3D MODELLING FROM MULTI-VIEWS IMAGES FOR CULTURAL HERITAGE IN WAT-PHO, THAILAND

N. Soontranon P. Srestasathiern and S. Lawawirojwong

Geo-Informatics and Space Technology Development Agency (Public Organization)
120, The Government Complex (Building B), Chaeng Wattana Road, Laksi, Bangkok 10210, Thailand
narut@gistda.or.th, panu@gistda.or.th, siam@gistda.or.th

KEY WORDS: 3D modelling, Cultural heritage, 3D reconstruction, Open source, Multi-view images

ABSTRACT:

In Thailand, there are several types of (tangible) cultural heritages. This work focuses on 3D modelling of the heritage objects from multi-views images. The images are acquired by using a DSLR camera which costs around $1,500 (camera and lens). Comparing with a 3D laser scanner, the camera is cheaper and lighter than the 3D scanner. Hence, the camera is available for public users and convenient for accessing narrow areas. The acquired images consist of various sculptures and architectures in Wat-Pho which is a Buddhist temple located behind the Grand Palace (Bangkok, Thailand). Wat-Pho is known as temple of the reclining Buddha and the birthplace of traditional Thai massage. To compute the 3D models, a diagram is separated into following steps; *Data acquisition, Image matching, Image calibration and orientation, Dense matching and Point cloud processing.* For the initial work, small heritages less than 3 meters height are considered for the experimental results. A set of multi-views images of an interested object is used as input data for 3D modelling. In our experiments, 3D models are obtained from MICMAC (open source) software developed by IGN, France. The output of 3D models will be represented by using standard formats of 3D point clouds and triangulated surfaces such as .ply, .off, .obj, etc. To compute for the efficient 3D models, post-processing techniques are required for the final results e.g. noise reduction, surface simplification and reconstruction. The reconstructed 3D models can be provided for public access such as website, DVD, printed materials. The high accurate 3D models can also be used as reference data of the heritage objects that must be restored due to deterioration of a lifetime, natural disasters, etc.

1. INTRODUCTION

Thailand has various cultural heritage sites of the world. One of the most tourists attractions is Wat-Pho, the official name being Wat Phra Chetuphon Vimolmangklararm Rajwaramahaviharn, located behind the Grand Palace at the center of Bangkok. Wat-Pho is also known as temple of the reclining Buddha and the birthplace of traditional Thai massage. Regarding to recent technologies, 3D models are able to obtain from 3D data acquisition (laser scanner: active device) or 2D images (camera: passive device) reconstruction based on photogrammetry. In this work, 3D modellings of many objects in Wat-Pho are reconstructed and presented by using the image-based approaches with photogrammetry techniques. The images are acquired by using a DSLR (Digital Single-Lens Reflex) camera which costs around $1,500 (camera and lens). Comparing with a 3D laser scanner, the camera is cheaper and lighter than the 3D scanner. The image-based approaches also provide more realistic 3D models and potential to acquire a wide range of objects or scenes (El-Hakim et al., 2004). It is also able to use images obtained from a crowdsourcing system (Snavely et al., 2006, Snavely et al., 2008). The objective of this work is to basically visualize and to use as 3D data archives which are required for more accurate 3D models. The sample images and directory of Wat-Pho are shown in figure 1.

Generally, a set of multi-view images acquired from the digital camera can be used to compute for a 3D model. There are several software developments in a decade. For the commercial software, they can be listed for example Agrisoft (`www.agrisoftdg.com/`), Pix4D (`https://pix4d.com/`), Smart3DCapture (`http://www.acute3d.com/`), 3DF Samantha (`http://www.3dflow.net/`), etc. For the open source software, it also provides efficient and accurate productions of 3D models. Structure from Motion (SfM) is a technique used for processing 3D structure from multi-view of 2D images (Wu, 2013, Olsson and Enqvist, 2011). OpenMVG

(a) Heritage objects in Wat-Pho captured by the camera.

(b) Wat-Pho directory (from `www.watpho.com/`).

Figure 1: Wat-Pho images and directory.

