Image based 3D Reconstruction in Cultural Heritage Preservation

Alessandro Cefalu, Mohammed Abdel-Wahab, Michael Peter, Konrad Wenzel and Dieter Fritsch
Institute for Photogrammetry, University of Stuttgart, Geschwister-Scholl-Straße 24D, Stuttgart, Germany

Keywords: 3D Reconstruction, Structure-from-Motion, Dense Image Matching, Cultural Heritage.

Abstract: Documentation of the current state of an object is often the first crucial step in cultural heritage preservation. Especially for large scale objects as buildings this task becomes complex and time consuming. Hence, there is a growing interest in new, more efficient techniques, which ease the process and reduce the financial impact of surveying actions. In case of façade restoration, experts need to map damages and plan the corresponding measures, before the actual restoration can take place. Here, two-dimensional CAD drawings, depicting each single stone, serve as a basis. Traditionally these plans are derived from classical surveying. Often a photogrammetric approach is chosen to reduce the efforts on site. But still image processing, including image registration and point measurements, is carried out punctually and manually. Since about three years, our institute supports the introduction of modern image processing tool chains to the application field of heritage preservation. Recently we participated in the restoration of the tower facades of the St. Martin dome in Rottenburg/Neckar, Germany. We combined laser scans, terrestrial imagery and images captured from a UAV platform, incorporating structure-from-motion, dense image matching, point cloud registration and production of orthographic projections, from which the CAD drawings could be derived.

1 INTRODUCTION

Cultural heritage preservation has come to more attention in the last years, as there is a growing awareness of the need to maintain monuments and other artefacts for future generations. As an initial step most preservation actions include a documentation of the current state of the object of interest. In case of façade restoration actions, standard surveying methods are usually chosen to observe the façade's geometric appearance. Enough points need to be measured to allow a mapping of each single stone and other details important for the restoration task. From these measurements 2D CAD drawings are derived, which again enable civil or structural engineers and architects to map damages, plan corresponding counter measures and estimate costs.

Tachymetry as a classical surveying method provides very accurate but only punctual measurements which need to be triggered manually. Thus laser scanning is often preferred due to its' high measurement density and fast acquisition rate. Photogrammetry also provides fast on-site acquisition. When used in a modern, highly automated manner, incorporating techniques as structure-from-motion and dense image matching, it can provide results comparable to laser scanning to a much lower price of hardware.

Certainly, it has the drawback of being a triangulating measurement technique and thus a point in space must be observed from more than one station, but a camera station can be changed without much effort. In fact the camera can be mounted on a moving platform as a crane or a hoisting platform and reach areas of a building which are not accessible from the ground. Due to its' relatively low weight it can also be carried by small UAVs, improving the approach's flexibility even more.

Since both, laser scanning and dense image matching, observe arbitrary points in space with a high density, producing orthographic projections (orthophotos) of the data seems an adequate basis for the final CAD drawings.

Our institute promotes the introduction of modern image processing strategies into the branch of heritage preservation through practical application of internal and external developments since more than three years. Within the presented project, which aims at the restoration of the facades of the St. Martin dome in Rottenburg am Neckar, Germany, we took over the task of status-quo-documentation.

Cefalu A., Abdel-Wahab M., Peter M., Wenzel K. and Fritsch D..
Image based 3D Reconstruction in Cultural Heritage Preservation.
DOI: 10.5220/0004475302010205
In *Proceedings of the 10th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2013), pages 201-205
ISBN: 978-989-8565-70-9

2 DATA ACQUISITION

The dome is located in the city's historical centre and is mostly surrounded by dense and irregular housing. The dome's tower rises from the dome's southern roof surface and reaches roughly 70m in height. The facades have a breadth of roughly 7,50m. Regardless of the used measurement technique, the rare possible terrestrial survey stations can mostly be found in the surrounding alleys and always suffer from either high distance or bad viewing angle or occlusion. Although there is a market square west of the dome, which provides a lot of space, the dome's entrance strongly occludes the view to the west side of the tower (Figure 1).

Figure 1: View at the west side of the dome from the market square.

As indicated before, dense image matching was to be used for as many parts of the tower as possible. The orientation of the images has been computed without any a priori information by means of structure-from-motion algorithms. To guarantee high coverage and quality of the derived point clouds, several aspects need to be taken into account. First, the image data set needs to provide enough overlap between the images to enable a stable orientation of the camera stations. Here usually a minimum of 60% is chosen. Second, this overlap needs also to fit the requirements of dense image matching. Here a higher overlap and hence greater similarity of the images is preferable. Of course, this results in a trade-off between matching completeness and triangulation accuracy, which gets worse for smaller intersection angles. We usually try to provide an overlap of 80%, which is a good compromise. Additionally, we make use of the given redundancy by applying multi-view-stereo triangulation to our data sets.

Figure 2: The tower helmet's rich details which should be collected from an elevated position.

