Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Elevating Construction Monitoring: A Comprehensive UAV-Based Data Collection System for Outdoor Construction Dynamics

Dayou Mao[1], Yuchen Lin[1], Yan Song Hu[1], Mihir Gupta[2], Maxwell Lou[2],
Farzad Jalaei[1], SeyedReza RazaviAlavi[1], Ashkan Ebadi[1],
Alexander Wong[2], Yuhao Chen[2]

[1] National Research Council Canada
[2] Vision and Image Processing Group, Systems Design Engineering, University of Waterloo

ABSTRACT: The construction industry is crucial to the global economy, yet project management and progress monitoring often suffer from delays and cost overruns. While many studies have examined RGB cameras, LiDAR scanners, and Computer Vision (CV) algorithms for tracking physical progress, most are project-specific, and a widely adopted data collection and processing system, along with a standard benchmark dataset, remains unavailable. Unmanned Aerial Vehicles (UAVs) have significantly advanced the methodology of construction site maps creation, activity recognition, and progress tracking, especially for indoor and static outdoor scenes. However, capturing dynamic outdoor environments over time with rich visual detail and processing large-scale data presents ongoing challenges, raising concerns about safety, quality, and consistency. To address these issues, this paper proposes a systematic and consistent data collection approach for outdoor construction sites, capturing RGB images, videos, and point clouds twice weekly starting September 27, 2024. To the best of our knowledge, we are the first to present a high-quality, safe, and repeatable workflow; our site mapping pipeline using DJI and CloudCompare; and alignment techniques using Umeyama and COLMAP. Key challenges include moving objects on site, occlusions causing incomplete point clouds, and acute weather conditions. Our dataset totals 20 successfully recorded and processed site maps by the time this paper was written. We also study modern 3D reconstruction baseline algorithms for building complementary digital site representations in addition to our processed point clouds and provide experimental setups details and benchmark results, demonstrating the effectiveness of 3D Gaussian Splatting (3DGS) models.

1. INTRODUCTION

The construction industry plays a pivotal role in the global economy, contributing significantly to infrastructure development and urbanization. Nevertheless, construction projects frequently encounter delays and cost overruns (Hamledari, et al., 2018), attributed to various factors like adverse weather, material shortage, and labor constraints. Effective project management and progress monitoring are paramount to address these challenges, enabling meticulous and timely adjustments to keep projects on schedule and within budget (Deng, et al., 2020). However, this endeavour is complex, time-consuming, and often plagued by inconsistency among management teams. A persistent challenge in construction project management is the recognition of deviations between the planned objectives and actual construction progress in a timely manner, which is crucial for informed decision-making and devising recovery plans. Despite this, most practices continue to rely on manual repeated inspection of construction progress by

construction managers, which imposes high burden on labor costs and is prone to human errors (Deng, et al., 2020).

To tackle the repetition, human errors, subjectivity, and extensive time consumption issues, numerous studies have ventured into the realm of reality capture technologies including high-resolution RGB cameras and high-precision Light Detection and Ranging (LiDAR) scanners to create 3D digital representations of the construction site (Deng, et al., 2020; Hamledari, et al., 2017; Roh, et al., 2011; Lin & Golparvar-Fard, 2017; Han, et al., 2018). With accurate and rich 3D digital representation, Artificial Intelligence (AI) and Computer Vision (CV) algorithms can be employed to automate the analysis of the captured data (e.g., images, videos, and point clouds), integrate it within 4D (3D with time) Building Information Modeling (BIM), and perform real-time progress monitoring, quality checks, activity recognition, and so on (Hamledari, et al., 2018; Han, et al., 2018; Deng, et al., 2020; Amer & Golparvar-Fard, 2018; Han & Golparvar-Fard, 2017; Shang & Shen, 2018; Tang, et al., 2020; Lin & Golparvar-Fard, 2017).

However, these works typically focus on comparing as-built and as-planned models at a single point in time and rely on case-specific data collection and preparation methods. Recent advancements in Unmanned Aerial Vehicles (UAVs) have largely boosted the efficiency in data collection and gained wide applications for construction site map creation (Guan, et al., 2022). However, the use of UAVs also presents unique challenges and remains underexplored due to safety constraints, seasonal changes, unpredictable weather conditions, and operational complexities when it comes to periodic data collection extended over half a year or longer in an outdoor environment (Guan, et al., 2022; Siebert & Teizer, 2014; Ibrahim, et al., 2017; Lin, et al., 2021). A proven safe, cost effective, systematic, and replicable data collection workflow that consistently produces high quality site maps is still lacking from the outdoor construction literature. Further, research addressing change detection and progress monitoring on dynamic, long-term evolution of outdoor construction sites conducted on open-source visually rich benchmark dataset remains limited and underexplored.

