Pixel-based 3D Object Reconstruction

A. Somogyi, A. Barsi
Department of Photogrammetry and Geoinformatics,
Budapest University of Technology and Economics, Budapest, Hungary
somogyi.arpad@epito.bme.hu, barsi.arpad@epito.bme.hu

***Abstract*— The photogrammetric object reconstruction seems to find its modern Holy Grail in the pixel-based technology. The great offer of sophisticated software packages prove that the formerly developed computer vision and photogrammetric functions can efficiently be combined to support an accurate object reconstruction. In the paper the general theory and workflow of the methodology is discussed followed by seven in-house projects. These projects show up big variety in the application: ancient wood and current stone sculptures, medieval cathedral and fortress or famous tourist places can be found as supporting case studies. The authors wanted to demonstrate also the independency from the image capture, which was proven the applied image sources: internet-based image data base, UAV-mounted action camera, high quality mirrorless and digital single-lens reflex (DSLR) camera, but also smartphone camera or YouTube video stream. The attached figures intentionally illustrate also the three processing steps: camera alignment, reconstruction and visualization.**

I. Introduction

It is no wonder that the Hollywood film *Avatar* from James Cameron is so famous and popular. Visual effect companies, like *Industrial Light & Magic* or *Weta Digital* use computer techniques based on the reality. Their technology is an implementation of the methodology currently used by the photogrammetry and computer vision community. Because this methodology is rapidly developing and spreading we do want to draw a sketch about its essence.

The very brief summary of the methodology is simple: only a series of pictures must be taken about an object and this technology can deliver the real spatial 3-dimensional model. The black-box technology has been realized in several commercial and free software packages, where the users can or need to have basic knowledge about the workflow. In the paper a draft overview is given followed by examples taken from our practice.

II. Methodology

A. 3D reconstruction workflow

The exact spatial reconstruction of an object is a very hard challenge. Images with suitable amount and quality must be captured in acceptable recording geometry. This means that convergent imagery must be shot about the object, possibly at high geometric and radiometric resolution. The more overlapping images cover the object, the better will be the reconstructed model.

An overview about the whole theoretic procedure can be seen in Fig. 1. The first step is the loading of these images into the processing environment. All of the images must be processed searching for matching features. Such features can be points (or corners) delivered by interest operators like Harris, Förstner etc. but can be complex ones being invariant against scaling or rotation, like HOG (Histogram of Oriented Gradients), SIFT (Scale Invariant Feature Transform) or SURF (Speeded Up Robust Features) [1], [2], [3]. These features are rugged enough to be found in more images and being able to connect them.

So the following step is the matching of images using these features. One can imagine that by increasing the number of images, the number of the features and much more the amount of the possible combinations increase. To cope this problem effective search strategies are deployed.

These strategies are either locally or globally “effective”. The first group is best represented by the Normalized Cross Correlation (NCC) technique, while the second one by a cost-function based optimization, e.g. Dynamic Programming (DP). The most known algorithms are of the global group: Graph Cut (GC), Dynamic Programming (DP), Belief Propagation (BP) and Semi Global Matching (SGM) [4], [5]. These global algorithms deliver esthetically and computationally better reconstruction.

