AUTOMATED 3D ARCHITECTURE RECONSTRUCTION FROM PHOTOGRAMETRIC STRUCTURE AND MOTION: A CASE STUDY OF THE "ONE PILLA" PAGODA, HANOI, VIETNAM

T. To [a,*], D. Nguyen [b], G.Tran [c]

[a] Geology, Space Technology Institute Vietnam Academy of Science and Technology, Vietnam – tttu@sti.vast.vn
[b] Department of Photogrammetry and Remote Sensing, Hanoi University of Mining and Geology, Vietnam – nguyenbaduy@humg.edu.vn
[c] Department of Cartography, Hanoi University of Mining and Geology, Vietnam – tranthihuonggiang@humg.edu.vn

KEY WORDS: 3D modelling, SFM, Culture Heritage

ABSTRACT:

Heritage system of Vietnam has decline because of poor-conventional condition. For sustainable development, it is required a firmly control, space planning organization, and reasonable investment. Moreover, in the field of Cultural Heritage, the use of automated photogrammetric systems, based on Structure from Motion techniques (SfM), is widely used. With the potential of high-resolution, low-cost, large field of view, easiness, rapidity and completeness, the derivation of 3D metric information from Structure-and-Motion images is receiving great attention. In addition, heritage objects in form of 3D physical models are recorded not only for documentation issues, but also for historical interpretation, restoration, cultural and educational purposes. The study suggests the archaeological documentation of the "One Pilla" pagoda placed in Hanoi capital, Vietnam. The data acquired through digital camera Cannon EOS 550D, CMOS APS-C sensor 22.3 × 14.9 mm. Camera calibration and orientation were carried out by VisualSFM, CMPMVS (Multi-View Reconstruction) and SURE (Photogrammetric Surface Reconstruction from Imagery) software. The final result represents a scaled 3D model of the One Pilla Pagoda and displayed different views in MeshLab software.

1. INTRODUCTION

Reconstructing 3D models of real-world is an exciting topic with a large variety of possible applications. In recent years, reconstruction has gained a lot of attention due to new capturing methods. Photogrammetric tools enable an intuitive and cost-efficient way of capturing scenes as point clouds. 3D reconstruction is an important tool for maintaining cultural heritage. In some cases, where physical monuments are endangered, digital preservation might be the only possibility. Having 3D models of important cultural heritage allows the presentation to more people than would be possible at the original site. Furthermore, non-invasive reconstruction methods enable the detection of small details which would not be possible otherwise. Having accurate 3D models of buildings provides advantages for many applications. 3D information is a valuable input for planning modifications of buildings, or for planning the placement of furniture in interior rooms. 3D models provide information for security applications, e.g. fall detection of elderly people. Digital map services, such as Google maps or Bing maps, have gained great success with including 3D models into their virtual city maps.

The large variety of applications produces different requirements on reconstruction systems. For some reconstructions it is necessary to have very exact measurements. Other applications might tolerate approximations in order to generate complete models. The reconstruction techniques have to be easy-to-use in order to be accessible for non-experts and thus being used for a wide range of applications. This means, that the algorithms are either fully automatic or that they have an intuitive user interface (Reisner-kollmann, 2013).

As state-of-the-art geodetic measuring methods photogrammetric multi-image techniques and, increasingly, terrestrial laser scanning, as a standalone system or in combination with other methods, are used for precise 3D data acquisition of complex objects. Requirements for the generation of 3D models are often very high with respect to level-of-detail, completeness, reliability, accuracy (geometrical and visual quality), efficiency, data volume, costs and operational aspects, but the priority order depends upon the object to be recorded. However, in recent years, real alternatives to classical systems and methods are presented by the large number of digital cameras on the market, which can be efficiently and successfully used as passive low-cost sensors when combined with appropriate algorithms such as structure-from-motion (SfM) and/or dense image matching for different 3D applications (object reconstruction, navigation, mapping, tracking, recognition, gaming, etc.). Due to the very low costs and current approval for open-source methods such systems (sensors in combination with appropriate algorithms) are very popular in many application fields. Nevertheless, the metrological aspect should not be neglected, if these systems are to be acknowledged as serious measuring and modelling procedures. Therefore, clear statements about the accuracy potential and efficiency of such systems must be empirically investigated through appropriate testing. In this context the 3D modelling results must be also analysed and compared with respect to reference data (Lehtola, Kurkela, & Hyyppä, 2014).