(Open Multiple View Geometry) is also a development tool based on SfM technique (Moulon et al., 2012, Moulon et al., 2013). Using the photos acquired by community, a collection of tourism photos in the internet can also be used for 3D reconstruction (Snavely et al., 2006, Snavely et al., 2008). MICMAC (Pierrot-Deseilligny and Cléry, 2011, Pierrot-Deseilligny and Paparoditis, 2012) is a tool for obtaining 3D models developed by National Institute of Geographic and Forestry Information (IGN), France. It can be used to process high resolution images from different platforms such as satellite, aerial, ground images. In this work, MICMAC is initially used for obtaining the 3D point clouds from multi-view images. Then, based on 3D point/mesh processing approaches, the point clouds are computed for the efficient 3D model.

The rest of the paper is organized as follows. An overview diagram is proposed in Section 2. Results of heritage objects are shown in Section 3. Conclusion and future works are discussed in Section 4.

2. OVERVIEW DIAGRAM

Given the multi-view images, 3D reconstruction software can be used to obtain set of point clouds in order to compute for a 3D model. Referring to traditional workflow, an overview diagram is provided in figure 2. It consists of data acquisition, image matching, image calibration and orientation, dense matching and point cloud processing.

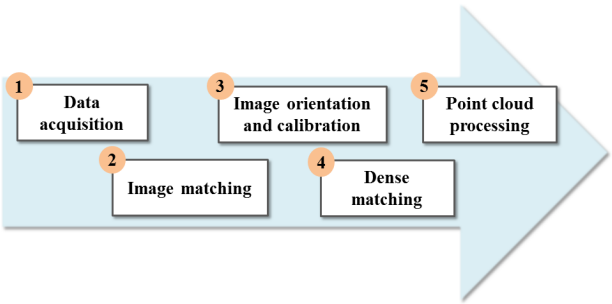


Figure 2: An overview diagram of 3D reconstruction.

2.1 Data Acquisition

Images are acquired from heritage objects in Wat-Pho by using DSLR cameras. The image are recorded at a good quality (less compression in jpeg format) and high resolution ($4,608 \times 3,072$ pixels) with a fixed focal length for each object. Regarding to the 3D reconstuction softwares, the stereo images should have approximately 60–70% of overlapped regions for obtaining satisfied results of image matching (see section 2.2). In terms of viewpoint changes, a short baseline stereo is generally required for the matching algorithms. However, a wide baseline stereo can be computed by using some efficient algorithms (Wu et al., 2008, Morel and Yu, 2009, Soontranon et al., 2013).

2.2 Image Matching

Image matching is an intial step of 3D reconstruction. Keypoints (interest points) are detected from the image. Then, the keypoints are computed for descriptors. Comparing between two sets of descriptors from stereo image, they will be matched for obtaining corresponding points. Regarding to image matching algorithms, SIFT is a very well-known and efficient algorithm (Lowe, 2004). It is used in several implementations. Other image matching approaches are listed as MSER–Maximally Stable Extremal

Figure 3: Image acquisition, neighbor images are recommended to capture at 60% (or more) for overlapped region.

Regions (Matas et al., 2002), SURF–Speeded Up Robust Features (Bay et al., 2008), HOG–Histogram of Oriented Gradients (Dalal and Triggs, 2005), GLOH–Gradient Location and Orientation Histogram (Mikolajczyk and Schmid, 2005), etc. However, many algorithms are limitted for obtaining point matches between wide baseline stereo. A-SIFT or Affine-SIFT (Morel and Yu, 2009) is an algorithm for solving wide baseline stereo, which has a tradeoff in computational time. A robust matching approach required 2D image and 3D data, Viewpoint Invariant Patches (VIP) is used for 3D scene alignment and large scale scene reconstruction (Wu et al., 2008). In order to efficiently deal with wide baseline stereo, a modification of SIFT by using 2D image and 3D mesh was proposed in (Soontranon et al., 2013) called SIFT with conformal images.

