Optimizing Structure from Motion Parameters for Volume Estimation of Construction and Demolition Waste

Ashwani Jaiswal [1] **Nikhil Bugalia** [1] **and Quang Phuc Ha** [2]

[1]Department of Civil Engineering, IIT Madras, India
[2]Faculty of Engineering Information and Technology, UTS, Australia

ce22d045@smail.iitm.ac.in, nbugalia@civil.iitm.ac.in, quang.ha@uts.edu.au

Abstract -

Reliable quantification of localized construction and demolition waste (CDW) stockpiles is crucial for promoting sustainable waste management and recycling practices. While prior Structure-from-Motion (SfM) research predominantly focuses on methodological parameters, comparatively limited research has addressed the role of environmental and capturing factors in small-scale reconstructions. This study investigates the accuracy of SfM-based 3D reconstructions for volumetric estimation of localized CDW stockpiles, benchmarking against a consumer-grade handheld LiDAR device. A comprehensive workflow was developed to achieve volume estimation from video sequences, encompassing data acquisition, pre-processing, processing, and post-processing steps. Unlike previous studies that typically prioritize high-resolution images and overlook camera positioning and lighting conditions, this study experimentally assesses these critical parameters. Results indicated minimal influence of lighting conditions, with optimal camera positioning identified as 1.5 times the stockpile height, a frame extraction rate of 1 FPS, and 720p resolution providing the most accurate reconstructions. Analysis of 30 stockpile samples demonstrated a strong correlation (R^2 = 0.929) between SfM and LiDAR volumes, demonstrating that SfM achieves volume accuracy comparable to LiDAR data while offering a faster (3.8 times) and cost-effective solution for the given experimental setup.

Keywords -

Construction and Demolition Waste; Structure from Motion (SfM); Waste Management; Point Cloud; Optimization.

1 Introduction

The global generation of over 10 billion tons of construction and demolition waste (CDW) annually presents a critical environmental and economic challenge, particularly when not effectively managed [1]. Accurate quantification of the localized CDW stockpile, which is often deposited illegally along the street, isolated land areas, and water bodies (an example shown in Figure 1), is essential for enabling efficient and improved waste management and reverse supply chain (RSC) [2]. Inefficiencies in waste collection operations, such as unknown quantity, location, and composition on the upstream side of the RSC, bring logistical complexities and a decline in the quality and quantity of recycled materials [3]. The unknown volumes of localized CDW stockpiles present a critical bottleneck among these challenges. Recent research has increasingly emphasized optimizing the RSC through technological interventions and digitization [4]. However, the existing research has primarily focused on minimizing waste generation, and the literature on technical interventions for other critical processes like waste collection and transportation is relatively scarce.

Figure 1. Examples of localized CDW.
(Source: Authors)

Accurately capturing a three-dimensional (3D) scene is vital to reliably quantifying the volume of a stockpile [5]. Point clouds have emerged as a widely adopted data format for representing 3D environments, providing real-world details, including spatial coordinates and color attributes. Recent advancements have leveraged diverse data acquisition devices, such as Terrestrial Laser Scanners, Unmanned Aerial Vehicles, Simultaneous Localization and Mapping, and RGB-D sensors [6]. However, these acquisition devices are often accompanied by operational challenges, extended processing, and high costs, limiting their applicability for small-scale applications like localized CDW stockpiles

and underscoring the urgent need for portable and scalable solutions.

Recent advancements in consumer-grade handheld LiDAR-equipped devices, such as the iPhone, have established their reliability and versatility as practical solutions for small-scale applications [7]. In addition to handheld LiDAR, Structure from Motion (SfM) is another significant technology that can be employed for small-scale applications [5]. SfM is a pivotal computer vision technique widely employed to reconstruct (3D) structures from two-dimensional (2D) image sequences and is frequently used in scientific research and engineering [8]. A wide range of SfM tools is available. In particular, the interface of Meshroom can be easily customized as it covers very minute details from a research point of view [9]. Generally, SfM is effective for 3D reconstruction. However, it is shaped by various influencing factors, like methodological, environmental, and capture parameters [10]. Among these, methodological parameters include factors related to the process of a tool, which stands out as the most refined, given the extensive research and advancements achieved over recent decades [8].