In this paper, we design our data collection workflow based on hands-on experience on test fields and real-world construction sites, explore efficient methods for processing the raw sensor data into high-quality construction site maps, and conduct baseline experiments to benchmark existing 3D reconstruction algorithms from the computer vision community to complement the LiDAR point cloud representation. By the time of this paper was written we have 20 successfully recorded and processed site maps, spanning across 5 months.

Our contributions are at least four-fold:

1. We present a comprehensive system aimed at outdoor construction site mapping encompassing hardware, data collection, and map reconstruction. The presented system is useful for both research benchmark curation and deployment of computer vision algorithms in construction applications.
2. We present our continuously expanding dataset of high-resolution digital 3D construction site maps with RGB images, point clouds, and video streams, recording the construction of a university student residence building from inception to completion. Data quality assessment and dataset characteristics are summarized.
3. We conduct empirical study on existing 3D reconstruction algorithms and benchmark their quality as complementary 3D digital representations of the construction site map.
4. We outline the remaining challenges in data collection for dynamic outdoor environments, highlighting areas for future research.

The remainder of this paper is organized as follows. We first outline our data collection setup in Section 2, including the hardware equipment and construction site information. In Section 3 we describe the three main design considerations: safety, quality, and consistency. Section 4 describes the series of steps we used to process our data and curate a dataset. We discuss the challenges we encountered and the solutions we took in Section 5. Section 6 reports the 3D reconstruction benchmark results on our dataset, and finally we conclude our paper with future research directions in Section 7.

2. DATA COLLECTION SETUP

2.1 Hardware Equipment

Based on the best available options prior to the start of our data collection work that provides the highest possible image and video resolution and the highest precision LiDAR recording devices (DJI Enterprise, 2025), we choose the following high-end equipment offered by DJI Enterprise. The hardware equipment that we use includes a drone (Matrice 350 RTK), a payload (Zenmuse L2), and a Real-Time Kinematics (RTK) positioning system (D-RTK 2 Mobile Station). The details and robustness are summarized in the remainder of this sub-section. Accessories for the drone include a controller (DJI RC Plus), a battery station (BS65 Intelligent Battery Station), and flight batteries (TB65 Intelligent Flight Battery).

The Zenmuse L2 camera can capture all three modalities of RGB image, point clouds, and video streams at very high resolution. The camera provides 5280×3956 (4:3) resolution images. During point cloud recording, the camera periodically captures RGB images at a fixed period of every 2 seconds. This has been the best available camera offered by DJI Enterprise prior to the start of our project. We also conduct video recording flights as additional visual feature enrichment for our dataset. The start and stop of point cloud and video recordings can be programmed into the sequence of actions at any waypoint during the mission.

The Matrice 350 RTK drone was chosen for the only compatible drone with the Zenmuse L2 camera. It supports three different types of automated missions: Waypoint Route, Area Route, and Linear Route. Due to safety concerns, as we explain more in Section 3.1, we only use Waypoint Route in our workflow. In the Waypoint Route, the drone follows an ordered sequence of predefined waypoints (points in 3D space) and performs a sequence of predefined actions upon reaching each waypoint. Different waypoint types can also be chosen: “Coordinated turn. Skips waypoint”, “Straight route. Aircraft stops”, “Curved route. Aircraft stops”, and “Curved route. Aircraft continues”. The Matrice 350 RTK drone operates under a maximum of 42km/h wind speed.

The D-RTK 2 Mobile Station positioning system offers centimeter-level positioning data with real-time differential correction. The RTK information is required to process point clouds using DJI Terra software (more in Section 4.1).

2.2 Construction Site

The construction site is for a student residence, located beside the Parking Lot A, University of Waterloo. As communicated with the construction team, the planned construction timeline spans across 3 years. The parking lot was used for construction vehicles for the construction project. The construction site was next to a train track, which limits the feasible flying region of the drone, as described in further detail in Sections 3.1 and 5.3. Multiple light poles were in the parking lot, which became vital for our point cloud registration process, as described in further detail in Section 4.1.

