Automatic image-based reconstruction of historical buildings from Ayutthaya

Krisada Chaiyasarn[1*], Bhakapong Bhadrakom[1]

[1] *Department of Civil Engineering, Faculty of Engineering, Thammasat University, Bangkok, Thailand*

Abstract

Traditional close-range photogrammetry involves the process of manually identifying control points in 2D images in order to match 3D control points, typically obtained from a total station. This process can be laborious and prone to error when control points cannot be installed on an interested object. This paper presents an automatic 3D reconstruction technology that allows a 3D model to be created from images without the need to manually identify control points. The technology relies on smart keypoints or markers in 2D images, which can be automatically detected and matched between images, and 3D coordinates can then be reconstructed. In this paper, the reconstruction technology is applied on images of historical buildings from Ayutthaya, which are suitable candidates, as no labels are allowed on these buildings. It is shown that the reconstruction technology can successfully create 3D models of the buildings and can be used as an alternative to traditional close-range photogrammetry methods.

Keywords: VisualSFM, Photogrammetry, SIFT, Structure From Motion, Historical Buildings

1. Introduction

Historical buildings are sensitive to damages due to ageing and nearby activities. This paper presents a study of historical temples from Ayutthaya, where many of the temples have been tilting possibly due to nearby road activities or ground subsidence. Bhadakom et al. [1] have conducted a study on some Ayutthaya temples to measure tilt angles using traditional photogrammetry techniques, which require laborious procedure in identifying control points manually. Furthermore, a full 3D model of a building can only be constructed from multiple 3D sub-models, which requires 3D registration. This process is troublesome and induces inaccuracy.

This paper presents a 3D reconstruction technique that is fully automated and only requires a set of images as input. The technique is based on Structure from Motion, which does not require a manual process. The models from this technique are then compared with other 3D modelling methods, including laser scan and photogrammetry methods, to assess its accuracy to determine if it is sufficient for assessing damages in historical buildings.

The paper is organised as follows, section 2 presents previous studies and a background of automatic image-based 3D reconstruction software. Section 3 explains how 3D models from each method are obtained and how the models from different methods are registered to measure accuracy. Section 4 explains how the models can be used to measure tilt angles of buildings. Section 5 presents the results and discussion, and conclusion is presented in Section 6.

2. Literature Review

2.1 Close-range photogrammetry

Close-range photogrammetry has been used to reconstruct 3D models of historical sites for archiving and for damage assessment purposes [2-3]. Traditional photogrammetry data collection process requires manual procedures in identifying control points, which can be impractical in real sites [4]. In recent years, automatic image-based 3D reconstruction software

* Corresponding author

packages based on computer vision algorithms have been freely available and are now used to create 3D models of historical buildings for archiving purposes [5-6]. The software rely on automatic control point detection algorithms, such as a Scale Invariant Feature Transform (SIFT) [10] algorithm. Structure from Motion algorithms then use these smart control points to reconstruct a 3D model. The automatic software enable data collection to become much easier than traditional photogrammetry methods [7], as images can be taken randomly. The accuracy of models created from the software has been reported for archiving purposes [9], although a limited number of studies is reported for damage assessment purposes, which is highlighted in this paper.

2.2 Background of automatic image-based 3D reconstruction software

Automatic image-based 3D reconstruction software can create a 3D point cloud together with camera poses from uncalibrated photographs. The software is based on two modules: point track generation; and Structure From Motion (SFM).

In the first module, interest points are detected in each image, and then matched across multiple images by matching algorithms to remove bad matches and outliers. These matches are collected and concatenated forming 2D trajectories over the multiple frames called *tracks*. An interest point is an image point whose neighborhood displays distinctive features, which are invariant to perspective transformations, illumination variations and noise. This invariance allows keypoints to be matched across multiple images. The Scale Invariant Transform Features (SIFT) is an example of a smart keypoint detection and matching algorithm [10].

The second module is a large-scale optimization problem. The tracks from the first module are used to initialize the optimizer by a Structure From Motion (SFM) algorithm, which gives initial estimates of a 3D point cloud and camera poses. The estimation is then refined by a Bundle Adjustment (BA) algorithm, which minimizes the sum of re-projection errors between the reconstructed 3D points and 2D tracks, see [8] for details.

