PFG 2012 / 4, 0421–0432
Stuttgart, August 2012

Article

Orientation and Dense Reconstruction from Unordered Wide Baseline Image Sets

Jan Bartelsen, Helmut Mayer, Neubiberg, Heiko Hirschmüller, Oberpfaffenhofen, Andreas Kuhn & Mario Michelini, Neubiberg

Keywords: automatic image orientation, dense 3D reconstruction, unordered image sets, wide baseline, semiglobal matching

Summary: In this paper we present an approach for detailed and precise automatic dense 3D reconstruction using possibly unordered image sets from consumer cameras. Opposed to other approaches we focus on wide baseline image sets. We have combined and improved several methods for robust matching and parameter estimation, particularly, least squares matching, RANSAC, scale-space maxima and bundle adjustment. Point correspondences and the five-point algorithm lead to relative orientation without a need for approximate values. Due to our robust matching method we can orient images under much more unfavourable conditions, for instance concerning illumination changes or scale differences, than it would be possible based on commonly used operators such as SIFT. For dense reconstruction, we use our orientation as input for semiglobal matching (SGM) resulting in dense depth images. The latter can be fused into a 2.5D model for eliminating the redundancy of highly overlapping depth images. However, some applications and acquisition scenarios have a need for full 3D modelling, for which preliminary results are presented. Using small unmanned aerial systems (micro UAS), it is possible to acquire images which have a similar perspective as terrestrial images and can thus be combined with them. Such a combination is useful for almost complete 3D reconstruction of urban scenes. We have applied our approach to blocks of several hundred aerial and terrestrial images, generating detailed 2.5D and 3D models of urban areas.

Zusammenfassung: *Orientierung und dichte Rekonstruktion aus ungeordneten Bildverbänden mit großer Basis*. In diesem Beitrag wird ein Ansatz für die detaillierte und genaue automatische dichte 3D Rekonstruktion auf Grundlage von möglicherweise ungeordneten Bildverbänden, welche mit Consumer Kameras aufgenommen wurden, vorgestellt. Im Gegensatz zu anderen Ansätze zielt der vorgestellte Ansatz auf Bilddatensätze mit großer Basis ab. Dafür wurden verschiedene Methoden, insbesondere Kleinste Quadrate Zuordnung, RANSAC, Maßstabsraum Maxima und Bündelausgleichung, für robuste Zuordnung und Parameterschätzung kombiniert und verbessert. Punktkorrespondenzen und der Fünf-Punkt Algorithmus führen zur relativen Orientierung ohne Bedarf für Näherungswerte. Die verwendete robuste Zuordnungsmethodik ermöglicht es, Bilder unter sehr viel ungünstigeren Bedingungen, z.B. bezüglich Beleuchtungsbedingungen und Maßstabsunterschieden, zuzuordnen, als häufig verwendete Operatoren, wie z.B. SIFT. Für die dichte Rekonstruktion wird die berechnete Orientierung als Eingabe für Semiglobal Matching (SGM) verwendet, mit dessen Hilfe dichte Tiefenbilder bestimmt werden. Diese können, um die Redundanz in den oft hochgradig überlappenden Tiefenbildern zu eliminieren, in 2,5D Modellen fusioniert werden. Einige Anwendungen und Aufnahmekonfigurationen benötigen aber eine volle 3D Modellierung, wofür erste Ergebnisse vorgestellt werden. Mit kleinen Drohnen / Unmanned Aerial Systems (Micro UAS) wird es möglich, Bilder zu erfassen, die eine ähnliche Perspektive auf die Szene haben wie terrestrische Bilddaten und daher mit diesen kombiniert werden können. Eine solche Kombination ist für eine fast vollständige 3D Rekonstruktion von städtischen Szenen sehr hilfreich. Der entwickelte Ansatz wurde auf Blöcke mit hunderten von Bildern aus der Luft und vom Boden angewandt und damit detaillierte 2,5D und 3D Modelle von Siedlungsbereichen erzeugt.

