



# Automated and Accurate Orientation of Large Unordered Image Datasets for Close-Range Cultural Heritage Data Recording

MOHAMMED ABDEL-WAHAB, KONRAD WENZEL & DIETER FRITSCH, Stuttgart

**Keywords:** automatic image orientation, 3D reconstruction, bundle adjustment, close range, cultural heritage

**Summary:** Reconstruction of image orientations and geometry from images is one of the basic tasks in photogrammetry and computer vision. A fully automated solution of this task in terrestrial applications is still pending in case of large unordered image datasets especially for close-range and/or low-cost applications. Current solutions require high computational efforts for image networks with high complexity and diversity regarding acquisition geometry. Unlike the methods suitable for landmark reconstruction from large-scale Internet image collections we focus on datasets where one cannot reduce the number of images without losing geometric information of the dataset. Within the paper, an automated pipeline for the reconstruction of reliable and precise camera orientation from unordered image datasets is presented. Results for a close-range cultural heritage application, the example of the Amsterdam project, are shown to demonstrate the performance of the presented pipeline for applications with low cost and high accuracy requirements.

**Zusammenfassung:** *Automatische und hochgenaue Orientierung von großen, ungeordneten Bildverbänden bei der photogrammetrischen Aufnahme von Weltkulturdenkmälern.* Die Rekonstruktion von Kameraorientierungen und Objektstrukturen aus Bildern ist eine der Hauptaufgaben der Photogrammetrie und der Computer Vision. Eine vollautomatische Lösung für terrestrische Anwendungen mit unregelmäßig angeordneten Bildverbänden, unabhängig vom Kamerasystem (professionell oder Amateuraufnahmen), steht noch aus. Gegenwärtige Lösungen erfordern einen hohen Rechenaufwand für komplexe Bildkompositionen. Im Gegensatz zu Ansätzen zur Landmarkenrekonstruktion mittels Bilddatensätzen des Internets will der vorliegende Beitrag alle verfügbaren Bilder nutzen, um wertvolle geometrische Details nicht zu verlieren. Aus diesem Grund stellen wir einen automatischen Workflow für die Rekonstruktion von reproduzierbarer und präziser Geometrie aus ungeordneten Bildkompositionen vor. Dieser wurde für eine spezielle Anwendung bei der Rekonstruktion der beiden Tympana des Königlichen Palastes in Amsterdam entwickelt und getestet. Die Ergebnisse belegen die Leistungsfähigkeit des Gesamtkonzepts wie auch der im automatischen Workflow realisierten Einzelpakete hinsichtlich niedriger Kosten und hoher Genauigkeit.

## 1 Introduction

In the past few years, close-range and/or low-cost photogrammetry has become a focus of research especially since cameras enable data acquisition at very low prices, but with high geometric and radiometric quality. Therefore, low-cost multi-camera systems for efficient

dense point cloud recording from imagery are particularly suitable for cultural heritage applications, where the requirements regarding acquisition efficiency, flexibility, but also spatial resolution and precision are high. For high resolution data recording in cultural heritage applications, the use of a rig with multiple cameras is beneficial. With one shot mul-

multiple views enable the dense reconstruction of the object surface using dense image matching methods, such as semi global matching (HIRSCHMÜLLER 2008). Thus, the high similarity between the imagery is exploited to generate one high resolution point cloud with a low amount of occlusions for each shot. However, for multiple shots the imagery requires an automatic registration method. Therefore, a structure-from-motion (SfM) reconstruction method was developed in the context of an example project of data recording using five industrial cameras. They are mounted on a compact square shaped rig, which enables handheld recording of complex objects. However, for large scale data recording the derivation of accurate orientation for the high number of imagery is a key problem to be solved.

The aim of this paper is to report a pipeline for fully automatic derivation of image orientations by using a divide-and-conquer strategy to speed up the SfM process from general imagery networks without initial orientation values. SfM was originally developed to estimate geometry and camera motion from multiple images of a scene. It is used for the determination of initial values for the final and global bundle adjustment step in our pipeline. Most SfM methods are starting with a small reconstruction, i.e. a pair or triplet of images, and then expanding the bundle incrementally by adding new images and 3D points as in SNAVELY et al. (2008). Here, each pose estimation and point triangulation is followed by an outlier rejection and a bundle adjustment. Other approaches increase the bundle hierarchically by merging smaller reconstructions (FARENZENA et al. 2009). Unfortunately, both approaches require multiple intermediate bundle adjustment results and rounds of outlier removal to minimize error propagation as the reconstruction grows due to the incremental approach. This can be computationally expensive for large datasets. This issue is considered to be solved partially in FARENZENA et al. (2009) by the introduction of a local bundle adjustment procedure and in SNAVELY et al. (2008) by optimizing the system over a graph to order the images and remove obsolete images from the dataset. However, we focus on datasets where one cannot reduce the number

of images dramatically without losing a substantial part of the model.