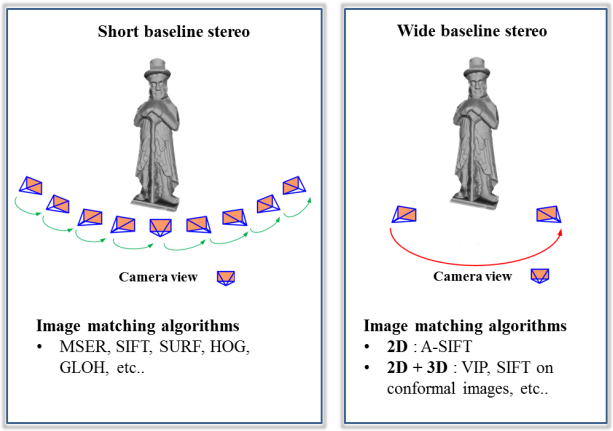


Figure 4: Short or wide baseline stereo with the matching algorithms.

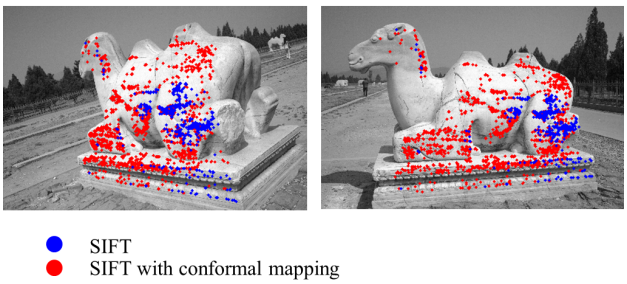


Figure 5: Comparison of the point matches obtained from SIFT (blue) and SIFT on conformal images (red). The figure is referred to (Soontranon, 2013).

The image matching algoritms for short or wide baseline stereo are listed in figure 4. The short baseline stereo is able to process by (Matas et al., 2002, Lowe, 2004, Bay et al., 2008, Dalal and

Triggs, 2005, Mikolajczyk and Schmid, 2005). The wide baseline stereo can be computed by using (Morel and Yu, 2009, Wu et al., 2008, Soontranon et al., 2013). Given the wide baseline stereo, based on existing development, a comparison between SIFT and SIFT on conformal images is shown in figure 5. In this case, SIFT with conformal images is more efficient than standard SIFT for obtaining the point matches but it is necessary to have 3D mesh and more computational time. It should be noted that some condition of the image acquisition (e.g. obtaining less number of images/viewpoints, crowdsourcing based on tourist photos) is required the matching algorithms for solving the wide baseline stereo.

2.3 Image Calibration and Orientation

The point matches will be used to calibrate and align the images in 3D coordinates. The points can be used for automatically obtaining camera's viewpoints. Based on camera parameters, RANdom SAmple Consensus (RANSAC) is a well-known algorithm used to remove unreliability point matches called outliers (Fischler and Bolles, 1981). A bundle adjustment used to compute the camera parameters is referred to (Pierrot-Deseilligny and Cléry, 2011). In general, the efficient image calibration and orientation are based on the quality and quantity of the point matches. The camera viewpoints of a 3D object is shown in figure 6 obtained from MICMAC software.

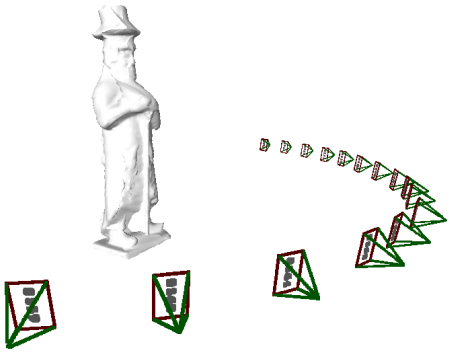


Figure 6: Camera viewpoints after image calibration and orientation were computed.

2.4 Dense Matching

After the oriented images were organized in 3D coordinates, the 3D points of overlapped pixels can be reconstructed by dense matching. An area of interest (AOI) is required to select from the reference image which is generally used to remove the object from the background. For an automatic procedure, AOI can be obtained by using image segmentation algorithms. Regarding to the dense matching algorithms (Scharstein and Szeliski, 2002, Hirschmuller, 2008), corresponding pixels will be matched and reconstructed for 3D point clouds. The results of the initial 3D points are shown in figure 7.

2.5 Point Cloud Processing

The initial point clouds typically consist of some noise which will be removed and reduced by the point cloud processing techniques. The processing steps are as follows; triangulated surfaces, noise filtering and surface simplification (Rusu and Cousins, 2011). The surface reconstruction from the point clouds is generally called mesh. For the visualization via a web service, the suitable resolution of 3D mesh can be found by evaluating results in the next section. Based on our strategies, the initial 3D meshes

Figure 7: Point clouds obtained from multi-view images using MICMAC software.

are removed the noise (filtering) and reduced to a suitable size (simplication), the results are shown in figure 8.