DOI: 10.1127/1432-8364/2012/0128 1432-8364/12/0128 $ 3.00

1 Introduction

Pollefeys et al. (2000) have demonstrated that sets of images from consumer cameras in combination with dense 3D reconstruction form a good basis for photo realistic visualization. Pollefeys et al. (2002) presented one of the first approaches for relative orientation for a larger number of images in a general configuration, i.e., without known approximate values such as for aerial images. By employing uncalibrated images, for which the camera constant / principal distance etc. is not known, the approach is very flexible, yet, on the other hand, reliant on sufficient 3D structure in the scene for the determination of calibration parameters. More recently, Pollefeys et al. (2008) have reconstructed 3D structures from sequences with more than one hundred thousand calibrated images.

For the work above, the overlap of the images is assumed to be known, either implicitly in form of the order in a sequence, or explicitly, e.g. in the form of an aerial flight plan. Schaffalitzky & Zisserman (2002) proposed one of the first methods which can automatically determine the overlap of images in unordered image sets. While Havlena et al. (2010) have proposed an approach which works efficiently for thousands of images based on graph optimization. Agarwal et al. (2009) and Frahm et al. (2010) have recently presented approaches which can deal with hundred thousands or even millions of unordered images from Community Photo Collections from the Internet to model urban areas. A major difference between the two is that the former runs on a cloud, the latter on just one multi-GPU (graphics processing unit) PC system. While both approaches are impressive, one has to note that they are based on certain characteristics of the data and several assumptions to make them tractable:

- At tourist attractions many images are taken from nearly the same spot. Thus, many similar images can be found even when extremely downsampling the images.
- The restricted goal is to reconstruct the obvious 3D structure. This leads to impressive 3D reconstructions of highlights, such as the Colosseum in Rome. Yet, there might be images, possibly with wider baselines, which are not linked, but could be used to extend the geometrical coverage or connect the tourist attractions. They are not considered, as it would mean a detailed comparison of many more images.

Opposed to the above approaches, we aim at a detailed and more complete modelling by making use also of wide baseline image sets. We assume that we know the calibration of the cameras, e.g. from the Exif (Exchangeable image file format) tags of the images in combination with a database about different cameras.

In section 2 we present the methods for point detection, matching and robust parameter estimation that we have improved and combined for orientation of possibly unordered wide baseline image sets. For dense reconstruction, the results of our orientation procedure are used as input for semiglobal matching – SGM (Hirschmüller 2008), which we are about to extend to 3D (section 3). We found that due to our precise relative orientation, very good depth estimates were possible also for wider baselines.

Section 4 gives results. We have processed blocks consisting of hundreds of images acquired from small unmanned aircraft systems (micro UAS). For a combination of UAS images with terrestrial images we have generated a preliminary dense 3D reconstruction of a building comprising the roof as well as the facades. In addition, we present preliminary results for dense 3D surface reconstruction from terrestrial images only. Finally, section 5 gives conclusions and discusses future work.

2 Orientation of Unordered Image Sets

This section is split into two parts: In the first part, we describe our orientation procedure. Contrary to state-of-the-art approaches, it relies on given information on overlap between images, e.g. in the form of a sequence. Yet, it is suitable for wide baseline image sets and it provides very accurate results by consequent propagation of covariance information through all steps.

In the second part, we present a preliminary approach for overlap detection for unordered image sets. Due to restrictions of the fast matching approach (Wu 2007) it is based on, it can only deal with short baselines for the time being.

2.1 *Orientation of Wide Baseline Image Sets*

With the scale invariant feature transform (SIFT) Lowe (2004) has presented a powerful solution for the estimation of point correspondences mainly for short baselines. Yet, reliable matching of points for wide baseline images is much harder and, thus, there is a need for improved matching methods. Our approach to wide baselines is based on scale invariant point matching, least squares matching and robust bundle adjustment (Fig. 1).

The developed approach for point matching produces reliable results even in case of major scale differences as well as viewpoint and illumination changes. It is based on normalized cross correlation (NCC), which is relatively invariant against the latter, but only weak compared to the former two. NCC is weak against scale changes, because in this case image patches with the same size in pixels around corresponding points contain different scene parts. To deal with scale differences, we make use of the work of Lindeberg (1994) and determine points in the form of scale space maxima for Differences of Gaussians (DoG). With the information on scale, we downsample image patches with higher resolution, so that they match to the same scene part. Potential correspondences are refined by least squares matching using an affine geometric model (Grün 1985). This results in subpixel relative point positions. Matched points in two and more images and their covariance information are employed for relative orientation of pairs, triplets and image sets.

