

# Research on Three Dimensional Reconstruction of the Ancient Building Based on Images

Yingfeng Hu

School of Railway Tracks and Transportation,  
East China Jiaotong University, Nanchang, China, 330013  
huyingfeng86@sina.com

**Abstract.** Three dimensional reconstruction of ancient building using the theory of computer vision is one of the main ways to protect, promote and disseminate the value of ancient building. Due to there exist some shortcomings in the current digitization process of ancient building, such as large-scale scene reconstruction algorithm is complex and the accuracy is low, we present a three dimensional reconstruction approach to point cloud of ancient building based on multiple view images. The proposed approach uses an ordinary digital camera to capture multiple images of the ancient building to be reconstructed, takes advantage of affine decomposition principle to calculate the depth information of SIFT feature point on the ancient building with no damage and convenient operation. Through three dimensional reconstructing of the shape of the Tengwang Pavilion lying in Nanchang city, one of the three famous pavilions in the south of Yangtze River in China, to verify the validity of the proposed approach.

**Keywords:** ancient building, three dimensional reconstruction, multiple view images, point cloud data, affine decomposition.

## 1 Introduction

China is an ancient country with long history of civilization, and has vast territory and abundant resources. Our forefathers carried down a rich heritage of ancient buildings to us, the ancient buildings have a long history of tradition and great achievement. However, different from the ancient stone buildings of the European and Arabia countries, the ancient buildings of China are mostly wooden structure, that supplemented with painted and sculpture, and usually are very complex. Wood characteristics of China's ancient buildings make it faces huge challenges on its protection. In history, many ancient buildings in China were destroyed in the war and natural disaster. For example, the Old Summer Place, resulting in long time labor and wisdom by ancient people and managing carefully by 6 dynasties emperors for more than 150 years in China, was destroyed seriously by the allied forces in 1860, and the Tengwang Pavilion that lies in Nanchang City Jiangxi Province, ranking first among the three famous pavilions in the south of Yangtze River, was burned down in the war in 1926, and so on. In recent years, people pay attention to protect the ancient buildings by utilizing the

theory of machine vision to digitize the ancient buildings with the development of economy, and the auxiliary technologies related to the protection of the ancient buildings have become a hot issue. Digitizing the ancient building of China is a powerful method in the protection, promotion, and dissemination the value and the design concept of the ancient buildings. Thus, three dimensional digitization preservation of ancient buildings of China will have an important impact in terms of design and innovation.

## 2 Related Work

As a kind of special buildings, the approach of three dimensional reconstruction of the ancient buildings can learn from the approach of three dimensional reconstruction of buildings in general. Zuxun Zhang etc. introduces several representative three dimensional reconstruction approaches, as well as the basic principle of the approaches, for ordinary buildings from the aspects of geometry modeling and texture mapping [1]. And Jia Hu etc. proposed three dimensional reconstruction approach for ordinary building based on graph understanding [2].

Different from buildings in general, the ancient buildings usually have complex structure and rich texture, which contains a lot of sculpture and painting. As far as the ancient building is concerned, Fei Deng etc. take advantage of the auxiliary equipment, the large three dimensional laser scanner, to reconstruct the ancient building [3]. But the large three dimensional laser scanner is not easy to install and not easy to calibrate the cameras, its operation is limited for huge ancient buildings. According to the basic elements of the ancient buildings, Ru Wang presents the geometric approach that studied the basic elements from different viewpoints of a single image [4]. Zhang Feng etc. present the automatic three dimensional reconstruction approach to distant scenes from wide baseline images [5].

The three dimensional reconstruction of the ancient buildings is also a kind of the reconstruction of big scene, the approach to three dimensional reconstruction of big scene have important reference meaning for it. As far as the three dimensional reconstruction of big scene is concerned, Carlo Tomasi and Takeo Kanade develop a factorization approach to reconstruct scene from a stream of images based on orthography projection camera model [6]; Noah Snavely etc. represent structure-from-motion and image-based rendering algorithms for image sequence [7]; Pollefeys M etc. present a system for automatic three dimensional reconstruction from video of scenes captured by handheld video camera; and Changchang Wu introduces a fast three-dimensional reconstruction approach from motion images that required only  $O(n)$  time complexity [9].