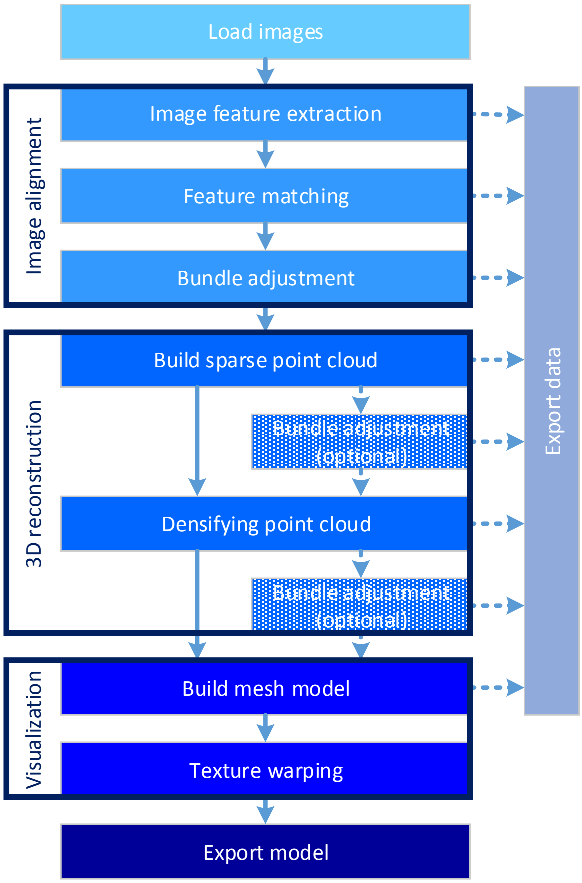

Figure 1. Flowchart of the spatial reconstruction methodology

After getting the matchings among the images, all of their pose, i.e. the positions of their projection centers and orientations can be computed. Because the matching may have some anomalies, the result of these orientation isn't error-free. The quality of the orientation is generally improved by sophisticated bundle adjustment. These last three steps belong to the Image Alignment phase of the procedure.

The 3D object reconstruction can be started now: first a rapid and draft solution is given, which results a sparse point cloud. In several cases an optional bundle adjustment can refine the orientation parameters prior the densification of the point cloud. These lastly mentioned step is the core of the Multi View Stereo (MVS) systems, because the reality-near point density of the object is achievable just by this step. Because the densifying requires very much computing power (significantly more than image alignment – see examples!), different realizations have been developed: Clustering Views for MVS (CMVS), Patch-based MVS (PMVS) or more sub-versions of them. (The visual effect companies in Hollywood are strongly interesting in these methods, because the film production can efficiently be supported by them.) [6], [7]

The result of the sparse and dense point cloud building is adequate for modeling the spatial relation of the object. For example, the surface model of a statue can be the output, which can be transferred into visualization modules. Sometimes the reconstruction software serves with visualization functionality, then point cloud meshing and texturizing is also included. The procedure finishes by the exportation of the digital object model for example in *obj* or *ply* format. This model can be then used for 3D printing or in on-line visualization embedded into a web portal or pdf-document. The users usually have the option to export all computed results (e.g. camera pose parameters, rotation matrices, essential and fundamental matrices, bundle adjustment data, point clouds). [8]

B. Reconstruction based on still images

The above described general method is implemented in the commercial and free software packages. Agisoft Photoscan, VisualSFM, 3DF Samantha, Colmap or Autodesk Remake were the mostly applied software tools. To test the applicability and the packages we have conducted several projects. Two main sources were used to feed the software packages: still images and frames taken from video streams. The most important descriptors are collected in Table 1.

TABLE I.
RECONSTRUCTION PROJECT PROPERTIES

Project name	Number of images	Number of points in dense model	Number of mesh elements
BME sculpture	11	638.588	1.272.362
Parliament	63	117.267	24.353
Heroes' Square	1626	663.268	n.a.
Ócsa aerial photos	15	2.310.794	462.148
Sirok	1042	13.544.267	2.609.077
Ócsa aerial video	583	795.463	90.919
Madonna	1804	2.667.782	533.556

n.a. – not available

The next part of the chapter will introduce the reconstruction projects. The *BME sculpture* has a stone sculpture next to the main entrance of the university's main building. It is a sitting woman figure allegorizing the civil engineering discipline. The sculpture is 3.3 m high and weights more than 9 tons. The sculpture images were taken by a Sony Nex-7 camera under clouded weather circumstances. The resolution of the images was 3376 × 6000 pixels. The rendered textured model can be seen in Fig. 2.