The difference between automatic image-based 3D reconstruction and traditional photometry methods is the fact that the automatic software does not require manual control points as input, which makes them more attractive and easier to collect input data.

3. 3D Modelling and Registration

In this study, three different methods of 3D modelling are compared, namely (1) automatic image-based 3D reconstruction (**VisualSFM**), (2) **Photogrammetry** method, and (3) laser scan method (**LIDAR**). The data is collected from three sites, situated in Ayutthaya historical park, namely **Wat Yai Chai Mongkol**, a **small chedhi** near Wat Yai Chai Mongkok, and **Wat Lang Kha Khaow**, as summarised in Tab. 1. Due to the availability of time and equipment, each of the sites does not contain data for each of the three modelling methods.

Section 3.1 explains data collection procedure and how 3D models are created for each modelling method. Section 3.2 explains how 3D models from different methods are registered together to assess accuracy and the results of registration are also shown.

Tab. 1 – a summary table of all data sets.

Data Set	#pictures	LIDAR #points	Photogrammetry #points
Wat Yai Chai Mongkol	489	21,763,543	N/A
Small Chedhi	231	8,516,481	N/A
Wat Lang Kha Khoaw	172	N/A	220,557

3.1 3D Modelling

Image-based 3D Reconstruction (VisualSFM)

Images collected by this method are generally not strict. The rule of thumb is to ensure that an overlap between two consecutive images is at least 50%. In the dataset presented in this paper, pictures were taken at a distance where an entire building was visible, and then a subsequent image was taken at approximate 3-4 metres from the previous image and the procedure was repeated to obtain images covering an entire building. An example of an image is shown in Fig. 1. Fig. 2 shows a 3D model created from this method and all camera poses. In this paper, a Canon 550D with a Sigma lens was used and settings are set as Auto. The images were collected for all three sites as summarised in Tab. 1, in the column #pictures.

The image data is then input to a 3D modelling software called **VisualSFM**, which is an open-source automatic image-based 3D reconstruction software. It only requires uncalibrated images as input, and the output is a 3D sparse point cloud, a dense 3D surface model, and all camera calibration parameters. As shown in Fig. 2 and 3, the output from VisualSFM shows a sparse point cloud and the locations of where each image was taken. It can also give a 3D surface model, in which the point cloud is denser and provides more realistic visualization for a building.

Fig. 1 – an example image from Wat Yai Chai Mongkol

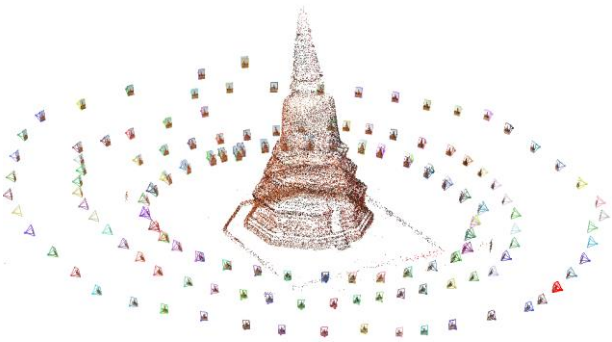

Fig. 2 – an output from VisualSFM from Wat Lang Kha Khao, the output shows a sparse 3D point cloud and the positions of 170 images that were used to create the model.

Fig. 3 – a dense 3D model of Wat Lang Kha Khao from VisualSFM

Photogrammetry

Close-range photogrammetry data was collected from Wat Lang Kha Khao. A pair of images, *i.e.* left and right, was taken to form a dataset for one of the directions, and the process was repeated to obtain data for a total of four directions (8 photos in total). Next, the data was collected using a total station to determine 3D control points, 40 coordinates were collected as reference. These control points were used to match with 2D coordinates from images to determine camera parameters for each pair of images. The control points were identified in images manually by visually recognizing them. The control points in both 2D and 3D were then input into the photogrammetry software to obtain camera parameters and a 3D model of a building. Four 3D sub-models were created for each direction as shown in Fig. 4.