With the five-point algorithm (Nistér 2004), one can directly compute the relative orientation from calibrated image pairs. That is, no approximate values for the orientation are needed regardless of the geometric configuration of the two images during acquisition. We have embedded a version of the five-point algorithm into RANdom SAmple Consensus – RANSAC (Fischler & Bolles 1981). This allows us to deal with the large proportion of outliers of possibly more than 80 % typical when matching wide baseline images. By using the geometric robust information criterion (GRIC) (Torr & Zisserman 1997), we compare models not only based on the number of

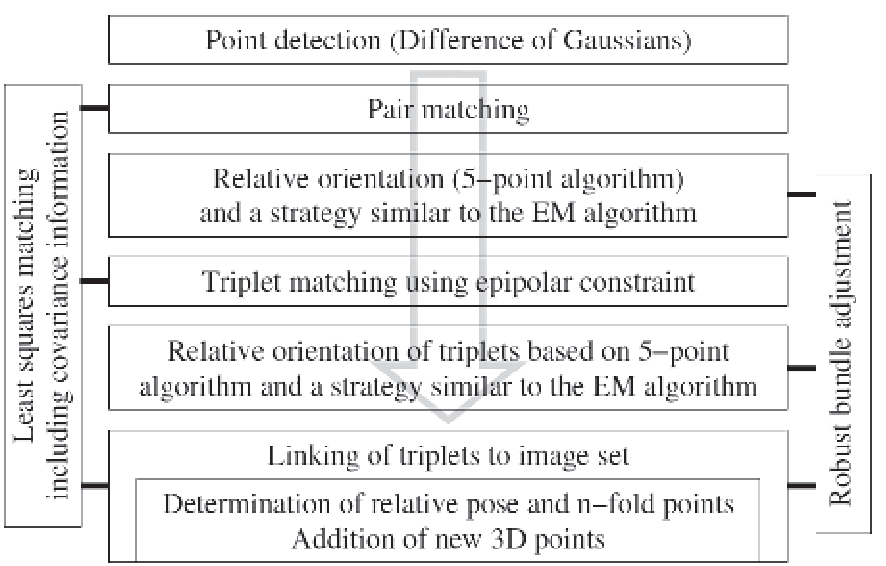

Fig. 1: Image orientation based on scale invariant point matching, least squares matching and robust bundle adjustment.

inliers as in standard RANSAC, but we also take the distance from the ideal solution, i.e., the epipolar line for image pairs, as well as the estimated covariance into account.

As we empirically found that incorrect models can be evaluated very similarly as correct models even when using GRIC, we refine the solution based on a strategy similar to the expectation maximization (EM) algorithm (Bartelsen & Mayer 2010). Partial solutions which initially stem from RANSAC are extended by alternating between robust bundle adjustment with all current inliers (maximization) and the determination of possibly new inliers for the adjusted solution (expectation). This strategy is employed for pairs as well as triplets, with robust bundle adjustment using the covariance information from image matching and reweighting in the form of an M-estimator (Huber 1981) at its core.

For the following reasons, we use triplets as basic geometric building block for image sets:

- While pairs of points can only be checked in one dimension by means of their distance from their respective epipolar lines, triplets permit an unambiguous geometric checking. This does not only result in much more reliable points, but also leads to improved, more reliable information for the cameras.
- Triplets can be directly linked into larger sets by determining their relative pose (translation, rotation and scale) from two common images.

When determining matching points for triplets, we make use of the information derived for image pairs by restricting the matching to a corridor around the epipolar lines. The relative orientation for triplets is computed by taking one image of a triplet as reference, twofold application of the five-point algorithm for the reference image and the other two images, and finally robust determination of the relative scale between the two pairs.

The construction of the relatively oriented image set starts with selecting one triplet. Triplets are linked to the image set based on common image pairs. One common image allows the propagation of translation and rotation. The distance to the second image provides scale. Matched points are propagated from the image set to the linked triplet by means of least squares matching leading to n-fold points, i.e., points that can be seen in $n > 3$ images. New 3D points from the linked triplet are added. Finally, robust bundle adjustment is employed to improve the accuracy, but also the reliability by eliminating wrong matches that could not be detected in the triplets due to the limited redundancy.