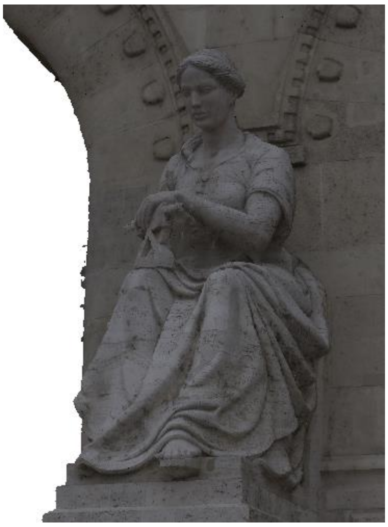

Figure 2. Textured model of the sitting woman sculpture at BME main building

The *Parliament* and the *Heroes' Square* projects had an unconventional image source: the necessary images were downloaded by a self-developed software tool from the Flickr image data base. Flickr belongs to Yahoo Group and its data base contains photos taken by anybody, who uploaded them. The downloader tool could get the images with a highest resolution of 1024 × 1024 pixels. Unfortunately, the image size was mostly much lower and the quality was very heterogeneous. The biggest problem was the automatic selection of the downloaded photos covering outdoor scenes, instead of indoor event contents. Wrong image tags, poor light circumstances, blurred images have to be removed before starting the reconstruction. The camera centers in Heroes' Square are shown in Fig. 3. One can notice that most tourists capture the view from the square's main axis; it seems that people are very sensitive on the symmetry in their pictures. The viewing angle is mostly about 40° or less, which is not a perfect configuration, only a single side can be reconstructed using these photos. On the Heroes' Square the farthest camera position is roughly 140 m from the colonnade. Because of the lower quality of dense point cloud, no mesh was obtained.

In case of the Parliament the nearest photos were taken from the Pest side, whilst the nearest pictures from Buda were ~ 400 m far from the building. There were more photos taken from a distance of 700-800 m. Pictures of longer distance were dropped by the software because the object was too small in the images (do not forget the image size!).

Because the input images have low geometric resolution, the derived model is quite noisy. As an illustration of the power of the spatial reconstruction even under so hard conditions (poor resolution and low quality), the Danube view of the Parliament building was achieved (Fig. 4). The resulting model has enough detail to derive sections from e.g. Cupola [9]. The given number of mesh elements in Table 1 are also only from cupola.

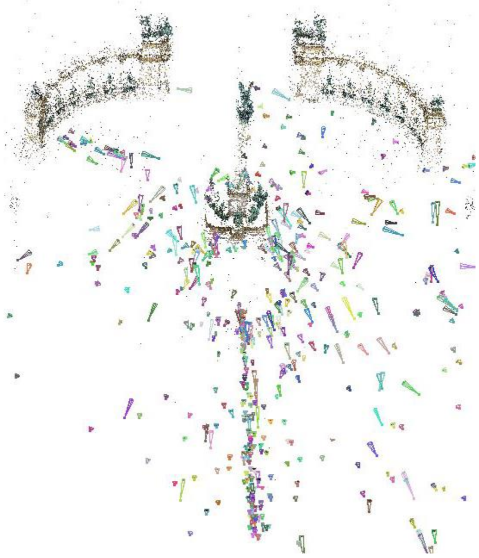

Figure 3. Computed camera centers with orientation in the Heroes' Square project

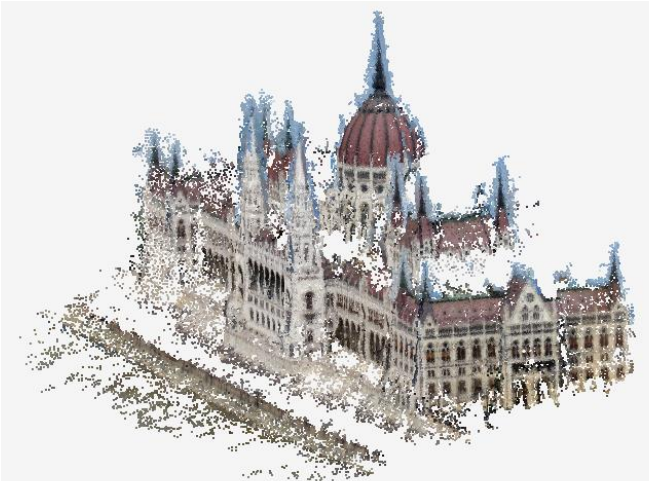

Figure 4. Reconstructed Danube view of the Parliament building from crowd-sourced images

The project *Ócsa aerial photos* was based on the excellent camera shots from an airplane. The involved images were capture by a Fuji FinePix S3Pro camera with a resolution of 4256 × 2848 pixels. This project has been started newly; the goal is to test the combination of terrestrial and aerial images with terrestrial laser scanning. The object is a medieval cathedral, which was part of a Premonstratensian monastery. The construction of the cathedral was started in the 13th century in roman style. It has a layout forming a Latin cross: nave, aisles, transept and 3 apses. The ownership of the cathedral was changed: since the 16th century it belongs to the Calvinistic church. Between 1986 and 1992 the cathedral was totally renovated. The cathedral is about 40 m long, 25 m wide and the towers are roughly 25 m high.

The photo projection centers were computed during the alignment phase giving a feed-back that the used imagery has nice uniform distribution around the object.