While significant progress has been made in improving SfM methodology, less attention has been given to two other key aspects: environmental and capture-related parameters. Environmental factors, like lighting and shadows, affect the scene's conditions. In contrast, capture-related factors, including the number of images, their quality, and camera setup, directly impact the accuracy of the SfM-generated point cloud [11]. A notable study focused on point clouds in tunnel environments, specifically examining the influence of lighting conditions and the number of images [10]. While it highlighted essential aspects, the study primarily addressed tunnel-specific scenarios and lacked consideration of broader applications, such as small-scale reconstructions. The existing body of literature recommends including more images and a higher percentage of overlap, which is necessary for effective 3D reconstruction [8]. However, these recommendations often lack specificity and fail to consider critical variables such as scene complexity and camera model used. In addition, image quality and resolution are rarely addressed in detail, particularly for small-scale, such as those involving localized CDW [11]. While high-quality images are generally assumed to be integral to the reconstruction process, the specific influence of these factors remains insufficiently explored. This highlights the pressing need for more precise, context-specific guidelines to inform 3D reconstruction practices across various applications.

To address the research gap, this study aims to address the following research question: *"How do environmental and capture parameters impact the quality and accuracy of SfM-based 3D point clouds for estimating localized CDW stockpile volumes from video sequences?"* By exploring this question, the study aims to extend an understanding of the influence of parameters on 3D reconstruction while offering practical understandings for optimizing volume estimation workflow. Accordingly, the study's primary objective is to investigate environmental and capture parameters affecting SfM-based point cloud accuracy for volume estimation of localized CDW stockpile compared with handheld LiDAR-based point cloud.

The structure of the paper is as follows. Section 2 outlines the proposed methodology. Section 3 presents the experimental design focusing on parameter influences. Section 4 presents the experimental results, discussion, and conclusions in Section 5.

2 Methodology

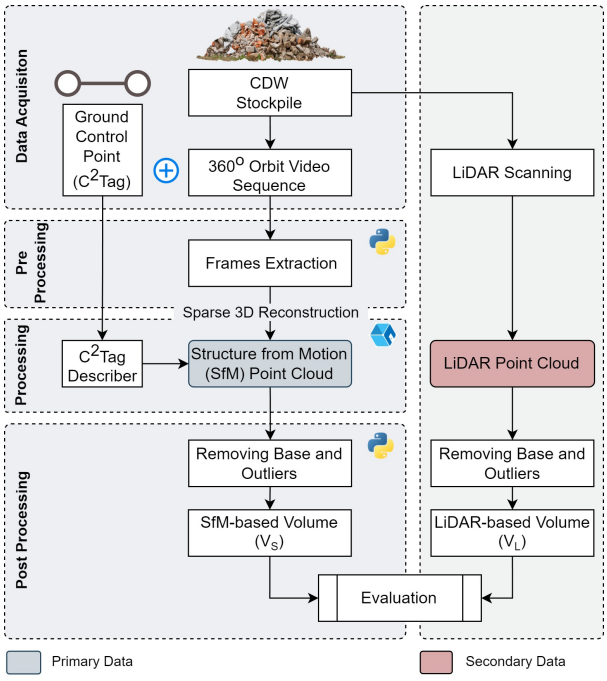

Figure 2. The overall workflow for CDW volume estimation using smartphone video sequence.

Figure 2 illustrates the proposed workflow for comparing the volume of CDW stockpiles from the SfM-based point cloud to the LiDAR-based point cloud. The process begins with a video sequence of the stockpile, captured using a smartphone camera as input. This video is transformed into a precise volume estimate through a systematic step sequence. Initially, individual frames are extracted from the video to facilitate sparse 3D reconstruction. Subsequently, post-processing techniques are employed to eliminate base and outliers, as well as

the final volume estimation of the stockpile. A detailed breakdown of each step is provided in the sections that follow.

2.1 Data Acquisition

Data acquisition in this study involves the formation of CDW stockpiles within a laboratory setup. The laboratory was illuminated using a three-point lighting system, providing sufficient brightness to ensure optimal experimental conditions. The primary smartphone camera for capturing visual data was an iPhone 15 Pro with a 48 MP rear camera. Video sequences were recorded in 4K (2160p) resolution at 60 frames per second. They typically lasted around 30 seconds, covering a complete orbit around the stockpile while maintaining walking speed, as shown in Figure 3a. The stockpile was positioned at the center of the setup, and a 360-degree orbital video sequence was captured to ensure comprehensive coverage, keeping the stockpile consistently framed at the center throughout. To maintain consistency, lighting conditions were standardized at 1000 lux for all experiments except those explicitly testing shadow and lighting parameters.

