# HORIZONTAL AND VERTICAL ACCURACY OF STRUCTURE FROM MOTION OF TERRESTRIAL IMAGERY

*Hewit Leo J*

M. Tech. Geospatial Engineering
IIT Roorkee ,Uttarakhand

*Udhay Kiraan K H*

M. Tech. Geospatial Engineering
IIT Roorkee, Uttarakhand

## ABSTRACT

This research presents a three-dimensional reconstruction workflow of a campus building using two-dimensional imagery acquired via a mobile device. A sequence of 472 image frames was extracted from a video recorded around the Geomatics Engineering Department at IIT Roorkee. The reconstruction was executed through the Structure from Motion (SfM) approach, where distinctive image features were detected and aligned using the Scale-Invariant Feature Transform (SIFT) algorithm. The photogrammetric processing was conducted using COLMAP, a freely available SfM and Multi-View Stereo (MVS) software. To enhance the spatial accuracy of the model, Ground Control Points (GCPs) were acquired using the Trimble DA2 GNSS receiver and were placed on checkerboards in the survey area. These GCPs enabled proper georeferencing of the final output, which included a dense point cloud and textured mesh. Horizontal and vertical positional accuracy were quantified by comparing model-derived coordinates with GCP measurements, and the discrepancies were expressed as percentage errors. The findings underline the efficiency of integrating mobile-based image acquisition with robust photogrammetric processing and ground referencing techniques for producing accurate 3D spatial datasets. To validate the model accuracy, several portions of the building were measured on-site using a tape and compared with corresponding dimensions obtained from the 3D model using the measurement tool in CloudCompare software. The model achieved an average horizontal accuracy of 99.28 percentage and vertical accuracy of 96.63 percentage. Due to limitations in image coverage, the reconstruction resulted in three separate models; the most detailed model—capturing the rear side of the building—was selected for analysis. The use of exhaustive matching in COLMAP, though computationally intensive, significantly contributed to improved feature correspondence and reconstruction completeness.

***Index Terms—*** 3D Reconstruction, SFM, SIFT

## 1. INTRODUCTION

Three-dimensional (3D) reconstruction from two-dimensional (2D) images has become an increasingly valuable technique in geospatial engineering, architectural modeling, and heritage preservation. Among the various photogrammetric approaches, Structure from Motion (SfM) has gained prominence due to its ability to generate accurate 3D models using only overlapping images taken from multiple viewpoints [1]. This technique allows for the reconstruction of complex structures without the need for expensive equipment or dense manual surveying. In this study, a mobile phone video was used as the data source to reconstruct the structure of the Geomatics Department building at IIT Roorkee. The video frames were extracted and processed through a pipeline involving feature extraction, feature matching, and both sparse and dense 3D reconstruction using COLMAP—a robust, open-source SfM and Multi-View Stereo (MVS) software. The exhaustive matching mode in COLMAP was chosen to ensure that all images were compared pairwise, improving reconstruction completeness and reliability despite the higher computational cost. The generated 3D outputs included a sparse point cloud, a dense point cloud, and a surface mesh using Poisson reconstruction. Due to some limitations in image coverage, the reconstruction resulted in three separate models. The model with the most point detail—capturing the back side of the building—was selected for further processing.

## 2. 3D RECONSTRUCTION

3D reconstruction is the process of creating a three-dimensional digital representation or model of a real-world object or scene. This is frequently accomplished using collections of images that show the subject from different viewpoints. Common methods for image-based 3D reconstruction include Structure-from-Motion (SfM) and Dense Multi-View 3D Reconstruction (DMVR) or Multi-View Stereo (MVS), which work by using overlapping images to determine the 3D structure. The outcome of this process can be a point cloud or a mesh model, often with texture applied [1]. Image-based

3D reconstruction is often highlighted as a cost-effective and efficient approach when compared to methods such as terrestrial 3D range scanning.

