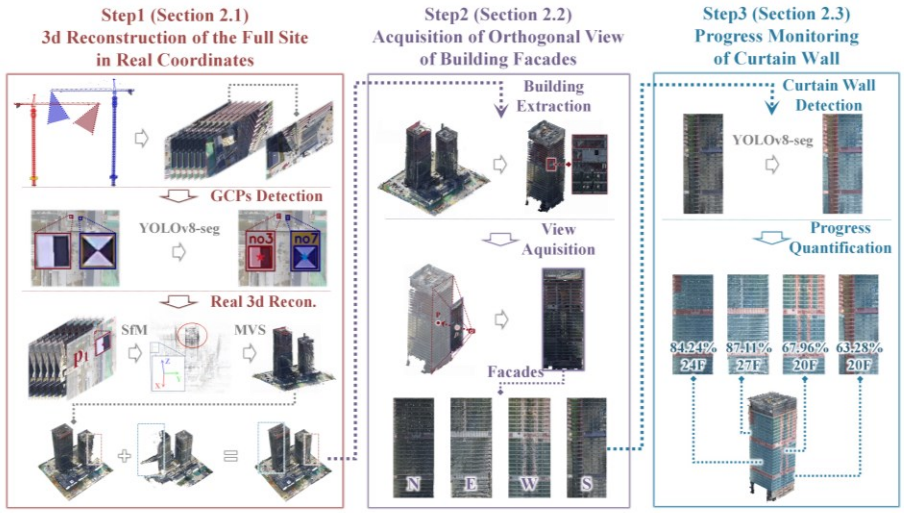


Figure 1: Curtain wall progress identification framework using multi-view 3D reconstruction

2 METHOD

2.1 3D Reconstruction of Construction Site with GCPs Integration

Cameras are mounted at the front ends of tower crane booms, enabling the capture of multi-view images from multiple tower cranes to generate a comprehensive site point cloud model. This model serves as a foundational dataset for extracting visual information related to the curtain wall. Ground Control Points (GCPs) are integrated into the process by identifying their 2D coordinates and using the COLMAP 3D reconstruction algorithm to accurately determine their corresponding 3D coordinates.

(1) Detection of GCP Pixel Centers

GCPs with known real-world coordinates are utilized to establish a coordinate transformation between the image coordinate system and the real-world coordinate system. The process begins by obtaining the pixel coordinates of the GCPs in each multi-view image. These GCPs are fixed markers strategically positioned at various heights across the construction site, each featuring a clearly defined geometric center. To determine the distribution of GCPs within the images, the YOLOv8-seg model is employed. The centroid of the corresponding image mask is then calculated to identify the precise GCP center (see Figure 2).

The YOLOv8-seg model, a segmentation-specific variant of the YOLOv8 model, is optimized for image segmentation tasks. Its architecture integrates a multi-detection head and an anchor-free design, enabling it to effectively detect target objects of varying sizes. The model consists of three main components: the backbone, neck, and head. The backbone extracts multi-scale features, the neck aggregates and bridges features across scales, and the head predicts object categories and masks using its multiple detection heads. Finally, by leveraging the symmetrical geometric shape of the GCPs, the geometric centroid of each object mask is computed to determine the precise pixel coordinates of the GCPs.

