

AUTOMATED HIGH RESOLUTION 3D RECONSTRUCTION OF CULTURAL HERITAGE USING MULTI-SCALE SENSOR SYSTEMS AND SEMI-GLOBAL MATCHING

J. Wohlfeil^{a,*}, B. Strackenbrock^b, I. Kossyk^b

German Aerospace Center (DLR), Robotics and Mechatronics Center

^a Institute of Optical Information Systems, Dept. of Data Processing for Optical Systems, Rutherfordstr. 2, 12489 Berlin, Germany

^b Institute of Robotics and Mechatronics, Dept. of Perception and Cognition, Münchner Str. 20, 82234 Wessling, Germany
(juergen.wohlfeil, bernhard.strackenbrock, ingo.kossyk) @dlr.de

KEY WORDS: Structure from Motion, Object Reconstruction, 3D Scan, Cultural Heritage Documentation, Robotics, Photogrammetry

ABSTRACT:

3D surface models with high resolution and high accuracy are of great value in many applications, especially if these models are true to scale. As a promising alternative to active scanners (light section, structured light, laser scanners, etc.) new photogrammetric approaches are coming up. They use modern structure from motion (SfM) techniques, using the camera as main sensor. Unfortunately, the accuracy and resolution achievable with the available tools is very limited. When reconstructing large objects with high resolution an unacceptably high laborious effort is another problem. This paper shows an approach to overcome these limitations. It combines the strengths of modern surface reconstruction techniques from the remote sensing sector with novel SfM technologies, resulting in accurate 3D models of indoor and outdoor scenes. Starting with the image acquisition all particular steps to a final 3D model are explained. Finally the results of the evaluation of the approach at different indoor scenes are presented.

1 INTRODUCTION

The ability to measure entire surfaces of objects and build up a true to scale 3D model of them is of importance in many different fields. One field is the documentation of cultural heritage, reaching from the scale of ancient buildings down to the scale of interiors and sculptures. Other fields are the survey and inspection of tunnels, mines, and ships, as well as the environmental 3D mapping for robotics.

Since many years several technologies have been developed to perform this task. Classically such models are generated via active scanners (light section, structured light, laser scanning, etc.) and textured with photographed color images, if required. When modeling large objects with high resolution this method results in a high effort due to the high number of scans required in order to achieve a high resolution. In addition, the images for the texture have to be taken with up to the same resolution.

When the texture images have to be taken anyway, it obviously is a good idea to generate the 3D model directly from the images, using a photogrammetric approach. Moreover if there is no other data available or the effort to collect it is too high, the approach we present in this paper - often called 'structure from motion' (SfM) - can be of great use. In the last years, several novel solutions and tools appeared, such as Bundler (Snavely et al., 2008) and VisualSFM (Wu, 2013) / Rome in a day (Agarwal et al., 2011). By combining novel feature matchers with robust stitching techniques, they even cope with the fact of having no initial information about the intrinsic and extrinsic parameters of the cameras and the scene.

With respect to the reconstruction of surfaces, these approaches still lag far behind the state of the art in the remote sensing sector. One major reason for this is that the accuracy of the deter-