### 2.1. STRUCTURE FROM MOTION (SfM)

Structure from Motion (SfM) is a photogrammetric technique used for 3D reconstruction or modeling. It operates on the fundamental principle that the three-dimensional structure of a scene can be determined from a series of overlapping, offset images. A key characteristic that differentiates SfM from traditional photogrammetry is its ability to automatically solve for the geometry of the scene, as well as the camera positions and orientation, without requiring prior knowledge of camera locations or the use of pre-defined ground control points [2]. This process typically involves automatically identifying and matching corresponding features across the multiple images, often utilizing algorithms like SIFT, and solving for camera parameters and scene geometry simultaneously through techniques such as iterative bundle adjustment. The initial output of SfM is usually a sparse 3D point cloud, which is frequently processed further using methods like Dense Multi-View Stereo (MVS) or Dense Multi-View 3D Reconstruction (DMVR) to generate a more detailed, or "dense," point cloud. SfM is recognized as a low-cost and effective tool, made accessible by the use of mass-market devices like digital cameras and smartphones [3], along with the availability of commercial and free/open-source software. Its applications span numerous fields, including geosciences, archaeology, and computer vision.

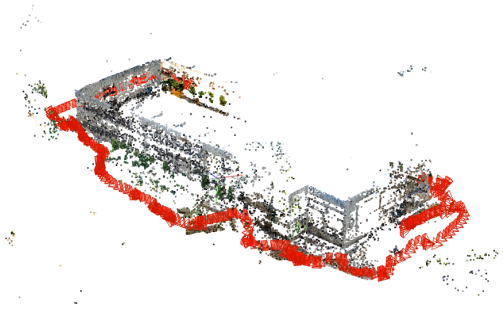

**Fig. 1**. Camera poses and sparse point clouds

### 2.2. SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

The Scale Invariant Feature Transform (SIFT) is a widely used feature detection and description algorithm, particularly essential in Structure from Motion (SfM) workflows. Introduced by David G. Lowe, SIFT is designed to automatically identify distinct regions in an image, called keypoints, that are likely to correspond in multiple images taken from different viewpoints. The core strength of SIFT lies in its ability to detect features that are invariant to changes in scale and rotation, and partially invariant to variations in illumination and 3D viewpoint. This is achieved through a multistep process. First, SIFT builds a Gaussian pyramid by progressively blurring the original image at various scales using different sigma ($\sigma$) values. Then, it subtracts adjacent layers in this pyramid to produce Difference-of-Gaussian (DoG) images. These DoG layers highlight edges and textures, helping the algorithm identify key points as local maxima or minima in both spatial and scale dimensions. This approach ensures that the detected features remain stable across different image resolutions. Once keypoints are detected, each one is assigned a feature descriptor based on local image gradients [4]. These descriptors are distinctive and robust, allowing for accurate matching of corresponding features across large image sets, even in the presence of geometric or photometric changes. In the SfM pipeline, matched SIFT features play a critical role in estimating camera parameters such as position and orientation, and in reconstructing the 3D coordinates of features. This results in a sparse point cloud that represents the geometry of the scene. The effectiveness of this process depends heavily on the quality of the input images, including factors such as texture richness, sharpness, resolution, and feature density.

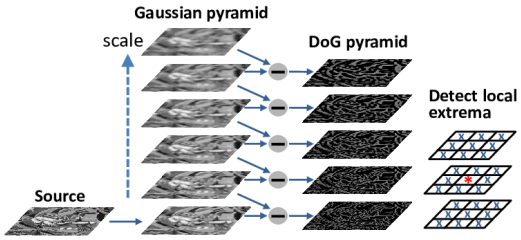


**Fig. 2**. SIFT keypoint detection algorithm showing one octave with 6 Gaussian image layers [4]

## 3. STUDY AREA

The Geomatics Department building at IIT Roorkee is situated in Uttarakhand, India, at coordinates $29°51'44.91''$N, $77°54'0.49''$E. This building was used for our 3D reconstruction project.