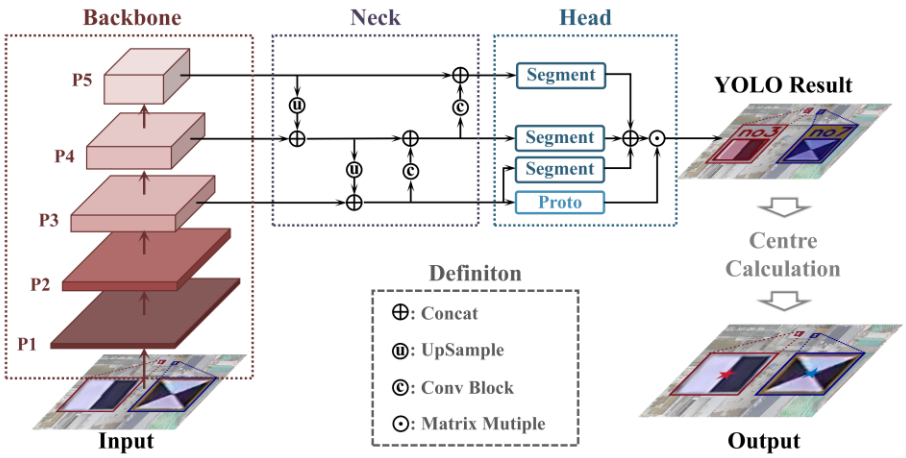


Figure 2: Process for identifying GCP pixel centers using the YOLOv8-seg model

(2) Framework for COLMAP-based 3D Reconstruction with GCPs Integration

COLMAP (Schonberger and Frahm 2016) is a well-established open-source framework for 3D reconstruction from multi-view images and has demonstrated strong performance across various outdoor scenarios. However, since the modeling process relies on relative pose estimation between images, it lacks absolute coordinate information, including real-world scale and orientation. To address this limitation, we propose a COLMAP-based 3D reconstruction framework with GCP integration (see **Figure 3**) to incorporate real-world coordinates and recover precise 3D information. The original COLMAP framework consists of two key components: sparse reconstruction (SfM) and dense reconstruction (MVS). Sparse reconstruction estimates camera poses through feature matching, triangulation, and bundle adjustment, while dense reconstruction generates the point cloud model via depth estimation and point cloud fusion. We integrate GCP information during the sparse reconstruction phase to recover accurate poses by triangulating corresponding GCPs to obtain normalized 3D coordinates P_n (as shown in Equation (1)). The transformation T between the normalized coordinates P_n and real-world coordinates P_r is estimated (see Equation (2)) and mapped onto the camera pose E_i (see Equation (3)) to recover the real-world coordinates of both the cameras and the point cloud model. With each local point cloud model reconstructed from individual tower crane cameras aligned within the real-world coordinate system, a complete site model is formed through the direct aggregation of these local models, as shown in **Table 1**.

$$P_n = \tau\left(p_a, p_b, E_a, E_b, I_a, I_b\right) \tag{1}$$

where, a and b are two images capturing the same GCP point P. The pixel coordinate of this point in image i is denoted as p_i, and E_i and I_i represent the extrinsic and intrinsic parameters of image i respectively. We use the DLT method (Robust 2020) for triangulation τ.

$$T = P_r^* \cdot \left(P_n^*\right)^{-1} \tag{2}$$

where, P^* denotes the homogeneous coordinates of the 3D point P.

$$E_r = T \cdot E_i \tag{3}$$

where, E_r represents the camera pose in the real-world coordinate system.

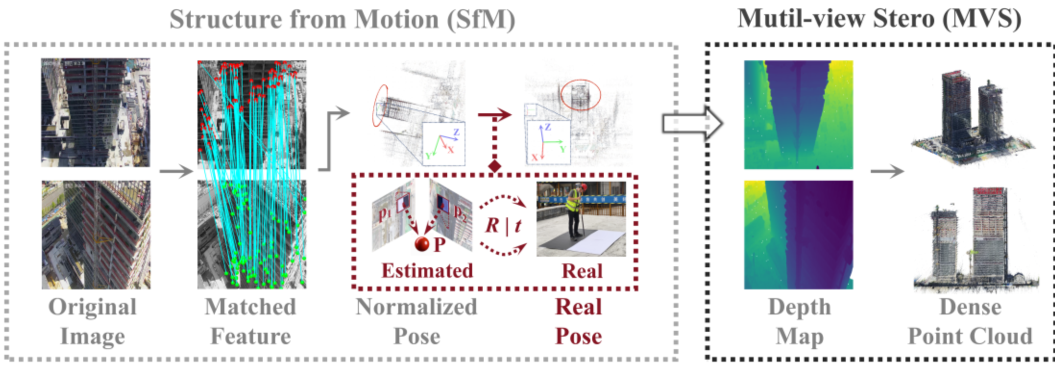


Figure 3: COLMAP 3D reconstruction framework with GCP integration, where red sections indicate the introduced module for recovering real-world poses