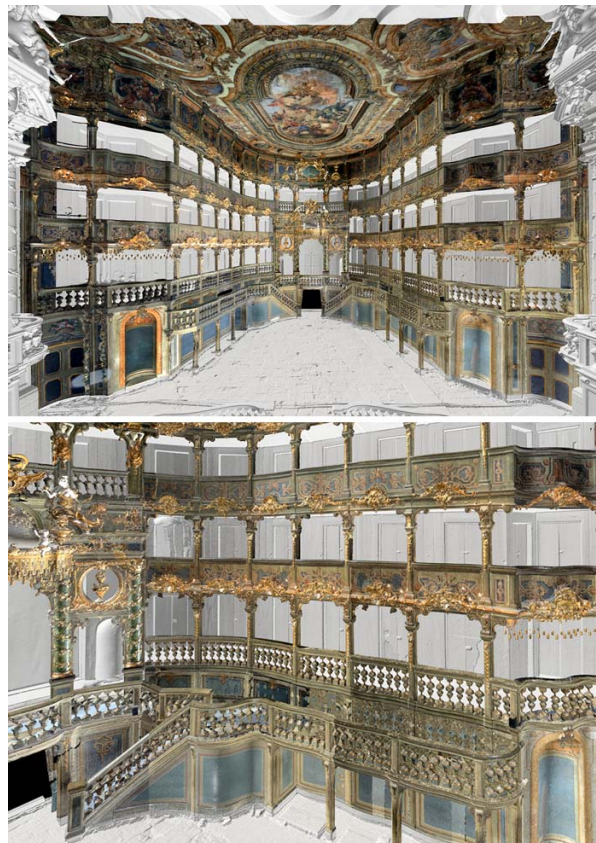


Figure 1: Overview of the margravian opera house with modeled and textured areas, manually cut out and cleaned

mined extrinsics and intrinsics is not sufficient to meet the requirements of high quality dense matching techniques like semi-global matching (SGM) (Hirschmüller et al., 2012).

In this paper it is explained how the experience from the remote sensing sector can be transferred to the terrestrial and indoor sector in order to achieve high quality and high resolution 3D models from such scenes as well. This means, that some boundary conditions for photogrammetric evaluation have to be met, concerning both, camera hardware and software/algorithms. It is shown that it is worth to do so, if the requirements on accuracy and quality of the results are higher than they are for photo tourism, for example.

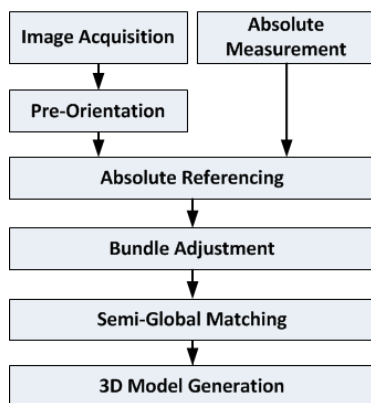


Figure 2: Block-Diagram of the presented approach

Starting with the image acquisition the particular steps to a final 3D model with sub-millimeter resolution are explained (figure 2). First, the hardware and techniques for the acquisition of the images are outlined in section 2. In section 3 it is shown how the rough geometry of the scene can automatically be estimated with little or even without any initial knowledge using well known standard SfM software. As an essential prerequisite for high-quality dense stereo matching the exterior and interior orientation of all images is determined with high accuracy using an own approach. Based on the initial estimates this is performed with an own approach for the bundle adjustment, which is explained in section 4. For dense reconstruction semi-global matching is used and it is shown in section 5 how redundant stereo information can be used to automatically filter matching errors and retrieve high-quality surface models. In section 6 the matching results are assembled to very large reconstructed objects with minimal manual effort using a small number of absolute reference measurements. The images are then re-used for the automatic texturing of the model. Finally the results of the evaluation of the approach at different indoor scenes is presented in section 7.

2 DATA ACQUISITION

The quality of a photogrammetric reconstruction is mainly dependent on the quality of the acquired images. This means that both, the choice of the right camera hardware as well as the technique of capturing the images, have a high impact on the results. In order to interpret the images in a photogrammetric manner, it is vital to be able to keep the geometry of the camera rigid over a certain time. This requirement concerns the whole camera -

the camera body as well as the optics. It implies that most consumer cameras are unsuitable for such tasks. Especially cameras with automatically extending optics, optical image stabilization, rolling shutters, or instable lenses, should not be used due to their obvious lag of rigidity.

Experience has shown that only solid metal optics (Leica M9 and hyper primes from the film industry) with fixed focal length and without auto focus are appropriate. It is also vital that the sensor is physically fixed to a constant position with respect to the optics. Many cameras are able to move the image sensor in order to clean it from particles or to allow the usage of optics from different manufacturers. Unfortunately, such mechanisms reduce the rigidity of the whole camera geometry and should be avoided. Cameras without anti-aliasing filters in front of the sensor allow for a sharper image and are therefore preferable for 3D reconstruction.