**Fig. 3**. Geomatics Department Building

## 4. METHODOLOGY AND SOFTWARES

### 4.1. Image Processing and 3D Reconstruction

The 3D reconstruction of the Geomatics building was performed using COLMAP, an open-source Structure-from-Motion (SfM) and Multi-View Stereo (MVS) software. A total of 472 images were extracted from a video captured using a OnePlus 8 Pro smartphone. These images were processed in COLMAP's exhaustive matching mode, where features were detected and matched across all image pairs using the SIFT (Scale-Invariant Feature Transform) algorithm. The reconstruction began with sparse reconstruction, which estimated camera poses and generated a sparse set of 3D points from the matched features. This was followed by dense reconstruction, producing a high-resolution, detailed point cloud. Finally, a mesh model was generated from the dense point cloud to represent the building's surface geometry.

#### *4.1.1. COLMAP Tool and Cloud Compare*

COLMAP is an open-source Structure-from-Motion (SfM) and Multi-View Stereo (MVS) pipeline for reconstructing 3D scenes from unordered image collections. The process begins with SIFT-based feature extraction, identifying keypoints in scale-space using Gaussian blurring and Difference-of-Gaussian (DoG) localization, followed by the generation of 128-dimensional descriptors for robust feature matching. Matching is accelerated through vocabulary trees and GPU support for handling large datasets efficiently. Sparse reconstruction uses incremental SfM with bundle adjustment to estimate camera poses and 3D point coordinates, starting from two-view geometry and refined through iterative optimization. For dense reconstruction, COLMAP implements a GPU-accelerated PatchMatch Stereo algorithm to compute per-pixel depth and normal maps, which are fused into dense point clouds and meshes using visibility-aware fusion. Surface reconstruction supports both Poisson reconstruction for smooth outputs and Delaunay triangulation to preserve sharp features. COLMAP structures its data modularly, separating features and matches into an SQLite database and storing sparse model outputs in binary files, allowing flexible and partial workflow restarts. Final outputs are saved in standard formats such as PLY and OBJ, and camera intrinsics can be estimated from EXIF metadata or refined using calibration data for greater accuracy [5].

CloudCompare is an open-source 3D point cloud processing software designed for visualization, editing, and analysis of dense point clouds and triangular meshes. It supports a wide range of file formats (e.g., LAS/LAZ, PLY, OBJ), offers robust tools for point-to-point and point-to-mesh distance computation, and includes advanced registration algorithms such as ICP. With its intuitive GUI and extensible plugin system, CloudCompare enables efficient georeferencing, segmentation, and statistical analysis of spatial data. In this study, we used CloudCompare to georeference our SfM-derived models and to extract precise measurements for accuracy assessment.

### 4.2. Georeferencing with Ground Control Points

To assign real-world coordinates to the 3D model, georeferencing was performed using five Ground Control Points (GCPs) collected with a Trimble DA2 GNSS receiver. These GCPs were placed on checkerboards visible in the scene. The georeferencing process was carried out in CloudCompare: first, the mesh model (exported in PLY format from COLMAP) was imported, followed by the GCP coordinates collected with the GNSS receiver. Using CloudCompare's "Point Picking" tool, corresponding points were manually identified on the 3D model to match the physical locations of the checkerboard GCPs. The "Align (point pairs picking)" tool was then used to compute and apply a rigid transformation matrix, aligning the model to the real-world UTM coordinate system while preserving its geometric accuracy and high-resolution texture.

### 4.3. Accuracy Assessment and Validation

The accuracy of the georeferenced 3D model was assessed by comparing manual field measurements with model-derived measurements. Horizontal and vertical dimensions of key building elements were measured manually using a tape measure. These measurements were then cross-checked against corresponding dimensions extracted from the 3D model using CloudCompare's "Point picking" and "Distance measurement" tools. This validation process confirmed the model's dimensional accuracy, ensuring that the reconstructed geometry aligned precisely with real-world measurements.