During the first four weeks, it is an L-shaped region, occupying half the space of the parking lot next to it for construction vehicles driving. It had a non-convex shape. Starting from the sixth week, the construction area evolved to adjust to the construction project needs. The fence surrounding the construction site was reshaped into a near rectangle region, occupying the other half of the parking lot for construction vehicles driving. Re-definition of the waypoint mission is needed every time the construction area reshapes. Moreover, starting from the fourth data collection week, a crane had been setup in the construction site, which increases the lower limit of flight altitude (relative to ground level) to 80 meters for safety reasons, as the crane is around 65 meters high.

3. WORKFLOW DESIGN

Our data collection approach is guided by three primary goals: safety, quality, and consistency. In this section, we describe the detailed considerations for each of the three goals.

3.1 Safety Considerations

Safety is our top priority, and it has a significant impact on the way we design our data collection workflow, especially flight trajectory design. We prioritize absolute safety over high quality and high consistency. We make sure that we clear all potential safety threats in every flight.

Weather Condition. According to the drone's official user guide, it is not recommended to fly under rain, thunder, or any acute weather conditions. Therefore, we avoid such weather conditions when selecting our data collection day every week.

Safety Regulations. According to construction site coordination and the City of Waterloo policy (University of Waterloo, 2025), we are not allowed to fly in the space above the construction area, buildings, or train tracks. Collision with any object or human inside or outside the construction area is strictly prohibited. Approvals from the construction site manager and local police are mandatory for every flight.

Trajectory Planning. To obey the safety regulations, we take the following design. To avoid collision with objects and humans inside the construction area, we set the drone take-off point outside the construction fence, and outside the radius of the crane. The operating altitude of the drone is above all structures in the construction site, including the crane. We use the Waypoint Route and "Straight route. Aircraft stops" waypoint type throughout our data collection process as they give precise control over the flight path. Waypoints are set above the boundary of the construction area. In addition, attention should be paid to the space needed for automatic Inertial Measurement Unit (IMU) calibration before and after the flight. During IMU calibration the drone flies back and forth along a straight line, which takes 30m obstacle clearance at the start and end point of the autonomous mission, according to the Zenmuse L2 user manual. To avoid collision with nearby objects and humans, we set up safety cones around the take-off point and the drone always returns to the same location after flight.

3.2 Quality Considerations

In this section, we describe our flight configurations that optimize for the best data quality, mainly around the following four aspects: image and video resolution, image overlap, point cloud density, and site coverage (point cloud completeness). Based on our site experimental flights, these aspects are mainly affected by the following flight configurations: flight altitude, flight speed, LiDAR scanning, and gimble rotation, as we will be describing below in detail. Quantitative measurements of data quality are detailed in Section 4.2.

Flight Altitude. Flight altitude is one of the highest impact factors on image and video resolution and point cloud completeness. For the first three scenes, the construction site was mostly empty and hence imposed no constraints on flight altitude. We flew at roughly 30m altitude for the best resolution. For the 4th (date 2024/10/18) scene, a crane has been setup in the construction site, which was about 65m high. To ensure safety, we flew at 70m altitude and further increased altitude to 80m starting from the 2024/10/30 flight. Starting from date 2024/11/16 we schedule additional flights at 30m altitude on weekends when the construction will be paused and the entire site remains static to accommodate for the lost in best video resolution and point cloud completeness.

Flight Speed. Flying speed is directly correlated to the overlap of RGB images collected in point cloud recording missions and motion blur in the video recording missions. Since the L2 camera captures images at a fixed period of 2 seconds during point cloud recording missions, we get more densely sampled frames as we decrease the average flying speed and increase flight duration. For the first 4 weeks we set the maximum speed to be 15m/s to maximize data collection efficiency, despite deceleration and acceleration before and after waypoints making the average speed less than 15m/s. Later, we decided to increase the

richness of the visual features available in our dataset by providing more RGB images in a denser manner. Therefore, we decreased the maximum flying speed to 1m/s starting from the 6th data collection week. This roughly triples the number of images we collect for a scene. Motion blurring effects in the video streams were concentrated at the waypoints when the aircraft was making a self-rotation quickly. This quick turn has been unavoidable in our data collection workflow so far. Since motion blur was negligible between waypoints, we kept the video recording flying speed at 15m/s to maximize data collection efficiency.