2.2 Acquisition of Rectified Facade Images

Building appearance images are captured using the site model to obtain geometric information about the curtain walls. These images must meet three essential criteria: unobstructed views, minimal distortion, and an accurate scale aligned with real-world dimensions. Adhering to these criteria ensures reliable identification of completed and incomplete sections of the curtain wall. To achieve this, the building structure is first extracted from the site model, and a corresponding appearance model is generated. Rectified facade images of the building are then created using projection principles, ensuring alignment with real-world geometry.

(1) Acquisition of the Building Visual Model

The building structure is extracted from the site model using point cloud projection, and a mesh model is generated using the Poisson surface reconstruction algorithm (Kazhdan, Bolitho et al. 2006) (see Figure 4). The mesh model, composed of numerous triangular patches, creates a continuous surface with realistic visual attributes, serving as the foundation for feature analysis. The process begins with horizontally projecting the site point cloud to calculate the horizontal point density distribution (see Figure 4(b)). Next, density gradients along the X and Y axes are computed, where significant changes highlight object edges (see Figure 4(c)). Point clustering is then applied to distinguish individual objects, identifying the structure with the largest horizontal area as the building (see Figure 4(d)). Subsequently, the convex hull of the edge points is calculated, followed by vertical cropping to isolate the building structure's point cloud. Finally, the Poisson algorithm is applied to reconstruct the mesh model from the processed point cloud.

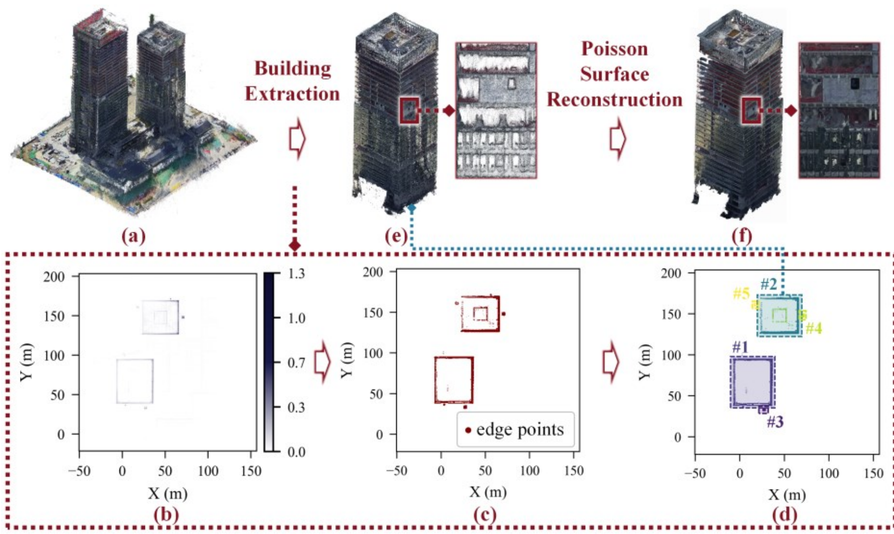


Figure 4: Workflow for generating the visual model of the building structure, where (a) represents the complete site point cloud, (e) represents the building structure point cloud, and (f) represents the mesh model of the building structure

(2) Acquisition of Building Facade Images

Virtual cameras are configured to capture rectified facade images of the visual model (see Figure 5). As described in Section 2.2 (1), the height H and width B of each building facade are determined from the geometric dimensions of the point cloud. Each pixel value in the image is calculated as follows: the optical center of the camera and each pixel p form a spatial ray $\vec{r}_w$ (see Equation (4)), which intersects the mesh model of the building at point P. The RGB value at the intersection point P is assigned as the pixel value of p. For the external camera parameters R and t, the viewing direction is aligned with the normal of the façade. The optical center is positioned at half the height $H/2$, and the distance D is set to the width of the facade. For the internal camera parameters I, the aspect ratio of the image is matched to the height-to-width ratio of the building facade, and the focal length f is set to the pixel width I_u of the image. According to the pinhole camera model, the camera is positioned to capture the entire facade (see Equation (5)). This process produces rectified facade images for each building structure, which serve as the data foundation for subsequent curtain wall progress quantification.