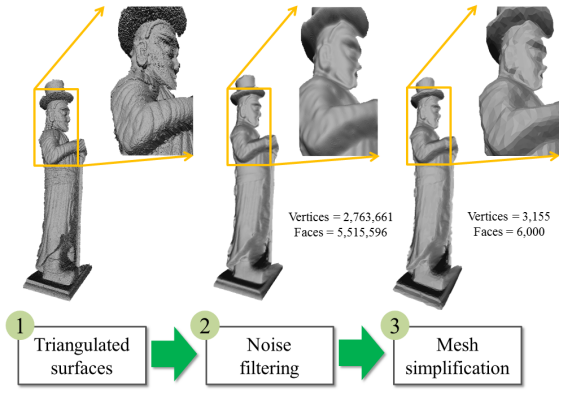


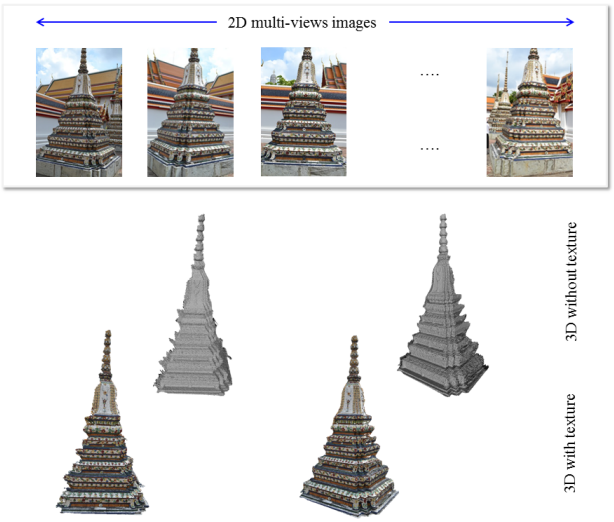
Figure 8: Mesh processing: triangulated surfaces from point clouds, noise reduction and mesh simplification.

3. EXPERIMENTAL RESULTS

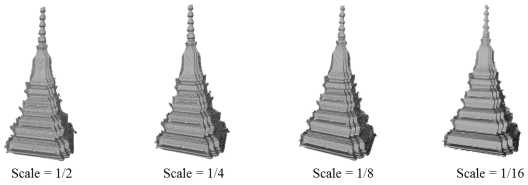
The images used for 3D modelling were obtained from heritage objects in Wat-Pho (Bangkok, Thailand) such as sculpture, pagoda, etc. Spending two days for the acquisition process in about 10% of the total area (total = 20 acres), approximate 20 objects were acquired and used for the experiments. Given each set of 2D multi-view images, 3D model is reconstructed and represented by non-texture, texture, different resolutions of 3D point clouds. The 3D models of non-texture are shown for presenting geometry of the reconstructed objects. The different resolutions of 3D point clouds are computed for observing distortion effects. For publishing on the media, a suitable resolution of the (satisfied) 3D model will also be evaluated. The sample 3D models are summarized in table 1, which consists of number of acquired images (*# Image*), object height (*Approx height*) and presented figure (*Fig*). The parameters of 3D models are shown in table 2.

Table 1: Summary of 3D models presented in this section.

Heritage obj.	*# Image*	*Approx height*	*Fig*
Pagoda	20	3 meters	9
Giant	22	3 meters	10
Buddha footprint	5	0.7 meter	11
Rahoo sculpture	7	0.5 meter	12



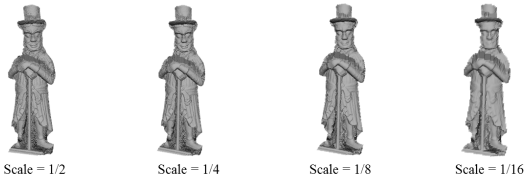
(a) 3D model of Pagoda (non-texture and texture).



(b) 3D point clouds with different resolutions.