A third solution is the so called partitioning method, which we follow in this paper, as presented in GIBSON et al. (2002). Here the key idea is reducing the problem to smaller and better conditioned sub-problems. The main advantage of these methods is not only the equalized error distribution on the entire dataset but also a speed up of the processing time. Recently, KLOPSCHITZ et al. (2010) presented a robust and flexible SfM pipeline where they used the image triplets as a base to reason about feature track compatibility and image connectivity. Within the following section the example of the cultural heritage data recording project in Amsterdam is briefly introduced. A description of the reconstruction of orientation and structure is given in section 3. In section 4 the clustering process of the global graph is described, where the large dataset is split into smaller clusters. Within section 5 the incremental reconstruction being performed for each cluster is discussed. Section 6 contains the stitching process of the multiple clusters to one cluster. Furthermore, the global bundle adjustment of the resulting cluster is explained. Experimental results for the presented dataset are shown in section 7, followed by the conclusions in section 8.

## 2 The Amsterdam Project

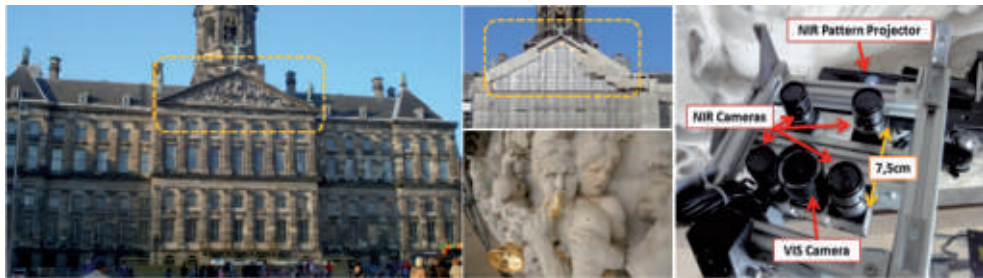
In March 2011 the Research Group "Photogrammetry and Computer Vision" of the Institute for Photogrammetry (University of Stuttgart) had an industrial contract to collect photos for a very dense 3D point cloud generation of the two Tympana of the Royal Palace in Amsterdam (Fig. 1). Each tympanum covers a triangular shape area of about 25 m in width by 5 m in height containing a relief with complex surface geometries such as statues. For this purpose it was planned to use a multi-camera system incorporating a fully automated pipeline for image orientation and dense matching methods. Thus, a method capable of processing very large image datasets with high accuracy and sufficient time was required here. The first comprehensive reports about the project are presented in FRITSCH et

al. (2011) and WENZEL et al. (2011). The sensor, as shown in Fig. 1 right, consists of four cameras used for the dense image matching and one camera with a larger field of view for the registration of multiple shots. The four cameras for the dense image matching have a resolution of 5 Megapixels and are equipped with lenses with a focal length of 8 mm. They are arranged in a square with the size of 7.5 cm by 7.5 cm on a solid aluminum bar to provide a stable relative orientation. The fifth camera with a resolution of 2 Megapixels, equipped with a lens with 4.7 mm focal length, is installed between the lower two matching cameras. An aluminum frame is surrounding the cameras in order to protect them from damage. Several mounts for the connection to tripods and arms for a flexible use are installed at this frame. The Microsoft Kinect is attached at the top with the pattern projector at the same height like the cameras to minimize occlusions.

In order to derive a point cloud with a sampling of 1 mm on the object and sub-mm accuracy, the presented sensor was employed for the data recording on scaffolding. Within 9.5 days about 2,000 stations were acquired leading to a total amount of about 10,000 images. First, to achieve a complete coverage, the images were acquired in nadir direction in a meandering pattern within each of the three levels of scaffolding. Then convergent shots have been captured to complete surfaces which were occluded or not covered. Ground control points measured by tachymetry provided the transformation to the global coordinate system.