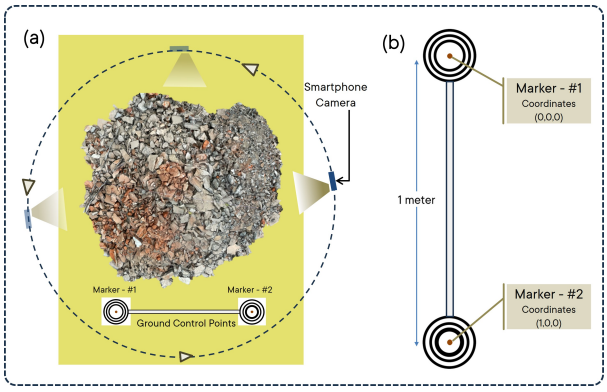

Figure 3. (a) Laboratory setup for data acquisition (b) C^{2}Tag calibration stick developed for this study.

In addition to primary data collection through video sequences captured with the iPhone camera, secondary data were obtained by scanning the stockpiles using the integrated LiDAR sensor of the iPhone 15 Pro. According to the literature, the iPhone's LiDAR sensor has been proven to rapidly produce precise and detailed 3D models, with its applications increasingly documented across various scientific research fields [7]. This secondary dataset was used as a validation reference, ensuring the robustness and accuracy of the primary data derived from video sequences. Monocular SfM fundamentally faces challenges with scale uncertainty, as it reconstructs scenes up to an arbitrary scale factor due to the absence of direct depth information from a single camera view [8]. This limitation necessitates external references, such as Ground Control Points (GCPs), to ensure geometric accuracy and enhance the georeferencing of 3D models. To address the challenge of scale accuracy during 3D reconstruction in this study, C^{2}Tag markers were employed as GCP, facilitating precise calibration and reconstruction [12]. A calibration stick was customized and utilized for accurate scale reference during scanning. The stick, measuring 1 meter in length, featured C^{2}Tag markers affixed to each end, as illustrated in Figure 3b. All the video sequences were obtained by placing the calibration stick.

2.2 Pre-Processing

The pre-processing step employed Python's OpenCV library to extract images from video sequences [13]. A custom Python script was developed using the CaptureVideo function of OpenCV to perform this operation. The script was designed to be alterable, allowing for modifying parameters such as frame extraction rate and resolution to tailor the output based on specific experiment requirements.

2.3 Processing

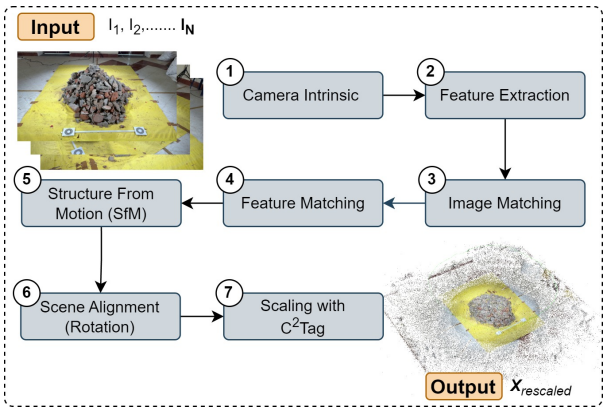

Figure 4. 3D reconstruction pipeline in Meshroom, including scaling with scene alignment and C^{2}Tag Node.

For 3D reconstruction from the extracted images, this study utilized AliceVision Meshroom 2023.3.0, an open-source 3D reconstruction pipeline known for its robust processing [9]. The reconstruction was performed in an HP Workstation Z2, with an Intel i9 processor, 64 GB RAM, and an Nvidia A4000 GPU. The detailed pipeline adopted for transforming images into a point cloud for this study is depicted in Figure 4. Each node in the pipeline is numerically labeled from 1 to 7. The pipeline initiates with N input images. Initially, in Node 1, the camera intrinsic matrix (K) is generated, which

is used to interpret image geometry accurately. In Node 2, features are extracted from each image using multiple feature detection describers like DSP-SIFT, AKAZE, and C^{2}Tag. Node 3 focuses on pairwise image matching to establish relationships between images. Node 4, feature matching uses Euclidean distance to align features across the matched images accurately. A detailed explanation of Node 5-7 is mentioned below.