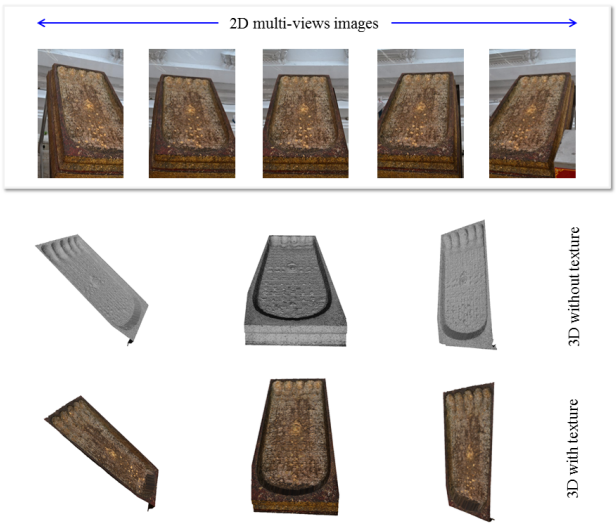
Figure 9: Pagoda

(a) 3D model of Giant (non-texture and texture).

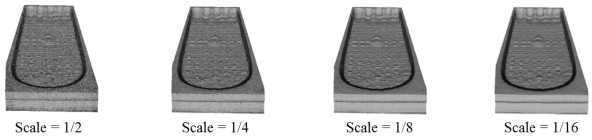


(b) 3D point clouds with different resolutions.

Figure 10: Giant

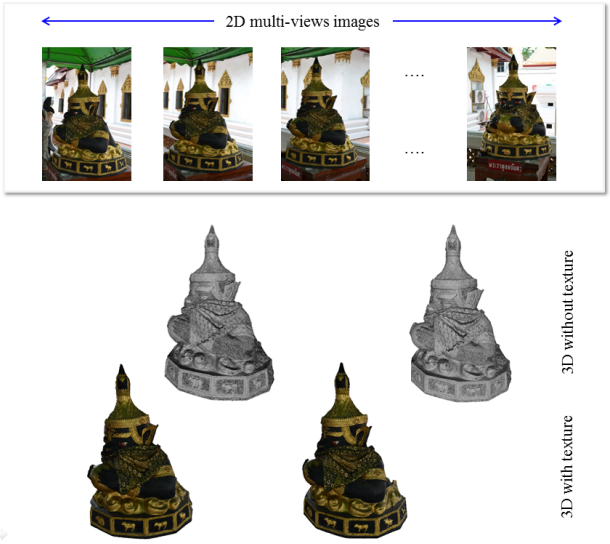


(a) 3D model of Buddha footprint (non-texture and texture).

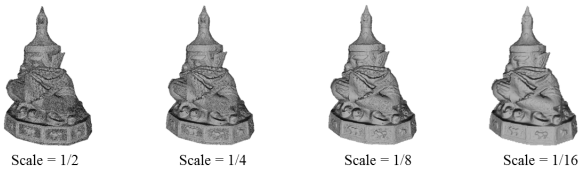


(b) 3D point clouds with different resolutions.

Figure 11: Buddha footprint



(a) 3D model of Rahoo sculpture (non-texture and texture).



(b) 3D point clouds with different resolutions.

Figure 12: Rahoo sculpture

Table 2: 3D parameters: vertices, faces and file size.

Heritage obj.	*Scale*	*Vertice*	*Face*	*Size (KB)*
Pagoda	1/2	320 K	638 K	11,852
	1/4	80 K	159 K	2,953
	1/8	20 K	39.2 K	733
	1/16	5 K	9.6 K	181
Giant	1/2	172 K	342 K	6,365
	1/4	43 K	85 K	1,583
	1/8	10.7 K	20.8 K	392
	1/16	2.7 K	5 K	96
Buddha	1/2	2,296 K	4,585 K	85,117
footprint	1/4	576 K	1,149 K	21,342
	1/8	144 K	286 K	5,326
	1/16	36 K	71 K	1,334
Rahoo	1/2	1,306 K	2,606 K	48,406
sculpture	1/4	328 K	654 K	12,154
	1/8	82 K	162 K	3,029
	1/16	20 K	40 K	760