#### *4.3.1. Accuracy Assessment Formulas*

#### *4.3.2. Absolute Difference*

The raw numerical difference between a ground measurement (tape/GNSS) and the corresponding 3D model value at the same location, calculated as $\Delta = X_{\text{field}} - X_{\text{model}}$. A positive $\Delta$ indicates model underestimation, while negative $\Delta$ indicates overestimation [6].

$$\Delta = X_{\text{tape}} - X_{\text{model}} \tag{1}$$

Where:

- $\Delta$ = Difference between measurements
- $X_{\text{tape}}$ = Field measurement with tape
- $X_{\text{model}}$ = Model-derived measurement

#### 4.3.3. Relative Error Percentage

Relative error percentage is a measure of how much a measured value deviates from the true or reference value, expressed as a percentage of the reference value. It indicates the accuracy of the measurement relative to the actual value [6].

$$Error\% = \left( \frac{|\Delta|}{X_{\text{tape}}} \right) \times 100 \tag{2}$$

#### 4.3.4. Accuracy Percentage

It indicates how close a measured or calculated value is to the true or accepted value, expressed as a percentage. Higher accuracy percentage means the measurement is more reliable and closer to the actual value [6].

$$Accuracy\% = 100 - Error\% \tag{3}$$

#### 4.3.5. Average Accuracy

Average accuracy is the mean of multiple individual accuracy values, typically calculated over a series of measurements or observations. It provides an overall indication of how consistently close the measured values are to the true or reference values across the dataset [6].

$$\bar{A}\% = \frac{1}{n} \sum_{i=1}^{n} A_i\% \tag{4}$$

The tables below summarize the calculations performed to evaluate both horizontal and vertical accuracy by comparing manual tape measurements with distances and heights extracted from our 3D reconstruction model.

**Table 1**. Horizontal distance measurements: tape vs. model.

| No | Tape (m) | Model (m) | Diff (m) | Err % | Acc % |
|---|---|---|---|---|---|
| 1 | 3.31 | 3.34 | -0.03 | 0.91 | 99.09 |
| 2 | 7.29 | 7.34 | -0.05 | 0.68 | 99.32 |
| 3 | 4.09 | 4.11 | 0.02 | 0.48 | 99.52 |
| 4 | 3.32 | 3.30 | 0.02 | 0.60 | 99.40 |
| 5 | 3.27 | 3.24 | 0.03 | 0.91 | 99.09 |
| **Average Accuracy** | | | | | **99.28** |

**Table 2**. Vertical distance measurements: tape vs. model.

| No | Tape (m) | Model (m) | Diff (m) | Err % | Acc % |
|---|---|---|---|---|---|
| 1 | 1.22 | 1.18 | 0.04 | 3.28 | 96.72 |
| 2 | 2.41 | 2.42 | -0.01 | 2.41 | 97.59 |
| 3 | 1.19 | 1.113 | 0.02 | 1.68 | 98.32 |
| 4 | 2.33 | 2.29 | 0.04 | 1.71 | 98.29 |
| 5 | 1.20 | 1.107 | 0.093 | 7.75 | 92.25 |
| **Average Accuracy** | | | | | **96.63** |

