Buildings Built in Minutes - Structure from Motion (SfM)

Jonathan Crespo
University of Maryland
College Park, USA
jcrespo@umd.edu

Naga Kambhampati
University of Maryland
College Park, USA
saigopal@umd.edu

Gaurav Upadhyay
University of Maryland
College Park, USA
Ugaurav@umd.edu

Abstract

The construction industry is witnessing a paradigm shift with the advent of advanced digital technologies. Among these, Structure from Motion (SfM) has emerged as a revolutionary technique, enabling the rapid creation of detailed 3D models of buildings. This paper explores the transformative potential of SfM in the context of architectural and construction practices. SfM utilizes a sequence of overlapping photographs taken from various angles to reconstruct precise three-dimensional structures, significantly reducing the time and cost associated with traditional building methods. By automating the modeling process, SfM offers unprecedented efficiency and accuracy, facilitating real-time visualization and analysis. This technology not only enhances the precision of architectural planning and execution but also supports the integration of augmented reality (AR) and virtual reality (VR) in construction workflows. The implications of deploying SfM extend to improved project management, enhanced safety, and sustainability in building practices. This study highlights case studies and practical applications, demonstrating how buildings can be virtually constructed within minutes, setting a new benchmark for speed and efficiency in the construction industry.

1. Introduction

Structure from Motion (SfM) is an advanced computational technique used to estimate the three-dimensional (3-D) structure of a scene from a collection of two-dimensional (2-D) images. This process leverages multiple views of a scene to reconstruct its 3-D geometry, offering significant potential in various fields, particularly in architecture and construction. The ability to build accurate 3-D models of structures rapidly and efficiently can transform traditional building processes, reducing time and costs associated with design and prototyping.

In this paper, we explore the practical implementation of SfM techniques in the rapid construction of building models. By simulating different case scenarios, we aim to understand the effectiveness and limitations of this technology in real-world applications. Our attempts involved capturing extensive image datasets of building exteriors and interiors, processing these images using SfM algorithms, and evaluating the resulting 3-D models for accuracy and completeness.

2. Algorithm and Pipeline

The Structure from Motion (SfM) algorithm is a computer vision technique used to reconstruct the three-dimensional structure of a scene from a collection of two-dimensional images. Here's a simplified description of the algorithm:

2.1. Image extraction and process

Figure 1. Original Image.

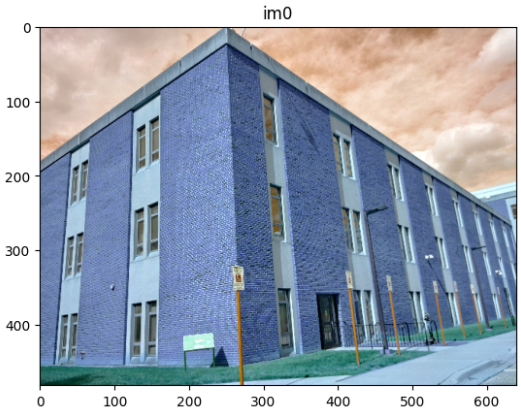

Figure 2. Downsampling Image BRG color scale.

2.2. Feature extraction

The algorithm begins by detecting distinctive features in each image, such as corners, edges, or keypoints. These features serve as reference points for matching corresponding points across different images. Common feature detection algorithms include SIFT (Scale-Invariant Feature Transform) or ORB (Oriented FAST and Rotated BRIEF)

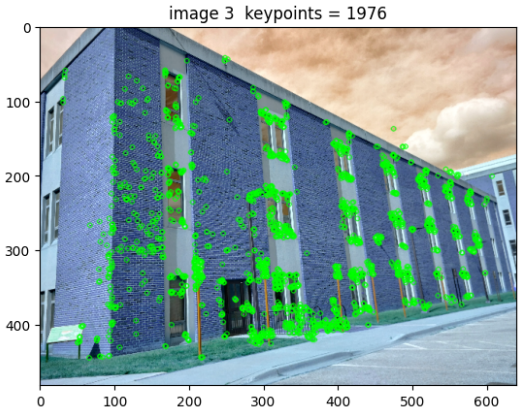

Figure 3. Keypoints for a Image.