Node 5, which implements Structure from Motion [9], first estimates the camera's orientation and position (R_t) in 3D space, commonly called the extrinsic matrix, by leveraging epipolar geometry. Matched features from Node 4 are fused into tracks, where each track corresponds to a point in space visible from multiple camera perspectives through triangulation and bundle adjustment. This process allows for the projection and reconstruction of the 3D scene. The relationship between 3D points (X) and their corresponding 2D pixel (x) is mathematically expressed in Equation (1). Z represents the depth of the point in the camera coordinate, which is determined using triangulation.

$$X = Z \cdot (K \cdot R_t)^{-1} \cdot x \tag{1}$$

Upon examining several point clouds generated through the pipeline mentioned above, using images captured by the iPhone 15 Pro camera, it was observed that its orientation was misaligned (see Figure 5a). A manual transformation is applied to resolve the misalignment by introducing Node 6, named Scene Alignment. Specifically, the point cloud is rotated 90 degrees clockwise around the x-axis by multiplying the entire point cloud X with the rotation matrix R_x:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Thus, the aligned point cloud X_{aligned} is (see Equation (2)).

$$X_{\text{aligned}} = R_x \cdot X \tag{2}$$

As discussed earlier, SfM suffers from inherent scale ambiguity. To address this, a C^{2}Tag calibration stick was placed during the data acquisition process, and subsequently, Node 7, named Scaling with C^{2}Tag, was introduced in the pipeline to scale the aligned point cloud (see Figure 5b). To achieve this, a scaling factor S is computed as (Equation 3):

$$S = \frac{d_{\text{real}}}{\left| X_{\text{aligned}(1)} - X_{\text{aligned}(2)} \right|} \tag{3}$$

Where, $X_{\text{aligned}(1)}$ and $X_{\text{aligned}(2)}$ are points in the point cloud corresponding to the same markers 1 and 2 in the real world, respectively. d_{real} is the calibration stick length. The scaling factor S is then applied to the aligned point cloud X_{aligned}, resulting in the rescaled point cloud (see Equation (4)).

$$X_{\text{rescaled}} = S \cdot X_{\text{aligned}} \tag{4}$$

This entire Meshroom pipeline ensures the 3D reconstructed scene is correctly oriented and accurately scaled, facilitating further analysis and visualization.

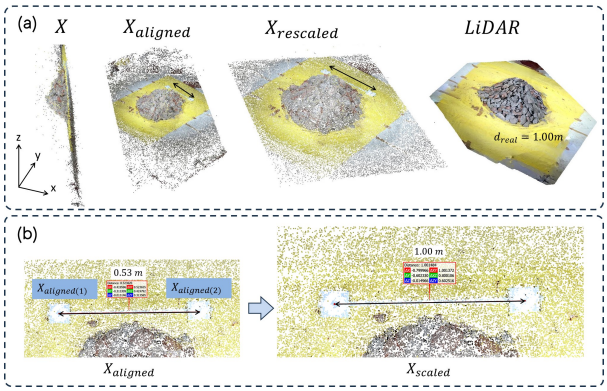

Figure 5. (a) Transformation of the point cloud from Node 5 to Node 7, (b) visual illustration of scaling with C^{2}Tag.

2.4 Post-Processing

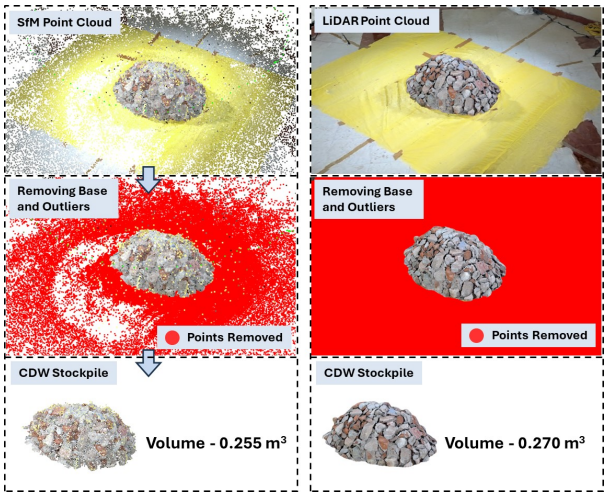

Figure 6. Visual representation of the post-processing step to estimate the SfM and LiDAR point cloud volume.

To ensure precise volume calculation of the CDW stockpile, a yellow sheet was placed beneath to establish a reference base plane (see Figure 3a). Following the

point cloud generation, the subsequent step involved refining the point cloud by eliminating the base plane and extraneous outliers (see Figure 6). Specifically, all points corresponding to the yellow sheet were identified and assigned as the base plane, ensuring that the calculated volume exclusively represents the stockpile material above this plane. The volume estimation was then conducted using the Delaunay Triangulation.