The dense point model can be directly visualized in the web by Potree, see Fig. 5.

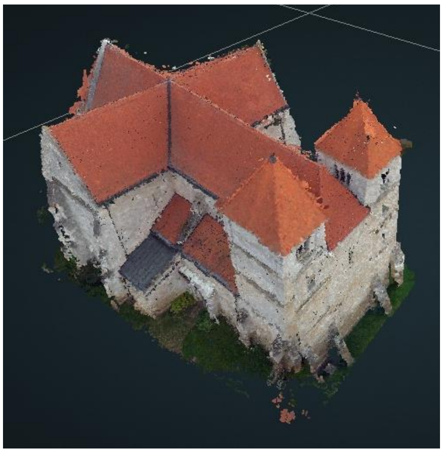

Figure 5. Potree viewer with the point cloud about the ancient cathedral in Ócsa

To illustrate the computing time, the following data can be given. The derivation of feature vectors and pairwise matching took 1 minute, the alignment of all image pairs, the connected bundle adjustment and the creation of the sparse point cloud 1 second, while the production of the dense point cloud was 9 seconds followed by a processing time of 2 minutes 46 seconds for meshing and further 1 minute for building texture.

In the *Sirok* project a GoPro Hero3+ Black Edition camera was mounted on a DJI Phantom platform and was flown around the ruins of the ancient castle. The fortress stands on a 296 m high hill; the first citations are from 1320 AD. Since that time it was continually developed till 1713, when the Habsburgs have destroyed. Now the ruins are in an area of almost 3000 m^2. The width of the remained wall segments is between 1.5 and 2.5-3.5 m in the upper and the lower castle respectively. There were 4 m deep and 4.5 m wide pits along the northern wall as well as 25-30 m high cliffs.

The resolution of the taken photos was 4000 × 3000 pixels. Figure 6 shows the full size meshed and textured terrain model about the working area. [10]

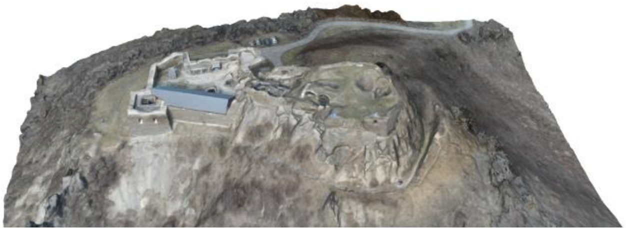

Figure 6. Textured model about Sirok

The Sirok project was the largest of the evaluated ones. The processing times are therefore interesting, even in comparison to the previously presented project. The feature vector extraction and pairwise matching required 10 hours 12 minutes, the basic alignment with sparse point cloud 2 hours

10 minutes, while the dense point cloud needed 1 whole day and 17 hours processing time on our dedicated workstation! The mesh step took then 7 minutes and the texturizing 3 hours 32 minutes respectively.

C. Video-based reconstruction

The processing of video streams has great similarity to the technology of still image processing. The main difference is the frame extraction, which must be performed by a sampling step. Each recorded movies must be played and every nth image frame must be cut off and save as still image in a dedicated directory. There can be quality differences between image streams, so the sampling frequency is depending mainly from the recording camera.

The two presented cases are therefore interesting in the capturing aspects. The *Ócsa aerial video* project was focused on the crowd-source, namely a YouTube video stream was download. The stream has a length of 9 minutes and 42 seconds. The framerate is 25 fps, the recording is full-HD (1920 × 1080 pixels). The supposedly action cam captured video was taken from a UAV. After executing all camera alignment and 3D reconstruction steps it is really worthy to go through all visualization phases, because the final model is very sculpturesque, “almost touchable” (Fig. 7).