Especially in indoor scenes with difficult lightning conditions high-end industrial cameras with a precise temperature control of the sensor can help to improve the image quality noticeably. If using consumer cameras the gain and the ISO settings should remain in their default settings in order to avoid unnecessary image noise. If the natural or artificial lightning is insufficient, professional LED lights or flashes proved to be clearly preferable to conventional discharge lamps due to the stability of their color. If they are connected to the camera trigger they consume very little energy and provide high flash rates, if needed.

As mentioned earlier, it is important to know that every change of the focus, the aperture, or any other change affecting the optical beam path results in a changed camera geometry. From a photogrammetric point of view a change of the camera geometry means creating a different (photogrammetric) camera. As it is practically impossible to restore the camera geometry after changing it, it is useful to capture as many images as possible with a constant geometry, then change mandatory settings (e.g. the focus to get closer to the object) and then capture another series of images, and so on. As a result, for every image within a series the same (photogrammetric) camera applies. It is useful to keep the number of (photogrammetric) cameras as low as possible as the intrinsic parameters for each of these cameras have to be determined. This can be performed via classical calibration methods (e.g. using a calibration board, etc.) or on the fly (e.g. as explained in section 4).

The best technique for image acquisition depends on the specific scenario. Going from large-scale to small-scale, the best choices for different scenarios are explained below.

For large-scale terrain or facade models images are typically captured from an aerial platform. Regular image mosaics with a constant overlap are captured from a manned or unmanned aerial vehicles for this purpose. For best results an image overlap of 80% to 90% is recommended. Good results have been achieved with UAS (Unmanned aerial systems) that can be programmed to capture the required images systematically from certain positions. We recommend this approach for the 3D reconstruction of vast terrain in particular. However in the context of automated indoor 3D reconstruction the usage of UAS for image acquisition is not feasible in many cases.

For the reconstruction of complex building interiors a very high resolution is often required. While for the documentation of the

architecture of a room a resolution of the model in the range of 2 mm is sufficient analysis and evaluations from artists and restorers has shown that for 3D documentation of furniture, decoration and sculptures a resolution down to 0.5 millimeter can be mandatory (Eckstein, 2008). For such multi-scale scenes images with different image resolutions have to be taken. As a result the focus of the camera has to be changed, meaning that the whole scene includes more than one (photogrammetric) camera. Hence, the images have to be assigned to the correct (photogrammetric) cameras and, if the camera is not calibrated explicitly after each change, an on the fly calibration is very useful for such scenes.

After every change of the camera geometry oblique images of the object should be taken. Such images are not useful for dense matching but support the geometric integrity of the image mosaic as they represent a link between several remote images. This, in turn, increases the absolute accuracy of the model in areas between absolute measurement, especially if the cameras are calibrated on the fly.

In order to ensure the required absolute accuracy of the 3D model a small number of reference points has to be measured with a laser scanner or tacheometer. The points should be well distributed in the scene and are picked manually in the images (although we are working on a solution to select them automatically by matching reflectance images of the laser scan with the images captured by the camera).

To support the initial image orientation, the angular difference of orientation of the cameras of subsequent images should be kept below 15 degrees. Changes of the image scale, which can occur either changing the distance to the object or using another (physical) camera or optics for some images, should be kept below 30% to prevent a loss of reliability of the feature matcher.

In order to avoid negative influences of lossy compression and to make use of the full dynamic range of the camera the raw images should be used instead of the 'optimized' and compressed output of most consumer cameras whenever possible.

3 BASIC SCENE GEOMETRY

In order to determine the initial values for the extrinsic and intrinsic parameters of the scene standard SfM software is used. The intrinsic parameters can be determined by an a-priori calibration of the camera, e.g. using the calibration utility Callab (Strobl et al., n.d.). These parameters are then used as a fixed calibration setting for the SfM software package VisualSfM (Wu, 2013). If cameras with significantly different geometries were used in one scene the intrinsic parameters cannot be specified for VisualSfM. In this case they are estimated automatically starting with the nominal values given in the EXIF tags of the image files. Even if these values are not present VisualSfM can determine a good estimate. Unfortunately, this increases the risk of fatal errors in the reconstruction process due to the ambiguity of geometric realities.