LiDAR Scanning. For the highest possible point cloud density, we set LiDAR scan mode to Penta return and 240KHz sampling rate. We chose non-repetitive over repetitive scanning for larger recording coverage over the construction site. From our manual qualitative inspection using the CloudCompare software (GPL Software, 2021), we do not observe differences in data quality with repetitive scanning.

Gimble Rotation. Gimble rotation needs to be carefully tuned according to flight route to ensure maximum recording coverage of the construction site. As the drone flies clockwise along the construction area boundary, we set the gimble yaw to $+90°$ so that the camera points perpendicular to the flying direction and points to the construction area throughout the entire flight. We set the gimbal pitch to $-80°$ to ensure that it has enough vision towards the interior of the construction site when flying at the boundary.

3.3 Consistency Considerations

To deliver data capturing the dynamic change in outdoor construction sites, we maintain a regular recording schedule to capture the construction site at a uniform frequency across the construction timeline. Considering the significance of the visual and geometrical changes observed onsite and the availability of our drone pilot team's effort, we decided on a frequency of data collection flights once every week on weekdays at 80m altitude, and one additional flight on weekends at 30m altitude. We make our best effort in scheduling with our pilot and the construction team and avoiding acute weather conditions including low visibility, strong wind, heavy rain or snow, and thunder. However, maintaining a regular schedule remains challenging due to unforeseeable issues with all parties and weather conditions.

4. DATASET CURATION

In this section, we first outline our procedure for processing the raw sensor data into high-quality construction site maps in Section 4.1; and then provide quantitative assessments of data quality and dataset characteristics in Section 4.2.

4.1 Data Preprocessing

Point Cloud Processing. We use DJI Terra software (DJI Enterprise, 2025), default parameters. We select the option to output merged point clouds to avoid manual merging. It requires a NVIDIA GPU with at least 4 GiB of RAM (DJI Enterprise, 2025). Example processed construction site maps represented using 3D point clouds are shown in Figure 1. Significant construction progress from Sep 27 to Oct 18 includes earth moving, crane setup, and construction of the ground base. We do not include that for all scenes for presentation conciseness.

Point Cloud Registration. Even with precise localization technologies provided by the D-RTK 2 Mobile Station, our processed point clouds still possess translations between different runs of our data collection workflow. To register all collected point clouds into a common coordinate frame, we ran the Iterative Closest Point (ICP) algorithm (Besl & McKay, 1992) on the collected point clouds to obtain transformation matrices between point clouds from different dates. Due to noise in the point clouds and changes in the scene across time, our initial attempt running the naïve ICP on the full scene point clouds failed to produce valid registrations. To solve this problem, we first reduce the amount of noise input to the ICP algorithm and eliminate construction changes by cropping out the points of the light poles from each scene as reference objects, as we know they are rigid objects and we assume that they will remain invariant throughout our study, until otherwise. Then we opt for the surrogate task of registering clean and static subsets of the full

scene point clouds by run ICP on the light poles only. Then the transformation obtained using light poles was used as the transformation between the full point clouds. Instability persists due to unavoidable minor noise around the light poles and sensitivity to initialization of the ICP algorithm. Therefore, we make multiple trials until an accurate light pole alignment is obtained. Ground-truth registration labels were provided by two members in our team following identical procedure, independently, and the better one was selected for each scene. The transformation matrices from each scene to the 2024/09/27 scene will be released as part of our dataset.

Figure 1: Selected renderings of point cloud processed by DJI Terra. Left: 2024/09/27. Right: 2024/10/18.

Camera Intrinsics and Extrinsics. 3D reconstruction algorithms take as input a set of RGB images with their camera poses (extrinsics), and the camera intrinsics (Mildenhall, et al., 2021; Muller, et al., 2022). For 3DGS the sparse point cloud from Structure from Motion (SFM) algorithms is also needed as input (Kerbl, et al., 2023). RGB images are directly collected by the DJI Zenmuse L2 camera. As the gimble rotations provided by DJI were not accurate enough, we rely on COLMAP (Schonberger & Frahm, 2016) to obtain accurate camera poses for each image, together with camera intrinsics, resulting in a relative coordinate system. Using relative coordinates (without absolute coordinates) is sufficient for conducting 3D reconstruction algorithms.

