Combined aerial and terrestrial images for complete 3D documentation of Singosari Temple based on Structure from Motion algorithm

Husnul Hidayat and A B Cahyono

Department of Geomatics Engineering, Institut Teknologi Sepuluh November, Sukolilo, Surabaya, 60111, Indonesia

hidayat@geodesy.its.ac.id

Abstract. Singosari temple is one of cultural heritage building in East Java, Indonesia which was built in 1300s and restored in 1934-1937. Because of its history and importance, complete documentation of this temple is required. Nowadays with the advent of low cost UAVs combining aerial photography with terrestrial photogrammetry gives more complete data for 3D documentation. This research aims to make complete 3D model of this landmark from aerial and terrestrial photographs with Structure from Motion algorithm. To establish correct scale, position, and orientation, the final 3D model was georeferenced with Ground Control Points in UTM 49S coordinate system. The result shows that all façades, floor, and upper structures can be modeled completely in 3D. In terms of 3D coordinate accuracy, the Root Mean Square Errors (RMSEs) are RMSE$_X$=0.041 m; RMSE$_Y$=0.031 m; RMSE$_Z$=0.049 m which represent 0.071 m displacement in 3D space. In addition the mean difference of length measurements of the object is 0.057 m. With this accuracy, this method can be used to map the site up to 1:237 scale. Although the accuracy level is still in centimeters, the combined aerial and terrestrial photographs with Structure from Motion algorithm can provide complete and visually interesting 3D model.

1. Introduction

As civilization evolves, it leaves many things as its evidences. One type of these evidences is cultural heritage building. These buildings are very important in the context of archaeology and architecture because they show the footprints on the development of art and construction technology. But many of these buildings are threatened. The heirs of these buildings may dissolved over the time, leaving these buildings without any proper maintenance and preservation. On the other hands, these buildings are also threatened for example by fire, earthquake, war, or erosion. Therefore, documenting these buildings becomes a fundamental task for restoration and reconstruction in the future [1]. One of cultural heritage building in Indonesia is Singosari Temple. This temple is located at Singosari District, Malang Regency, East Java. The archaeologist predicts that this temple was built in 1300s. In 1930s, the Dutch colonial government restored this temple [2].

Documenting cultural heritage building can be done by using terrestrial laser scanner. This device can record million point cloud data of the building quickly. In early stage of development, laser technology has surpassed traditional close range photogrammetry. But in past years by development of digital photogrammetry and computer vision, the extraction of 3D point data from images can be done...
automatically. Nowadays the generation of detailed 3D models from images is a fact, eroding the superiority of laser scanning [3]. The Structure from Motion (SfM) algorithm which based on multiview stereovision enables the extraction of 3D point cloud from unordered image sequences from low cost camera. This novel method is low cost, effective, and can be an alternative to traditional surveying or photogrammetric method [4].

Nowadays, existing terrestrial photogrammetry technique utilizes this algorithm and greatly improves the result and productivity. The building façades can be modeled perfectly while the roof and upper structure are difficult to be modeled only by using terrestrial images [5]. By the advent of low cost UAV system, nowadays aerial photography can be done easily. Combining this technology with existing terrestrial photogrammetry will give more complete coverage data of building structure, especially the roof and upper structures. Therefore a complete model of cultural heritage site and its surrounding environment can be modeled completely without any missing parts. Several projects have shown that employing UAVs for cultural heritage documentation gives satisfactory result [6,7,8]. UAV can obtain fast overview of the site, being a standard tool for excavation, or even to do 3D modeling which make it a universal tool for archaeologists [9]. This research aims to make a 3D model of Singosari Temple by combining aerial and terrestrial photographs as shown in figure 1. The site was selected because currently there is no digital 3D archive of this landmark. The incomplete top structure becomes a good case study on how aerial images play important roles in cultural heritage modeling.

