CROWDSOURCING BASED 3D MODELING

A. Somogyi*, A. Barsi, B. Molnar, T. Lovas

Budapest University of Technology and Economics (BME), Dept. of Photogrammetry and Geoinformatics, Hungary – (somogyi.arpad, barsi.arpad, molnar.bence, lovas.tamas)@epito.bme.hu

Commission V, WG V/3

KEY WORDS: structure-from-motion, web-based photo album, object reconstruction, modeling

ABSTRACT:

Web-based photo albums that support organizing and viewing the users' images are widely used. These services provide a convenient solution for storing, editing and sharing images. In many cases, the users attach geotags to the images in order to enable using them e.g. in location based applications on social networks.
Our paper discusses a procedure that collects open access images from a site frequently visited by tourists. Geotagged pictures showing the image of a sight or tourist attraction are selected and processed in photogrammetric processing software that produces the 3D model of the captured object. For the particular investigation we selected three attractions in Budapest. To assess the geometrical accuracy, we used laser scanner and DSLR as well as smart phone photography to derive reference values to enable verifying the spatial model obtained from the web-album images. The investigation shows how detailed and accurate models could be derived applying photogrammetric processing software, simply by using images of the community, without visiting the site.

1. INTRODUCTION

Picture or photo sharing is not only trendy, but also has enormous potential to collect information about the environment, especially about frequently visited sights. Tourists, but also local people take images with their cameras, which are later uploaded to image sharing sites, such as Flickr, Instagram, Panoramio, Picasa, Pinterest and others. These sites can be considered as image databases, from where experts can download image series about interesting places that can be the input data of photogrammetric procedures.
One of the great cutting edge theories and their software realization is the dense 3D solutions with structure from motion (SFM) methodology mathematically based on semi-global matching (SGM) (Hirschmuller, 2008).Thanks to this technique arbitrary images can be loaded into the suitable software, followed by the orientation and adjustment without having tie point and control point sets manually selected. As a potential output, the workflow can be completed by object reconstruction, hence the results can be object meshes or even high quality and valuable object models. Both commercial software (like Agisoft PhotoScan (Agisoft, 2016) or Pix4Dmapper Enterprise (Pix4D, 2016)) and open-source development software (e.g. VisualSFM (VisualSFM, 2016)) are available for the society.
In the paper we are presenting a methodology starting from the download of an adequate image set about a specific site, followed by its processing steps and presenting the results. The latter contains also a comparison to other data capture and processing results, such as homogeneous high resolution camera images taken by a single camera (vs. multiple camera types) and terrestrial laser scanning. During the processing we have discovered some interesting issues about photography styles; these are also discussed in the paper.

2. COLLECTING CROWDSOURCED IMAGERY

Flickr was created by Ludicorp in 2004 for image and video hosting and sharing. The service was acquired by Yahoo one year later, and in 2013 it already had 87 million registered members and more than 6 billion images. The daily upload is more than 3.5 million new photos (FlickrWikipedia, 2016).
The uploaded photos have not only the pictures, but also metadata, like Exif-info, geolocation data or different tags, people names, license information, etc. Flickr offers a bunch of application programing interfaces (APIs) in App Garden, where developers can find codes for mainly non-commercial purposes e.g. uploading, replacing, retrieving, commenting images (FlickrAPI, 2016). These APIs enable to download geotagged photos, which are suitable for experimenting photogrammetric techniques.
Based on the basic idea of Tamara Berg and James Hays (Hays and Efros, 2016) we developed an image query and download script in Matlab. The query requires basic data from Flickr: by the use of Flickr maps (Flickrmap, 2016), one can specify the center point of the area of interest by storing its coordinates. The map can furthermore inform about the amount of available photos, which was crucial in our tests (Fig. 1).

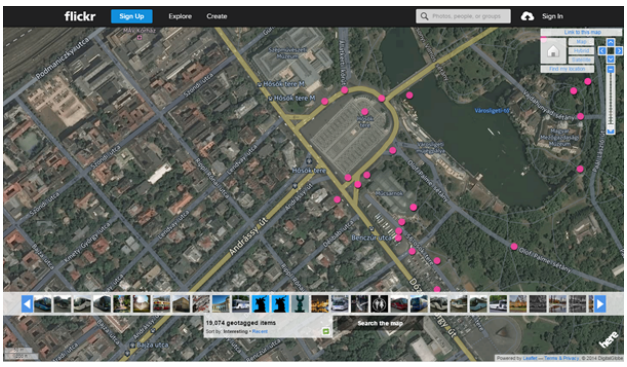

Figure 1. The Flickr map query to study the available uploaded photos

The App Garden contains also a search option (Flickrsearch, 2016), which is a form managing the query fields, e.g. tags, dates, formats. The location (latitude and longitude) based queries can also be run within this form, returning an XML-list as a result. Study-