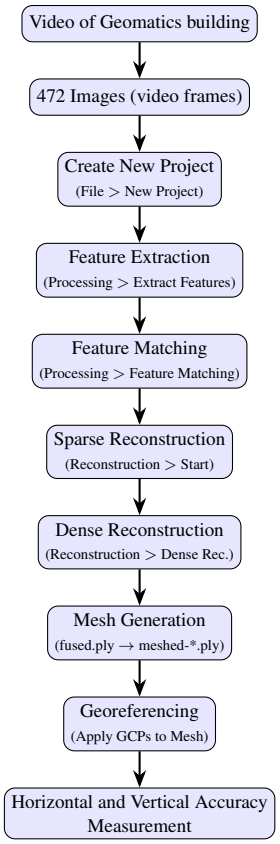


**Fig. 4**. COLMAP 3D reconstruction pipeline.

## 5. DATA ACQUISITION AND PROCESSING

For the purpose of 3D reconstruction, a continuous video of the Geomatics Department building at IIT Roorkee was recorded using a OnePlus smartphone. From this video, 472 high-overlap frames were extracted using the VLC media player software and used as input images for the Structure from Motion (SfM) process. To accurately georeference the model, five Ground Control Points (GCPs) were established using a Trimble DA2 GNSS receiver. These GCPs were placed on checkerboard targets positioned around the building to ensure precise spatial alignment. Executing the exhaustive feature matching and dense reconstruction steps in COLMAP on our 472-image dataset proved extremely resource-intensive, consuming over ten hours on our high-end workstation. Although this lengthy processing yielded a highly detailed and complete point cloud, it underscored the importance of exploring more efficient matching algorithms or leveraging hardware acceleration to reduce runtime in future projects. The radial distortion in the COLMAP model was modeled using the OpenCV distortion model (ID 1).

The image dimensions are $3831 \times 2154$ pixels, with a focal length of $f_x = f_y = 3665.0$ pixels. The principal point is located at $(c_x, c_y) = (1915.5, 1077)$. The distor-

tion coefficients used are as follows: $k_1 = 0.1$, $k_2 = -0.05$, $p_1 = 0.001$, $p_2 = 0.0$, and $k_3 = 0.0$. This information is gave by the COLMAP itself, no need to do the seperate camera calibration.

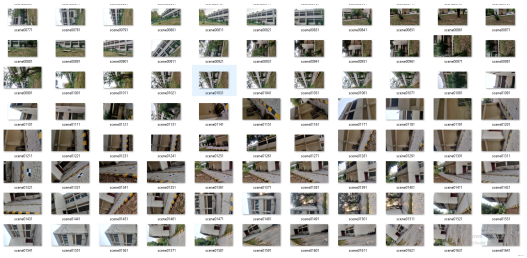

**Fig. 5**. Sample frames from the video

**Fig. 6**. GCP points using the checkerboard

## 6. RESULTS

The Geomatics Department building's 3D reconstruction was accomplished with COLMAP. A comprehensive reconstruction workflow in the software produced both point cloud and Poisson mesh models. The processing time increased because of the exhaustive matching mode, which matches every image with every other image. However, the outcome was a more thorough and accurate reconstruction, which was particularly appropriate for the 472-image dataset.

The model was divided into three parts during the reconstruction process, most likely as a result of overlapping image limits or alignment problems. Since the model with the most points offered the most comprehensive geometry, it was selected for additional examination. However, this model primarily reconstructed only the back side of the Geomatics building, where feature density and image coverage were greater. To validate the model, real-world horizontal and vertical measurements were collected using a tape at selected building sections. These measurements were compared with corresponding values from the 3D model using the CloudCompare measuring tool. The model showed a horizontal accuracy of 99.28 persentage and a vertical accuracy of 96.63 persentage, demonstrating its geometric reliabilAdditionally, five Ground Control Points (GCPs) were collected using the Trimble DA2 GNSS receiver, placed on checkerboards for visibility. These were used to georeference the model within COLMAP and CloudCompare, ensuring real-world coordinate alignment. This georeferencing step was essential for accurate scale, positioning, and for applications such as mapping, documentation, and engineering analysis.