Once 3D models from each direction were created, they were combined into a single 3D model using the 3D control points. This process is done through the photogrammetry software. Tab. 2 contains a summary of statistics when combining the sub-models for each direction, #mesh is the number of triangular mesh created for each 3D model, RMS is the least-square error obtained from the minimization of re-projection errors between 3D and 2D coordinates of the control points.

Tab. 2 – statistics obtained from traditional photogrammetry software for each of the 4 directions.

Direction	#points	#mesh	RMS (mm)
West	54,468	108,903	0.074
East	45,611	91,202	0.137
South	21,969	43,918	0.038
North	26,819	53619	0.131

Fig. 4 – 3D models of Wat Lang Kha Khoaw from four directions.

Fig. 5 – examples of 3D models: left – Wat Lang Kha Khaow from the photogrammetry method; Right – small chedhi from the LIDAR method

Laser Scan (LIDAR)

The data from a laser scanner was collected from two sites, Wat Yai Chai Mongkol and a small chedhi as shown in Tab. 1. The 3D laser scanner used in this work is FARO Focus3D. This system requires at least 3 physical control points to be visible for the laser scan, which can be troublesome to find suitable locations for these control points. Once the locations of control points were setup, the laser scan can then collect the data. The data was collected from 8 locations for each temple in order to obtain point clouds that cover entire buildings. Then the software *Faro Scene* was applied to combine or register the point clouds from each location to form a single point cloud for each temple.

3.2 Registration

In this section, the comparison between image-based method and the other 3D modelling methods is made to determine its accuracy as described below.

VisualSFM vs. LIDAR

In this section, the comparison is made between models from VisualSFM and LIDAR. The comparison is conducted on Wat Yai Chai Mongkol and a small chedhi. The 3D models from VisualSFM and LIDAR are registered in the software called *CloudCompare* using an iterative closest point (ICP) algorithm. Since the models from VisualSFM are not of real scale, different scaling factors are applied manually and incrementally in order to register with LIDAR models accurately. It is assumed that the deformation at the base is negligible and, also the base of a building does not move, hence the base is used as a reference for registration. The 3D points from the LIDAR models were also downsampled to be off a similar number to the VisualSFM points to help with accuracy and reduce computational costs. Fig. 6 exemplifies registration between VisualSFM and LIDAR point clouds. The summary of the results are shown in Tab. 3.

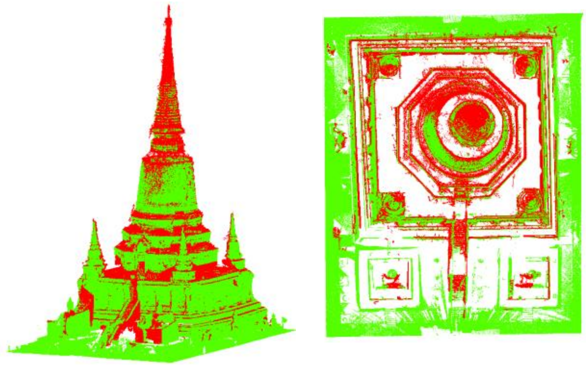

Fig. 6 – an example of registration between VisualSFM and LIDAR model, red is VisualSFM point cloud, and green is LIDAR data.

Tab. 3 – a summary table of registration results between VisualSFM and LIDAR

Data	RMS	LIDAR #points	VisualSFM #points	#point Ratio
Small stupa	0.1145	1,392,587	1,947,378	0.715
Wat Yai Chai Mongkol	0.6964	5,000,000	3,970,488	0.8

VisualSFM vs Photogrammetry

The comparison is made between 3D models from VisualSFM and Photogrammetry, only data from Wat Lang Kha Khao is available. The scaling factor of the 3D model from VisualSFM can be adjusted by using a real measurement from the temple as shown in Fig. 7, where two measurements are used to scale the 3D model from VisualSFM. Then the models from VisualSFM and Photogrammetry are registered together using the ICP algorithm in CloudCompare as shown in Fig. 8. Since the orientation between the two models is different, the ICP algorithm is applied iteratively until RMS errors between the models are constant, which is taken as the final RMS error. Initial alignment between two models must be specified as a reference in order to register the models accurately. In this study, the base of the temple was used as the initial alignment. Tab. 4 shows the results of registration between two 3D models.