For processing with VisualSfM the raw images have to be converted to JPEG format. The lowest possible compression (if any) should be used. The standard settings for image resolution and image count have to be modified in the configuration file of VisualSfM, else the images are rescaled, which can lead to significant

degradation of the accuracy of the results. After bundle adjustment with VisualSfM, the extrinsic camera parameters - namely camera position and orientation - only describe the relative geometry within the scene up to an unknown scaling factor. The absolute extrinsic parameters are determined via control points from reference measurements. For this purpose control points can be entered in to VisualSfM and picked manually in the images.

The result is a set of extrinsic parameters and individual intrinsic parameters for every image which is stored to an NVM file. At this point it becomes obvious whether or not the images were captured adequately in terms of sufficient overlap, stereo angles, lightning, etc. If the geometry of the scene, which can be visualized in the software, contains big deformations, fragmented sections or missing parts then the quality of the input images should be evaluated. If the geometry appears to be reasonable the accurate geometry of the scene can be determined in the next step.

4 REFINING BUNDLE ADJUSTMENT

According to (Hirschmüller, 2008), the relative orientation between every pair of images being matched has to be very accurate in order to perform semi-global matching. It has to be good enough to predict the epipolar lines with at most half a pixel of accuracy. Due to several reasons the standard SfM software (Bundler, VisualSfM, etc.) is not designed to determine highly accurate relative orientations of the images, as needed for optimal dense matching results.

One reason for that is that they allow either fixed intrinsic parameters for all images within one scene or variable intrinsics for every single image. As mentioned in section 2, the scenes should be captured in series of images, each taken by one (photogrammetric) cameras. This implies that it is known which images share which camera but this useful information cannot be used in standard software, including most of the available bundle adjustment software libraries.

As a result the intrinsics of every image is calculated separately. This is a problem because the determined intrinsic parameters typically differ to a small degree between one image and another, which is mainly caused by small errors in the used keypoint detectors and feature descriptors. If not all images are perfectly interlinked (by features) with each other then these small differences sum up to large discrepancies of the intrinsics of images taken with knowingly the same (photogrammetric) camera. This problem becomes obvious when modeling long corridors or tunnels. But it already reduces the accuracy of the model in scenes where not every image contains the whole object being modeled.

Another reason why standard SfM software is inappropriate for accurate reconstruction is that the used camera models are not particular sufficient to describe the camera's geometry precisely. Usually only one radial symmetric distortion parameter is used. For most optics it is not sufficient to describe their radial symmetric distortion with sub-pixel accuracy. On the other hand, increasing the number of parameters bears a high risk of over-fitting the model and it increases the number of unknowns even more. The unnecessary high number of unknowns leads to an increased difficulty of outlier¹ detection. Most outliers cannot be detected by

¹Outliers are incorrectly matched features

analyzing the feature descriptors or the corresponding image regions, e.g. when they are caused by repetitive structures, moving objects, changing shadows, reflecting surfaces. etc. Fortunately they can be detected during the bundle adjustment steps, as explained later. But a higher number of unknown parameters implies a higher risk that a set of parameters is found that does not only explain the correct features but also the outliers in a plausible way.

In order to achieve the best possible absolute accuracy of the generated model it is mandatory to make the bundle adjustment as stable as possible. In particular, this is vital in large scenes because a less stable bundle adjustment leads to a higher amount of absolute measurements required in order to provide a high absolute accuracy.

Due to these reasons and the high degree of automation needed to process scenes with hundreds and thousands of images a customized bundle adjustment software has been developed. It allows to take into account the prior knowledge available on the scene, including control points as well as the definition of image-to-camera associations. The current implementation is based on Christopher Zach's Simple Sparse Bundle Adjustment code (Zach, 2011). Other than most of the available open source bundle adjustment libraries its design allows the optimization of global parameters of the scene together with image-specific parameters. Thanks to this design feature it is possible to implement the optimization of intrinsic parameters of many cameras, each assigned to several images.

Starting with the initial values for the intrinsic and extrinsic parameters as determined by the SfM software and the control points provided by the absolute measurement, the bundle adjustment is initialized. It optimizes these parameters with respect to minimize mean residuals of all feature points.

Incorrect feature points (outliers) are detected and eliminated iteratively: After every optimization of the unknown parameters the residual of each single feature point is evaluated. The points with the highest residuals (> 2 standard deviations) are eliminated and the bundle adjustment is repeated without these points.