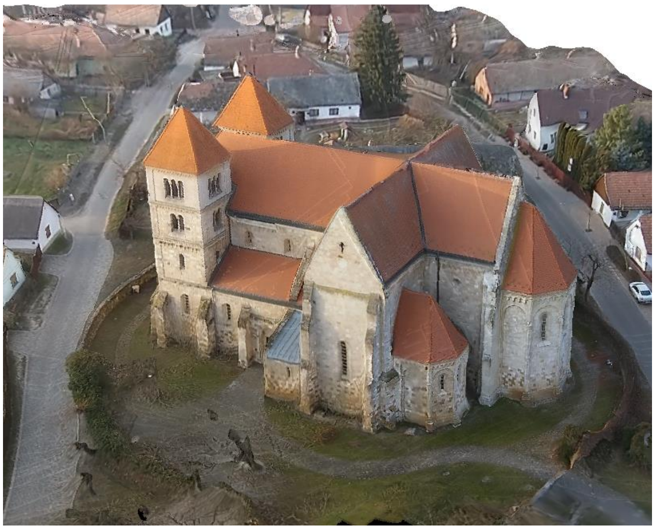

Figure 7. The Ócsa cathedral model after all reconstruction steps

In the *Madonna* project a Samsung Galaxy S5 mobile phone was used as a camera. The wood sculpture was made by an unknown Italian master from Fabriano in the 14th century. The 153 cm high painted wood Madonna holds the child Jesus in her hand. On the backside of the sculpture there is a roughly 20 cm big padded niche presumably for keeping valuable relic.

The stream was recorded as full-HD video. Every 10th frame was sampled and saved for later processing. Two trajectories were applied around the head of the Madonna sculpture: 31 s and 34 s long streams were stored. More details about the modeling and results, see [11]. Fig. 8 illustrates the two trajectories around the reconstructed Madonna-head. It isn’t necessary to mention that the dense projection centers strongly correlate with the camera movement, so the image capture can easily be checked. In this project the very end was reached by the 3D printing of the head model.

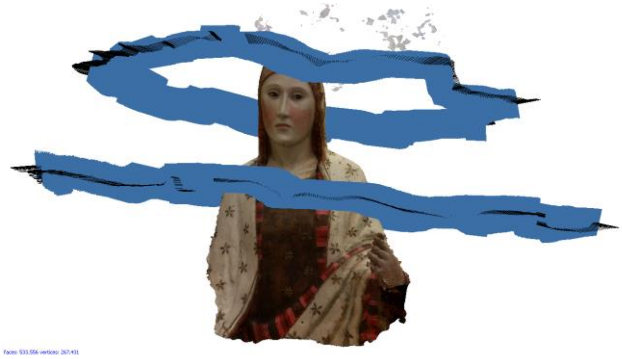

Figure 8. Madonna-head with the camera poses in two circular trajectories

The presented projects were computed on a computer with double Intel Xeon E5-2650 @ 2.0 GHz CPU, 64 GB RAM and NVidia Quadro K5000 video card with CUDA support.

III. CONCLUSIONS

In this paper the big picture of the spatial object reconstruction was drawn. The main components of the methodology have been presented followed by a collection of divergent case studies. These cases aimed to show that the technology can be involved universally for images captured from smaller indoor objects (like a wood art treasure) to a bigger outdoor objects (like a fortress on a hill). The origin of the images can have also great variety, the image alignment and point cloud production can achieve the expected outputs.

REFERENCES

[1] http://en.wikipedia.org

[2] MathWorks Matlab, Computer Vision Toolbox

[3] Barsi, Á. (2013): Application of image segmentation techniques in remote sensing [in Hungarian], Geomatikai Közlemények, XVI, pp. 83-88

[4] Unger, Ch. (2013): Contributions to Stereo Vision, PhD thesis, Technical University Munich, p. 152

[5] Hirschmüller, H. (2008): Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 2, pp. 328-341

[6] http://www.di.ens.fr/cmvs/

[7] http://www.di.ens.fr/pmvs/

[8] http://ccwu.me/vsfm/

[9] Somogyi, A., Barsi, A., Molnar, B., Lovas, T. (2016): Crowdsourcing based 3D modeling, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B5, pp.587-590

[10] Török, A., Bögöly, Gy., Czinder, B., Görög, P., Kleb, B., Vásárhelyi, B., Lovas, T., Barsi, Á., Molnár, B., Koppányi, Z., Somogyi, J.Á. (2016): Terrestrial laser scanner aided survey and stability analyses of rhyolite tuff cliff faces with potential rock-fall hazards, an example from Hungary, In: Reşat Ulusay; Omer Aydan; Hasan Gerçek;Mehmet Ali Hindistan; Ergün Tuncay (eds.) Rock Mechanics and Rock Engineering: From the Past to the Future: Eurock 2016, pp. 877-881

[11] Kapitany, K., Somogyi, A., Barsi, A. (2016): Inspection of a Medieval Wood Sculpture Using Computer Tomography, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B5, pp.287-291