Alignment between DJI and COLMAP Coordinates. To support registration of all images and point clouds from different scenes onto the same coordinate system, the relative coordinates obtained from COLMAP would not suffice, because accurate camera rotation in absolute coordinates is still lacking. We utilize the Umeyama method (Umeyama, 1991) to estimate a transform, which is composed of rotation, translation, and scaling components, between the set of camera positions from DJI and those from COLMAP. The Umeyama method was chosen for its wide applications in the literature (Fischler & Bolles, 1981; Besl & McKay, 1992), and its effectiveness due to the facts that it is a closed-form formula, and we have perfect 1-1 correspondence between the points, and that we have no outlier points. Then we apply the obtained transformation on the accurate but relative camera poses produced by COLMAP to accurately align them to the point cloud processed by DJI. The accurate camera poses (position and rotation) for each image, aligned to the same coordinate system as the point clouds, will be released as part of our dataset.

4.2 Data Quality Assessment

We use Ground Sampling Distance (GSD) and point cloud density as our current metrics for assessing data resolution. Developing other quality assessment metrics to address image overlap, point cloud density, and site coverage is left as a future work.

Ground Sampling Distance. GSD is defined to be the linear distance between two consecutive pixels in the image. Let d denote the GDS, h the altitude of the drone, w the sensor width, f the focal length, and p the image pixel width, then GSD can be calculated using Eq. 1.

$$[1]\ d = \frac{hw}{fp}$$

In our setting we have $w = 35mm$, $f = 24mm$, and $p = 3956$. We summarize the flight altitude and corresponding GSD at different altitudes in Table 1. The first 6 data collection dates were selected as representative dates as the following collections were carried out in a similar fashion as the first 6 scenes. In total, we have successfully recorded and processed 20 site maps by the time this paper was written.

Table 1: Dataset statistics and quality assessment (ground sampling distance). All dates are in year 2024. Only those for the first 6 data points were included for conciseness. The full list will be released as dataset metadata.

Data Collection Date	09/27	10/02	10/09	10/18	10/30	11/06
Flight Altitude	30m	30m	30m	70m	80m	80m
Ground Sampling Distance	1.11cm	1.11cm	1.11cm	2.58cm	2.95cm	2.95cm

Point Cloud Density. We define point cloud density as the average number of points in a 3D ball neighborhood for each point in the point cloud. Mathematically, we define neighborhood $\mathcal{N}(p, \mathcal{P}, r)$ and density $\mathcal{D}(\mathcal{P}, r)$ in Eq. 2 and Eq. 3.

$$[2]\ \mathcal{D}(\mathcal{P}, r) \coloneqq \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} |\mathcal{N}(p, \mathcal{P}, r)|.$$

$$[3]\ \mathcal{N}(p, \mathcal{P}, r) \coloneqq \{p' \in \mathcal{P} : p' \neq p, d(p, p') < r\}.$$

Intuitively, $\mathcal{N}(p, \mathcal{P}, r)$ is the set of points within r distance away from p and d is the Euclidean distance. We visualize the density of all point clouds using different radius values in Figure 2.

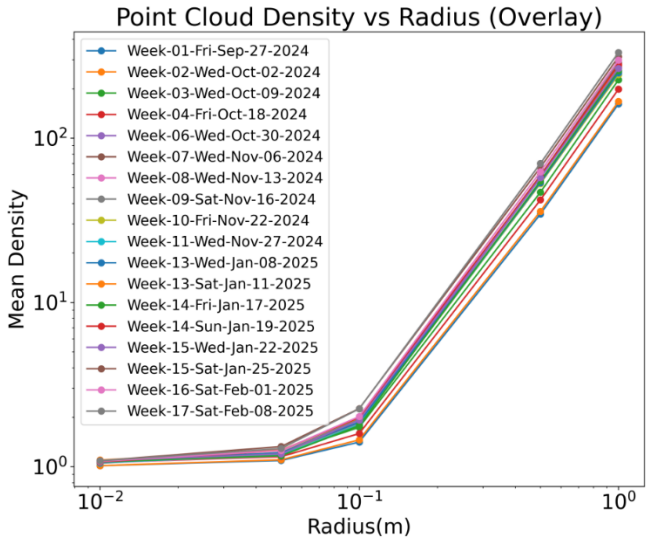

Figure 2: Point cloud density under various radius for all scenes for all 16 scenes collected on weekdays, overlaid in the same figure. Both axes are in log scale.

5. CHALLENGES AND SOLUTIONS

We describe the challenges in quality and consistency we encountered so far. They come from the following factors: flight scheduling, moving objects, and sub-optimal vision. Flight planning introduces challenges for regular data collection schedules and hence affects dataset consistency, while moving objects and sub-optimal vision can potentially decrease our data quality. Solving or mitigating these challenges is left for our future work.