*Corresponding author

ing these lists we developed our downloader script, containing a valid API-key (e.g. *"255702f37957715afc964fc416b9791d"*), latitude-longitude data in degrees (e.g. *47.514979*°, *19.077933*°), and search radius in kilometers (e.g. *0.2 km*). After setting some further parameters (e.g. web access time out) an automatic URL-address generation was executed, which pointed to a valid photo on the Flickr-server farm (*static.flickr.com*). The routine collects the list of the available photos, with their access codes as well, then downloads them automatically in the requested format. Unfortunately, the best available photo size is about 1024 × 1024 pixels. The developed routine was able to collect 6301 photos about one of our selected historical sites, the Heroes Square (Budapest, Hungary) in about 2.5 hours. (This performance obviously highly depends on the band width of the network. The downloads were executed at the university with about 100 Mbit/s download rate.)
There were three test sites in Budapest, where sufficient number of available photos could be found in the database:

- Heroes Square (6301 photos),
- Gresham Palace and the Lions of the Chain Bridge (6815 photos),
- and the Parliament building (6938 photos).

3. REFERENCE MEASUREMENTS

The reference for our tests was created by three, independent remote sensing techniques. The first one is the terrestrial laser scanning (TLS), which was done by a *Faro Focus 120* scanner. The Lion on the Chain Bridge was acquired from two positions, merged by ICP (Iterative Closest Point) technique. Each station was captured by 3 mm on 10 m resolution and without color information. The captured point cloud contained 27 million points from the surrounding area and 216 thousand points from the sculpture (Fig. 2).
The second technique was capturing high resolution images by a *Canon EOS 760D* DSLR camera, followed by the dense 3D object reconstruction. This technique was also used for the Lion, since its size enabled to test and compare all three selected technologies. There were 25 photos taken with 6000 × 4000 pixel resolution from roughly 180° around the sculpture. The dense reconstruction was carried out by the Agisoft PhotoScan and resulted about 28 million points (Fig. 3).
To simulate images taken by tourists or residents, a third measurement was performed by a *Xiaomi Redmi note 2* smart phone. There were 19 photos taken with 2368 × 4028 pixel resolution from roughly 100° around the sculpture. The dense reconstruction was carried out by VisualSFM and resulted about 120 thousand points (Fig. 4)

4. PROCESSING IMAGES

The crowdsourcing based images have to be preprocessed before reconstruction. The downloaded image sets contained a lot of redundant photos therefore automated algorithms were applied by dupeGuru Picture Edition (dupeGuru, 2016). After that step the number of images were reduced to less than half of the original amount. Many pictures were "misgeotagged"; filtering of these image sets was a time-consuming manual work.
After filtering the images, the input data for the reconstruction were as follows:

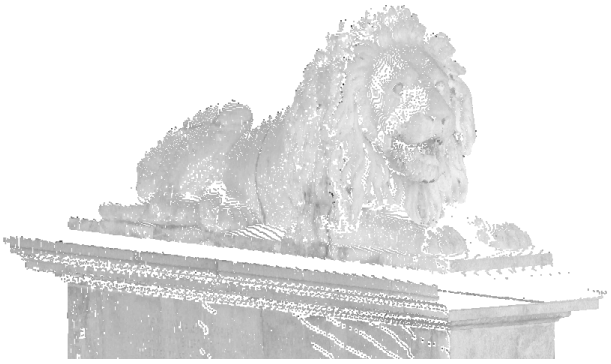

Figure 2. Reconstruction from TLS

Figure 3. Reconstruction from DSLR camera images

Figure 4. Reconstruction from smart phone images

- Heroes Square (1626 photos)
  - Pairwise matching: 55.697 minutes; 3D reconstruction: 104 seconds; bundle adjustment: 5 seconds; dense reconstruction: 278 seconds
- Lion on the Chain Bridge (63 photos)
  - Pairwise matching: 86 seconds; 3D reconstruction: 57 seconds; bundle adjustment: 7 seconds; dense reconstruction: 22 seconds
- The Parliament building (173 photos)
  - Pairwise matching: 14.583 minutes; 3D reconstruction: 102 seconds; bundle adjustment: 7 seconds; dense reconstruction: 113 seconds

The reconstruction was carried out by VisualSFM (VisualSFM, 2016), which implemented four basic steps to generate dense

point cloud. First, it used SiftGPU algorithm to find correspondence between the images. Then it created sparse point cloud based on the calculated positions and key points. In the next step bundle adjustment should be applied to achieve greater accuracy. The final task to create dense cloud is the 3D reconstruction, which used Clustering Views for Multi-View Stereo (CMVS) algorithm (Furukawa, 2016).