### 3 Orientation Reconstruction Pipeline Overview

Our 3D reconstruction pipeline intends to automatically and accurately process unordered sets of images to determine relative image orientations and a sparse point cloud of tie points without prior knowledge of the scene. The pipeline, as shown in Fig. 2, mainly consists of four processing steps: (1) Employ fast image indexing to avoid costly matching of all possible image pairs, which dominates computational complexity along with the multiple bundle adjustment steps. (2) Generate tie points by means of feature extraction and matching where the required automatic measurements are realized at maximum accuracy and reliability. (3) Building and optimizing a geometry graph based on the image network, whereby the dataset can be split into reliable clusters of neighbouring images that can be processed independently and in parallel within the reconstruction step. (4) Merge all clusters and then finally adjust the full model with integrating the ground control points. A detailed description of the individual processing steps is given in the following sections. In general, camera calibration parameters are not strictly necessary for Euclidean 3D modeling, since self-calibration methods exist. However, if a stable camera with a fixed focal length is used and the values for the interior orientation are determined a priori by standard calibration methods robustness and accuracy are usually greatly improved. Furthermore, also an increase in processing speed is achieved due to the lower dimensionality of the prob-



**Fig. 1:** East tympanum of the Royal Palace of Amsterdam; left: from distance; upper middle: with the scaffolding; lower middle: a DSLR colour image of the scene shows the relief containing whole statues (size of visible control point target is  $4 \times 3 \text{ cm}^2$ ); right: sensor design overview.

lem. Pursuant to that, we prefer to use interior calibration parameters for high accuracy applications where these values can be considered to be stable.

### 3.1 Initial Network Geometry Analysis

This step is designed to accurately and quickly index unordered collections of photos. A *connectivity* matrix is the output of this step and it reveals singleton images and small subsets that should be excluded from the dataset. Finally, it is used to guide the process of pairwise matching (section 3.2). Recent developments regarding this analysis can be distinguished into two major categories according to the type of image representation. Local feature based approaches use quality measures of matched local descriptors while global feature based approaches utilize matching histograms of full images visual words (ALY et al. 2011). In fact, both categories represent the same approach with varying degrees of approximation to improve speed and/or storage requirements. Generally, the first category provides superior recognition performance and the dimensionality is not an issue when only several thousands of images need to be processed. Consequently, we utilise a local feature based method in the pipeline presented in this paper.

For local feature-based indexing, we follow an approach adapted to the method presented in BROWN & LOWE (2003) and FARENZENA et al. (2009). The first step is the extraction and description of local invariant features from each image by using the SURF (BAY et al. 2006) operator on a downsampled image, e.g. using images with 2 Megapixels resolution. Then all descriptors are stored in a randomized forest

of kd-trees, which represents the search dataset, to improve the effectiveness of the representation in high dimensions. Each descriptor in the query image is matched to its nearest neighbours in feature space (we used 10). For that purpose we used the fast library for approximate nearest neighbours FLANN (MUJA & LOWE 2009) and the kd-tree implementation in the VLFeat library (VEDALDI & FULKERSON 2008).

Thus, each feature in a query image is initially matched to 10 features of the search dataset. Then we indicate the outliers which have a feature distance more than a certain threshold. We used twice the standard deviation of the distances as threshold value. That gives us a statistical information about the number of matched features between the query image and the remaining images which we can store in a 2D histogram. For more efficiency we used a weighted 2D histogram where the inverse of the distances between each matched feature pair are used as weights. Furthermore, we introduce additional quality measures for possible connections between the query image and the remaining images such as the approximate area of overlap derived from the convex hull of the matched feature points. Finally, the quality measures and the 2D histograms are normalized and summarized to one single quality 2D histogram, which is stored in the *index* matrix (Fig. 3a). Then this *index* matrix is binarized to the *connectivity* matrix, using three thresholds determined empirically, to determine initial probable connections as shown in Fig. 3b. Here any image pair with a quality value more/less than first/second threshold is indicated as connected/disconnected pair. The number of connected images will be compared to the third threshold, which

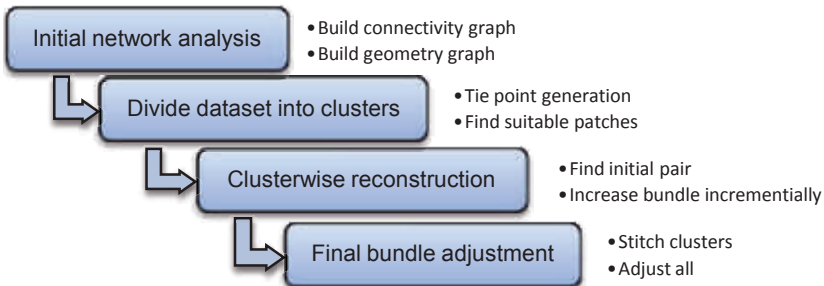
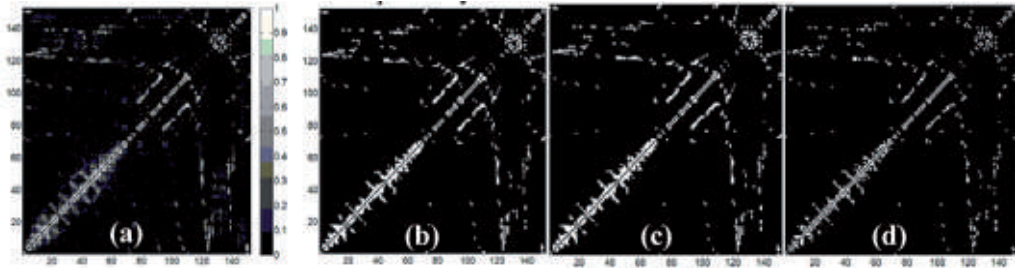


Fig. 2: Flowchart of the presented pipeline.