As a criterion for termination of the outlier detection iteration and as a quality measure a special epipolar error measure is introduced. The epipolar error is defined to be the RMS of the intersection errors between the predicted epipolar lines and the actual corresponding points in the image. As there are no independent check points, the feature points are taken for the calculation of the epipolar error. It is important to take into account that the feature points - selected with SIFT - are not matched with sub-pixel precision. Their positions are rounded to full pixel positions, which means that a partial error is caused by quantization noise. Due to the high number of features this uncorrelated noise does not influence the solution but affects the magnitude of the error, which has to be considered.

In terms of practical use the measure of the epipolar error has turned out to work adequately. Nevertheless it is statistically questionable and there are ways to improve this method, as discussed in section 8.

5 DENSE MATCHING

As the result of the refining bundle adjustment step the prerequisites are given for dense matching. We chose SGM (Hirschmüller, 2008) for this step as it turned out to achieve better results in many cases than other stereo matching methods (Hirschmüller and Bucher, 2010) and other technologies, e.g. laser scanning (Gehrke et al., 2010). Moreover, stereo matching can be performed more economically than its technological alternatives, as it does not require additional sensors. Despite of its relatively high computational complexity the computation time can be handled very well by parallelization and/or optimization for special hardware like graphic cards (Ernst and Hirschmüller, 2008) and FPGAs (Gehrig et al., 2009).

SGM requires two rectified² images as input. These can be generated using the precisely known relative orientation of the images. Not every pair of images is useful for SGM and due to the high number of images it is practically impossible to choose the suitable pairs of images manually. Thanks to the information available about the geometry of the scene, including intrinsics, extrinsics and homologous points (matched features), these pairs can be found automatically. In order to achieve this all available pairs of images are rated with respect to the size of the image area, e.g. where the image contents overlap and the stereo angles are suitable for semi-global matching (around 3-30). Based on this rating the best pairs are chosen for each image up to a reasonable maximum number of images, according to the available computation power. A number in the order of 10 pairing images has shown to be a good compromise between performance and quality.

The result of SGM is a disparity map for each image pair. Using the relative orientation of the images again, the disparities can be transformed into the coordinates of the object point corresponding to each pixel (excluding the occluded points and mismatches), which leads to a large and dense point cloud, representing the object's surface. Nevertheless a point cloud is not very useful to describe a surface. Moreover, it may contain many outliers caused by mismatches. In the following section we describe how we handle this issue.

6 SCENE RECONSTRUCTION

The goal of 3D scene reconstruction is to generate a complete and textured mesh, which describes the surface of the objects in the scene down to pixel scale. At this point the photogrammetric process finishes and the computer aided design (CAD) process begins. At first, a partial mesh is generated from the dense matching results of each image. Then the partial meshes are combined to one large mesh of the whole scene. Optionally the mesh can be smoothed, reduced and manually cleaned. For these steps the software of David laser scanner system (Bauer, 2013) proved to give the best results. For all other steps in this part the software framework Scanbox from the ForBAU Projekt (Hirzinger and B.Strackenbrock, 2011) is used.

The model is then textured by the same images which were used for SGM processing. Larger objects with hundreds of thousands

²rectified in terms of stereo rectification, ensuring that every point in the first image can be found in the same image row in the second image

of images have to be split up into several parts before the scene reconstruction steps can be performed. Finally, the scene (or all its parts) can be loaded in commercial tools like 3D Studio for further processing.



Figure 3: Untextured and textured 3D model with sub-millimeter accuracy using the presented approach

7 TEST AND APPLICATION OF THE APPROACH

The presented approach was improved and tested within the scope of the project MuSe Bayreuth, where special cultural heritage objects are modeled using different cameras and laser scanners. The first object is the UNESCO world heritage of the margravian opera house in Bayreuth. As one of Europe's few still existing theaters of the Baroque period the extensive variety of figures is subject of being documented with a resolution of 0.5 mm.