Based on the highly reliable relative orientation thus derived, we have shown in (Bartelsen & Mayer 2010) how to calculate the absolute orientation from unreliable and imprecise GPS data of low cost sensors, e.g. in a GPS camera, also in areas with strong occlusions, e.g. cities, similarly to Strecha et al. (2010). Taking other information for absolute orientation, such as ground control points, into account is even easier, as one does not have to deal with a possibly larger number of gross errors as for GPS cameras.

2.2 *Overlap Determination for Unordered Image Sets*

To deal with unordered image sets, we employ automatic overlap detection consisting of the following steps:

- determination of similarity between images,
- construction of a two-view graph and
- construction of a three-view graph.

A fast GPU implementation (Wu 2007) of SIFT is used for detecting points and correspondences by pair-wise image matching. The two-view matching graph consists of images as nodes, whereas its edges connect similar images. The weight of an edge, i.e., the image similarity, is assumed to be given by the number of correspondences between connected images. Images with the number of correspondences below a threshold will be considered as dissimilar and no edge is inserted in the matching graph. To reduce the complexity, available GPS information in the Exif tags of the images is used to derive the distance between images and thus to sort out unlikely pairs.

Once the similarities between the images have been derived, we determine a connected image set by constructing the maximum span-

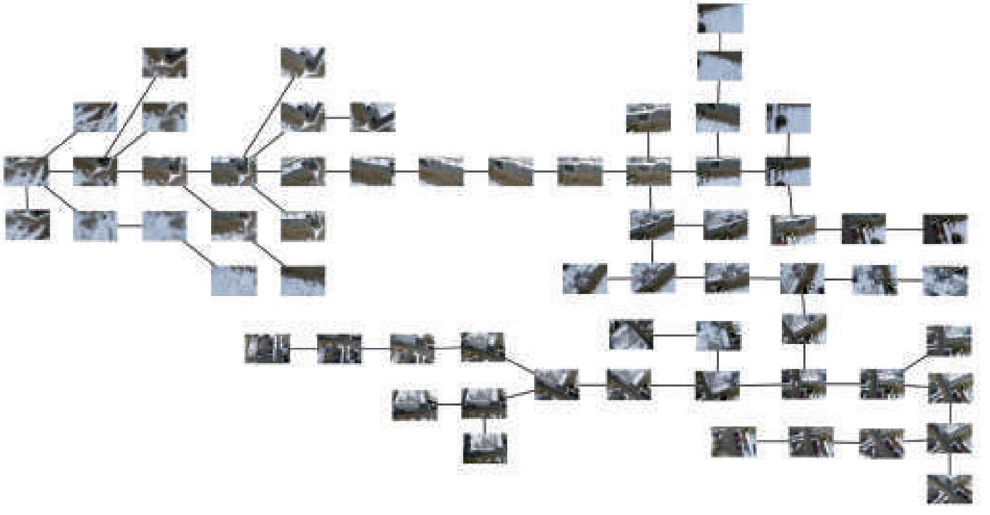

Fig. 2: Minimum spanning tree (MST) for image pairs.

ning tree (MST) of the two-view matching graph using the modified algorithm of PRIM (1957). Fig. 2 is an example, for which results are presented further below in Fig. 6.

Finally, triplets are derived from the connected image set by iterating through the MST using the depth-first traversal algorithm. We discard triplets with the images having a number of correspondences or a normalized overlap area below a threshold (Fig. 3). For the determination of the overlap area between the images of triplets, we calculate the convex hull of correspondences between all three images using the algorithm of SKLANSKY (1982).

The state concerning unordered image sets is still preliminary and, thus, it is only used for the example presented in Fig. 6. Many pairs which can be oriented by our robust matching method described in section 2.1 are not found due to the limited capability of the employed fast matching method (WU 2007). Therefore, whereas we can orient wide baseline combinations of images from the air and from ground, the detection of overlapping images has to be conducted manually at the moment in this case. The integration of our capabilities for wide baseline matching into the overlap determination for unordered image sets is our most important goal for the future.

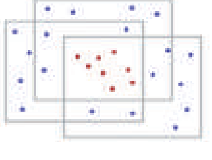

Fig. 3: Number of common points (red – left) and area (green – right) for overlap determination between three images.

3 Dense Reconstruction

For dense reconstruction semiglobal matching – SGM (HIRSCHMÜLLER 2008) is employed. It is based on

- mutual information (MI) or the census filter for cost computation and
- the substitution of a 2D smoothness term by a combination of 1D constraints (semi-global).