In most cases the point clouds were contaminated with errors, typically with noise. To improve quality, the datasets were loaded into Geomagic Studio (Studio, 2016). Semi-automatic noise reduction algorithm was used to create the final point clouds. In case of the lion, this workflow converted 63 photos into about 22 thousand points.

In practice, mesh models are needed for several further working phases e.g. 3D printing, documentation with 3D PDF files or generating continuous sections around the object. A surface model could be wrapped around the point cloud by Geomagic Studio (Fig. 9).

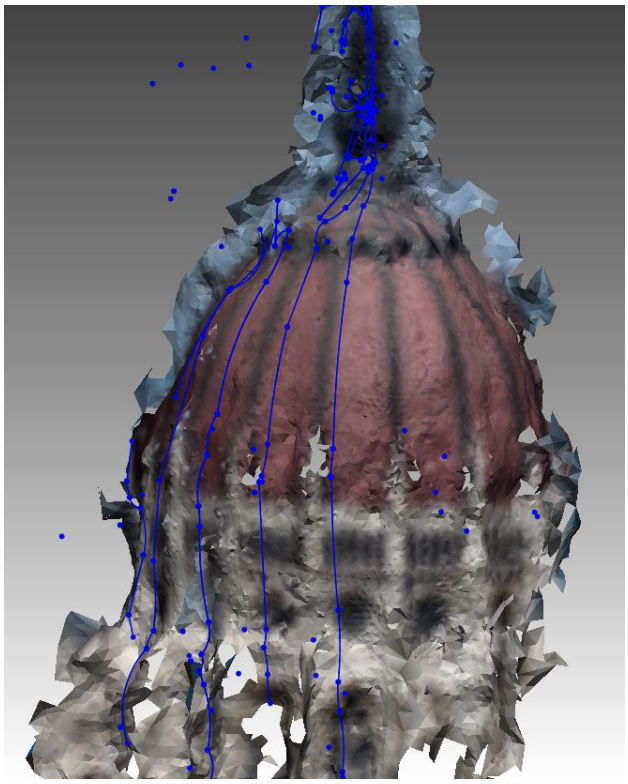

Figure 5. Mesh model of the Parliament's cupola with some section lines

The dense reconstruction with Agisoft is quite similar to the method of VisualSFM. The differences are in the process of the key point based image matching and the densifying. Both algorithms aim to achieve the most accurate model even at the cost of increase in computational time.

5. RESULTS

As already has been mentioned, point clouds can be one of the results of the reconstruction. To validate the model from crowdsourced image dataset, all derived object models have to be registered in a common system. Only the terrestrial laser scanning can produce true-scale model, so all other data have to be scaled before registration. Manual *n*-point based alignment was used to calculate the coarse transformation parameters, followed by the final position calculation with ICP-method in Geomagic Studio

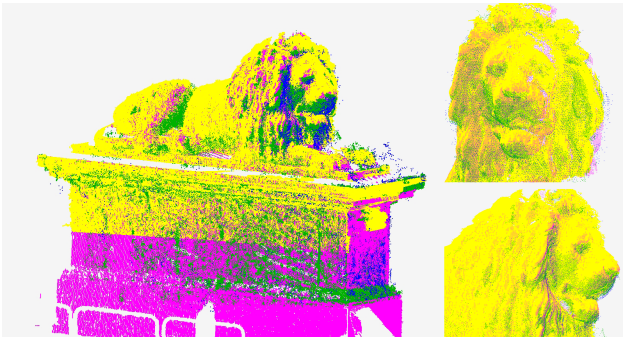

Figure 6. Point clouds of the lion on the Chain Bridge (yellow: DSLR camera images; magenta: terrestrial laser scanning; green: smart phone images; blue: crowdsourced imagery)

(Fig. 6).

The web-based photo albums could be used to create maps that can display tourist sights in photorealistic manner (Fig. 9). To develop these types of infographics, the photos' approximate location data is required. Although geotag information is assigned to the images, some visual and date/time features can improve their exact localization. (Tammet et al., 2013), (Crandall et al., 2009). These maps can represent the most interesting places even at city level all around the world. Performing the necessary processing steps, as a secondary output we can obtain additional information on peoples' photographing habits: the adjusted projection centers can be put in a situation map together with the camera orientation. In some cases, these center maps show the limitations, from where the objects can be captured, e.g. the lion on the Chain Bridge can be taken only from the sidewalks (see the concentrated image locations on Fig. 7). The other more fascinating fact is that people have very strong demand on symmetry; this is a potential explanation of the experienced projection center positions in the Heroes Square, where the central standing column with the symbolic angel figure on the top must be in the center of peoples composition and the columns with the kings' sculptures should be positioned symmetrically on the right and left side (Fig. 8). The camera orientation is "position-dependent": the closer the photographer was to the column, the steeper the photo was taken – it seems the angel figure was mostly in the focus of the camera.

