

THREE DIMENSIONAL RECONSTRUCTION OF LARGE CULTURAL HERITAGE OBJECTS BASED ON UAV VIDEO AND TLS DATA

Z. Xu^{a,b}, T. H. Wu^c, Y. Shen^d, L. Wu^{e1}

^aCollege of Geoscience and Surveying Engineering, China University of Mining & Technology, Beijing 100083, China -

^bKey Laboratory of Environmental Change & Natural Disaster of MOE, Beijing Normal University, Beijing 100875, China -
hzxurs@mail.bnu.edu.cn

^cSatellite Surveying and Mapping Application Center, NASG, Beijing, 101300, China -
wupengtianhao92@126.com

^dCollage of Information Engineering, China University of Geosciences, Wuhan 430074, China-
yonglinshen@gmail.com

^eSchool of Environment Science & Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China-
awulixin@263.net

Special Session, CIPA

KEY WORDS: 3D reconstruction, cultural heritage, video image sequences, terrestrial laser scanning, unmanned aerial vehicle

ABSTRACT:

This paper investigates the synergetic use of unmanned aerial vehicle (UAV) and terrestrial laser scanner (TLS) in 3D reconstruction of cultural heritage objects. Rather than capturing still images, the UAV that equips a consumer digital camera is used to collect dynamic videos to overcome its limited endurance capacity. Then, a set of 3D point-cloud is generated from video image sequences using the automated structure-from-motion (SfM) and patch-based multi-view stereo (PMVS) methods. The TLS is used to collect the information that beyond the reachability of UAV imaging e.g., partial building facades. A coarse to fine method is introduced to integrate the two sets of point clouds UAV image-reconstruction and TLS scanning for completed 3D reconstruction. For increased reliability, a variant of ICP algorithm is introduced using local terrain invariant regions in the combined designation. The experimental study is conducted in the Tulou culture heritage building in Fujian province, China, which is focused on one of the TuLou clusters built several hundred years ago. Results show a digital 3D model of the Tulou cluster with complete coverage and textural information. This paper demonstrates the usability of the proposed method for efficient 3D reconstruction of heritage object based on UAV video and TLS data.

1. INTRODUCTION

3D reconstruction of culture heritage objects is increasingly required for different purposes, e.g., preservation (Remondino, 2011), reconstruction (El-Hakim et al., 2004), and as-built documentation (Yastikli, 2007). There are several instruments can be used to construct high-quality 3D point clouds. Among others, terrestrial laser scanner (TLS) systems and unmanned aerial vehicles (UAV) have been increasingly used for 3D reconstruction of heritage objects to date. Good examples can be found in literatures, where 3D reconstructions of culture heritage objects were obtained from TLS data and UAV photogrammetry (Remondino, 2011). With the fact that no single sensor can acquire complete information by performing one or several multi-surveys to cultural object reconstruction, some studies have been done by integrating these two platforms for 3D reconstruction of large and complex scenes (Remondino and Rizzi, 2010; Xu et al., 2014). In addition, the rapid development of advanced image-based reconstruction techniques enables to obtain 3D reconstructions from both images and video frames. As a result, it brings diverse combinations of multi-scale data for the reconstruction. Nevertheless, few studies have considered the viewing differences between UAV imaging and TLS scanning in the combined designation. To solve that, this work introduces a coarse-to-fine registration method using terrain invariant

regions for combining multi-source point clouds. With that, this paper provides a pipeline for the design of efficient 3D reconstruction of culture heritage objects based on UAV video and TLS data

2. STUDY AREA AND DATA ACQUISITION

2.1 Study Area

The methodology was applied to the area in Fujian province of China. More than 3000 Tulou clusters built several hundred years ago are located in the study area, which were viewed as great cultural heritage. In this study, we focused on the Zhencheng TuLou (Figure 1), that built in 1912, which was considered as the best preserved and most typical one. It was well designed with a combination of two concentric rings in the center and two allocated units aside. It contained four floors and covered the area of around 5000 m².

2.2 UAV Data Acquisition

An octocopter named BNU D8-1 (Xu et al., 2014) was employed for aerial video shooting. It worked at an average height of 80 m above the ground with manually control. A digital video camera Sony PJ 790 was used to perform video shooting over the Zhencheng Tulou. In total of 234 frames were continuously extracted from the video, with size of 1920×1080 pixels

¹ Corresponding author



Figure 1 Example video frame of the study area in Longyan, Fujian Province, China

2.3 TLS Data Acquisition

In this case study, the close range of TLS is too short-range to measure a huge building. Thus, we used the long-rang scanner Riegl VZ-4000 to capture the point cloud of the Zhencheng Tulou. The building exterior facade information can be captured by the scanner, while the roof information which can be complemented by UAV data is beyond Riegl VZ-4000's reachability. In total of seven stations were captured around the Tulou. The multi-scan registration procedure was performed using the RiSCAN PRO 1.1.7 software.