Inspired by the approaches of three dimensional reconstruction of big scene, we propose a three dimensional reconstruction approach to point cloud of the ancient building based on multiple view images. Firstly, many images are captured from multiple views by using an ordinary Single Lens Reflex (SLR) camera. Secondary, the common SIFT features on the images of the ancient buildings are searched and matched. Finally, the depth information of SIFT feature points on the ancient building are calculated that used the affine factorization principle, by those step we can obtain the point cloud data of the ancient building.

### 3 Approach to Three Dimensional Reconstruction of the Ancient Building

#### 3.1 The SIFT Feature Points Matching

When using a regular SLR camera to take pictures of the ancient building, it is vulnerable to affect by some negative factors that lead to match the feature pixel points on multiple images difficultly, such as external ambient light and lens distortion etc. We use the SIFT algorithm to match the feature pixel points.

SIFT is a similarity invariant, it stays the same when images scale and rotate. When the two corresponding digital images of the ancient buildings are transforming, rotating, or affine transforming etc., SIFT feature matching algorithm can match the strong texture feature points accurately. Even for the images that captured under random angles, it can work stably. In order to distinguish with the general feature points, it is noted that David G. Lowe called the feature points as key points, which maintain the scale and orientation constant. The detailed steps of SIFT algorithm are as follows [10, 11]:

- (1) Build scale-space and detect extrema: Through exploiting a difference-of-Gaussian function, the potential key points are effectively identified which are invariant to scale and orientation. This is the first stage of computation searches over all scales and image locations.
- (2) Locate key points: For each candidate key points, their location and scale are determined by simulating 3D quadratic function. At the same time, the nonstable bordering key points are deleted.
- (3) Set the direction vector of key points: Based on local image properties, each key point is assigned a consistent orientation, then the key point descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.
- (4) Generate the local key point descriptor: By computing the gradient magnitude and orientation at each image sample point in a region, such as 4x4 or 2x2 window, around the key point location, the key point descriptor is created, namely the feature vector of SIFT algorithm.

The Euclidean distances between the feature vectors are exploited to measure the similarity of key points in different digital images after the SIFT feature vectors of the key points are created. A feature point on one image is randomly chosen, and then another two feature points are found, having the shortest and sub shortest Euclidean distances, by traversing all the feature points in another corresponding images. Among these two feature points, if the divisor between the shortest and sub shortest Euclidean distances is less than a threshold value, then they are judged to be paired homologous feature points.

#### 3.2 The Point Cloud Reconstruction Based on Affine Factorization

According to the theory of computer vision, the general affine camera model is used when mapping the three dimensional space point  $P$  with coordinate values  $(x, y, z)$  to two dimensional pixel point  $p$  with coordinate values  $(u, v)$ . The relationship between the space point  $P(x, y, z)$  and the pixel point  $p(u, v)$  is as follows:

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

In equation (1), the parameter  $s$  is scale factor, and the matrix at right is projection matrix, denoted as  $M$ , the elements  $m_{ij}$  ( $i = 1, 2, 3; j=1, 2, 3, 4$ ) in the projection matrix  $M$  is the inside and outside parameters of the camera. If the image coordinate and the spatial coordinate are transformed to generalized coordinates, equation (1) can be converted to equation (2):

$$s p' = M P' \quad (2)$$

We get  $m$  vision images from different perspective for the ancient building, each image has  $n$  SIFT feature points, and those feature points have been matched through the SIFT algorithm, namely give a three dimensional space point  $P$  with unknown coordinates values through different projection matrices project to  $m$  pixel points on  $m$  images. Fig.1 gives a toy example that shows the space points  $P$  projects to three matched pixel points  $p_1, p_2$ , and  $p_3$  on three images.