The first attempt was to make about 150 panoramic laser scans using a ZF 5006h in the mode 'very high resolution', 'low noise quality'. These scans have been co-registered via manually selected homologous points using a regression calculation, achieving an accuracy of about 2 mm. For the textures about 750 images using a 28 mm NIKON 800E camera and another 600 images with a 12 mm Sony Next7 with flash light have been captured. These images were co-registered to the laser-scans and used to texture the 3D data, resulting in a very high manual effort.

For the presented photogrammetric approach the images were divided into two blocks and processed as explained in this paper. A few homologous points have been selected in the laser scans and the images and are used as control points for absolute reference in the bundle adjustment.

After the refining bundle adjustment an epipolar error (measured at independently and manually picked check points in the original images for the evaluation of the test) of 0.687 pixels was achieved. Due to the Bayer mosaic of RGB imaging sensors SGM is run on images which were reduced to half resolution³. This means that the effective epipolar error is half the error at full resolution and, hence, indicates a suitable accuracy of the relative orientation.

The evaluation of the absolute accuracy of the model was performed using control points extracted from various laser scans (ZF 5006h) by comparing their positions with the corresponding points of the model and revealed an absolute accuracy of 3 mm RMS. The final model is displayed in figure 1.

Since 2013 the opera house is restored, therefore the scaffold is providing the opportunity to take photos of the artwork and masks very closely. As a first test of the reconstruction with very high resolution, a mask has been photographed with a NikonF800E from a distance of 150 cm. The resulting model is shown in figure 3 demonstrating the high resolution of fine structures. The determination of the accuracy of the model is difficult as there is no reference data available with a higher accuracy than the photogrammetric reconstruction itself. Due to fact, we chose the method 'partially free adjustment' (Kotowski, 1996) of the commercial photogrammetric tool CAP (Hinsken, 1989), which allows evaluating the accuracy of various 3D points of the model while taking the low precision of the control points into account. These control points were taken from a close laser scan (ZF 5006h) with a local precision of about 1 mm. The analysis showed that the accuracy of the model is 0.4 mm RMS within the laser scanner coordinate system, which confirms the visual impression of the restorers.

The third experiment has been performed with the royal slide of Ludwig II., exhibited in the Marstallmuseum Schloss Nymphenburg. The golden surface of the object is very challenging due to its high reflectance - even for most other 3D reconstruction techniques. According to the approach presented in this paper, 90 images were taken with a Sony Nex-7 using rigid 12 mm optics and a LED flash light system from a distance of about two meters. Along with ten reference tacheometer measurements the data was processed as described. The resulting model has a resolution of 2 mm and is shown in figure 4. The absolute accuracy of the model is roughly 10 mm, which still is slightly too high. Unfortunately the reason for this was not particularly clear when this paper was written. Nevertheless, already the current approach gives good results regarding the challenging surface properties of the slide.

8 CONCLUSIONS AND OUTLOOK

With the presented approach the strengths of modern surface reconstruction techniques common in remote sensing have been combined with novel SfM technologies, resulting in accurate 3D models of indoor and outdoor scenes. It turned out to be a very promising and common-sense solution in terms of labor and computational effort, even for challenging high resolution 3D reconstruction tasks of large objects and challenging object surfaces / material properties. The test presented in this paper emphasizes

³As the sensor of the camera has a bayer mosaic the true resolution of each color band is half the image resolution in each image dimension