**Fig. 3:** Top cluster of the east tympanum dataset. The axes refer to the image identifiers in the dataset. (a) Index matrix according to the method presented in section 3.1 with 1457 edges, colour-coded between one/zero indicating connected/disconnected pair; (b) adjacency connectivity matrix before geometry verification, where the number of edges is reduced to after binarization; (c) adjacency connectivity matrix after geometric verification and only 2 edges removed (section 3.2). A white entry in a connectivity matrix indicates that the image pair is connected. (d) Connectivity matrix after the graph optimization step with 396 edges (section 5.1).

refers to the minimum number of images connected to the query image. If this threshold is not met, images with values between the first and second threshold will be added in descending order until the condition is satisfied. We used 0.7, 0.3 and 10 as the threshold values respectively.

### 3.2 Pairwise Feature Matching

Matching each connected image pair is accomplished using the *connectivity* matrix obtained during the previous section. Corresponding 2D pixel measurements are determined between all connected image pairs. Afterwards a weighted undirected graph, we call it *geometry* graph  $G_E = (V, E)$  where  $V$  is a set of vertices and  $E$  is a set of edges, is constructed. Thus, two view relations are encoded such that each vertex refers to an image while each weighted edge presents the overlap between the corresponding image pairs. The weights of the edges are stored according to the number of their common points,  $w_{ij}^p$ , and the overlap area,  $w_{ij}^a$ , between view  $i$  &  $j$ . For the computation we follow the approach of FARENZENA et al. (2009) where a set of candidate features is matched using a kd-tree procedure based on the approximate nearest neighbour algorithm. This step is followed by a refinement of correspondences using an outlier rejection procedure based on the noise statistics of correct/incorrect matches. The results are then filtered

by a standard *RANSAC* based geometric verification step, which robustly computes pairwise relations. Homography and fundamental or essential, in the calibrated case, matrices are used with an efficient outlier rejection rule

**Tab. 1:** Pseudo code for the clustering approach.

<b>Input:</b>	Geometry graph $G_E$
<b>Output:</b>	Collection of clusters graph
1.	Set new empty graph (cluster) $G_c = \{\}$
2.	Determine most reliable edge $E_{ij}$ in $G_E$
3.	Add the vertices $V_i, V_j$ of this edge into $G_c$
4.	Set $E_{ij} = 0$ in $G_E$
5.	$\forall V_k$ in $G_E$ connected at least with two vertices $V_n, V_m$ in $G_c$ If $w_{nk}^p$ & $w_{mk}^p \geq \max\left(\frac{1}{2}w_{ij}^p, 200\right)$ & $w_{nk}^a$ & $w_{mk}^a \geq \frac{1}{2}w_{ij}^a$ <ul style="list-style-type: none"> <li>• Add <math>V_k</math> into <math>G_c</math> and set <math>E_{nk} = E_{mk} = 0</math> in <math>G_E</math></li> </ul>
6.	Add edges in between inlier vertices in $G_c$
7.	Set all these edges = 0 in $G_E$
8.	Repeat steps 5, 6 and 7 until <ul style="list-style-type: none"> <li>• if <math>V_k = 0</math> in step 5</li> <li>• or if size of <math>G_c =</math> predefined value</li> </ul>
9.	Store $G_c$ and repeat all steps until all edges in $G_E = 0$

called *X84* (HAMPEL et al. 1986) to increase reliability and accuracy. The final output of this step is the *geometry* matrix or graph as illustrated in Fig. 3c. For an in-depth discussion see FARENZENA et al. (2009) and SNAVELY et al. (2008) and references therein.