3 Experimental Design

A series of systematic experiments were conducted to evaluate the impact of environmental and capture parameters on 3D reconstruction and volume estimation. These experiments were carefully designed to identify optimal conditions for achieving efficient and reliable outcomes. Among capture parameters, camera height and distance were selected due to their direct impact on the field of view [14]. Shadows and lighting were examined, as they significantly affect feature detection [8]. Frame rate extraction and resolution were also investigated to balance temporal resolution and computational efficiency, ensuring capturing sufficient details without introducing unnecessary computational overhead [10]. The investigations were conducted in the following order: (1) camera height and distance, (2) shadows and lighting, (3) frame rate extraction, and (4) frame resolution. To ensure comprehensive analysis, the experiments were conducted on three separate CDW stockpiles in a controlled environment, with results reported based on the average values across these trials.

3.1 Camera Height and Distance

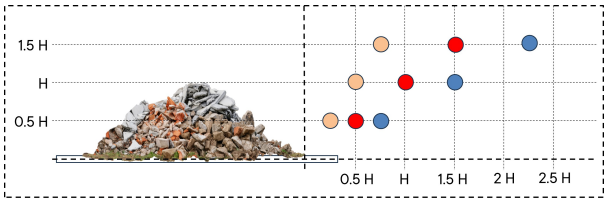

Figure 7. Visual representation of camera height and distance experiment matrix.

To assess the influence of camera height and distance of capture on the 3D reconstruction of a stockpile, an experiment was conducted with the height of the stockpile (H) as a reference parameter. Three camera heights were selected relative to the stockpile: 0.5H, H, and 1.5H to account for varying localized CDW stockpile sizes and ensure diverse camera positions. For instance, 0.5H represents scenarios where the stockpile height is twice the camera height. Similarly, the distance from the stockpile's edge was determined using three distance-to-height (D/H) ratios: 0.5, 1, and 1.5, ensuring that the camera distance scaled proportionally to the stockpile height. This setup resulted in 9 unique camera positions, as shown in Figure 7.

3.2 Shadow and Lighting

This experiment evaluated the effect of shadow and lighting with a design intended to replicate field environmental lighting scenarios where shadows are often unavoidable. This experiment aimed to determine whether variation in lighting conditions during repeated scans of the same stockpile affects reconstruction accuracy and volume accuracy. Figure 8 illustrates the three lighting conditions tested to evaluate their impact on 3D reconstruction:

Case 1: 10 lux lighting (complete darkness).

Case 2: 1000 lux with a single-point light source, which created a shadow on the opposite side of the stockpile.

Case 3: 1000 lux with a three-point light setup, ensuring even illumination and minimizing shadows.

Figure 8. (Case 1) Experimental setup for 0 lux lighting, (Case 2) 1000 lux with a single-point light source, and (Case 3) 1000 lux with a three-point light source.

3.3 Frame Extraction Rate

In this experiment, the number of frames extracted (e.g., FE03 for three frames extracted per second) was varied to identify the optimal requirement for reconstruction and volume estimation, independent of the video length. For this experiment, the custom Python script developed for the pre-processing step was used to perform the frame extraction from video sequences. While increasing the number of extracted frames typically enhances reconstruction precision, it also prolongs the processing time, potentially hindering the requirement to perform quick volume estimation of localized CDW. To balance these factors, nine different frame extraction rates were tested, including FE30, FE10, FE06, FE04, FE03, FE2, FE01, and FE0.5, to determine the optimal frame extraction value.

3.4 Frame Resolution

The literature indicates that frame resolution is typically expected to be used at the raw quality captured. This experiment investigated the effect of reduction in frame resolution on reconstruction to determine the optimal value. The resolution of frames extracted from the video sequence was progressively reduced from the original raw image quality to 2160p, 1080p, 720p, 480p, and 360p using the custom Python script described in the pre-processing step.

3.5 Evaluation Criteria

This study employs two primary evaluation criteria to identify the optimal values. The first criterion focuses on the accuracy of the volume estimation when compared to secondary data, calculated using Equation (5):

$$\text{Volume Accuracy (\%)} = \left(1 - \frac{|V_S - V_L|}{V_L}\right) \times 100 \quad (5)$$

where V_S and V_L are the SfM-based and LiDAR-based volume estimates obtained after the post-processing step. The second criterion assesses the points generated per unit time, representing the total number of points generated during the entire 3D reconstruction process divided by the reconstruction time. A higher value indicates an optimized reconstruction with a shorter processing time, reflecting the optimal value, provided that the volume accuracy remains high for the same condition.