* Corresponding author.

Moreover, unmanaged tourism development and the resultant degradation of the environment results on the damage of Vietnam culture heritage sites. The challenge for tourism development at Vietnam is how to retain or revitalise the richness, complexity and creativity of the traditional heritage and how to maintain cultural authenticity and communicate an appreciation of this to the visitor – both foreign and domestic-privileged enough to experience it. Recently, with the development of 3D construction, low-cost automated photogrammetric system is used widely in the field of Cultural Heritage. In particular, it is applied for the study and for the documentation of the ancient ruins.

2. THE STUDY SITE

The One Pillar Pagoda, a historic Buddhist temple, is located in Hanoi, the capital of Vietnam. The temple was built by Emperor Ly Thai Tong, who ruled from 1028 to 1054. One Pillar Pagoda is not spectacular but distinctive and unique in term of architecture (Figure 1).

One Pillar pagoda is also called "Lien Hoa Dai" because it looks like a lotus flower emerging from the water's surface. The pagoda was made of wood in square shape with each side of 3m and a curved roof on a single stone pillar 1.25 m in diameter. The stone pillar includes two blocks which are connected together skilfully and look like one. This stone pillar is approximately 4m high (excluding the underground section). The upper part of the pillar has 8 wooden petals which are similar to lotus petals. The pagoda's roof is decorated with dragons flanking the moon. In the pagoda, the bodhisattva Avalokiteshvara is placed on a wooden and red-lacquered Buddha's throne at the highest position. One Pillar pagoda contains a great value of religious culture that opposite to its small size to ensure the architecture of a lotus.

Figure 1. "One pilla" pogoda, Hanoi, Vietnam (modification from http://www.panoramio.com/photo/11899993)

Experiencing decades, One Pillar pagoda was repaired and restored many times in Tran, Post-Le, Nguyen dynasties. In 1954, French colonists used explosives to destroy the pagoda before withdrawing from Vietnam. In 1955, Ministry of Culture had One Pillar pagoda rebuilt by Nguyen Ba Lang architecture.

3. DATA COLLECTION AND APPLIED SOFTWARE

3.1 Data collection

The work is carried out through digital camera Cannon EOS 550D, CMOS APS-C sensor 22.3 × 14.9 mm, with a fixed focal length lens of 20 mm. In the specific case, a close-range block of 307 images has been used while the shooting has been carried out with a distance of from 10 meters to the monument and about 50 centimetres between an image and the other for both directions horizontal and vertical (Figure 2).

Figure 2. Figure placement and numbering

Our system requires accurate information about the relative location, orientation, and intrinsic parameters such as focal length for each photograph in a collection, as well as sparse 3D scene geometry. Some features of our system require the absolute locations of the cameras, in a geo-referenced coordinate frame. Some of this information can be provided with GPS devices and electronic com-passes, but the vast majority of existing photographs lack such information. Many digital cameras embed focal length and other in-formation in the EXIF tags of image les. These values are useful for initialization, but are sometimes inaccurate.

In our system, we do not rely on the camera or any other piece of equipment to provide us with location, orientation, or geometry. Instead, we compute this information from the images themselves using computer vision techniques. We first detect feature points in each image, then match feature points between pairs of images, and finally run an iterative, robust SfM procedure to recover the cam-era parameters. Because SfM only estimates the relative position of each camera, and we are also interested in absolute coordinates (e.g., latitude and longitude), we use an interactive technique to register the recovered cameras to an overhead map. Each of these steps is described in the following subsections.