$$\vec{r}_w = R \cdot \vec{r} + t,\ \vec{r} = \frac{\left[(i-c)/f \quad (j-c)/f \quad 1\right]^T}{\sqrt{(i-c)^2/f^2 + (j-c)^2/f^2 + 1}} \tag{4}$$

where, i and j denote the pixel's 2D coordinates; R and t represent the camera's external parameters; f and c denote the focal length and optical center, respectively.

$$U = 2 \cdot \frac{I_u \cdot D}{2f} = 2 \cdot \frac{I_u \cdot B}{2I_u} = B \tag{5}$$

where, U is the actual width in the camera's field of view, I_u denotes the pixel width of the image, D represents the distance between the optical center and the object, and f denotes the focal length of the camera.

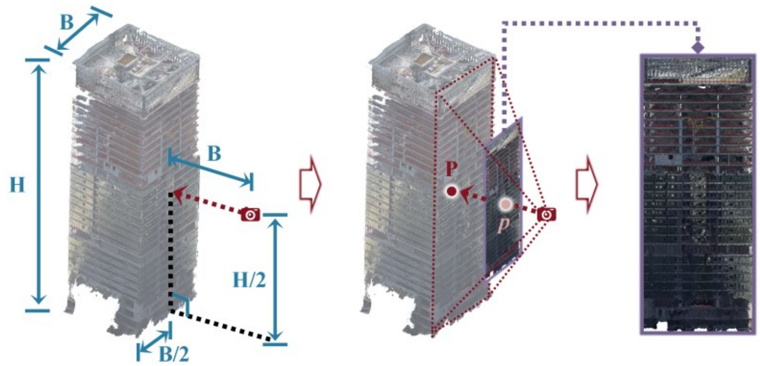


Figure 5: Workflow of rectified facade image generation

2.3 Construction Progress Analysis of Curtain Wall

(1) Curtain Wall Distribution Identification

The YOLOv8-seg model is employed to identify the distribution of installed curtain walls on the building facade, enabling precise analysis of construction progress. By leveraging the distinct visual differences between curtain walls and structural elements, the image segmentation algorithm effectively delineates their boundaries. The recognition results are further refined using geometric priors of the facade. First, since the actual boundaries between curtain walls and structural elements are either horizontal or vertical, we straighten the detected edges to align with these orientations (see Figure 6(c)). Second, since only curtain walls and structural elements are present in the facade image and they do not intersect, the mean value of their adjacent boundaries is calculated and used as the actual dividing line (see Figure 6(d)).

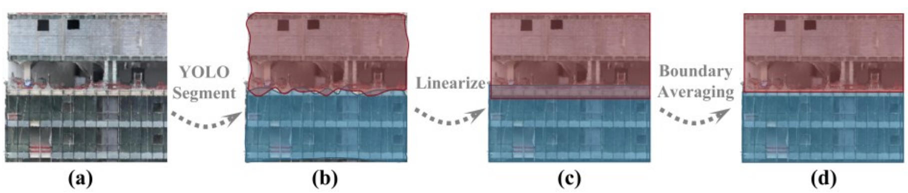


Figure 6: Post-processing diagram of curtain wall recognition results

(2) Curtain Wall Construction Progress Calculation

The overall progress, floor-specific progress, and the identification of the standard floor under construction are calculated as key indicators of curtain wall installation progress. Since the image-to-real-world scale is established during the generation of rectified facade images, pixel positions in the image can be accurately mapped to their real-world coordinates (see Equation (6)). Consequently, the geometric data in the image reflects real-world dimensions with precision. The overall curtain wall progress is defined as the ratio of the total curtain wall area to the facade area. Using the elevations of standard floors, boundary lines for each floor are determined within the image. The progress for each floor is calculated as the ratio of the curtain wall area on that floor to the total area of the floor. Finally, the boundary where significant progress differences are observed from top to bottom is identified as the standard floor currently under construction.

$$[X,Y]^T = [i,j]^T \cdot I_u / U \tag{6}$$

where, X and Y are the real-world facade coordinates, and i and j are the pixel coordinates.