4 Results and Discussion

4.1 Results of Camera Height and Distance

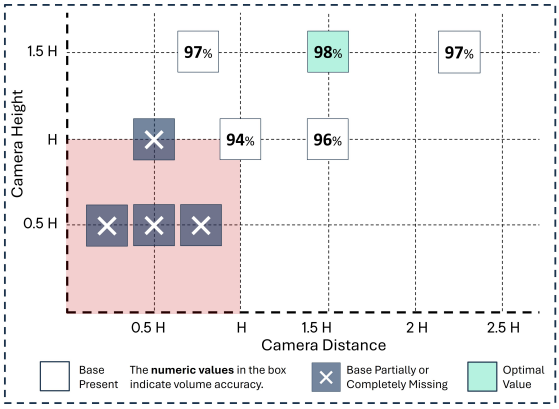

Figure 9. Result matrix of camera height and distance experiment.

The findings from this experiment reveal that the camera positions below the height and distance threshold of H resulted in incomplete reconstruction, with the reference base partially or entirely missing, hampering the post-processing step. (see Figure 9). In contrast, camera positions above the height and distance threshold of H enabled complete reconstruction, allowing for the effective removal of the base and outliers, which ultimately led to successful volume estimation. Among the camera positions, a height and distance of 1.5H proved optimal, achieving the highest volume accuracy of 98%. The points generated per unit of time remained consistent at approximately 330 for camera positions above the H threshold. However, the points generated per unit of time were inconsistent for positions below the threshold, where the reconstruction was incomplete. This suggests that for localized CDW volume estimation, a threshold of camera height and distance above H can be effectively utilized.

4.2 Results of Shadow and Lighting

This experiment was conducted using the optimal camera height and distance, fixed at 1.5H. As shown in Figure 10a, the volume accuracy remained consistently close to 97% across all tested conditions. However, a noteworthy observation is that the point generation rate per unit time in Case 1 was significantly lower than in the other two cases, with the total number of generated points being approximately half that of Case 3. This variation is attributed to reduced feature matching in Case 1, highlighting the key role of lighting in reconstruction. Despite these variations, no significant changes in volume accuracy were observed, indicating that lighting has minimal impact on localized CDW volume estimation.

4.3 Results of Frame Extraction Rate

Figure 10b illustrates that accuracy remained consistent despite a decreased frame extraction rate from FE30 to FE01. However, a marked drop in volume accuracy was observed for FE0.5. Simultaneously, the number of points generated per unit of time steadily increased until reaching FE01, beyond which it sharply decreased. These results suggest that FE01 represents the optimal parameter for maximizing both reconstruction efficiency and volume accuracy, as it is associated with the lowest reconstruction time of 93 seconds and a 98% volume accuracy.

4.4 Results of Frame Resolution

This experiment was carried out using the optimal frame extraction rate of FE01 while varying the frame resolution. The results shown in Figure 10c indicate that reducing the frame resolution below 720p caused a significant decline in volume accuracy. However, this reduction in resolution also decreased reconstruction time from 93 seconds to 56 seconds. Therefore, 720p was identified as the optimal

resolution, balancing volume accuracy and reconstruction efficiency. In contrast, the iPhone LiDAR exhibited much slower performance, requiring an average of 350 seconds per scan.