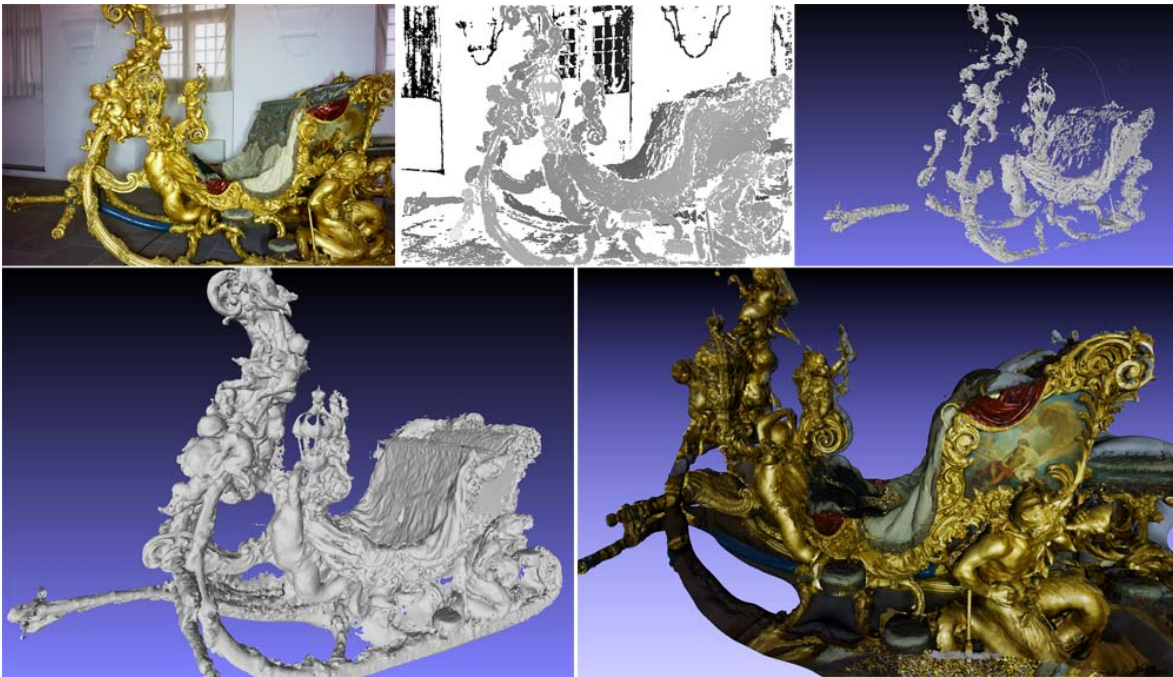


Figure 4: Different stages towards the 3D model of the royal slide: Upper left: original image. Upper center: depth map from SGM. Upper right: Partial 3D model coming from this image (and pairing images). Lower left and right: Untexturized and textured 3D model

this experience. It demonstrates that the accuracy of the models meets requirements for a documentation of cultural heritage and has high potential to be used in many other fields.

More extensive tests are subject of further investigation. In particular, an evaluation of high-resolution reconstruction should be repeated using a test object with a shape known with sub-millimeter accuracy.

The mentioned standard SfM software greatly supports the presented approach enabling the estimation of the rough intrinsics and extrinsics of every image even without any initial values. However, experience has shown that this process can fail due to various reasons. Suboptimal poses and angles, changes in lighting conditions and loosely connected image mosaics are only some of the reasons that can cause the procedure to fail. Another limitation is the computation time. In many cases it is necessary to match all images with all other images resulting in exponential computational complexity. This fact practically limits the number of images to a couple of hundreds, using normal PCs, or few thousands when using a cluster computer.

Due to these problems there is work in progress to combine an Integral Positioning System (IPS) (Griessbach et al., 2012) with a rigidly connected high resolution camera. The IPS system tracks the positions and orientations of the camera, using the stereo camera pair to ensure the correct scale of the model. It is impossible to determine the scale without absolute reference measurements when using only one camera. The tracked positions and orientations can then be used as initial parameters for the camera's extrinsics. This initial estimate simplifies and speeds up feature matching, and increases the robustness of the initial reconstruction step. It is expected that the initial values from the IPS can directly be used as initial values for the refining bundle adjustment.

Another goal is a more efficient implementation of bundle adjustment. The SSBA-based implementation only runs on one CPU core. Parallelized implementations for the bundle adjustment problems exist, e.g. Multicore Bundle Adjustment (Wu et al., 2011), and are subject of further evaluation.

Even though proven to be feasible, the current approach for the determinations of outliers (as described in section 4) bears the risk to eliminate the wrong points. If a set of parameters found by bundle adjustment causes the points in some regions of the image to fit perfectly with the models, while the residuals of points in other regions are high, then these points are eliminated systematically, although they are not outliers. The typical RANSAC approach is difficult to apply as the number of feature points, and so the number of potential outliers, is very high. A practicable solution for this problem is subject of further research.

Also the calculation of the epipolar error will be improved. One could use a part of the matched points for the bundle adjustment and another part as independent check points to evaluate the quality of the result. Another option is the independent comparison of the image's correlation around the predicted epipolar lines at control points that are used in the bundle adjustment.