3 RESULTS

3.1 Experiment Site Overview

Experiments were conducted at a super high-rise construction site in China, with data collected weekly over four months. During this time, primary construction activities included curtain wall installation, secondary structure installation, and interior decoration. The site consisted of two super high-rise commercial buildings, standing 180 meters and 135 meters tall, respectively. Two attached tower cranes facilitated material transport and installation: Tower #1 was positioned on the north side of Building #A, while Tower #2 was located on the west side of Building #B (see Figure 7). Pan-tilt-zoom cameras were mounted at the front ends of both crane booms to capture a panoramic view of the entire site. Additionally, seven Ground Control Points (GCPs) were placed around the site to integrate real-world coordinates into the 3D data. A total of 600 images were collected for GCP identification, while 800 images were used for curtain wall recognition.

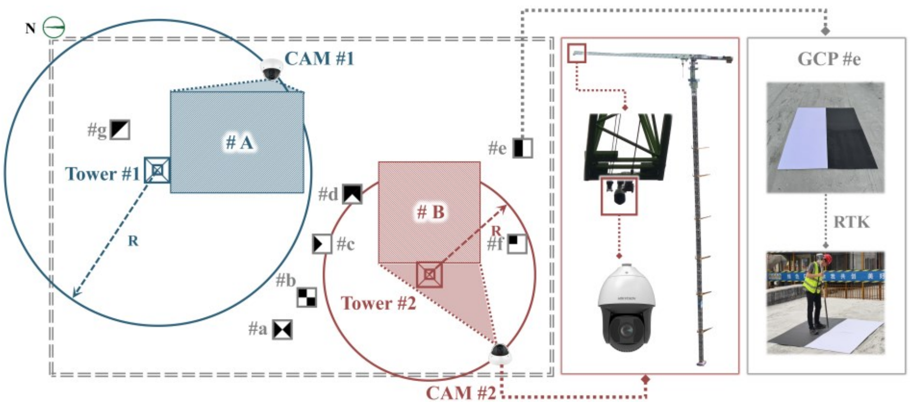
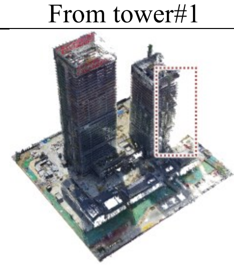
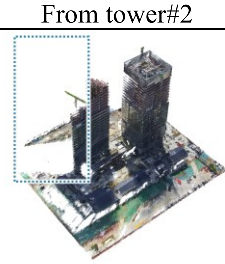
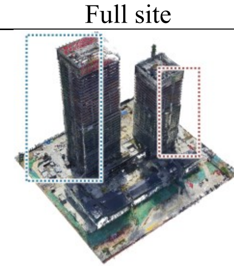


Figure 7: Experimental setup and hardware configuration

3.2 The Results of Integrated 3D Reconstruction

A comprehensive 3D reconstruction of the construction site was achieved through integrated modeling using dual tower crane cameras. When compared to the point cloud model reconstructed from drone flights, which served as the ground truth, the proposed model demonstrated an average error of 0.18 meters. This level of accuracy effectively reflects the actual conditions of the construction site, meeting the precision requirements for progress monitoring.