The processing of web-based images with VisualSFM doesn't provide accurate geometry, and the point density compared to a professional camera measurement is significantly lower. However, an approximate point cloud may be appropriate to present a variety of tourist attractions. Because these point clouds have less amount of reconstructed points, the results are also easy to be published online.

6. CONCLUSION AND FURTHER WORKS

In the current research we presented a workflow, where crowdsourced web-based image albums were used as an excellent source for object reconstruction. The pilot sites were in Budapest, Hungary. The selected attractions had enough photos uploaded in the database, from where a script could download the geolocation-filtered images. These images were processed both by open-source and proprietary structure-from-motion software packages. Prior to the processing a manual control (selection) had to be done. The crowdsourced image data and their reconstruction were compared to three independent references, being homogenized into a single system. The obtained results can be used widely: from the desktop and web-based publication as simple visualization to the object surface modeling as engineering application. Another possible output can be the 3D printing and

Figure 7. Reconstruction of the Lion on the Chain Bridge with the camera positions and orientations

Figure 8. Reconstruction of the Heroes Square with the camera positions and orientations

Figure 9. Dense point cloud of the Parliament building just after the reconstruction

further sophisticated analysis steps. As it turned out from the experiments, the high geometric resolution of the input imagery is mandatory for achieving better outputs. Future work can focus on the involvement of pixel color information and mobile laser scanning data. The photographing habit of people is worth to study at more tourist attractions.

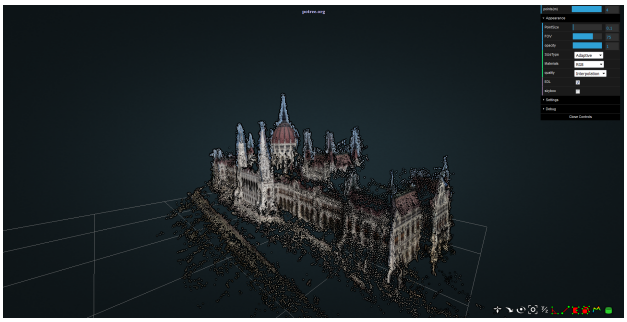

Figure 10. The cleaned error-free Parliament in a web browser environment (web.fmt.bme.hu/fmtpontfelho/parlament.html)

REFERENCES

Agisoft, 2016. http://www.agisoft.com/. Accessed: 2016-04-28.

Crandall, D., Backstrom, L., Huttenlocher, D. and Kleinberg, J., 2009. Mapping the world's photos. In: Proceedings of the 18th International Conference on World Wide Web, WWW '09, ACM, pp. 761–770.

dupeGuru, 2016. https://www.hardcoded.net/dupeguru. Accessed: 2016-04-28.

FlickrAPI, 2016. https://www.flickr.com/services/api/. Accessed: 2016-04-28.

Flickrmap, 2016. https://www.flickr.com/map. Accessed: 2016-04-28.

Flickrsearch, 2016. https://www.flickr.com/services/api/explore/flickr.photos.search. Accessed: 2016-04-28.

FlickrWikipedia, 2016. https://en.wikipedia.org/wiki/Flickr. Accessed: 2016-04-28.

Furukawa, Y., 2016. Clustering views for multi-view stereo (cmvs). http://www.di.ens.fr/cmvs/. Accessed: 2016-04-28.

Hays, J. and Efros, A., 2016. Im2gps: estimating geographic information from a single image. http://graphics.cs.cmu.edu/projects/im2gps/. Accessed: 2016-04-28.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), pp. 328–341.

Pix4D, 2016. https://www.pix4d.com/. Accessed: 2016-04-28.

Studio, G., 2016. http://www.geomagic.com/en/. Accessed: 2016-04-28.

Tammet, T., Luberg, A. and Järv, P., 2013. Information and Communication Technologies in Tourism 2013: Proceedings of the International Conference in Innsbruck, Austria, January 22-25, 2013. Springer Berlin Heidelberg, Berlin, Heidelberg, chapter Sightsmap: Crowd-Sourced Popularity of the World Places, pp. 314–325.

VisualSFM, 2016. http://ccwu.me/vsfm/. Accessed: 2016-04-28.