MI presents the conditional probability distribution for the intensities in the matching image given an intensity in the reference image, without resorting to a parametric model. Thus, MI can compensate a large class of global radiometric differences. Because the conditional probability is computed for the whole image, problems can arise for local radiometric changes, e.g. if materials with very different reflection characteristics exist in the scene or lighting conditions change.

In Hirschmüller & Scharstein (2009) the census filter was found to be the most robust variant for matching cost computation. It defines a bit string with each bit corresponding to a pixel in the local neighbourhood of a given pixel. A bit is set if the intensity is lower than that of the given pixel. Census thus encodes the spatial neighbourhood structure. A 7×9 neighbourhood can be encoded in a 64 bit integer. Matching is conducted via computing the Hamming distance between corresponding bit strings.

The smoothness term of SGM punishes changes of neighbouring disparities (operator $T[]$ is 1 if its argument is true and 0 otherwise):

$$E(D) = \sum_{\mathbf{p}} \Big(C(\mathbf{p}, D_{\mathbf{p}}) + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_1 \mathrm{T}[|D_{\mathbf{p}} - D_{\mathbf{q}}| = 1] + \sum_{\mathbf{q} \in N_{\mathbf{p}}} P_2 \mathrm{T}[|D_{\mathbf{p}} - D_{\mathbf{q}}| > 1] \Big). \quad (1)$$

- The first term consists of matching costs for all disparities of disparity image D.
- The second term adds a constant penalty P_1 for all pixels $\mathbf{q}$ from the neighbourhood $N_{\mathbf{p}}$ of $\mathbf{p}$ for which the disparity changes only slightly (1 pixel).
- The third term adds a larger constant penalty P_2 for larger disparity changes. Because it is independent of the size of the disparity change, it preserves discontinuities.
- As discontinuities in disparity are often visible as intensity changes, P_2 is calculated depending on the intensity gradient in the reference image (with $P_2 \geq P_1$).

Global minimization in 2D is NP complete for many discontinuity preserving energies $E(D)$. Opposed to this, in 1D, minimization can be done in polynomial time via dynamical programming. The latter is usually applied within image lines. Because the solutions for neighbouring lines are computed independently, this typically leads to streaking. For the semiglobal solution, 1D matching costs are computed in different, (practically 8) directions, which are aggregated without weighting. In the reference image, straight lines are employed, which are deformed in the matching image.

By computing the disparity images D for exchanged reference and matching image one can infer occlusions or matching errors by means of a consistency check. If more than one pair with the same reference image is matched, the consistency check is conducted for all pairs only once (Hirschmüller 2008).

With SGM, very dense disparity maps having one disparity per image pixel can be computed. Using the orientation parameters, all points can be projected into 3D space, leading to dense 3D point clouds. While the original work of Hirschmüller (2008) has shown how to derive 2.5D surface models, work on the derivation of 3D surface models by means of triangulation of the 3D points, dealing also with outliers, has been started only recently.

Because modelling large-scale scenes fully 3D can produce billions of points, efficient processing with regard to the computational and memory costs is a must. We consider octrees to be very suitable for this purpose. Hence, we use a triangulation based on balanced octrees for meshing (Bodenmüller 2009). Besides removing redundancy, octrees are particularly useful for visibility-checks in multiple-view geometry.

Before mesh generation, normal vectors are determined from the neighbours of a point, and points with uncertain normal vectors are eliminated. The triangle mesh is built incrementally. Iterating through all remaining points, the temporary mesh is projected on the tangent plane in a neighbourhood of a

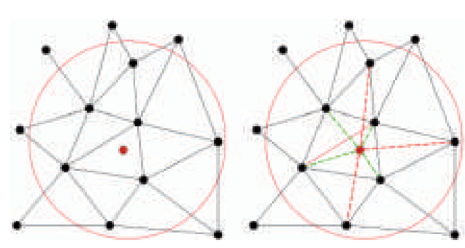

Fig. 4: Local update of triangulation by adding a new vertex to a temporary mesh. Left: New vertex v (red) and projected neighbourhood. Right: New candidate edges (dashed lines). Green lines were accepted and red lines were removed because of intersection with shorter edges.

new point (Fig. 4). The new point is connected with all vertices within the neighbourhood. If a new edge intersects an old edge in the plane, the longer one is removed.