Date	From tower#1	From tower#2	Full site
April 18th			

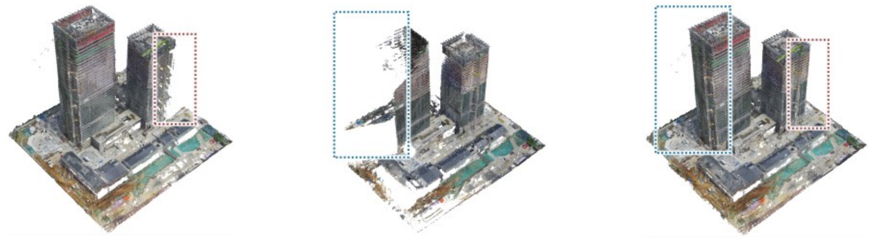
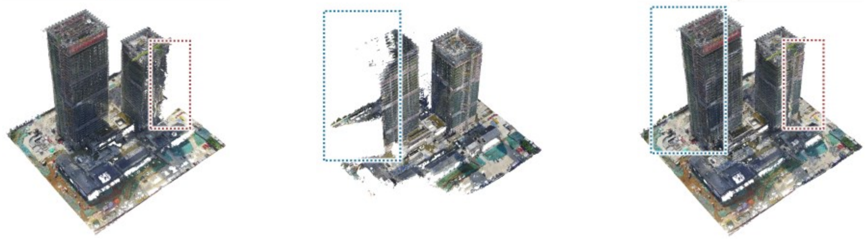
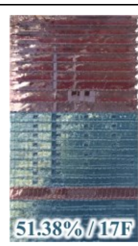
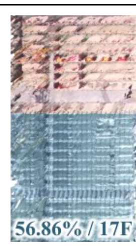
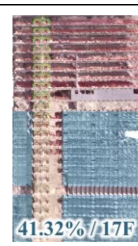
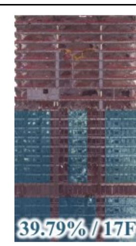
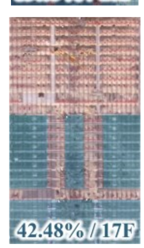
Date	Models
June 27th	
August 7th	

Table 1: Visual results of 3d joint modeling of the construction site

3.3 The Results of Curtain Wall Progress Quantification

Three time points were selected to illustrate the quantified results of curtain wall progress recognition (see **Table 2**). As of August 7th, the overall curtain wall installation progress had reached 75.65%. The remaining uninstalled curtain wall areas were primarily located at equipment floors, construction hoists, and tower crane attachment points, which will be addressed after the removal of these temporary structures. It is important to note that minor deviations in progress recognition, typically ranging from 1% to 2%, may occur due to partial surface feature blurring in the mesh model or slight inaccuracies in image recognition. For instance, the east facade showed a 2% recognition error between June 27th and August 7th, despite no actual progress occurring during this period.

Date	North	East	South	West	All
April 18th					
June 27th					

August 7th

Table 2: Curtain wall construction progress visualization, where blue sections represent curtain walls, and red sections represent structural components

4 DISCUSSION

4.1 Applicability of YOLOv8 in This Scenario

The YOLOv8-seg algorithm was employed for image recognition tasks, including the detection of GCPs and curtain walls. To assess its performance in detecting objects of varying sizes, YOLOv8 was benchmarked against two state-of-the-art algorithms, SOLOv2 and Mask2Former, using the mAP50 metric (see Table 3). The mAP50 metric evaluates detection accuracy based on a minimum 50% overlap between predicted and ground truth bounding regions. The results indicate that YOLOv8 achieves exceptional accuracy in detecting GCPs due to their distinct visual features, which enhances the precision of 3D modeling. For curtain walls, YOLOv8 attained a recognition accuracy of 95.9%, satisfying the precision requirements for construction progress monitoring. Furthermore, YOLOv8 outperformed SOLOv2 and Mask2Former, demonstrating superior accuracy and requiring fewer parameters, thus offering a more practical solution for large-scale construction monitoring.

Algorithm	Params	GCPs	Curtain wall	Structural part
SOLOv2	46.26M	0.958	0.910	0.836
Mask2Former	44.06M	0.930	0.890	0.815
YOLOv8	**27.24M**	**0.981**	**0.959**	**0.875**

Table 3: Comparison of segmentation accuracy of different algorithms

4.2 Outperformance of the Generated Rectified Images

Our contribution is the development of an efficient method for generating complete rectified facade images using tower crane cameras. Unlike traditional methods such as manual inspection, UAV photography, or standard crane camera captures, our approach produces rectified images that provide comprehensive facade coverage, remove distortions and obstructions, and accurately map pixels to real-world coordinates (as shown in **Figure 8**). This enables accurate progress recognition and precise localization of uninstalled curtain wall sections. Moreover, compared to methods that directly correct 2D images (Wei, Lu et al. 2023), our approach preserves a clear correspondence with real-world dimensions (see Equation (6)), offering more accurate scaling and a distinct advantage in floor-by-floor progress identification.