2.3. Feature Matching

Once features are detected in each image, the algorithm matches similar features between pairs of images. This is typically done by comparing the descriptors (feature representations) of keypoints and identifying matches based on similarity metrics like Euclidean distance or cosine similarity.

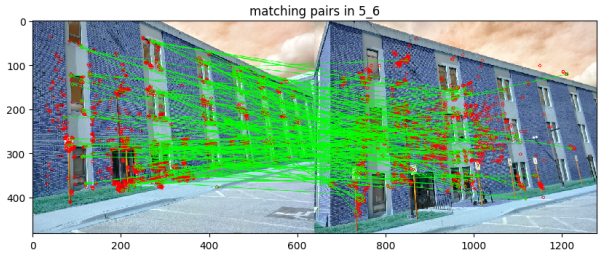

Figure 4. Matching points between set of images

2.4. Geometry Pose estimation

With the matched feature correspondences between images, the algorithm estimates the relative poses or camera positions between each pair of images. This involves determining the translation and rotation transformation that aligns the matched keypoints in one image with their corresponding keypoints in another image. Techniques like RANSAC (Random Sample Consensus) may be used to robustly estimate these transformations while handling outliers.

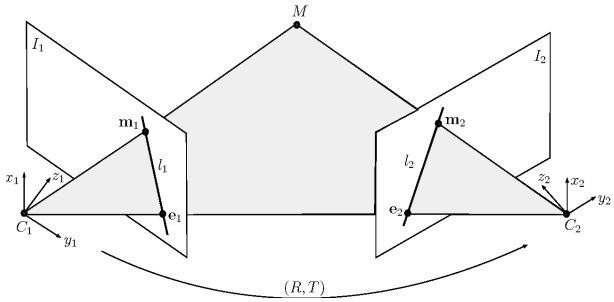

Figure 5. Getting the rotation and translation between cameras. https://d3i71xaburhd42.cloudfront.net/cc67dcbc111827905234f821006d4900c0b3379c/2-Figure1-1.png

2.5. Camera Calibration

Camera calibration using a chessboard involves a process of determining the intrinsic and extrinsic parameters of a camera system by analyzing images of a chessboard pattern captured from different viewpoints. The intrinsic parameters include the focal length, principal point, and lens distortion coefficients, which characterize the internal geometry of the camera. The extrinsic parameters describe the position and orientation of the camera relative to the chessboard pattern. By capturing multiple images of a chessboard placed at different positions and orientations in the camera's field of view, the corresponding corners of the chessboard squares are detected and matched. These image correspondences are then used to estimate the camera parameters through a calibration procedure, such as

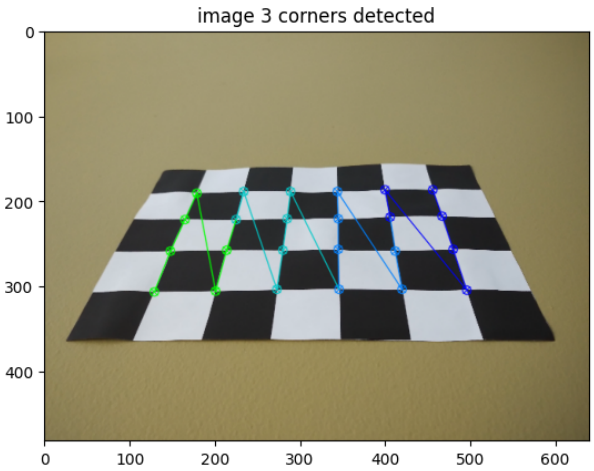

Figure 6. Chessboard patron used for calibration.

the Zhang's method or the OpenCV camera calibration toolbox, which iteratively minimizes the reprojection error between the observed image points and their corresponding 3-D world coordinates. Once calibrated, the camera can accurately project images onto the 3-D scene, facilitating tasks such as 3-D reconstruction, object tracking, and augmented reality.