## 4 Clustering of the Global Graph

In order to speed up the computation of the incremental reconstruction, we address a fast local optimization instead of a global optimization approach. We divide the dataset into  $n$  overlapping clusters, where each one contains a manageable size of images. Thus, a parallelizable process replaces the process of reconstruction of the whole scene at once. This is particularly important since for complex datasets the large number of iterations with the growing number of unknowns can lead to very high computation times for complex datasets. The idea is to start from the most reliable part and use three images as the basic entity to extend each cluster until a predefined size. In practice, we use the workflow as presented in Tab. 1 to identify reliable clusters with the highest mutual compatibility. The idea is to start each cluster  $G_c$  from image pair  $i$  and  $j$  with high overlap in order to ensure a reliable geometry. As shown in Tab. 1, we select an initial pair according to the most reliable edge being identified by its weights,  $w_{ij}^p$  and  $w_{ij}^a$  within the geometry graph as presented in section 3.2. The graph of this cluster is then extended by the neighbouring edges with a weight (common points) greater than the half of the weight of the initial edge of this cluster or a certain threshold (we used 200 matching points for the presented dataset). We repeat this process until a predefined cluster size is reached or until no more images have sufficient overlap with this cluster. For this we apply thresholds depending on the initial edge weights. Furthermore, only images overlapping with at least two images inside the cluster are considered. While the cluster graph is growing, each used edge is eliminated in the geometry graph. As soon as the cluster graph is finalized, the whole procedure is repeated to find the next cluster until all imagery is covered. The overlap between the final clusters

is ensured by considering and removing connections (edges in the graph) only instead of images. Thus, common cameras between the clusters remain.

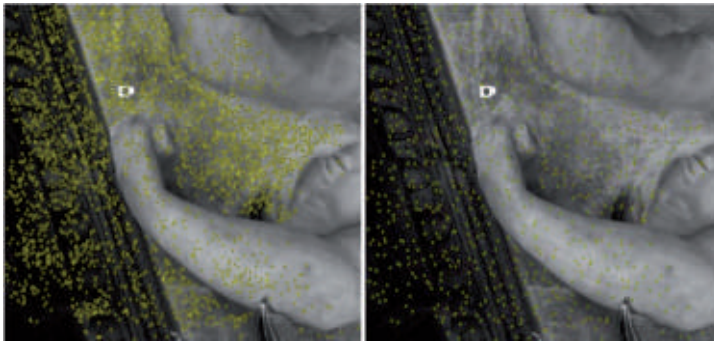
## 5 Cluster Reconstruction

Once the clusters are divided as described in the previous section, we can start the reconstruction process for each cluster as follows.

### 5.1 Optimization of Cluster Graph

For each cluster we track the keypoints only over images in this cluster (locally) and store the results in a visibility matrix, which depicts the appearance of points in the images. The results of this step will be the keypoints which have been correctly tracked in at least three images. For more efficiency, we apply a non-maximum suppression filtering approach (Fig. 4) for the tracked points to keep only the points with the highest connectivity. For each image we sort the keypoints in descending order according to their number of projections in other images. Then, the point with the greatest number of projections is visited, followed by an identification and rejection of all nearest neighbour points with a distance less than a certain threshold, e.g. 20 pixels. This step is repeated until the end of the points list. In order to maintain continuity, all points selected in an image must be considered as filtered (fixed) in the following filtering of other images. Filtering is done in order to increase the accuracy but also to reduce the number of obsolete observations. Consequently, the geometric distribution of keypoints is improved, which reduces the computational costs significantly without losing geometric stability.

Once correspondences have been tracked and filtered, we optimize the cluster graph such that we construct a weighted undirected epipolar graph for each cluster  $G_p$  containing common tracks. The weight  $w_{ij}$  of an edge represents the number of common points between the corresponding image pair. Then we build  $G_r$ , the edge dual graph of  $G_p$ , where every node in  $G_r$  corresponds to an edge in  $G_p$ . Two nodes in  $G_r$  are connected by an edge if



**Fig. 4:** Point distribution before and after filtering, 3395 & 819 points according to a filtering distance of 40 pixels.

and only if they are sharing an image and 3D points. Thus, each edge represents an image pair with sufficient overlap. Note that even if  $G_p$  is fully connected any spanning tree of  $G_p$  may be disconnected. This can happen if a particular pairwise reconstruction did not have 3D points in common with another pair. Thus, we use three images as basic geometric entity by using only points that were tracked in at least three images. These points are used to build the graph in order to guarantee full connection for any sub-sequential image. The maximum spanning tree (MST), which maximizes the total edge cost of the final graph, is then computed. The image relation retrieved as  $G_p^{\max}$  graph is used for the bundle adjustment. For example, Fig. 6 presents the results of the top cluster of the east tympanum where the previous process reduced the pairwise connection from 600 edges (Fig. 3d) to 396 to orient 150 images.

## 5.2 Camera and Geometry Recovery

Each cluster is processed individually beginning with an initial reconstruction for the two, most suitable images. After this step, orientations and tie points in object space are available for these two images where one image defines the local coordinate system. Within the incremental approach images are added to the existing bundle by triangulating new points, rejecting outliers and performing another iteration of the bundle adjustment. This incremental process is repeated until all images within the cluster are processed.