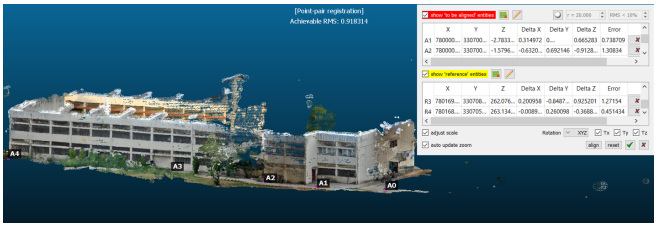

**Fig. 7**. Georeferenced mesh model

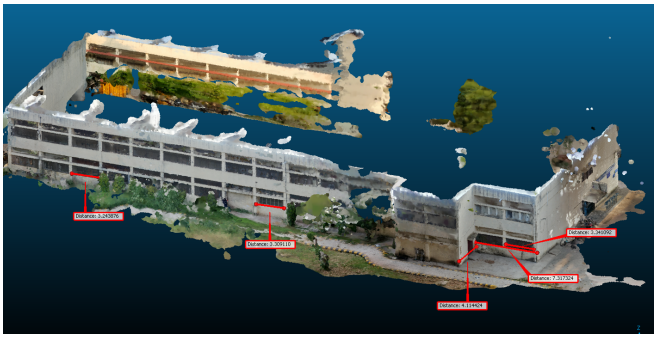

**Fig. 8**. Horizontal measurements from model

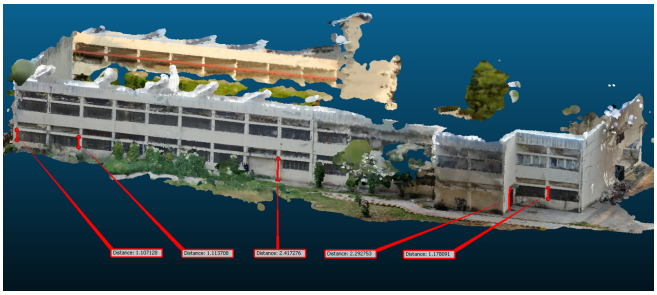

**Fig. 9**. Vertical measurements from model

## 7. CONCLUSION

This research confirms that combining smartphone-based image capture with an exhaustive SfM pipeline in COLMAP can yield precise 3D reconstructions of building facades. By georeferencing the dense point cloud and mesh with five GNSS-derived ground control points and validating against tape-measured dimensions, the model achieved mean horizontal and vertical accuracies of 99.28% and 96.63%, respectively. Although image overlap constraints split the reconstruction into three parts, the most complete segment of the Geomatics Department façade demonstrated dependable geometric fidelity. Future work will target more uniform image coverage

in occluded or shaded regions, enhanced feature matching strategies, and integration of complementary sensors—such as terrestrial LiDAR or UAV imagery—to further improve model completeness and accuracy.

# References


[1] M. R. James and S. Robson, "Straightforward reconstruction of 3d surfaces and topography with a camera: Accuracy and geoscience application," *Journal of Geophysical Research: Earth Surface*, vol. 117, F03017, 3 Sep. 2012, ISSN: 2169-9011. DOI: 10.1029/2011JF002289.

[2] A. Eltner and G. Sofia, "Structure from motion photogrammetric technique," in *Developments in Earth Surface Processes*. Elsevier B.V., Jan. 2020, vol. 23, pp. 1–24, ISBN: 978-0-444-64177-9. DOI: 10.1016/B978-0-444-64177-9.00001-1.

[3] M. J. Westoby, J. Brasington, N. F. Glasser, M. J. Hambrey, and J. M. Reynolds, "'structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications," *Geomorphology*, vol. 179, pp. 300–314, Dec. 2012, ISSN: 0169-555X. DOI: 10.1016/j.geomorph.2012.08.021.

[4] G. Wang, B. Rister, and J. R. Cavallaro, "Workload analysis and efficient opencl-based implementation of sift algorithm on a smartphone," in *2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP 2013)*, 2013, pp. 759–762, ISBN: 978-1-4799-0248-4. DOI: 10.1109/GlobalSIP.2013.6737002.

[5] COLMAP, *Colmap: Structure-from-motion and multi-view stereo*, https://colmap.github.io, Accessed: 2025-04-25.

[6] J. L. Mesa-Mingorance and F. J. Ariza-López, "Accuracy assessment of digital elevation models (dems): A critical review of practices of the past three decades," *Remote Sensing*, vol. 12, no. 16, 2020, ISSN: 2072-4292. [Online]. Available: https://www.mdpi.com/2072-4292/12/16/2630.

[7] R. Wrózyński, K. Pyszny, M. Sojka, C. Przybyła, and S. Murat-Błazejewska, "Ground volume assessment using 'structure from motion' photogrammetry with a smartphone and a compact camera," *Open Geosciences*, vol. 9, pp. 281–294, 1 2017, ISSN: 2391-5447. DOI: 10.1515/geo-2017-0023.