Within the MuSe project image acquisition will be performed with a PCO Edge camera system beginning from April 2014. Thanks to the high dynamic range of 20,000:1 of the sCMOS imaging technology and a linear 16 bit output signal for each color band this system is of great value, especially for challenging/reflecting surfaces of masks and artwork. It can be coupled with an LED flash light and frame rates up to 50 fps can be captured and processed with the camera. This means that the camera can be handled like a handheld movie camera in order to capture the backsides of figures and complex scenes with ease.

9 ACKNOWLEDGEMENTS

At this point we want to thank the Bavarian Research Foundation for support the MuSe Bayreuth project. Thanks also to the Bavarian administration of the state palaces, gardens and seas, enabling the recordings at the margravian opera house.

REFERENCES

- Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M. and Szeliski, R., 2011. Building Rome in a Day. *Communications of the ACM* 54(10), pp. 105–112.
- Bauer, N., 2013. Der DAVID-Laserscanner: Erprobung und Analyse geeigneter 3D-Messprinzipie zur berührungslosen optischen Messung anthropometrischer Daten. Bachelor-Thesis at Technische Universität Ilmenau.
- Eckstein, G., 2008. Empfehlungen für Baudokumentation, Landesdenkmalamt Baden-Württemberg.
- Ernst, I. and Hirschmüller, H., 2008. Mutual Information based Semi-Global Stereo Matching on the GPU. *International Symposium on Visual Computing, Las Vegas*.
- Gehrig, S., Eberli, F. and Meyer, T., 2009. A Real-Time Low-Power Stereo Vision Engine using Semi-Global Matching. *International Conference on Computer Vision Systems (ICVS)*.
- Gehrke, S., Morin, K., Downey, M., Boehrer, N. and Fuchs, T., 2010. Semi-Global Matching: An Alternative to LIDAR for DSM Generation? *Int. Arch. Photogr. Remote Sens. Spatial Inf. Sci.*
- Griessbach, D., Baumbach, D., Boerner, A., Buder, M., Ernst, I., Funk, E., Wohlfeil, J., and Zuev, S., 2012. IPS A system for real-time navigation and 3D modeling. *Int. Arch. Photogr. Remote Sens. Spatial Inf. Sci.*
- Hinsken, L., 1989. CAP - Ein Programm zur kombinierten Bündelausgleichung auf Personal Computern. *Zeitschrift für Photogrammetrie und Fernerkundung* 57(3), pp. 92–95.
- Hirschmüller, H., 2008. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), pp. 328–341.
- Hirschmüller, H. and Bucher, T., 2010. Evaluation of Digital Surface Models by Semi-Global Matching. DGPF Vienna.
- Hirschmüller, H., Buder, M. and Ernst, I., 2012. Memory Efficient Semi-Global Matching. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* I(3), pp. 371–376.
- Hirzinger, G. and B. Strackebrock, 2011. Digitale Baustelle - innovativer Planen, effizienter ausführen pp. 50–59.
- Kotowski, R., 1996. Photogrammetrische Bündeltriangulation zur 3D-Objektrekonstruktion und simultanen Sensorkalibrierung in der Nahbereichsphotogrammetrie. *VDV-Schriftenreihe VDV-Schriftenreihe*, pp. 22–31.
- Snavely, N., Seitz, S. M. and Szeliski, R., 2008. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision* 80(2), pp. 189–210.
- Strobl, K. H., Sepp, W., Fuchs, S., Paredes, C., Smisek, M. and Arbter, K., n.d. DLR CalDe and DLR CalLab.
- Wu, C., 2013. Towards Linear-time Incremental Structure from Motion. *Proc. of International Conference on 3DTV-Conference* pp. 127 – 134.
- Wu, C., Agarwal, S., Curless, B. and Seitz, S. M., 2011. Multicore Bundle Adjustment. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR, IEEE)*, pp. 3057–3064.
- Zach, C., 2011. Simple Sparse Bundle Adjustment (SSBA), <http://www.inf.ethz.ch/personal/chzach/opensource.html> (checked Oct. 2013).