### Reconstruction of the initial pair

The incremental reconstruction step begins with the reconstruction of orientation and 3D points for an initial image pair. The choice of this initial pair is very important for the subsequent reconstruction of the scene. The initial pair reconstruction can only be robustly estimated if the image pair has at the same time a reasonable large baseline for high geometric stability and a high number of common feature points. Furthermore, the matching feature points should be distributed well in the images in order to reconstruct a maximum of initial 3D structure of the scene and to be able to determine a strong relative orientation between the images. Therefore, suitable image pairs should be selected according to the following conditions: the number of matching points is acceptable and the fundamental matrix must explain the matching points far better than homography models. Here we employ the geometric robust information criterion (GRIC) scores to ensure that the criteria are met as used in FARENZENA et al. (2009). After that, relative orientation values for this initial pair are estimated by using Nister's implementation of the five point algorithm (NISTER 2004). A two-frame bundle adjustment starting from this initialization is performed to improve the reconstruction.

### Adding new images and points

After reconstructing the initial pair additional images are added incrementally to the bundle. The most suitable image to be added

is selected according to the maximum number of tracks from 3D points already being reconstructed. Within this step not only this image is added but also neighbouring images that have a sufficient number of tracks as mentioned in SNAVELY et al. (2007). Adding multiple images at once reduces the number of required bundle adjustments and thus improves efficiency. Next, the points observed by the new images are added into the optimization. A point is added if it is observed by at least two images, and if the triangulation gives a well-conditioned estimate of its location. This procedure follows the approach of SNAVELY et al. (2007).

### Sparse bundle adjustment

Once the new points have been added, a bundle adjustment is performed on the entire model. This procedure of initializing a camera orientation, triangulating points, and running bundle adjustment is repeated, until no images observing a reasonable number of points remain. For the optimization we employ the sparse bundle adjustment implementation "SBA" (LOURAKIS & ARGYROS 2009). SBA is a non-linear optimization package that takes advantage of the special sparse structure of the Jacobian matrix used during the optimization step in order to provide a computation with reduced time and memory requirements.

## 6 Stitching of Clusters and Global Adjustment

After the reconstruction of points and orientations for the overlapping clusters the results are merged. Since outlier rejection was performed within the previous steps, the available 3D feature points are considered to be reliable and accurate. Due to the overlap, the clusters have a certain number of points and image orientations in common which enables the determination of a seven-parameter transformation in order to align the clusters into a common coordinate system. The transformed orientations and points are introduced into a common global bundle adjustment of the whole block. If ground control point measurements

are available they can be used to improve the bundle stability and to enable georeferencing.

## 7 Experimental Results

For the derivation of the point cloud, the exterior orientations were derived first using the presented method, where each façade was automatically divided into 6 individually processed patches. Secondly, an additional dense image matching step followed using the obtained orientations. However, since the exposure of all cameras was not synchronized sufficiently, the relative orientation was not sufficiently stable. This is particularly important since the acquisition distance was short and the accuracy requirements were high. Thus, the relative orientation from the calibration was omitted and determined using the SfM process instead. In order to use the relative orientation directly, hardware triggering of the cameras should be used instead of software triggering.

Figs. 5 and 6 depict the results of the tympanum at the west façade where approximately 4000 images are oriented. The first row shows the reconstructed and stitched 6 clusters where the mean reprojection error before merging are around 1 pixel and it is reduced to 0.5 pixel after the final bundle adjustment step. The second row left shows the full sparse cloud of 1.1 million feature points in object coordinates and right the dense point cloud derived by a subsequent dense image matching step with about 1.1 billion points. As presented in section 2, the targets are distributed over the whole object and are used for georeferencing. These control points are captured in 12 independent clusters. The white circles within the targets are detected and measured automatically using an ellipse fit. These image measurements are considered to have an accuracy of about 0.1 pixels.

Up to this point the control points are not used in the bundle and thus do not impact the orientations to be evaluated. Consequently, they can be used to assess the quality of the relative orientations by evaluating the reprojection errors. These results are demonstrated in Tab. 2 rows 2 and 5, for each cluster of both tympana respectively. The root-mean-square