4 Results and Discussion

We have applied our approach for orientation and dense 2.5D and 3D reconstruction to several image sets. As our focus is on wide baseline scenarios, we have manually determined the overlap for all but the results in Fig. 6. For the latter, we demonstrate the potential of our

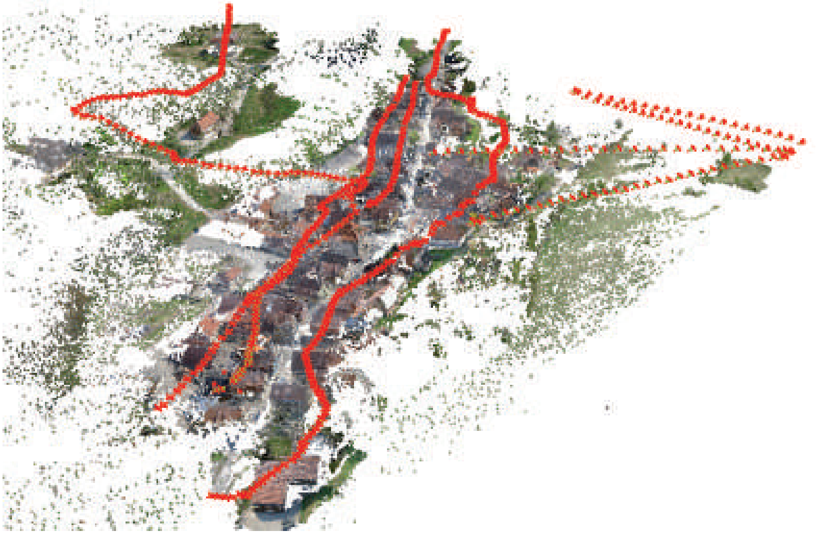

Fig. 5: 3D points and cameras (red pyramids) for a model generated from more than 600 images from a micro unmanned aircraft system (micro UAS).

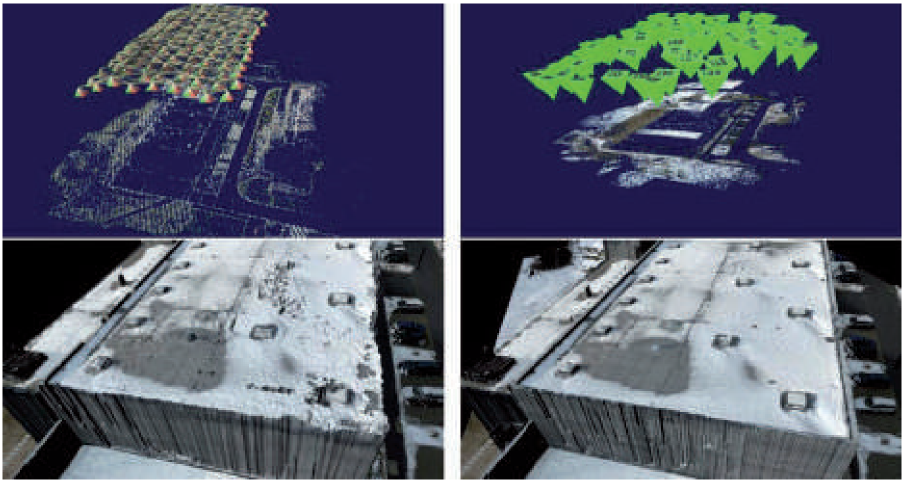

Fig. 6: Top row: 3D points and cameras (pyramids) for a set of 166 images of a large building in Wessling. Bottom row: Resulting 2.5D models. Left column: Result for Bundler (Snavely 2010). Right column: Result derived by our approach – the more precise relative orientation allows for a much more detailed scene reconstruction.

approach for unordered image sets introduced in section 2.2 also in comparison with Bundler (Snavely 2010). In all cases, the census filter (section 3) was used for cost computation for SGM.

Fig. 5 presents 3D points and camera positions (red pyramids) for more than 600 images acquired by a micro unmanned aircraft system (micro UAS) from about 50 m above the ground. Orientation was possible in spite of the lack of approximate values for exterior orientation of this highly non-regular flight configuration. Particularly, for overlapping areas of the flight strips, images with wide baselines could be matched, leading to a more stable geometry.