2. Methodology

2.1. Research location
The Singosari Temple is located at Candi Renggo Village, Malang Regency, 9 kilometers northward from Malang City (figure 2). The site is located in the middle of settlement, but still accessible by many vehicles since it is located only about 500 meters westward from national road. This temple is 14.10 high with the footprint measured about 8×8 meters [10]. Unlike other temples, this temple has very little reliefs. The top structure is also incomplete. These facts suggest that the construction process was not finished. Today, the temple is still used as a tourism object and sometimes the cultural events are held there.
2.2. Terrestrial image acquisition
Terrestrial image acquisition was done by covering the object from all directions. This is done to ensure that all façades of the object are well-captured as shown in figure 3. A 12-MP resolution consumer grade camera Canon Powershot SX260 HS with 4-mm focal length was used for this task. There are 39 terrestrial images in this project that were taken from all directions. The images were taken from roughly 15-meter distance. Since the images cover the object in 360-degrees, therefore the intersection angle between image’s optical axis is roughly 9 degrees. The short baseline ensures large overlap between images which the SfM algorithm requires [4].

![Figure 3. Research location (source: http://maps.google.com).](image)

2.3. Aerial image acquisition
The aerial image acquisition was done mainly to cover upper structures and building floor by employing the same camera used in terrestrial image acquisition. The camera was mounted on a lightweight quadcopter DJI Phantom 2. In order to ensure each façades covered completely, the vehicle moved in circular trajectory around the temple while the camera takes the images every 5 seconds. The camera was tilted about 45 degrees to the ground to take oblique imagery (figure 4). To cover the building completely from the air, the vertical aerial imagery was also taken (figure 5).
2.4. Ground control survey and measurement

The SfM algorithm is very efficient in shape reconstruction. The level of detail of 3D point cloud depends on image quality. It can generate comparable results with Terrestrial Laser Scanning does [11]. But the main drawback of SfM algorithm is it only recovers the shape of an object. The SfM algorithm cannot recover absolute scale, position, orientation of an object unless there are some control measurements. In order to achieve properly georeferenced model, the Ground Control Points (GCPs) were established by terrestrial surveying method. The reference frame consists of 4 benchmarks whose coordinates were measured using total station instrument. As the control, two of these benchmarks were surveyed using differential Global Positioning System. Once the reference frame have been established, the GCPs can be surveyed. There were 5 GCPs located at some parts of the temple. In order to check the accuracy of georeferencing, some Independent Control Points (ICPs) were also measured. Additional length measurements of some line segments of the temple were also conducted to check the geometric accuracy of the model.

2.5. Image matching and alignment

This is the first step in SfM algorithm pipeline which becomes a key for the next step. Three dimensional restitution from 2D images requires information about camera position and orientation in
3D space. The camera position and orientation in 3D space can be obtained from relative orientation between images. The relative orientation can be done by selecting corresponding tie points between the images. This process can be done automatically with image matching technique. As the correspondence between images has been established, then camera position during image acquisition can be estimated.

In this research, aerial and terrestrial image alignment were done separately for hardware efficiency. Beside that, between terrestrial and aerial images there are wide baselines. This causes the aerial and terrestrial image alignment cannot be done simultaneously with SfM unless there are some tie points that link both types of image blocks.

Terrestrial image alignment gives image position and orientation in arbitrary coordinate system. Beside that, this process also delivers sparse point cloud which reveals wall structures of the building. Since floor and top structures have lack visibility in terrestrial images, these structures are less revealed in sparse point cloud. To overcome this problem the aerial image alignment was done. This alignment forms photogrammetric block of aerial images. This block reveals floor and top structures although the wall structures were partially revealed. Combining these blocks could reveal whole structure of the temple which is represented in sparse point cloud (figure 6). Some tie points were manually marked on both image blocks, which further these block were joined together based on coordinate transformation using these tie points.

![Figure 7](image.png)

**Figure 7.** Terrestrial (a), aerial (b), and combined image block (c).

2.6. Dense reconstruction

After image alignment and sparse reconstruction, the next step is dense reconstruction. In this process the object geometry are revealed with higher detail. The dense reconstruction mainly relies on space intersection and digital image matching technique. The exterior orientation parameters of each image were used to estimate 3D position of point cloud. These parameters also play important role in digital image matching. Epipolar geometry of a point between many images can be generated from exterior orientation parameter. Having epipolar line between the images, the digital image matching process
can be done only by searching along the epipolar line instead searching in entire image. Therefore the matching process can be done faster [12].