3. METHODOLOGY

This work aimed at providing a practical solution for efficient 3D reconstruction of culture heritage objects based on both UAV video and TLS data. The viewing differences between image shooting and TLS scanning were considered in particular. Figure 2 shows the framework of the proposed methodology. The procedures contain three components: 1) 3D point cloud generation from UAV video image sequences, 2) coarse-to-fine registration of UAV-reconstructed point clouds and TLS data using invariant regions, and 3) textural model of the culture heritage object.

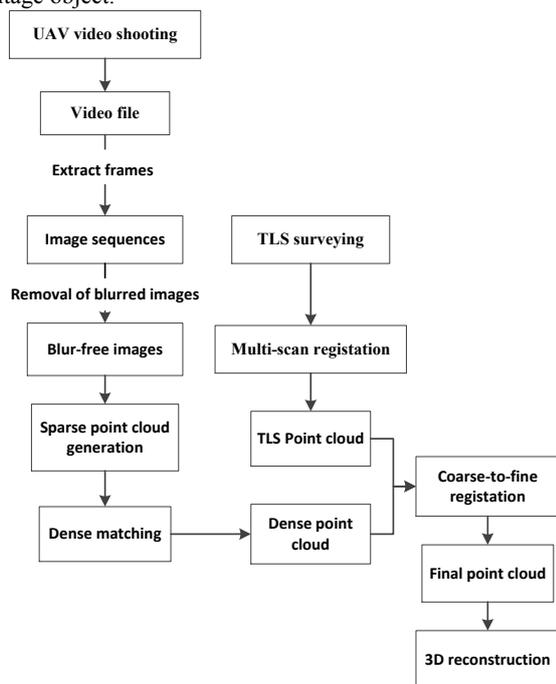


Figure 2 Framework of the proposed method

3.1 3D point cloud generation from UAV Video frames

This step started from extracting frames from videos, resulting in a set of highly overlapped image sequences. Following this, the removal of blurred images was performed to ensure the remaining images with high quality. Considering its efficiency and fast implementation, the method developed by Crete et al. (2007) was used to achieve the goal. The automated structure-from-motion (SfM) (Snavely, 2009) method was used to process the remaining set of blur-free images to generate a sparse point cloud. Further, the Patch-based Multi-view Stereo (PMVS2) (Furukawa and Ponce, 2010) was used to increase the density of SfM point cloud. Considering its fully automated capability the VisualSFM software package (Wu, 2013) that integrates both SfM and PMVS2 techniques was used. The final outputs include the coordinates of a set of dense points and corresponding normal as well as RGB color.

3.2 Coarse-to-Fine Registration of Multi-Source Point Clouds Using Invariant Regions

The aim of this step is to estimate the Euclidean motion between the two point clouds. To achieve this, we treated the one from TLS measurement as base and oriented the other one using a coarse-to-fine method. A group of common features that manually selected from the two sets of point clouds is used for estimating the transformation parameters in order to coarse registration. The well-known Bursa model was used in this procedure, as shown in Formula (1).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = kR \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T \quad (1)$$

where (x, y, z) and (x_0, y_0, z_0) represent the points in the TLS and image-reconstructed point clouds.

From a general point of view, the Iterative Closest Point (ICP) (Besl and McKay, 1992) or its variants can be used to fine-tune the Euclidean motion in the following sequences. Nevertheless, as the point clouds yielded using UAV image reconstruction and TLS measurement have different densities and with different views, directly performing the ICP algorithm may result in local minimization. Hence, rather than using the whole set of points for the estimation, we used only the potential invariant regions. In detail, the terrain points were viewed as invariant ones between the two clouds. We then used the estimated parameters to perform Euclidean transformation for the remaining set of points. For increased speed, we used the Iterative Closest Compatible Point (Bae and Lichti, 2008) algorithm for the fine registration, where the Euclidean motion between two point clouds is iteratively computed by minimizing the cost function C with a least square method, see Formula (2).

$$c = \min_{\substack{\mathbf{n}-\mathbf{n}_0 \in [-a, a] \\ \mathbf{n} \in \text{ann}(d)}} \left\| \sum_{i=1}^{m \in \text{ann}(d)} (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \right\| \quad (2)$$

where (x, y, z) and (x_0, y_0, z_0) represent the points from the two sets, \mathbf{n} and \mathbf{n}_0 represent the corresponding normal within the search radius d , $[-a, a]$ is the threshold of normal constraint.

3.3 Textural Model of the Culture Heritage Object

Note that the digitization was not the main focus of this study. Here, we followed the traditional methods on digital reconstruction. With the complete point cloud from both UAV

image reconstruction and TLS measurement, further steps in terms of digitalization, reconstruction, and texturing are implemented to obtain a high-resolution photorealistic 3D reconstruction.