3.2 Applied Software

For investigation of the automatic generation of 3D point clouds and 3D surface models from image data the following software

packages and/or web services were used: CMPMVS, VisualSFM (open-source software), SURE, and MeshLab.

3.2.1 CMPMVS

CMPMVS is an open source program that allows the input of a set of perspective images and markers, with camera positions and calibrations, and outputs a textured mesh of the rigid scene. CMPMVS works as a Structure from Motion (SfM) procedure for arbitrarily arranged imagery and was developed for Microsoft's Photo Tourism Project (Snavely et al., 2006). Feature extraction in the images is performed by the SIFT (Scale Invariant Feature Transform) algorithm from Lowe (2004). The software supports camera calibration data (focal length f from EXIF data, two radial distortion parameters k1 and k2), image orientations and a thin 3D point cloud (scene geometry) as results for any image blocks using a modified bundle block adjustment from Lourakis & Argyros (2004). The results of CMPMVS in order to generate a denser point cloud of non-moving objects by dense image matching. As well as the 3D coordinate each point additionally receives the colour value of the object taken from the images. As an example, after input of the images CMPMVS are executed automatically and the result is finally presented in MeshLab.

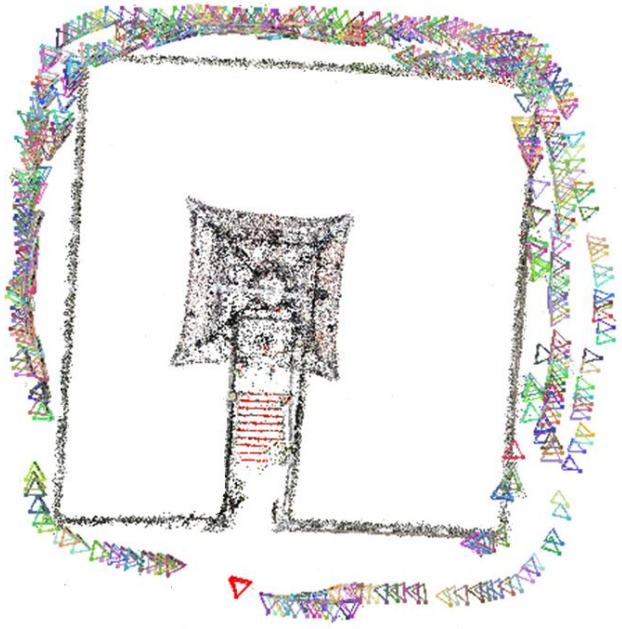

Figure 3. Figure placement and numbering from VisualSFM

3.2.2 VisualSFM

VisualSFM (Figure 3) is a GUI application for 3D reconstruction of objects from images using the SfM system, which was developed at the University of Washington (Wu, 2007). The software is a re-implementation of the SfM system of the Photo Tourism Project and it includes improvements by integrating both SIFT on the graphics processing unit (SiftGPU) and Multicore Bundle Adjustment (Wu, 2011). Dense reconstruction can also be performed through VisualSFM using PMVS/CMVS (Patch or Cluster based Multi View Stereo Software, Furukawa & Ponce, 2010). Changchang Wu is the author of this software. Basically, it is an upgraded version of his previous projects that is complemented by SiftGPU and Multicore Bundle Adjustment algorithms. Moreover, this software provides an interface to run tools like PMVS/CMVS, resp. it is able to prepare data for CMP-MVS software.