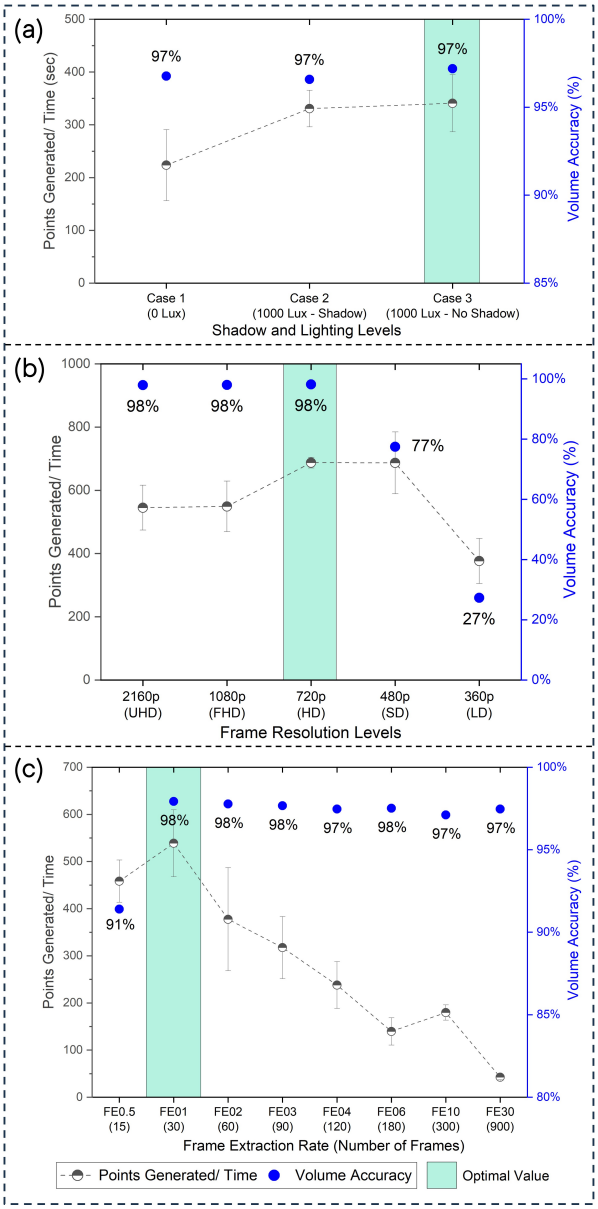

Figure 10. (a) Plot of shadow and lighting experiment, (b) frame extraction rate experiment, and (c) frame resolution experiment.

4.5 Accuracy Assessment

The optimal results derived from the experiments were subsequently employed to evaluate the SfM-based CDW volume estimation workflow. Using the optimized conditions, video sequences of 30 distinct CDW stockpiles were captured to evaluate the performance of the workflow. Figure 11 illustrates a scatter plot comparing volume estimates derived from the SfM-based method (V_S) to those obtained using LiDAR (V_L), yielding a high correlation with an R^2 value of 0.929. This high degree of correlation highlights the proposed SfM-based workflow's suitability for localized CDW volume estimation. Furthermore, the SfM-based workflow shows exceptional efficiency, achieving an average end-to-end time of approximately 97 seconds from data acquisition to post-processing. This time represents a performance that is 3.8 times faster than the 373 seconds required by the LiDAR-based volume estimation workflow, highlighting its operational efficiency. However, the SfM-based end-to-end time includes the processing time, which depends on the system specification used for reconstruction and is subject to change. It also highlights the practicality of the SfM-based workflow for on-site applications.

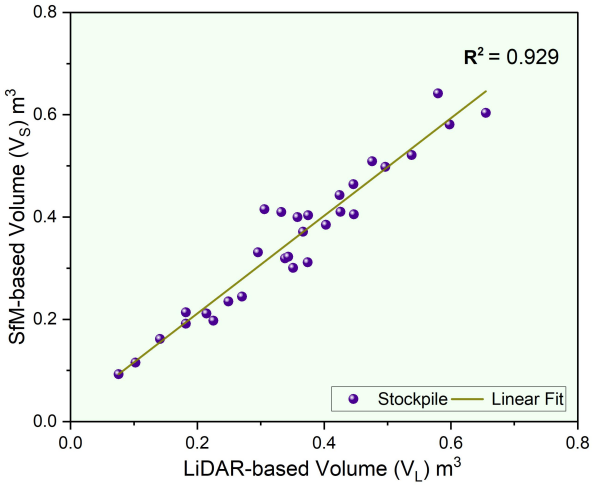

Figure 11. Scatter plot of SfM-based CDW stockpile volume comparison with the LiDAR volume.

4.6 Limitations and future research directions

The present study highlights the potential of SfM for quantifying localized CDW under controlled environments, specifically evaluating the influence of environmental and capture parameters. However, practical application in real-world field environments is limited by several factors, such as site layout, terrain complexity, material composition, spatial variability, and climate conditions. Future research should systematically address these factors to enhance the proposed workflow effectiveness and applicability in the real world. Another limitation of this study is its reliance on the iPhone as the representative smartphone camera, excluding alternative smartphone models. Preliminary investigation indicated that the SfM point clouds remains largely consistent across

different devices under identical capturing conditions, with the only significant change observed in terms of scene alignment. Consequently, Node 6 of reconstruction pipeline require adjustment when using another smartphone camera.