The result in Fig. 6 is based on 166 images, acquired by a micro UAS. Although the flight was controlled automatically, the obtained image set is not very well structured. Because of too small overlap, many triplets could not be matched. For this image set, we have compared our approach with Bundler (Snavely 2010). In particular, we found that the relative orientation produced by Bundler is not very precise and the 3D point cloud contains many obviously false points. Thus, SGM could only be applied in a meaningful way after downsampling the images to half the original resolution. In contrast, the relative orientation obtained by our approach is much more precise (Tab. 1) and could be used as basis for SGM on

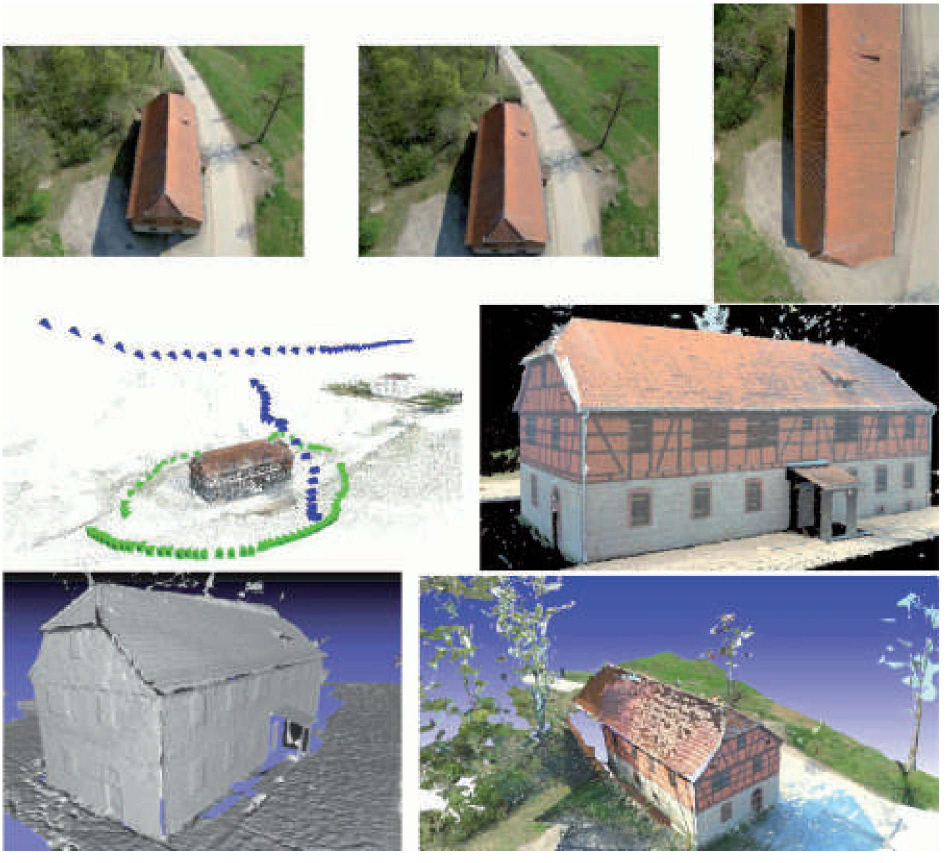

Fig. 7: Top row: Two images from the aerial and one from the 'ascending' sequence which could be matched and oriented. Middle row left: 3D points and cameras derived from a combination of images from the air (blue pyramids) and from the ground (green pyramids) of a building near Hammelburg. Middle row right: 3D point cloud generated by SGM from the flight and the terrestrial sequence. Bottom row: Preliminary result for dense 3D surface reconstruction – shaded with wire frame as well as textured.

the original resolution, leading to a more detailed and realistic 2.5D model.

Fig. 7 presents preliminary results for our new approach for 3D surface reconstruction based on a combination of images from UAS and from ground. The set of 205 images contains three different sequences (Fig. 7 centre row, left):

- The flight sequence was taken from about 20 m above ground.
- The terrestrial image-sequence was acquired around one building.
- The 'ascending' sequence connects images from the air and from the ground. The images change in small steps from the aerial to the ground perspective. This image configuration is only feasible for micro UAS, which can be flown very close to facades and roofs.