Combined aerial and terrestrial image block gives complete 3D point cloud of the temple. The point cloud of walls, floors, and top structures shows clearly the entire shape of the temple with intricate details (figure 7). But however the 3D point cloud only represents discrete points of continuous surfaces. Therefore in order to make continuous surfaces of the temple the meshing process must be carried out.

![Figure 8. 3D point cloud of the temple with high detail.](image)

2.7. Meshing and texturing
Meshing process was done to turn discrete point cloud data into continuous 3D surface of the building. Before beginning the process, some points which were considered as outliers were removed. These points are located far from predicted model surface which was caused by inaccurate dense matching process. The surfaces were constructed in the form of Triangulated Irregular Network (TIN).

After the surface have been constructed, the next step is texturing. In this process the texture from original images is projected to model surfaces. If there are some particular surfaces imaged at more than one image, then the projected texture is mixed from these images.

![Figure 9. 3D mesh and textured model.](image)

2.8. Georeferencing
Since the SfM algorithm only reveals 3D shape of an object with arbitrary coordinate system, it is obvious that georeferencing process is required. This process can be done by using Ground Control Points whose coordinate have been surveyed before. To check geometric accuracy of reconstructed 3D model, some Independent Check Point were also surveyed and marked (figure 9).
The Ground Control Points consist of 5 points in UTM Zone 49S with WGS 1984 ellipsoid. The elevation uses ellipsoidal height as reference. The Independent Check Points were placed at various parts of the temple. The ICP placement is only based on natural target that identifiable on the images. This was done since target placement on the upper parts of the temple were not allowed. The coordinate of Ground Control Points and Independent Check Points were acquired by terrestrial surveying.

![Ground Control Points and Independent Check Points distribution](image)

**Figure 10.** Ground Control Points position (red dots) and Independent Check Point (yellow dots) distribution.

### 3. Results and discussion

#### 3.1. A 3D model of Singosari Temple

From SfM pipeline applied in this research, a complete 3D model of Singosari Temple was reconstructed. This model not only shows the building, but also its surrounding environment. All façades were reconstructed with good detail except the rear façade. This was caused by backlighting which causes this façade was not seen clearly which affect the accuracy of 3D point clouds. Therefore the resulting façade has rough surface compared to other façades.

In terms of model completeness, the combination of aerial and terrestrial images gives satisfactory results. Terrestrial imaging could not provide complete 3D reconstruction because the lack of visibility in floor and roof structure. With the aid from aerial images, the floor and top structure can be modeled very good forming complete 3D model with its surrounding environment.
Figure 1. 3D model of Singosari Temple.
(a) North view. (b) South view. (c) Top view. (d) East view.

This 3D model also shows good details of top structure. From the model it has been shown that the upper structure were not formed completely as shown in figure 11.

Figure 2. Incomplete top structure of the temple.

3.2. Accuracy analysis

Accuracy analysis was done by comparing 3D coordinates of Independent Check Points (ICPs) and length of line segments with reference data obtained from terrestrial surveying. There are 12 ICPs which used for this accuracy test. The results show that the mean absolute errors of 3D coordinates are $dX=0.034$ m; $dY=0.026$ m; and $dZ=0.042$ m, with Root Mean Square Error (RMSE) values $RMSE_X=0.041$ m; $RMSE_Y=0.031$ m; and $RMSE_Z=0.049$ m. These RMSE values correspond to 0.071 m displacement in 3D space. The planimetric accuracy of this project tends to be better than its vertical accuracy. This is caused that most of control points lie on horizontal plane.
To check the geometric accuracy of 3D model, length measurements were also conducted. There are 43 line segments used for accuracy test. These lines were measured on the site using distometer. For the line segments that lie on the upper part of the temple their lengths were determined by well known intersection method in terrestrial surveying. The result shows that the mean absolute error of line measurements is 0.057 m with RMSE=0.065 m. This result is roughly equal to the RMSE of 3D position.

**Figure 13.** Error in 3D position.
According to Technical Specification of Topographic Mapping for 1:25000 scale by Bakosurtanal (now Geospatial Information Agency: BIG) [13], the cartometric accuracy for mapping is 0.3 mm. Based on largest achieved RMSE, i.e. 0.071 m, the 3D modeling by combining UAV aerial and terrestrial images with SfM processing pipeline can be used for mapping the building up to 1:237 scale.