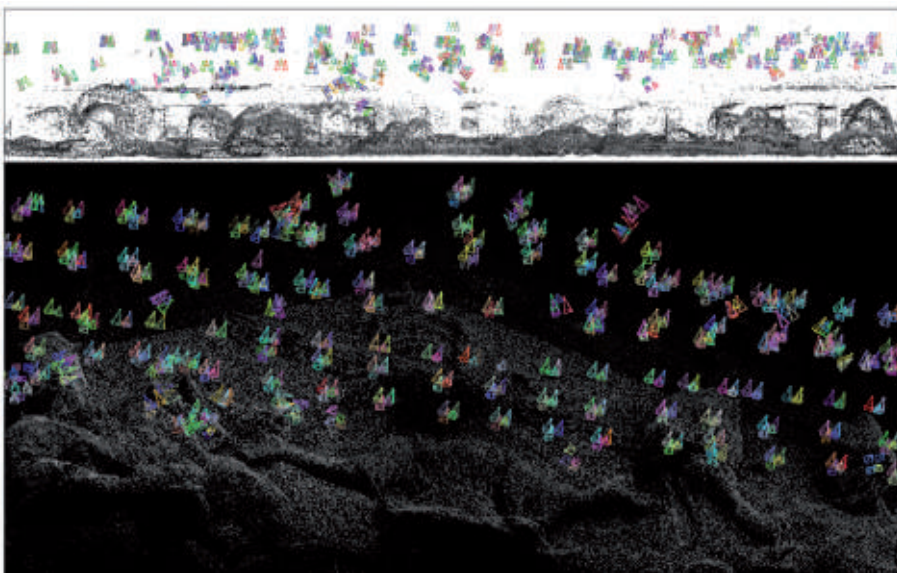


of the reprojection errors for each dataset is about 0.3 pixels. At 70 cm distance this corresponds to an error of approximately 0.2 mm

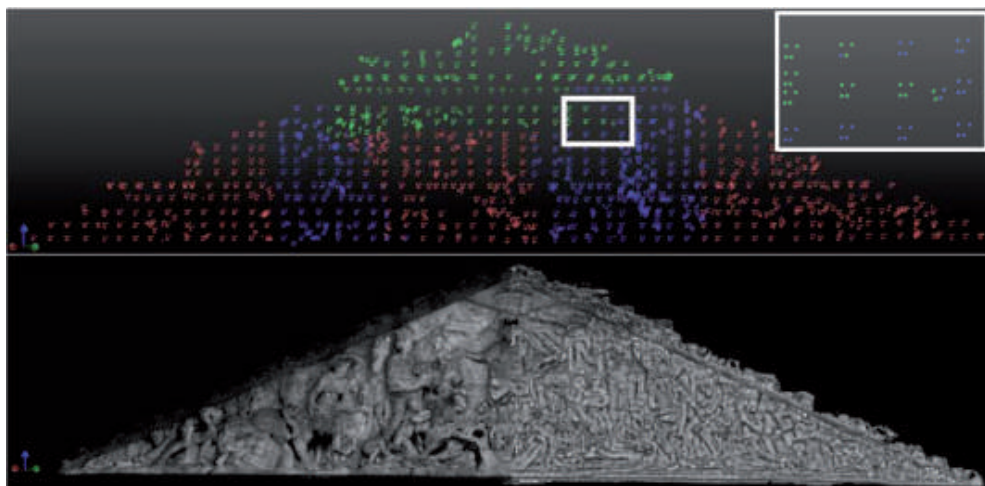
for the image scale of this dataset. This is considered to meet the requirements for the later dense surface reconstruction step, where

**Tab. 2:** Overview of the 6 clusters C1–C6 and their performance of the east and west tympanum. Time = runtime for each cluster, GC point / Projection = identified ground control points per overall count of projections in the image, RMS (pixel) = error of the reprojection, GC = number of ground control points, RMS (mm) = error of a ground control point in object space.

	Cluster Ids.	C1	C2	C3	C4	C5	C6	.
East tympanum	Images	909	526	440	682	412	460	
	3D points	517,360	386,708	300,799	420,020	276,534	258,184	
	Time (hrs)	3.48	1.62	0.61	3.08	0.69	0.78	Mean
	GC points/Proj.	20/163	12/63	11/52	13/80	5/17	9/55	11.67/71.67
	RMS (pix.)	0.39	0.18	0.26	0.26	0.36	0.21	0.28
	GC points	18	6	11	13	5	9	10.33
	RMS (mm)	2.63	2.16	2.18	3.25	2.97	3.21	2.73
West tympanum	Images	862	478	779	995	956	849	
	3D points	318,064	149,349	245,232	318,366	332,613	322,151	
	Time (hrs)	3.47	0.71	3.26	4.21	3.27	3.38	Mean
	GC points/Proj.	24/176	11/65	12/85	15/97	13/95	15/114	15/105.33
	RMS (pix)	0.26	0.26	0.26	0.31	0.26	0.26	0.27
	GC points	20	9	7	14	12	15	12.83
	RMS (mm)	2.24	1.59	1.36	2.06	1.13	1.90	1.71



**Fig. 5:** Reconstructed cameras and point clouds of the tympanum at the west façade. First row: geometry of imagery with close-up area is shown in upper right corner; second row: sparse point cloud resulted from SfM (left) and derived by dense image matching (right).