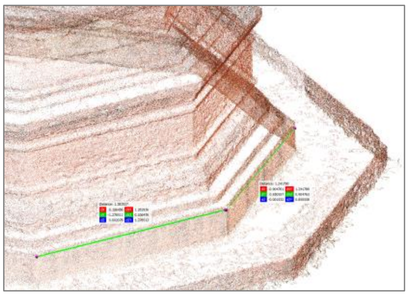

Fig. 7 shows two measurements are used to scale the model from VisualSFM

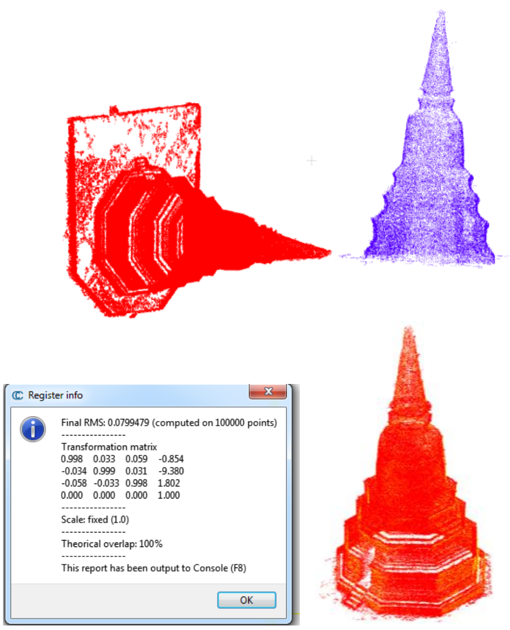

Fig. 8 shows registration of the models from Photogrammetry and VisualSFM, *top* – before registration, *bottom* – after registration

Tab. 4 – a summary table of registration results between VisualSFM and Photogrammetry

Data	RMS	Photogrammetry #Points	VisualSFM #points	#point Ratio
Wat Lang Kha Khaow	0.0799	220,557	675,518	0.33

4. Angle Measurement

The temples from Ayutthaya historical park have been tilting due to nearby road activities and subsidence. In this section, the results after registration from Section 3.2 are used to obtain cross-sectional profiles of the temples to evaluate tilt angles, similar to the work by Bhadrakom et al. [1]. The evaluation is conducted for Wat Lang Kha Khoaw only. As shown from Fig. 5(Left) the model from Photogrammetry contains errors at the joining locations due to meshing error. Therefore, a vertical plane is cut between the joining locations to obtain cross-sectional profiles from the Photogrammetry and VisualSFM models. Two planes perpendicular to each other are applied to the models to obtain cross-sectional profiles for North-South and East-West directions as shown in Fig. 9.

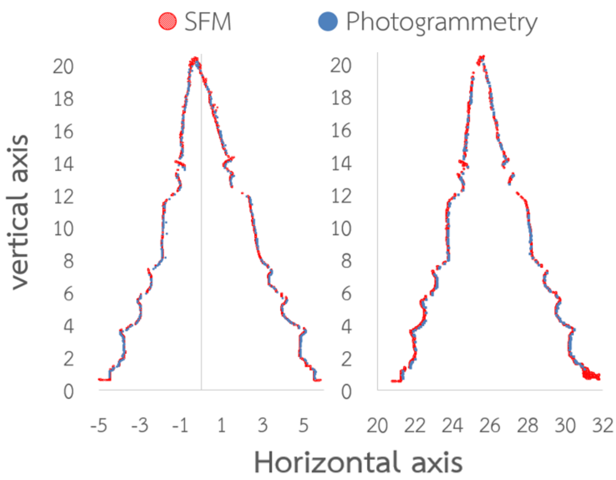

Fig 9. – cross-sectional profiles of models from Photogrammetry and VisualSFM, left – North and South, and right – East and West directions.