5 Conclusion

Quantifying localized CDW stockpiles is critical in enhancing waste collection and recycling practices. With the rapid advancements in SfM-based 3D reconstruction technologies, it is essential to understand how environmental and capture parameters influence small-scale reconstructions, particularly in the context of localized CDW quantification. This study systematically investigates and identifies, through rigorous experimentation, the optimal environmental and capture conditions required for efficient reconstruction and accurate quantification of CDW stockpiles. Key experimental findings suggest that environmental parameters, such as shadows and lighting, have an insignificant impact on accuracy but moderately affect feature detection during reconstruction, which aligns with the previous studies. However, no significant changes in volume accuracy were observed. In contrast, capture-related parameters, including camera position, frame extraction rate, and frame resolution, emerge as primary determinants of SfM-based 3D reconstruction performance. Notably, optimal point cloud generation and volume accuracy were achieved when the camera height and distance were 1.5 times the stockpile height. A frame extraction rate of FE01 and a resolution of 720p were identified as optimal settings for efficient reconstruction and accurate volume estimation, suggesting that high-fidelity images can be avoided to reduce reconstruction time. Using the optimal values derived from experiments, this study analyzed 30 CDW stockpile samples, demonstrating a high correlation (R^2 = 0.929) between the volume estimates derived from SfM and those obtained from LiDAR. These results showcase the optimization of key parameters, enabling efficient 3D reconstruction and accurate volume estimation of localized CDW stockpiles, making it particularly suitable for small-scale applications.

References

[1] K. Bhattacharjee, S. Chaudhary, A. Vishnoi, D. A. Patel, and N. Bugalia. Characterization of health and safety hazards of deconstruction activities. *American Journal of Industrial Medicine*. doi:10.1002/ajim.23652.

[2] Z. Wu, T. W. Ann, L. Shen, and G. Liu. Quantifying construction and demolition waste: An analytical review. *Waste management*, 34(9):1683–1692, 2014. doi:10.1016/j.wasman.2014.05.010.

[3] R. Brandão, D. J. Edwards, M. R. Hosseini, A. C. Silva Melo, and A. N. Macêdo. Reverse supply chain conceptual model for construction and demolition waste. *Waste Management & Research*, 39(11):1341–1355, 2021. doi:10.1177/0734242X21998730.

[4] A. Jaiswal, P. Roy, N. Bugalia, K. Varghese, and Q. P. Ha. Framework to identify directions for future construction and demolition waste management technologies. In *E3S Web of Conferences*. EDP Sciences, 2024. doi:10.1051/e3sconf/202449604005.

[5] R. Wróżyński, K. Pyszny, M. Sojka, C. Przyby la, and S. Murat-B lażejewska. Ground volume assessment using'structure from motion'photogrammetry with a smartphone and a compact camera. *Open geosciences*, 9(1):281–294, 2017. doi:10.1515/geo-2017-0023.

[6] A. Alsayed and M. R. A. Nabawy. Stockpile volume estimation in open and confined environments: A review. *Drones*, 7(8):537, 2023. doi:10.3390/drones7080537.

[7] G. Luetzenburg, A. Kroon, K. K. Kjeldsen, K. D. Splinter, and A. A. Bjørk. High-resolution topographic surveying and change detection with the iphone lidar. *Nature Protocols*, pages 1–22, 2024. doi:10.1038/s41596-024-01024-9.

[8] O. Özyeşil, V. Voroninski, R. Basri, and A. Singer. A survey of structure from motion*. *Acta Numerica*, 26: 305–364, 2017. doi:10.1017/S096249291700006X.

[9] C. Griwodz, S. Gasparini, L. Calvet, P. Gurdjos, F. Castan, B. Maujean, G. De Lillo, and Y. Lanthony. Alicevision meshroom: An open-source 3d reconstruction pipeline. In *Proceedings of the 12th ACM multimedia systems conference*, pages 241–247, 2021. doi:10.1145/3458305.3478443.

[10] R. García-Luna, S. Senent, R. Jurado-Piña, and R. Jimenez. Structure from motion photogrammetry to characterize underground rock masses: Experiences from two real tunnels. *Tunnelling and underground space technology*, 83:262–273, 2019. doi:10.1016/j.tust.2018.09.026.

[11] I. Nikolov and C. Madsen. Benchmarking close-range structure from motion 3d reconstruction software under varying capturing conditions. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection Proceedings, Part I 6*, pages 15–26. Springer, 2016. doi:10.1007/978-3-319-48496-9_2.

[12] L. Calvet, P. Gurdjos, and V. Charvillat. Camera tracking using concentric circle markers: Paradigms and algorithms. In *2012 19th IEEE International Conference on Image Processing*, pages 1361–1364. IEEE, 2012. doi:10.1109/ICIP.2012.6467121.

[13] G Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[14] A. Bisson-Larrivée and J.-B. LeMoine. Photogrammetry and the impact of camera placement and angular intervals between images on model reconstruction. *Digital Applications in Archaeology and Cultural Heritage*, 26: e00224, 2022. doi:10.1016/j.daach.2022.e00224.