5.1 Moving Objects

During the data collection session, there could be moving objects in the scene, including but not limited to humans, trucks, and the crane. This will create a blurry region in the point cloud processed by DJI Terra. An example is shown in Figure 3. To sum up, we have moving vehicles in the following 9 scenes: 2024-09-27, 2024-10-02, 2024-10-09, 2024-11-06, 2024-11-13, 2024-11-22, 2024-11-27, 2024-12-03, and 2025-01-08. The movement of the crane is also captured in the following 7 scenes: 2024-10-30, 2024-11-06, 2024-11-13, 2024-11-22, 2024-11-27, 2024-12-03, and 2025-01-08. In winter seasons, flying snow also caused noise in the DJI processed point clouds and is observed in the scene 2025-01-11.

5.2 Incomplete Point Cloud

We have incomplete point cloud scans due to complex occlusion issues of construction objects. To mitigate these issues, we conduct additional video recordings at low altitudes: 20m, 25m, and 30m, in each recording. For the ease of SFM algorithms, we fly the drone along horizontal lines at the three altitudes. The additional video recordings could be used for completing the point clouds in future research.

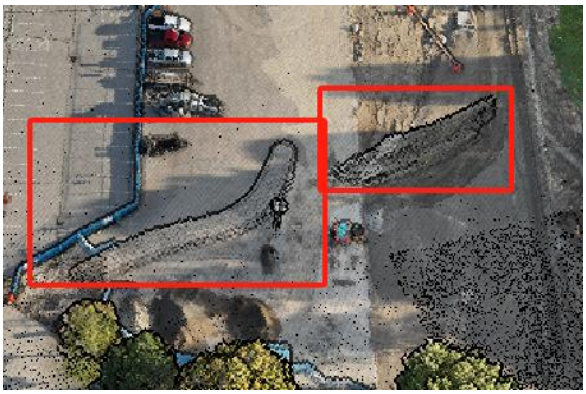

Figure 3: An example point cloud processed by DJI Terra from the 2024/09/27 scene, showing the effect of moving objects during the data collection session. Highlighted in red bounding boxes are regions containing driving vehicles.

5.3 Sub-Optimal Vision

Due to various reasons the 3D space in which the drone is allowed to fly will be limited and hence the drone's vision towards the construction site becomes sub-optimal, either at an acute angle or at a far distance. The current biggest challenge is that we increased our flight altitude to the sub-optimal configuration of 80m to avoid potential interference with the crane which is 65m high. This harms the resolution of our data as it causes a larger GSD.

5.4 Flight Scheduling

Collecting on a fixed day of every week is challenging. This is mainly attributed to the following two factors: acute weather conditions, including low visibility, strong wind, heavy rain or snow, and thunder, and flight scheduling. Unfortunately, we have a missing data point for collection week 05 (Sun Oct 20 - Sat Oct 26, 2024) due to a combination of these factors.

6. 3D RECONSTRUCTION BENCHMARKS

To complement the weakness of point cloud models that the representation is discrete in space, we also consider other 3D models including volumetric models, 3DGS (Kerbl, et al., 2023), and implicit models, NeRF (Mildenhall, et al., 2021) and Instant-NGP (Muller, et al., 2022). In this section, we provide the initial 3D reconstruction benchmark results on our dataset to demonstrate their effectiveness to be used as complementary approaches to model the 3D as-build construction site.

6.1 Experiment Setup

We have selected 3D Gaussian Splatting (Kerbl, et al., 2023), NeRF (Mildenhall, et al., 2021), and Instant-NGP (Muller, et al., 2022) for comparison, due to their popularity in literature and ease of implementation. We trained all the models using the NerfStudio (Tancik, et al., 2023) library. We follow the standard evaluation protocol as in (Mildenhall, et al., 2021) and report the PSNR, SSIM, and LPIPS scores. Throughout our experiments, we used 30,000 iterations. To reduce computation costs, we downscaled input images to 1920×1440 resolution for both training and evaluation.

The original 3DGS (Kerbl, et al., 2023) algorithm starts from the sparse point cloud produced during SFM. However, this creates training difficulty due to under determinism in the scene due to a special set of camera poses: we were only able to place the camera poses to be vertically above the construction site, surrounding the boundary, and approximately facing vertically downwards (top-view). To mitigate this issue, we use the dense point cloud collected by our DJI Zenmuse L2 LiDAR camera and processed by the DJI software as the initialization point of the 3DGS (Kerbl, et al., 2023) training algorithm.