4. RESULTS AND DISCUSSIONS

Figure 3 presents the reconstructed point clouds from UAV video image sequences. The results from the full set of 234 image sequences and from only 115 of blur-free images were respectively shown in Figure 3a and 3bb. Note that, the point cloud from blur-free images achieved more complete information than that from the full-set of image sequences and on the roof of the Tulou in particular. Moreover, with the fact that the complexity of SfM algorithm is $O(n^2)$, the time cost of using only 115 of blur-free images was dropped to a quarter of that yielded using the full set of 234 images. The experimental results demonstrated that the completeness of image-reconstructed point clouds depends rather than the number of images involved in but on the image qualities. This also confirmed the knowledge in literature (Alsadik et al, 2015). Hence, for increased speed, the removal of redundant images can make some sense (Alsadik et al, 2013; Alsadik et al, 2015). Furthermore, from our experimental experiences, decreasing the number of matched image pairs also contributes to saving time in image reconstructing process while without in loss the quality of 3D scene geometry (Xu et al, 2014). From image reconstruction point of view, the combination of decreasing both redundant images and matched image pairs are good issues for real-world applications.

Figure 4 shows the 3D reconstruction of the Zhencheng Tulou with complete and textural information. Although the digitization was out the scope of this paper, the experimental results demonstrated that the integrated points from UAV image reconstruction and TLS measurement met the demand of digitization and texturing. However, density differences should be considered in the integration procedure. Generally, the points yielded using only SfM reconstruction possess too sparse density to ensure enough common features for the reliable registration. Hence, further process should be considered to improve the density of SfM outputs in order to obtain a point cloud with comparable density to that of TLS measurement



Figure 3. 3D point clouds from UAV full-set video image sequences



Figure 4 3D point clouds from UAV blur-free video image sequences

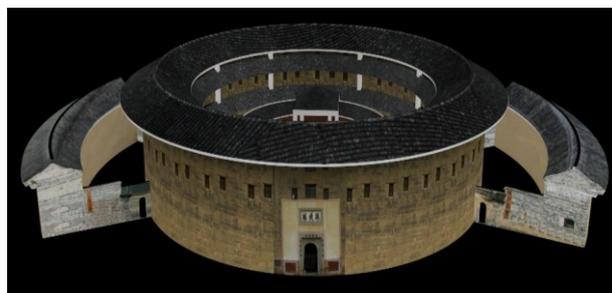


Figure 4. 3D reconstruction of Zhengyang Tulou with integrated point clouds from UAV photogrammetry and TLS data

5. CONCLUSIONS

This paper gave a framework for the design of 3D reconstruction of cultural heritage objects. The attention was focused on 3D reconstruction of huge or complex objects which possesses constraints for data acquisition with any single sensor. A novel registration method was introduced using invariant regions in a coarse-to-fine configuration. UAV video frames were used for image 3D reconstruction, where around a half of blurred images were removed from the full image set. The remaining set of images achieved more complete information than that of using the full image set. Moreover, the time cost of using the blurred-free images was decreased around 3 times that yielded using the full image set. Complete 3D model was obtained from the integrated points yielded using both UAV image-reconstruction and TLS measurement. Further work can investigate the use of invariant features for rapid and automatic feature matching that enable efficient point clouds registration.

ACKNOWLEDGEMENTS (OPTIONAL)

This work is partially supported by grants from the National Basic Research Program of China under Grant (grant # 2011CB707102), Fundamental Research Funds for Central University (grant # 105565GK) and China Scholarship Council (CSC).

REFERENCES

- Alsadik, B., Gerke, M., Vosselman, G., 2013. Automated camera network design for 3D modeling of cultural heritage objects. *Journal of Cultural Heritage* 14(6), 515-526.
- Alsadik, B., Gerke, M., Vosselman, G., 2015. Efficient use of video for 3D modelling of cultural heritage objects. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, 1-8.
- Bae, K.-H., Lichti, D.D., 2008. A method for automated registration of unorganised point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (1), 36-54.
- Besl, P.J., McKay N.D. Method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 1992, 14, 239-256.
- Crete, F., Dolmiere, T., Ladret, P., Nicolas, M., Year. The blur effect: perception and estimation with a new no-reference perceptual blur metric, *Proc. Electronic Imaging 2007*, pp. 649201-649201-64911.
- El-Hakim, S.F., Beraldin, J.A., Picard, M., Godin, G., 2004. Detailed 3D reconstruction of large-scale heritage sites with integrated techniques. *Computer Graphics and Applications*, *IEEE* 24 (3), 21-29.
- Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* 32 (8), 1362-1376.
- Remondino, F., 2011. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sensing* 3 (6), 1104-1138.
- Remondino, F., Rizzi, A., 2010. Reality-based 3D documentation of natural and cultural heritage sites—techniques, problems, and examples. *Applied Geomatics* 2 (3), 85-100.
- Snavely, K.N., 2009. Scene reconstruction and visualization from internet photo collections.
- Wu, C., Year. Towards linear-time incremental structure from motion, *Proc. 3D Vision-3DV 2013*, 2013 International Conference on, pp. 127-134.
- Xu, Z., Wu, L., Shen, Y., Li, F., Wang, Q., Wang, R., 2014. Tridimensional Reconstruction Applied to Cultural Heritage with the Use of Camera-Equipped UAV and Terrestrial Laser Scanner. *Remote Sensing* 6 (11), 10413-10434.
- Yastikli, N., 2007. Documentation of cultural heritage using digital photogrammetry and laser scanning. *Journal of Cultural Heritage* 8 (4), 423-427.