3.2.3 SURE

SURE is a software solution for multi-view stereo, which enables the derivation of dense point clouds from a given set of images and its orientations (Figure 4). Up to one 3D point per pixel is extracted, which enables a high resolution of details. It is based on LibTSgm (LibTSgm is a library for dense multi-view reconstruction), which implements the core functionality for image rectification, dense matching and multi-view triangulation, and provides a C/C++ API. SURE can handle image collected by various types of sensors and can be utilized for close range, UAV and large frame aerial datasets. It scales well to datasets with large image numbers and large image resolution (e.g. >200MP aerial imagery). Also imagery with high dynamic range (e.g. 16 Bit depth) can be used. Initial depth information of the scene is not required, which enables the use within fixed calibrated camera setups. The efficiency of processing is increased by utilizing parallel processing and hierarchical optimization. The input of SURE is a set of images and the corresponding interior and exterior orientations. This orientation can be derived either automatically (e.g. by Structure from Motion methods) or by using classical image orientation approaches. So far interfaces to common image orientation software, such as Bundler (.out), VisualSFM (.nvm), Trimble/INPHO Match-At (aerial imagery, .prj), are provided. The output of the software is point clouds or depth images. SURE is a multi-stereo solution. Therefore, single stereo models (using 2 images) are processed and subsequently fused. Within a first step images are undistorted and pair-wise rectified. Within a second step, suitable image pairs are selected and matched using a dense stereo method similar to the Semi Global Matching (SGM) algorithm (Rothermel et. al., 2013). In contrast to the original SGM method as published in (Hirschmüller, 2008) a more time and memory efficient solution is implemented. Within a triangulation step the disparity information from multiple images is fused and 3D points or depth images are computed. By exploiting the redundancy of multiple disparity estimations precision of single depth measurements is increased and blunders are efficiently filtered.

Figure 4. Point cloud from SURE software

3.2.4 MeshLab

MeshLab is an open-source, portable, and extensible system for the processing and editing of unstructured 3D triangular meshes developed at the Visual Computing Lab, which is an Institute of the National Research Council of Italy in Pisa (Cignoni et al., 2008).

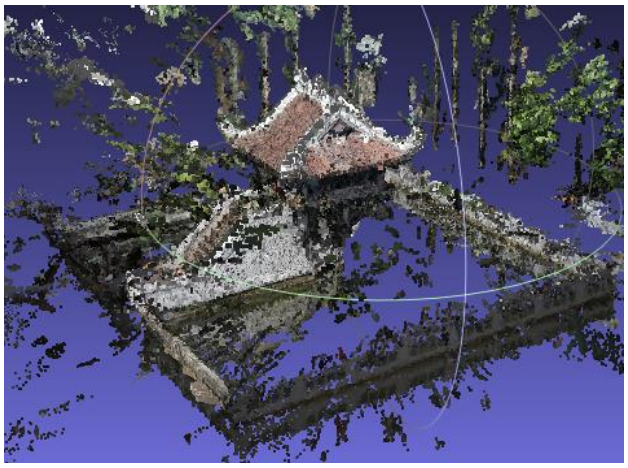

Figure 5. Point cloud from CMPMVS software

4. WORKFLOW

The general workflow for image-based 3D reconstruction using low-cost systems is illustrated in Figure 6. For photogrammetric object recording multiple photos are taken of the object from different positions, whereby coverage of common object parts should be available from at least three but preferably five photographs from different camera positions. After import of the images into the respective processing software the parameters for camera calibration (interior orientation) and (exterior) image orientations are automatically computed. The subsequent generation of 3D point clouds or 3D surface models is also carried out in full automatic mode. Only for the 3D transformation of the point cloud or the meshed model into a superordinate coordinate system must the user measure control points interactively.