6.2 Results Discussion

Baseline Comparisons. For conciseness of presentation, we only include results on the first 3 scenes. They are summarized in Tables 2. Results have shown that 3DGS (Kerbl, et al., 2023), with or without initialization from DJI, significantly and consistently outperformed both NeRF (Mildenhall, et al., 2021) and Instant-NGP (Muller, et al., 2022) under all evaluation metrics on all scenes. However, the benefit of using the DJI point cloud on initializing the 3DGS model was not clear from these evaluation scores. Instant-NGP (Muller, et al., 2022) significantly and consistently outperformed NeRF (Mildenhall, et al., 2021) under PSNR and SSIM on all scenes.

Table 2: 3D reconstruction benchmark results on scenes 2024/09/27, 2024/10/02, and 2024/10/09. The best score under each metric on each scene is highlighted in bold.

Method	2024/09/27			2024/10/02			2024/10/09		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
3DGS	**23.973**	**0.843**	**0.129**	18.157	0.609	0.329	18.100	0.590	0.358
3DGS-DJI	23.377	0.809	0.138	**18.768**	**0.629**	**0.29**	**21.704**	**0.774**	**0.174**
NeRF	16.890	0.325	0.475	15.932	0.320	0.573	15.810	0.311	0.579
Instant-NGP	20.834	0.555	0.473	18.557	0.459	0.575	18.634	0.459	0.575

Under-constraint Side Views. As the images were captured by the drone from a top-down view, there are no constraints imposed during training from the side views. Figure 4 shows that top-down views displayed high-resolution and low-noise renderings, but side views presented severe geometry errors.

7. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we described our system to safely and consistently capture the dynamics of outdoor construction sites, in high quality, over a long period of time using UAVs. We presented the hardware equipment, considerations on safety, quality, and consistency, our data collection workflow, and the challenges we encountered. We also presented how we preprocess and prepare the raw data captured by the drone into high-quality construction site maps. To the best of our knowledge, we are the first to initiate a systematic data collection workflow that produces such high-quality site maps. The larger impact of our work lies in providing a benchmark dataset with robust evaluation system and enabling fair comparison for computer vision algorithms solving the 3D reconstruction, point cloud registration, 3D change detection, and progress monitoring tasks, and robotics algorithms such as SLAM algorithms.

Figure 4: Top-view (Left) and side-view (Right) of the trained 3DGS model with DJI point cloud initialization of roughly the same region in scene 2024/09/27. Side-view Gaussians have severe geometric errors due to lack of constraint from the side-views.

We aim for the following extensions in the future.

1. We will be developing further collaboration with the university and the construction project, and our data collection efforts will be continued to expand the scale of our dataset.
2. Building Information Models (BIMs) will be made available if further collaboration with the university and construction company is successful. With accurate point cloud registration, it is feasible for construction experts to provide ground-truth annotations for construction-specific changes in point clouds over time. Collectively, these future efforts will support research on detecting changes in the as-built 3D models over time and performing comparison between as-built and as-planned 3D models.
3. Other improvements to be made include (1) investigating how we can effectively deal with scene-moving objects to reduce noise in data; (2) developing more quantitative quality assessment metrics; and (3) fixing geometry errors from the side-views of 3D reconstruction benchmarks.

ACKNOWLEDGEMENTS

This work was supported by the Construction Sector Digitalization and Productivity program of the National Research Council of Canada (NRC). Thanks to Bobby Byers and Bryan Popowich from MELLOUL-BLAMEY CONSTRUCTION INC. for coordination with our team on flight planning.

REFERENCES

Amer, F. & Golparvar-Fard, M., 2018. *Decentralized Visual 3D Mapping of Scattered Work Locations for High-Frequency Tracking of Indoor Construction Activities.* Reston, VA, Computing in Civil Engineering, pp. 454--462.

Besl, P. & McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume 14, pp. 239-256.

Deng, H. et al., 2020. Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision. *J. Constr. Eng. Manage.,* Volume 146, p. 04019095.

Deng, H. et al., 2020. Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision. *Journal of Construction Engineering and Management,* Volume 146, p. 04019095.