In terms of object coverage, it is obvious, that images from elevated stations were needed to obtain data for areas appearing occluded from below. Hence, we decided to hire an external company to collect the images for us, using an octocopter (Figure 3). We planned to cover the tower in twelve vertical flight routes, to guarantee horizontal linkage of the four facades. The camera was triggered every two seconds. Unfortunately, the GPS stabilisation of the octocopter was corrupted in lower areas by the surrounding housing, thus the pilot decided to steer the UAV completely manually. This again, was made difficult by wind shear around the tower. The strong and fast steering manoeuvres, together with the high payload, caused one of the rotors to break during the second vertical flight route. Fortunately, the emergency landing could be carried out without any people being hurt or material being damaged, except the octocopter itself. Nevertheless, it was not possible to complete the flight. Since there was only a short time left between the date of receiving the flight permission and the date of raising a scaffold around the tower, it was not possible to organize a crane or similar to reach the upper parts of the tower (Figure 2). Instead we decided to obtain as much terrestrial imagery as possible and collected very close images from the scaffold itself, accepting the predictable complications caused by non-optimal choice of camera stations. Also, scaffolds are usually built in a manner which occludes view from one layer to the next and the strong scaling differences between UAV / terrestrial images and the images captured from the scaffold, were expected to make an image based registration impossible. To overcome this challenge, we additionally laser scanned the tower, in order to perform a point cloud based registration of the data.

All in all, we collected laser scans from twelve stations, roughly 1900 terrestrial and UAV images and roughly 2400 very close images from the six uppermost scaffold levels.

Figure 3: The octocopter collecting images at the south-west side of the tower.

3 DATA PROCESSING

As mentioned above, our goal is to identify and practically apply given software solutions to heritage preservation. In this case, the tasks which we needed to perform range from image registration, dense point measurement and registration of point clouds to the production of orthographic projections and deriving CAD drawings from the latter. To our knowledge, there is no software package available which can handle all these topics in one pipeline, but there is a variety of packages which can provide solutions for single sub-tasks. For structure-from-motion, as well as dense image matching we have applied internal own developments. For point cloud processing we mostly used freely available packages (OpenSource).

3.1 Structure-from-Motion

This step appeared to be the most sensitive part of the process chain. We applied our internal development (Abdel-Wahab et al., 2012) and cross-checked it with results obtained from the free software VisualSFM (Wu et al., 2011). The basic work flow of both implementations consists of:

1. Keypoint detection and descriptor extraction (per image)
2. Keypoint matching and estimation of relative orientation (for each image pair)
3. Point tracking and connectivity computation
4. Iterative triangulation of points contained in oriented images and resection of new images, followed by a (small scaled) bundle adjustment
5. Final (global) bundle adjustment

For completeness, it should be mentioned, that in some cases optical flow is used instead of extraction of distinct features. However, optical flow is better suited for video streams, where parallaxes between consecutive images stay small.

The selection of images and assigning the initial image pair and the setting of program parameters has a high impact on the results. While both implementations yield comparable results, problems regarding connectivity or orientation accuracy occur at different parts of the dataset.

Both implementations use a 2-parameter radial distortion model and allow either assigning one camera model to all images or assigning a separate model to each image. While the latter is especially useful for cases in which images are obtained from internet data bases or with a zoom lens, the first option should be preferred when the images are taken with a single camera with fixed focal length. In our case four cameras have been used, partially with zoom lenses and partially with fixed focal length. Assigning separate models per image was the most practicable solution for us and delivered reasonable results. Nonetheless a more flexible assignment of camera models is a desirable feature for future developments, as it can be expected to make the structure-from-motion process more robust in some cases. The same might hold for embedding more complex distortion models.

However, the major drawback of many approaches to structure-from-motion is their iterative nature. Errors introduced in an early iteration are propagated through the process and can lead to a drifting behaviour. This is especially easy to observe in circular camera station configurations, where this circumstance can lead to the well-known loop-closure problem. The propagated errors lead to a rejection of the linkage between images introduced in an early iteration and images introduced in a later iteration. As mentioned this problem occurred with both software variants, but at different parts of the data. This might be explained by the different approaches to evaluating connectivity and accordingly, to a different behaviour in rejection of images and a different order of processing the images.

The connectivity of the images yields a network of relative orientations. While the described technique can be understood as an attempt to find optimal paths through this network, some recent developments aim at solving the issue as a global network optimization problem (Crandall et al., 2011). The basic idea is that a concatenation of relative transformations of a network cycle needs to result in an identity transformation. As we see a lot of potential in this approach, our future developments will aim in this direction.

Figure 4: Sparse point cloud and stations of terrestrial and UAV images (VisualSFM output).

However, it was possible to process all terrestrial and UAV images in one project and obtain a single connected scene (Figure 4) . As expected, it was not possible to successfully include the close-up images collected from the scaffold. Also, it turned out during the dense matching procedure, that the registration accuracy was insufficient. Instead, processing smaller subsets of the data delivered better results, which were co-registered in an extra step.