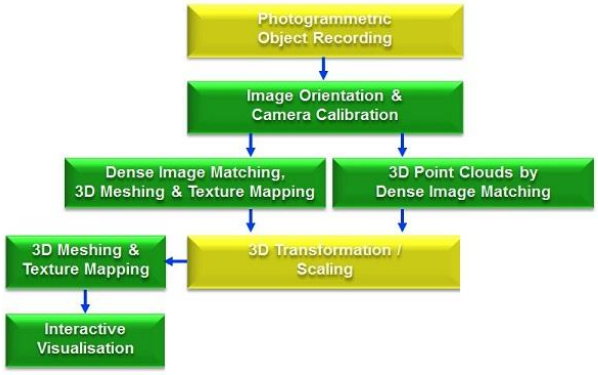

Figure 6. Workflow for image-based low-cost 3D object reconstruction procedures

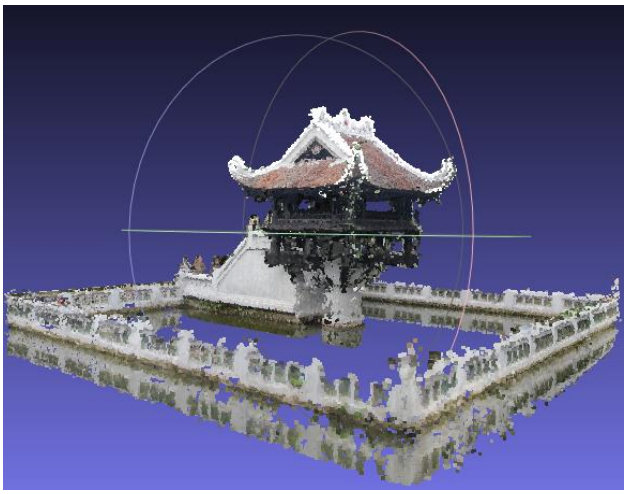

Figure 7. Point cloud processing with MeshLab software

The first step is to find feature points in each image. We use the SIFT key point detector (Lowe, 2004), because of its in variance to image transformations. A typical image contains several thousand SIFT key points. Other feature detectors could also potentially be used; several detectors are compared in the work of (Mikolajczyk, et al, 2005). In addition to the key point locations themselves, SIFT provides a local descriptor for each key point. Next, for each pair of images, we match key point descriptors between the pair, using the approximate nearest neighbours package of (Arya, et al, 1998), then robustly estimate a fundamental matrix for the pair using RANSAC (Fischler and Bolles, 1987). During each RANSAC iteration, we compute a candidate fundamental matrix using the eight-point algorithm (Hartley and Zisserman, 2004), followed by non-linear refinement. Finally, we remove matches that are outliers to the recovered fundamental matrix. If the number of remaining matches is less than twenty, we remove all of the matches from consideration. After finding a set of geometrically consistent matches between each image pair, we organize the matches into tracks, where a track is a connected set of matching key points across multiple images.

If a track contains more than one key point in the same image, it is deemed inconsistent. We keep consistent tracks containing at least two key points for the next phase of the reconstruction procedure.

5. RESULTS & CONCLUSION

In this section the results of the applied software packages CMPMVS, VisualSFM, SURE, and MeshLab, respectively, are presented for applications in “One Pilla” pagoda 3D model generation. The primary result is presented in Figure 8.

The construction of 3D models of the archaeological site of “One Pilla” pagoda (Hanoi, Vietnam) was discussed in this paper. Motivated by the requirement to generate a virtual reconstruction of this study site with sub-decimetre accuracy, a large number of images were taken. These images were taken from the ground and with the digital camera Cannon EOS 550D, CMOS APS-C sensor 22.3 × 14.9 mm. The images were processed in a SfM-based workflow and other open source software. A workflow was estimated, not only for 3D models generation but also is available in an interactive 3D viewer.

Further work is required to do the accuracy assessment for the result.

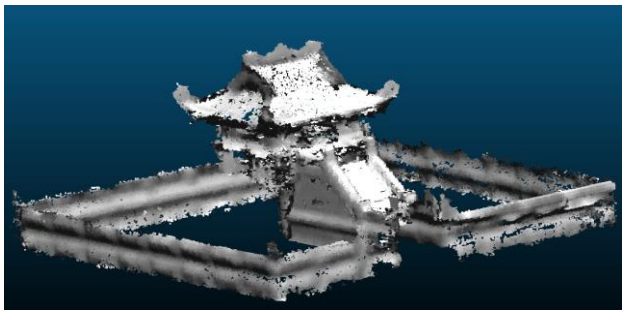

Figure 8. “One Pilla” pagoda 3D model result

ACKNOWLEDGEMENTS

The authors would like to thank to Space Technology Institute (STI) for providing us the best support for data collection and all the help to solved difficulties in the researching period. We are also very grateful to the NAFOSTED organization for supporting us to attend this conference.