2.6. Stereo Image Rectification and 3D projection

2.6.1 Stereo Undistortion and rectification

With the intrinsic parameters of the camera obtained rectification is applied, that's the projection of planes onto a common plane parallel to the line between camera centers in which also to reduce the potential re-projection error, undisortion of images is crucial.This process is excecuted for every pair of the set of images.

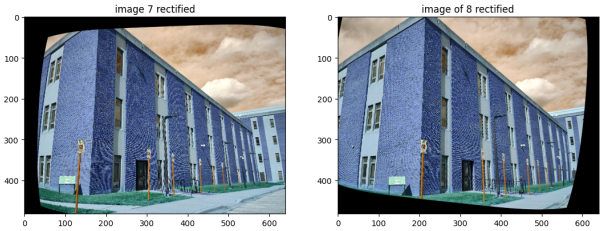

Figure 7. Undistorted and rectified images for stereo.

2.6.2 Disparity maps

With this set of couples of images, disparities maps are build to calculate after depth via triangulation as they give the displacements of conjugate points between pair of image in correspondence with the 3D point they represent.

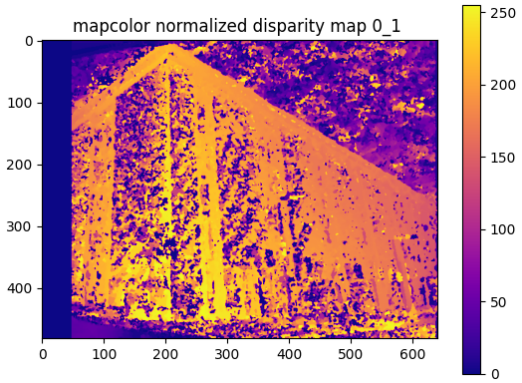

Figure 8. Disparity map.

2.6.3 Triangulation and 3-D

Once the camera poses are refined, triangulation is performed to determine the 3-D positions of the keypoints in the scene. By intersecting the rays projected from corresponding keypoints in different images, the algorithm computes the 3-D coordinates of these points in space. These reconstructed 3-D points represent the structure of the scene.

2.7. Dense Reconstruction

Finally, the reconstructed 3-D points can be visualized to create a representation of the scene's structure. This may involve rendering the 3-D points as a point cloud or mesh, which can be viewed from different perspectives to analyze the scene's geometry.

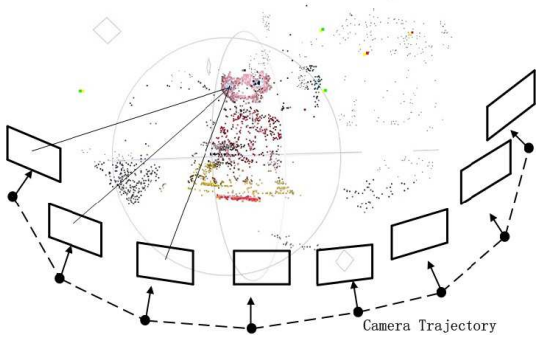

Figure 9. Dense reconstruction.

3. Results

The final result have to be limited due to the quantity of points generated. There a lot of outliers and the depth requires better parametrization or use of additional options to deliver best results, this can be inferred from the quality of the disparity maps.

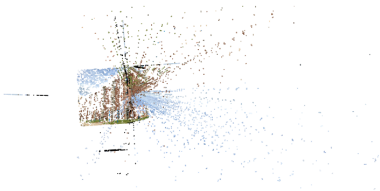

Figure 10. Dense reconstruction for 200k points.

Figure 11. Dense reconstruction 1 for 200k points.