DJI Enterprise, 2025. *DJI Terra - Make the World Your Digital Asset.* [Online] Available at: https://enterprise.dji.com/mobile/dji-terra [Accessed 12 April 2025].

Fischler, M. A. & Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM,* Volume 24, p. 381–395.

GPL Software, 2021. *CloudCompare.* s.l.:s.n.

Guan, S., Zhu, Z. & Wang, G., 2022. A Review on UAV-Based Remote Sensing Technologies for Construction and Civil Applications. *Drones,* Volume 6, p. 117.

Hamledari, H. et al., 2018. *UAV-Enabled Site-to-BIM Automation: Aerial Robotic- and Computer Vision-Based Development of As-Built/As-Is BIMs and Quality Control.* New Orleans, Louisiana, Construction Research Congress, pp. 336--346.

Hamledari, H., McCabe, B. & Davari, S., 2017. Automated computer vision-based detection of components of under-construction indoor partitions. *Automation in Construction,* Volume 74, pp. 78--94.

Han, K., Degol, J. & Golparvar-Fard, M., 2018. Geometry- and Appearance-Based Reasoning of Construction Progress Monitoring. *J. Constr. Eng. Manage.,* Volume 144, p. 04017110.

Han, K. K. & Golparvar-Fard, M., 2017. Potential of big visual data and building information modeling for construction performance analytics: An exploratory study. *Automation in Construction,* Volume 73, pp. 184--198.

Ibrahim, A., Golparvar-Fard, M., Bretl, T. & El-Rayes, K., 2017. *Model-Driven Visual Data Capture on Construction Sites: Method and Metrics of Success.* Seattle, Washington, Computing in Civil Engineering, pp. 109--116.

Kerbl, B., Kopanas, G., Leimkuhler, T. & Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.,* Volume 42, pp. 139--1.

Lin, J. J. & Golparvar-Fard, M., 2017. *Proactive construction project controls via predictive visual data analytics.* s.l., Computing in Civil Engineering, pp. 147--154.

Lin, J. J., Ibrahim, A., Sarwade, S. & Golparvar-Fard, M., 2021. Bridge Inspection with Aerial Robots: Automating the Entire Pipeline of Visual Data Capture, 3D Mapping, Defect Detection, Analysis, and Reporting. *J. Comput. Civ. Eng.,* Volume 35, p. 04020064.

Mildenhall, B. et al., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM,* Volume 65, pp. 99--106.

Muller, T., Evans, A., Schied, C. & Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG),* Volume 41, pp. 1--15.

Roberts, D., Torres Calderon, W., Tang, S. & Golparvar-Fard, M., 2020. Vision-Based Construction Worker Activity Analysis Informed by Body Posture. *J. Comput. Civ. Eng.,* Volume 34, p. 04020017.

Roh, S., Aziz, Z. & Pena-Mora, F., 2011. An object-based 3D walk-through model for interior construction progress monitoring. *Automation in Construction,* Volume 20, pp. 66--75.

Schonberger, J. L. & Frahm, J.-M., 2016. *Structure-from-motion revisited.* Las Vegas, NV, USA, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4104--4113.

Shang, Z. & Shen, Z., 2018. *Real-Time 3D Reconstruction on Construction Site Using Visual SLAM and UAV.* New Orleans, Louisiana, Construction Research Congress, pp. 305--315.

Siebert, S. & Teizer, J., 2014. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Automation in construction,* Volume 41, pp. 1--14.

Tancik, M. et al., 2023. *Nerfstudio: A modular framework for neural radiance field development.* s.l., ACM SIGGRAPH 2023 Conference Proceedings, pp. 1--12.

Tang, S., Roberts, D. & Golparvar-Fard, M., 2020. Human-object interaction recognition for automatic construction site safety inspection. *Automation in Construction,* Volume 120, p. 103356.

Tan, Y., Chen, P., Shou, W. & Sadick, A.-M., 2022. Digital Twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic BIM. *Energy and Buildings,* Volume 270, p. 112271.

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume 13, pp. 376-380.

University of Waterloo, 2025. *Remotely Piloted Aircraft Systems (Drones).* [Online] Available at: https://uwaterloo.ca/safety-office/occupational-health-safety/remotely-piloted-aircraft-systems-drones
[Accessed 12 April 2025].

Zhou, X., 2023. Computer Vision Enabled Building Digital Twin Using Building Information Model. *IEEE Trans. Ind. Inf.,* Volume 19, pp. 2684--2692.