3.2 Dense Image Matching

The undistorted images and their orientations were passed to our dense image matching software SURE (Rothermel et al., 2012), which incorporates a variant of semi-global matching, with a multi-view stereo triangulation step. The quality of the obtained point clouds has a large variety. As expected, the point clouds derived from the UAV images and those derived from the images collected from the scaffold (Figure 5), yield good to very good results. In contrast, the point clouds derived from terrestrial images yield a medium to bad quality. Nonetheless, redundancy and applying different filters to the point clouds delivered a sufficient overall data quality.

Figure 5: The six point clouds captured from the scaffold registered into a common coordinate system.

3.3 Point Cloud Registration

In a first step the point clouds obtained by laser scanning were registered using the ICP (iterative closest point) functionality provided by the free software package Meshlab. Here, by manually choosing a few tie-points, an initial rigid body transformation is applied, and is refined afterwards by the actual algorithm. The transformation can be set to keep one of the original coordinate systems fixed.

In a second step, a selection of the vast number of point clouds obtained from structure-from-motion and dense image matching was registered to the final laser data. Here, the available scaling option was activated to obtain metric results.

3.4 Orthophoto and CAD Drawing

In order to produce metric orthographic projections, a small software tool, which allows to define a local coordinate system for the projection and is capable of dealing with the relatively high amount of noise, had to be developed. The density of points is tracked along the projection ray and the first found maximum is chosen as the correct depth. Points in the neighbourhood of this solution are aggregated to further reduce noise. These simple features are usually not given in other tools. The orthophotos served as a basis for manually deriving the requested contour lines as a collection of 15 CAD drawings (Figure 6, right and Figure 7, right). Not only orthophotos depicting the colour information (Figure 7, left) of the images were used, but also orthoimages showing the normal vector directions (Figure 6, left) proved useful.

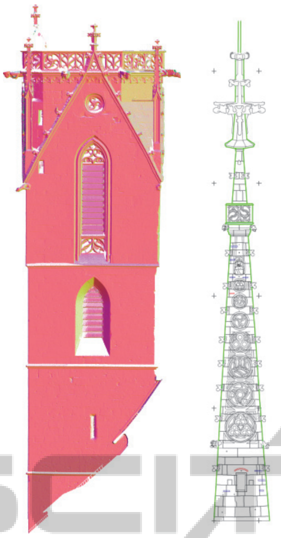

Figure 6: Left: Highlighting geometrical contours by color-coding normal vector directions (eastern facade). Right: Drawing of the south-western side of the tower helmet.

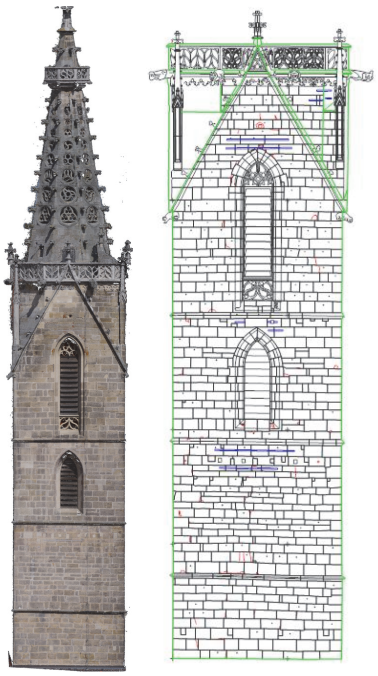

Figure 7: Left: Orthophoto of the tower's south façade, derived from UAV imagery. Right: The corresponding CAD drawing.

4 CONCLUSIONS

Although adverse conditions in the data acquisition step made processing of the data difficult in some points, we can state that in general it is possible to apply the described techniques to heritage applications. To make the procedures applicable for a broader range of users, there is a need to further increase robustness of the software solutions and to integrate a wider range of functionalities to the software packages. When processing point clouds, the usability is often decreased by the high amount of data. Space for improvements may be found in all parts of the processing chain, but are especially needed in structure-from-motion, since it stands in the beginning of the chain and influences all consecutive results. Nonetheless, we observe the given techniques to become more and more attractive to practical application.

REFERENCES

Abdel-Wahab, M., Wenzel, K., Fritsch, D., 2012. Automated and Accurate Orientation of Large Unordered Image Datasets for Close-Range Cultural Heritage Data Recording. *Photogrammetrie – Fernerkundung – Geoinformation (PFG),* Vol. 6, 2012, pp. 679-690.

Crandall, D., Owens, A., Snavely, N., Huttenlocher, D., 2011. Discrete-Continuous Optimization for Large-Scale Structure from Motion, In *Proceedings Conference on Computer Vision and Pattern Recognition (CVPR).* Colorado Springs.

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

Wu, C., Agarwal, S., Curless, B., Seitz, S. M., 2011. Multicore Bundle Adjustment. In *Proceedings Conference on Computer Vision and Pattern Recognition (CVPR).* Colorado Springs.