Figure 12. Dense reconstruction 2 for 200k points.

4. Challenges faced

Managing the substantial amount of information generated during the process of camera calibration and 3-D point generation poses a significant challenge, particularly for computational resources. With the complexity of modern camera systems and the increasing demand for high-resolution imagery, the computational burden of processing large datasets can be overwhelming. As a potential solution, leveraging High Performance Computing (HPC) infrastructure could expedite the calibration and point generation process, allowing for efficient handling of massive datasets and enabling faster analysis and visualization of results.

The fine-tuning of parameters plays a crucial role in determining the quality of the calibration and 3-D reconstruction outcomes. Even minor adjustments to parameters can have a substantial impact on the accuracy and reliability of the results. This is especially true in the case of stereo disparity mapping, where the disparity values between corresponding points in stereo image pairs are calculated. Inaccuracies or inconsistencies in parameter tuning can lead to frustrating discrepancies in the disparity map, affecting the overall quality of the reconstructed 3-D scene.

Undistorting photos as part of the camera calibration process is essential for correcting lens distortions and ensuring accurate geometric representation. However, this step can sometimes introduce noise into the images, particularly in regions with low contrast or texture. Dealing with noisy images poses a challenge during subsequent processing steps, as it can adversely affect feature detection, matching, and ultimately, the accuracy of the 3-D reconstruction. Developing robust noise reduction techniques or refining the undistortion process could help mitigate this issue and improve the overall quality of the calibration results.

Handling the multiple view stereo problem, which involves establishing dense correspondences between image pairs to reconstruct 3-D geometry, presents a significant challenge in computer vision. While dense correspondence methods aim to match every pixel in one image to its corresponding pixel in another image, achieving ideal results remains elusive due to factors such as occlusions, variations in lighting and texture, and geometric ambiguities. Addressing these challenges requires the development of more robust algorithms for dense matching and the integration of additional information, such as depth cues and contextual constraints, to improve the accuracy and completeness of the reconstructed 3-D scene.

When converting image coordinates to a reference coordinate system, it is crucial to consider the effects of stereo rectification. Stereo rectification transforms the images to ensure that corresponding epipolar lines are parallel, simplifying the process of matching points between stereo image pairs. However, failure to account for stereo rectification can result in errors in the matching process and lead to inaccuracies in the final 3-D reconstruction. Therefore, ensuring proper stereo rectification is essential for achieving accurate and reliable results in camera calibration and 3-D point generation tasks.

5. Improvements

5.1. Tuning of Parameters via Deep Learning or Machine Learning Algorithm

To address the need for parameter tuning, particularly in cases where each pair of images may require its own parameterization, future work could explore the application of deep learning or machine learning algorithms. These algorithms could be trained to automatically adjust the parameters of the Structure from Motion (SfM) pipeline based on characteristics of the input images and the desired output. For example, convolutional neural networks (CNNs) could learn to predict optimal feature extraction and matching parameters for different types of scenes, lighting conditions, or camera configurations. Similarly, recurrent neural net-

works (RNNs) or reinforcement learning algorithms could adaptively adjust parameters during the SfM process to optimize reconstruction accuracy and efficiency. By automating parameter tuning through machine learning, the SfM pipeline could become more robust and adaptable to diverse scenarios, ultimately improving the quality of visual reconstruction.

5.2. Implementation of Bundle Adjustment Techniques for Refining Visual Reconstruction

As the complexity of scenes increases, visual reconstruction algorithms may encounter challenges such as generating too many points that extend into infinity, leading to inaccuracies in the reconstructed 3-D model. To address this issue, future work could focus on implementing bundle adjustment techniques to refine the visual reconstruction process. Bundle adjustment is a nonlinear optimization method that iteratively refines the camera parameters and 3-D point positions to minimize reprojection errors across multiple views. By incorporating bundle adjustment into the SfM pipeline, redundant or outlier points that contribute to inaccurate reconstructions can be effectively identified and removed. Additionally, bundle adjustment can help improve the geometric consistency of the reconstructed 3-D model by ensuring that points lie within feasible depth ranges and that camera poses are globally consistent. This refinement process can result in more accurate and reliable reconstructions, particularly in scenes with complex geometries or challenging imaging conditions.