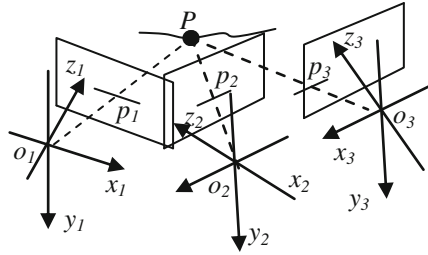


Fig. 1. A toy example of the space point projects to three images

The  $n$  space points with unknown coordinates values are denoted as  $P_i$  ( $i = 1, 2, \dots, n$ ), the  $j$ -th pixel point mapped to  $P_i$  is denoted as  $p_{ij}$  ( $j = 1, 2, \dots, m$ ), then we can have

$$s_{ij} p'_{ij} = M_j P'_i \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (3)$$

For all the space point  $P_i$  ( $i = 1, 2, \dots, n$ ) and their corresponding pixel points  $p_{ij}$  ( $j = 1, 2, \dots, m$ ), the following simultaneous equations is set up:

$$\begin{bmatrix} s_{11}P'_{11} & s_{12}P'_{12} & \dots & s_{1n}P'_{1n} \\ s_{21}P'_{21} & s_{22}P'_{22} & \dots & s_{2n}P'_{2n} \\ \dots & \dots & \dots & \dots \\ s_{m1}P'_{m1} & s_{m2}P'_{m2} & \dots & s_{mn}P'_{mn} \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \\ \dots \\ M_m \end{bmatrix} \begin{bmatrix} P'_1 & P'_2 & \dots & P'_n \end{bmatrix} \quad (4)$$

The left side of the equations (4) is a  $3m \times n$  order matrix, called measurement matrix, and it can be broken down into  $3m \times 4$  order movement of the camera matrix

and  $4 \times n$  order generalized coordinate matrix on the left side of the equation. Due to the rank of measurement matrix is equal to 4, we can calculate the unknown coordinates values of the space points by singular value decomposition and others matrix operations.

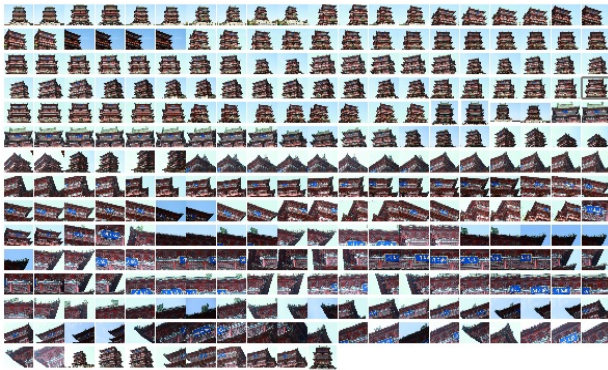
The point cloud obtained by the above approach may be a local view of the ancient building, therefore, in order to get the whole outline point cloud of the ancient building, we must repeat the above operations to get the point cloud of the other local view, and merge them together.

## 4 Experimental Results

To test the validity of the proposed approach, we reconstruct the shape of the Tengwang Pavilion lying in Nanchang city Jiangxi Province in China. The Tengwang Pavilion is one of the three famous pavilions in the south of Yangtze River, it is made of wood and is like the building style of the Song Dynasty. In the experiment, the front shape of the building has been reconstructed.

We use the ordinary SLR camera, Canon EOS 70D, with standard zoom lens Canon EF 24-105mm f/4L IS USM, and collect 291 images of the Tengwang Pavilion from different perspectives. In order to reduce noise that has nothing to do with the topic, we try to avoid the surrounding trees and tourists. Before processing, the original images scale are reduced to  $1800 \times 1200$  pixels, Fig.2 gives a brief view of all the collecting images.

Fig.3 gives the point cloud of the Tengwang Pavilion after three dimensional reconstruction. And Fig.4 gives the rendered model of the point cloud used the open source package, MeshLab1. Because there are many tourists wandered around at the bottom of the Tengwang Pavilion during taking pictures, the point cloud seems to be incomplete at this part. However, in addition to this part, the data remains intact.