(a) (b) (c) (d)

Figure 8: Comparison of the generated rectified image with common captures: (a) captured with a mobile phone, (b) captured by a UAV from a distance, (c) captured by a tower crane camera, and (d) our generated rectified image, the red dotted box marks the obstructed areas of the facade

5 CONCLUSION

Curtain wall progress recognition is a crucial aspect of curtain wall construction management. To address the challenges of inaccurate progress recognition and difficulty in localization for high-rise buildings, we propose a curtain wall progress recognition method based on joint 3D reconstruction from the aerial perspective of tower cranes. First, the real 3D scene is reconstructed using multi-view images captured by tower crane cameras. Next, rectified facade images are extracted from the 3D model. Finally, progress quantification and analysis are performed on the rectified images, achieving a modeling accuracy of 0.18 meters and a curtain wall recognition accuracy of 95.9%. With the introduction of real 3D information, our method enables precise calculation of curtain wall progress at any location and accurate localization of uninstalled curtain wall sections. This accurate progress information allows managers to track project status, assess individual task efficiency, and optimize material deliveries and scheduling, facilitating precise on-site management.

The contribution of this paper lies in proposing a spatial appearance analysis method for large-scale objects on construction sites, addressing distortion, lack of scale, and incompleteness inherent in traditional 2D images. This method provides a reference for analyzing facades, foundation pits, and other large-scale outdoor construction information. It also demonstrates the practical application of digital twins on construction sites. However, our method requires regular building facades to obtain distortion-free surface images. In the future, surface unfolding techniques can be explored to fully analyze the visual features of irregular high-rise buildings.

6 ACKNOWLEDGMENTS

The study was supported by the National Key Research & Developement Program of China (2022YFC3801700), Fundamental Research Funds for the Central Universities (2024-1-ZD-02).

References

Baduge, S. K., et al. (2022). "Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications." Automation in Construction **141**: 104440.

Eom, J. and Y. Kang (2022). "Curtain Wall Construction: Issues and Different Perspectives among Project Stakeholders." Journal of Management in Engineering **38**(5): 04022054.

Kazhdan, M., et al. (2006). Poisson surface reconstruction. Proceedings of the fourth Eurographics symposium on Geometry processing.

Kim, T., et al. (2021). "A simulation-based dynamic scheduling model for curtain wall production considering construction planning reliability." Journal of cleaner production **286**: 124922.

Lei, L., et al. (2019). "A CNN-based 3D patch registration approach for integrating sequential models in support of progress monitoring." Advanced Engineering Informatics **41**: 100923.

Lin, J. J. and M. Golparvar-Fard (2020). Visual and virtual progress monitoring in Construction 4.0. Construction 4.0, Routledge**:** 240-263.

Lu, Y., et al. (2023). "A deep learning method for building façade parsing utilizing improved SOLOv2 instance segmentation." Energy and Buildings **295**: 113275.

Masood, M. K., et al. (2020). "Multi-building extraction and alignment for as-built point clouds: a case study with crane cameras." Frontiers in Built Environment **6**: 581295.

Quesada-Olmo, N., et al. (2018). "Real-time high-rise building monitoring system using global navigation satellite system technology." Measurement **123**: 115-124.

Robust, A. (2020). "Revisiting Projective Structure from Motion: A Robust and Efficient Incremental Solution." IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE **42**(2).

Schonberger, J. L. and J.-M. Frahm (2016). Structure-from-motion revisited. Proceedings of the IEEE conference on computer vision and pattern recognition.

Wei, W., et al. (2023). "Augmenting progress monitoring in soil-foundation construction utilizing SOLOv2-based instance segmentation and visual BIM representation." Automation in Construction **155**: 105048.