6. Future Work

Future work may involve refining SfM techniques for enhanced integration with vSLAM, advancing 3-D scanning technologies, and innovating AR systems for more immersive experiences.

6.1. Integration with vSLAM Techniques

Future research can focus on further enhancing the integration between Structure from Motion (SfM) and Visual Simultaneous Localization and Mapping (vSLAM) techniques to improve real-time mapping and localization accuracy. This may involve developing more robust feature detection and matching algorithms, optimizing camera pose estimation methods, and exploring advanced sensor fusion techniques for comprehensive environmental perception.

6.2. 3-D Scanning Technologies

Continued advancements in Structure from Motion (SfM) for 3-D scanning applications can be pursued to enhance the efficiency, accuracy, and scalability of the reconstruction process. Future work may involve investigating novel algorithms for handling large-scale datasets, improving reconstruction quality in challenging scenarios (e.g., low-texture environments), and integrating SfM with complementary sensing modalities such as LiDAR or depth sensors for enhanced scene understanding.

6.3. Augmented Reality (AR)

Future developments in AR systems can leverage advancements in SfM techniques to enable more immersive and interactive augmented experiences. Research efforts may focus on refining SfM-based scene reconstruction algorithms to support real-time AR content placement and interaction, optimizing resource-efficient rendering techniques for AR devices, and exploring novel applications of AR-enabled by SfM, such as collaborative virtual environments and context-aware assistance systems.

7. Conclusion

This research thoroughly investigates the impactful role of SfM technology in reshaping the fields of architecture and construction. The study reveals that SfM provides significant enhancements in efficiency, precision, and cost savings, fundamentally changing the way buildings are designed and constructed.

Our applied research has shown that SfM significantly cuts down the time required to develop detailed architectural models. This acceleration not only speeds up the design process but also improves decision-making by providing accurate, real-time visualizations of architectural proposals. Furthermore, the combination of SfM with augmented reality (AR) and virtual reality (VR) brings a new depth of immersion and interaction to architectural exploration.

The discussion in this paper centers on the SfM algorithmic framework, focusing on feature extraction, matching, and accurate pose estimation. The use of sophisticated algorithms like SIFT and ORB for detecting features, along with RANSAC for pose estimation, is critical for achieving precise 3D reconstructions. Camera calibration methods, especially those utilizing chessboard patterns, play a vital role in enhancing the accuracy of these spatial measurements.The study also tackles the challenges of processing large datasets and managing intensive computational demands. High-performance computing emerges as crucial for handling the vast amounts of data generated during high-resolution imaging processes. Adjusting stereo image rectification and disparity mapping is highlighted as essential for improving 3D model quality and reducing errors in final reconstructions.

Future enhancements and research directions were also proposed. Optimizing the SfM process through machine learning could automate and refine parameter adjustments. Additionally, integrating bundle adjustment techniques might enhance visual reconstruction, ensuring more consistent and precise 3D models.

References

[1] MathWorks. (n.d.). MathWorks. Retrieved from Structure from Motion from Multiple Views: https://www.mathworks.com/help/vision/ug/structure-from-motion-from-multiple-views.html

[2] OpenCV. (n.d.). Camera Calibration and 3D Reconstruction. Retrieved from https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html

[3] Singh, C. D. (n.d.). Structure From Motion. Retrieved from Computer Vision: https://cmsc426.github.io/sfm/

[4] Satya. (2023, March 21). Medium. Retrieved from 3D Image Reconstruction From Multi-View Stereo: https://medium.com/@satya15july_11937/3d-image-reconstruction-from-multi-view-stereo-782e6912435b