**Fig. 2.** The 291 images captured from different perspectives of the Tengwang Pavilion

---

<sup>1</sup> <http://meshlab.sourceforge.net/>

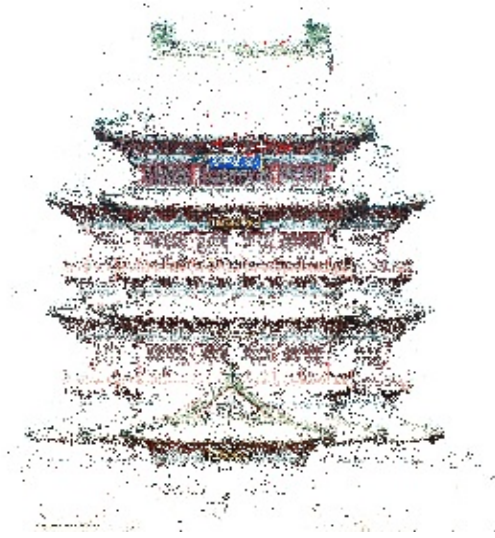


Fig. 3. The point cloud of the Tengwang Pavilion

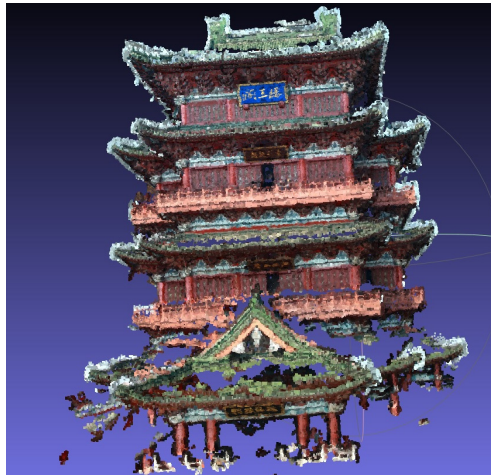


Fig. 4. The rendered model of the Tengwang Pavilion

## 5 Conclusion

In conclusion, a three dimensional reconstruction approach for ancient building based on multiple view images is proposed, and the results of reconstruction the front shape of the Tengwang Pavilion prove that the approach can be used to digitalize the ancient building, which will provide the technical support for the digital preservation of the ancient buildings.

**Acknowledgment.** This research has been funded by the Natural Science Foundation of Jiangxi Educational Committee of China under Grant No. GJJ13316.

## References

- [1] Zhang, Z., Zhang, J.: 3D Reconstruction and Visualization of House in Creating Digital City. In: Proceedings of ASIA GIS 2001- Collaboration Through GIS in the Internet Era, pp. 20–22 (2001)
- [2] Hu, J., Yang, R.-Y., Cao, Y., et al.: 3D Reconstruction Technology for Architectural Structure Based on Graphics Understanding. *Journal of Software* 13(9), 1873–1880 (2002) (in Chinese)
- [3] Deng, F., Zhang, Z., Zhang, J.: 3D reconstruction of old architecture by laser scanner and digital camera. *Science of Surveying and Mapping* 32(2), 29–30 (2007)
- [4] Wang, R.: Research on the key technique of digitalization of ancient building and 3D modeling. Thesis of Northwestern University (2010)
- [5] Zhang, F., Xu, Z., Shi, L., et al.: Automatic Reconstruction of Distant Scenes from Wide Baseline Images. *Journal of Computer-Aided Design & Computer Graphics* 22(2), 256–263 (2010) (in Chinese)
- [6] Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision* 9(2), 137–154 (1992)
- [7] Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the World from Internet Photo Collections. *International Journal of Computer Vision* (2007)
- [8] Pollefeys, M., Nistér, D., Frahm, J.M.: Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision* 78(23), 143–167 (2008)
- [9] Wu, C.: Towards Linear-time Incremental Structure from Motion. In: 3DV (2013)
- [10] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
- [11] Hu, Y.: Research on a three-dimensional reconstruction method based on the feature matching algorithm of a scale-invariant feature transform. *Mathematical and Computer Modelling* 54, 919–923 (2011)