5. Results and Discussion

5.1 3D modelling and Registration

In general, the 3D models from LIDAR give the highest number of points, although, for visualization purpose, the model from VisualSFM is sufficient as shown in Fig. 3. The model from Photogrammetry method contains some errors in the place where the sub-models are joined due to meshing error as shown in Fig. 5(Left). One benefit of VisualSFM over other methods is that a 3D model is constructed and registered using consecutive images and can be extended with more images, whereas the other two methods require registration in 3D in order to make the models larger. 3D alignment is difficult and inaccurate.

From Tab. 3 and 4, it can be seen that, generally, the RMS error of the registration between VisualSFM and Photogrammetry is less than the registration between VisualSFM and LIDAR. This suggests that the accuracy of VisualSFM and Photogrammetry is of similar order, whereas the LIDAR data will give the most accurate 3D model. Therefore, the models from VisualSFM can replace the models from Photogrammetry, although it may not be able to replace LIDAR if the required density of point cloud is high.

5.2 Angle measurement

From Fig. 9, it can be seen that cross-sectional profiles from the two methods are well-aligned. This suggest that the method from VisualSFM can replace the traditional photogrammetry since the profiles are almost identical. Hence the results from VisualSFM can be used to monitor tilt angles for these historical buildings in the future study.

6. Conclusion

The study presents an automatic image-based 3D reconstruction software, which can create 3D models from uncalibrated images without the need of manual control points, unlike traditional photogrammetry methods. It is shown that the output from the software provides comparable accuracy to LIDAR models and photogrammetry methods, and can be used to create 3D models for visualization purpose and for damage assessment purpose, such as assessing tilt angles from historical buildings.

References

[1] B. Bhadrakom, N. Poovarodom, U. Weesakul, Y. Fujii, K. Watanabe and S. Kawano. Application of Photogrammetry for Conservation of Historical Remains. *NCCE 18*, 2012.

[2] Y.Alshawabkeh and M. El-Khalili, “Detection and Quatification of Material Displacements at Historical Structures Using Photogrammetry and Laser Scanning Techniques”, *Mediterranean Archaeology and Archaeometry*, Vol. 13, No 2, pp. 57-67, 2013

[3] Y. Fujii, E. Fodde, K.Watanabe, K. Murakami. “Digital photogrammetry for the documentation of structural damage in earthen archaeological sites: The case of Ajina Tepa, Tajikistan”. *Engineering Geology*, 105, pp.124–133, 2008.

[4] H.M. Yilmaz, M. Yakar, S.A. Gulec, O.N. Dulgerler. “Importance of digital close-range photogrammetry in documentation of cultural heritage”. *Journal of Cultural Heritage*, pp.428-433, 2008.

[5] Th. P. Kersten and M. Lindstaedt. “Image-Based Low-Cost Systems for Automatic 3D Recording and Modelling of Archaeological Finds and Objects”. *Progress in Cultural Heritage Preservation Lecture Notes in Computer Science Volume 7616*, pp 1-10. 2012.

[6] F. Remondino, S.D. Pizzo, Th. P. Kersten, Salvatore Troisi. "Low-Cost and Open-Source Solutions for Automated Image Orientation – A Critical Overview". *Progress in Cultural Heritage Preservation 4th International Conference*, EuroMed 2012 Limassol, Cyprus, October 29 –November 3, 2012

[7] E. Andaroodi and Z. Taqipour. "Architectural Photogrammetry by Non-Metric Cameras: CAD-Based 2D Drawing of Facades from Rectified Photos". *Springer-Berlin Heidelberg*, 2012.

[8] N. Snavely, S. Seitz, and R. Szeliski. “Photo tourism: exploring photo collections in 3d.” *ACM Transactions on Graphics*, 2006, 25(3), 835–846.

[9] Brutto, M. Lo, and Paola Meli. "Computer vision tools for 3D modelling in archaeology.", *International Journal of Heritage in the Digital Era 1*, 1-6, 2012.

[10] Lowe, D. “Distinctive image features from scale-invariant keypoints.” *IJCV*, 60(2), 91–110, 2004.