**Fig. 6:** Image orientations, randomly coloured, and sparse point cloud from SfM for the queen area within the top cluster of the tympanum at the west façade.

a relative accuracy in image space of better than 0.5 pixels is required for a reliable image matching. The control points were measured by reflectorless tachymetry. The mean standard deviation determined by the network adjustment amounts to 1.4 mm in position and 1.6 mm in height. However, many points were occluded and thus could not or only once be measured. However, the RMS values, see the third and the sixth rows in Tab. 2, derived from the residuals after the georeferencing cannot be used for an absolute accuracy assessment. Since a reference measurement with an accuracy of one magnitude higher would be required. In contrast, the accuracy of the tachymetric measurements is in a similar range. Therefore, these values are only used to validate the reconstructed orientations in object space.

## 8 Conclusions

The presented pipeline for the reconstruction of orientations and surface information is specifically designed for the efficient processing of large datasets with high accuracy requirements. An initial network analysis is used along with other techniques to realize a reasonable processing time while adjusting a stable bundle containing information from a maximum number of images. Thus, it is spe-

cifically suitable for large scale photogrammetric applications at low costs. In order to complement the measures of computational efforts it was our goal from the beginning that the whole processing pipeline should run on standard PC environments (for example i3 processors).

## References

- ALY, M., MUNICH, M. & PERONA, P., 2011: Indexing in Large Scale Image Collections: Scaling Properties and Benchmark. – IEEE WACV 2011 Hawaii.
- BAY, H., ESS, A., TUYTELAARS, T. & VAN GOOL, L., 2006: SURF: Speeded Up Robust Features. – Computer Vision and Image Understanding: 346–359.
- BROWN, M. & LOWE, D., 2003: Recognizing panoramas. – 9th International Conference on Computer Vision 2: 1218–1225.
- FARENZENA, M., FUSIELLO, A. & GHERARDI, R., 2009: Structure and Motion Pipeline on a Hierarchical Cluster Tree. – ICCV Workshop on 3-D Digital Imaging and Modeling: 1489–1496.
- FRITSCH, D., KHOSRAVANI, A.M., CEFALU, A. & WENZEL, K., 2011: Multi-Sensors and Multiray Reconstruction for Digital Preservation. – FRITSCH, D. (ed.): Photogrammetric Week '11: 305–323, Wichmann Verlag, Berlin/Offenbach.
- GIBSON, S., COOK, J., HOWARD, T., HUBBOLD, R. & ORAM, D., 2002: Accurate Camera Calibration for Off-Line, Video-Based Augmented Reality.

- IEEE and ACM International Symposium Mixed and Augmented Reality: 37–46.
- HAMPEL, F., ROUSSEEUW, P., RONCHETTI, E. & STAHEL, W., 1986: Robust Statistics: The Approach Based on Influence Functions. – Series in probability and mathematical statistics, Wiley, New York.
- HIRSCHMÜLLER, H., 2008: Stereo Processing by Semi-Global Matching and Mutual Information. – IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2): 328–341.
- KLOPSCHITZ, M., IRSCHARA, A., REITMAYR, G. & SCHMALSTIEG, D., 2010: Robust Incremental Structure From Motion. – 3DPVT 2010.
- LOURAKIS, A. & ARGYROS, A., 2009: SBA: A Software Package for Generic Sparse Bundle Adjustment. – ACM TOMS **36** (1): 1–30.
- MUJA, M. & LOWE, D., 2009: Fast Approximate Nearest Neighbors with Automatic Algorithmic Configuration. – VISAPP 2009.
- NISTÉR, D., 2004: An efficient solution for the five-point relative pose problem. – IEEE PAMI **26** (6).
- SNAVELY, N., SEITZ, S. & SZELISKI, R., 2007: Modeling the World from Internet Photo Collections. – International Journal of Computer Vision **80** (2): 189–210.
- VEDALDI, A. & FULKERSON, B., 2008: VLFeat: An Open and Portable Library of Computer Vision Algorithms, <http://www.vlfeat.org/> (28.8.2012).
- WENZEL, K., ABDEL-WAHAB, M. & FRITSCH, D., 2011: A Multi-Camera System for Efficient Point Cloud Recording in Close Range Applications. – LC3D workshop 2011: Berlin.

Address of the Authors:

M.Sc.Eng. MOHAMMED ABDEL-WAHAB, Dipl.-Ing. KONRAD WENZEL and Prof. Dr.-Ing. DIETER FRITSCH, University of Stuttgart, Institute for Photogrammetry, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, Tel: +49-711-68584117, Fax: +49-711-68583297, e-mail: {mohammed.othman}{konrad.wenzel}{dieter.fritsch}@ifp.uni-stuttgart.de

Manuskript eingereicht: Mai 2012  
Angenommen: August 2012