4. CONCLUSIONS AND FUTURE WORK

The paper provides an overview diagram of 3D reconstruction from multi-view images. In the experiments, 3D models of heritage objects in Wat-Pho (Bangkok, Thailand) are computed and reconstructed. Regarding to the diagram, it consists of image acquisiton, image matching, image calibration and orientation, dense matching and point cloud processing. Using MICMAC (open source) software, the initial point clouds are obtained. The point clouds from MICMAC is efficient for the initial computation but the post-processing is still required. For visualization and collection as the data archive, the 3D point clouds will be processed for the better results. The steps of 3D point cloud processing are triangulated surfaces, noise filtering and mesh simplification. The efficient tools can be referred to PCL (Point cloud library) and Meshlab software. Based on the simplification algorithm (quadric edge collapse decimation), the different resolutions of 3D models are re-computed and compared for the memory sizing.

In the future, the reconstructed 3D models can be provided for public access such as website, DVD, printed materials. The high accurate 3D models can also be used as reference data of the heritage objects that must be restored due to deterioration of a lifetime, natural disasters, etc. The experiments investigate distortion effects (quality of data) comparing with different resolutions of 3D point clouds. A suitable resolution with good quality of 3D visualization is required for publishing on the website.

REFERENCES

Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding (CVIU)* 110(3), pp. 346–359.

Dalal, N. and Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 886–893.

El-Hakim, S. F., Beraldin, J.-A., Picard, M. and Godin, G., 2004. Detailed 3D Reconstruction of Large-Scale Heritage Sites with Integrated Techniques. *IEEE Computer Graphics and Applications* 24(3), pp. 21–29.

Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* 24(6), pp. 381–395.

Hirschmuller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)* 30(2), pp. 328–341.

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision (IJCV)* 60(2), pp. 91–110.

Matas, J., Chum, O., Urban, M. and Pajdla, T., 2002. Robust wide baseline stereo from maximally stable extremum regions. *British Machine Vision Conference (BMVC)* pp. 384–393.

Mikolajczyk, K. and Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 27(10), pp. 1615–1630.

Morel, J. and Yu, G., 2009. ASIFT, A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* 2(2), pp. 438–469.

Moulon, P., Monasse, P. and Marlet, R., 2012. Adaptive structure from motion with a contrario model estimation. *Asian Conference on Computer Vision (ACCV).*

Moulon, P., Monasse, P. and Marlet, R., 2013. Global Fusion of Relative Motions for Robust, Accurate and Scalable Structure from Motion. *IEEE International Conference on Computer Vision (ICCV).*

Olsson, C. and Enqvist, O., 2011. Stable structure from motion for unordered image collections. *SCIA 2011.*

Pierrot-Deseilligny, M. and Cléry, I., 2011. APERO, an Open Source Bundle Adjusment Software for Automatic Calibration and Orientation of a Set of Images. *The ISPRS Commission V Symposium, Image Engineering and Vision Metrology*, Trento, Italy.

Pierrot-Deseilligny, M. and Paparoditis, N., 2012. A multiresolution and optimization-based image matching approach: An application to surface reconstruction from SPOT5-HRS stereo imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36 (Part 1/W41).*

Rusu, R. B. and Cousins, S., 2011. 3D is here: Point Cloud Library (PCL). *IEEE International Conference on Robotics and Automation (ICRA).*

Scharstein, D. and Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision (IJCV)* 47(1-3), pp. 7–42.

Snavely, N., Seitz, S. M. and Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics (TOG)* 25(3), pp. 835–846.

Snavely, N., Seitz, S. M. and Szeliski, R., 2008. Modeling the world from Internet photo collections. *International Journal of Computer Vision (IJCV)* 80(2), pp. 189–210.

Soontranon, N., 2013. Matching between images of distant point of view by utilisation of depth map. *PhD Thesis, University of Picardie Jules Verne, France.*

Soontranon, N., Pierrot-Deseilligny, M. and Mouaddib, E., 2013. Photometric based 3D Registration. *IAPR Int. Conf. on Machine Vision Applications*, MVA'13, Kyoto, Japan.

Wu, C., 2013. Towards linear-time incremental structure from motion. *International Conference on 3D Vision (3DV).*

Wu, C., Clipp, B., Li, X., Frahm, J.-M. and Pollefeys, M., 2008. 3D model matching with Viewpoint-Invariant Patches (VIP). *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).*