REFERENCES

Barazzetti, L., Remondino, F., & Scaioni, M. (2009). Combined use of photogrammetric and computer vision techniques for fully automated and accurate 3D modeling of terrestrial objects: In *Proceedings of the Society of Photo-Optical Instrumentation Engineers* (Vol. 7447, p. 74470M–). doi:10.1117/12.825638

Bartelsen, J., Mayer, H., Hirschmüller, H., Kuhn, a., & Michelini, M. (2012). Orientation and Dense Reconstruction of Unordered Terrestrial and Aerial Wide Baseline Image Sets. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *I-3*(September), 25–30. doi:10.5194/isprsannals-I-3-25-2012

Kersten, T., Acevedo Pardo, C., & Lindstaedt, M. (2004). 3D adquisition, modelling and visualization of North German castles by digital architectural photogrammetry. In *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 126–132).

Kersten, T. P. (2006). Combination and Comparison of Digital Photogrammetry and Terrestrial Laser Scanning for the Generation of Virtual Models in Cultural Heritage Applications. *The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 207–214.

Lehtola, V., Kurkela, M., & Hyyppä, H. (2014). Automated image-base reconstruction of Building interiors: A case study. *The Photogrammetric Journal of Finland*, *24*(1), 1–13.

Reisner-kollmann, I. (2013). *Reconstruction of 3D Models from Images and Point Clouds with Shape Primitives*. Vienna University of Technology.

Remondino, F., & Menna, F. (2008). Image-based surface measurement for close-range heritage documentation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, *XXXVII*, 199–206.

Rothermel, M., Wenzel, K., Fritsch, D., Haala, N. (2012). SURE: Photogrammetric Surface Reconstruction from Imagery. *Proceedings LC3D Workshop*, Berlin ,December 2012

Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 328-341.

Wenzel, K., Rothermel, M., Fritsch, D., and Haala, N.: Image Acquisition and Model Selection for Multi-View Stereo, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*., XL-5/W1, 251-258, doi:10.5194/isprsarchives-XL-5-W1-251-2013, 2013.

Rothermel, M., Haala, N., 2011. Potential of Dense Matching for the Generation of High Quality Digital Elevation Models. *In ISPRS Proceedings* XXXVII 4-W19

Haala, N. & Rothermel, M., 2012. Dense Multi-Stereo Matching for High Quality Digital Elevation Models, in *PFG Photogrammetrie, Fernerkundung, Geoinformation*. Jahrgang 2012 Heft 4 (2012), p. 331-343.

Wenzel, K., Abdel-Wahab, M., Cefalu, A., and Fritsch, D.: High-Resolution Surface Reconstruction from Imagery for Close Range Cultural Heritage Applications, Int. Arch. *Photogramm. Remote Sens*. Spatial Inf. Sci., XXXIX-B5, 133-138, doi:10.5194/isprsarchives-XXXIX-B5-133-2012, 2012.

Martin A. Fischler and Robert C. Bolles. 1987. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *In Readings in computer vision: issues, problems, principles, and paradigms*, Martin A. Fischler and Oscar Firschein (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA 726-740.

Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International Journal of Computer Vision 60*, no. 2 (2004): 91-110

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Gool, L., 2005. A comparison of affine region detectors. *International Journal of Computer Vision 65*(1-2), pp. 43–72.

S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, 45, Nov. 1998.

Fischler, M.A., Bolles, R.C., 1987. Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: *Martin, A.F., Oscar, F. (Eds.), Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*. Morgan Kaufmann Publishers Inc., London, pp. 726-740.

Hartley RI, Zisserman A (2004) Multiple View Geometry in *Computer Vision*. Cambridge University Press