4. Conclusion
This research has found that combination of terrestrial and aerial UAV images with Structure from Motion processing method can be used to make a 3D model of Singosari Temple completely. All façades, structures, floors, and its surrounding environment can be modeled without any missing part contrary to the same method that relies only to terrestrial images. From the accuracy analysis it has been shown that the Root Mean Square Errors (RMSEs) are RMSE$_x$=0.041 m; RMSE$_y$=0.031 m; RMSE$_z$=0.049 m which represent 0.071 m displacement in 3D space. With this accuracy, this method can be used to map the site up to 1:237 scale. The use of natural target in this research can reach centimeter-grade accuracy. Finally it can be stated that UAV images are suitable for complete 3D documentation of cultural heritage site.

Acknowledgement
Authors would like thank to Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Institut Teknologi Sepuluh November for supporting the research funding, and also for Balai Pelestarian Cagar Budaya Jawa Timur, Mojokerto who has permitted field data collection at research area.

References
[1] Alshawabkeh Integration of laser scanning and photogrammetry for heritage documentation [dissertation on the internet]. [Stuttgart (Germany)]: Universität Stuttgart; 2006 [cited 2016 Sep 10]. Available from: http://www.ifp.uni-stuttgart.de/publications/dissertationen/phd_thesis_alshawabkeh.pdf
[2] Perpustakaan Nasional Republik Indonesia. Candi Singasari [Internet]. Jakarta: Perpustakaan Nasional Republik Indonesia; [cited 2016 Sep 10]. Available from: http://candi.perpusnas.go.id/temples/deskripsi-jawa_timur-candi_singasari

[3] Skarlatos D and Kiparissi S 2012 Comparison of laser scanning, photogrammetry, and SfM-MVS pipeline applied in structures and artificial surfaces ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I3

[4] Westoby M J, Brasington J, Glasser N F, Hambrey M J and Reynolds J M 2012 Structure from Motion photogrammetry: a low cost, effective tool for geoscience applications Geomorphology 179 300-14

[5] Hidayat H and Cahyono A B 2015 3D Reconstruction of Singosari Temple Using Terrestrial Images and Structure from Motion Algorithm Proc. of Int. Conf. of Indonesian Society for Remote Sensing (ICOIRS) (Surabaya: Department of Geomatics Engineering)

[6] Püschel H, Sauerbier M, Eisenbeiss H 2008 A 3D model of Castle Landenberg (CH) from combined photogrammetric processing of terrestrial UAV-based images The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences B6b 37

[7] Mozas-Calvache A T, Perez-Garcia J L, Cardenal-Escarcena F J, Mata-Castro E and Delgado-Garcia J 2012 Method for photogrammetric surveying of archaeological sites with light aerial platforms Journal of Archaeological Science Issue 2 39 521-30

[8] Lo Brutto M, Garraffa A, Meli P 2014 UAV platforms for cultural heritage: first results ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II 5

[9] Seitz C and Altenbach H 2011 Project Archeye-the quadcopter as the archaeologist’s eye The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38-1 C22

[10] Balai Pelestarian Cagar Budaya Jawa Timur. Candi Singosari [Internet]. Mojokerto: Balai Pelestarian Cagar Budaya Jawa Timur; [cited 2015 Mar 15]. Available from: http://kebudayaan.kemdikbud.go.id/bpcbtrowulan/2015/01/23/candi-singosari-2

[11] Mancini F, Dubbini M, Gattelli M, Stecchi F, Fabbri S and Gabbianelli G 2013 Using Unmanned Aerial Vehicles (UAV) for high resolution reconstruction of topography: the Structure from Motion approach on coastal environments Remote Sensing 5 6880-98

[12] Wolf P R and Dewitt B A 2004 Elements of Photogrammetry: with Application in GIS (Columbus: McGraw Hill)

[13] Bakosurtanal. SNI 6502.2:2010 - Spesifikasi Penyajian Peta Rupa Bumi – Bagian 2: Skala 1: 25000 [Internet]. 2010 [cited 2016 Sep 10]. Available from http://www.bakosurtanal.go.id/assets/download/sni/SNI/18.%20SNI%206502.22010%20Spesifikasi%20penyajian%20peta%20rupa%20bumi%2025.000.pdf