The combination of the flight and the 'ascending' sequence is quite difficult, because of major scale differences in combination with small overlap and perspective distortion. Fig. 7 top row shows an example for an image triplet which could be matched. The dense 3D point cloud (Fig. 7, centre right) was generated from images from the terrestrial and the flight sequences (with the relative orientation of both sequences determined using the 'ascending' sequence). It illustrates that roof and walls exactly fit to each other (please note the roof overhang) and thus, that our relative orientation is very precise. Finally, the bottom row of Fig. 7 shows our preliminary 3D surface reconstruction as a shaded view and also as a textured model. The colour variations of the roof texture are caused by the different lighting conditions during acquisition of the ground and the flight sequences.

Fig. 8 shows another preliminary result of our work for full dense 3D surface reconstruction. We have applied our approach to the image sets 'fountain-R25' and 'castle-R20'

Fig. 8: Top: Textured preliminary result for 3D surface reconstruction of the facade of Ettlingen castle. Bottom left: Dense 3D point cloud. Bottom right: Shaded 3D surface with wire frames. The results base on the image-sets 'fountain-R25' and 'castle-R20'of (Strecha et al. 2008).

of STRECHA et al. (2008). Due to our robust matching approach, a combination of both sets was possible.

Finally, Tab. 1 presents the average back projection errors σ_0 as well as computation times for all examples. For all experiments one PC with an Intel Xeon processor with four cores with 2.3 GHz and a 2GB Nvidia Geforce GTX 285 graphics card was used. All back projection errors for our approach are around or below 0.3 pixels, with the majority in the range of 0.15 pixels. Although Bundler (SNAVELY 2010) can produce quite accurate results too, it had severe problems with our challenging wide baseline dataset, reflected by the large reprojection error (0.44 pixels for half resolution) and leading to the inferior 2.5D model presented in Fig. 6.

The computation times for our approach for all but the first experiment are reasonable. Concerning the first experiment we note that it was computed with the sequential version of our orientation algorithm, where robust bundle adjustment is computed for every linked image triplet. This is not necessary and we are, thus, about to replace this by a hierarchical solution, where image sets are combined, which is much more efficient for large sets. For the dataset of Fig. 6, images downsampled to half resolution (i.e., 2.5 Mpixel instead of 10 Mpixel) were used for Bundler for reducing the number of SIFT features. Still, Bundler required longer to finish, because it simply tries to match each image with every other image. In contrast, we use the fast matching approach (WU 2007) for overlap determination and we only try to orient the rather limited number of image triplets for which sufficient overlap could be determined.

5 Conclusions and Future Work

In this paper, we have presented an approach for automatic orientation and dense 3D reconstruction from wide baseline image sets. As key characteristics, it aims at a high precision in every step from least squares matching to robust bundle adjustment. Currently, our approach for full 3D reconstruction does not maintain all the details that are available in the high resolution depth images of semi-global matching. We plan to take into account the uncertainty of the 3D points from different pairs as well as smoothness priors. How this can be effectively and efficiently done is subject of our current research. Our ability for very precise relative orientation is of fundamental importance for accurate 3D modelling.

Although our point matching approach is pretty robust against scale and illumination changes, it is still not robust enough concerning viewpoint changes. Currently, arguably the best known concept for matching which is robust concerning viewpoint changes is the approach of MOREL & YU (2009) simulating off-image-plane rotations. This concept has not yet been integrated into an approach for 3D reconstruction, also due to its prohibitive computational complexity. The improvement of our approach by a similar, yet more efficient procedure is also part of our future work.

Acknowledgements

We gratefully acknowledge the support of Dipl.-Ing. FLORIAN SEIBEL, Dr. rer. nat. PATRICK REIDELSTÜRZ and Prof. Dr.-Ing. PETER STÜTZ during the acquisition of the aerial image sets and we thank the reviewers for the helpful comments.

Tab. 1: Average reprojection error σ_0, number (#) images and computing times for orientation of the image sets in the given figures.

	σ_0 (pixels)	# images	Computing time (hours)
Fig. 5	0.12	603	170
Fig. 6: Our approach	0.31	166	5
Fig. 6: Bundler (half resolution)	0.44	166	8
Fig. 7	0.14	205	8
Fig. 8	0.11	47	2

References

Agarwal, S., Snavely, N., Simon, I., Seitz, S. & Szeliski, R., 2009: Building Rome in a Day. – Twelfth International Conference on Computer Vision: 72–79.

Bartelsen, J. & Mayer, H., 2010: Orientation of Image Sequences Acquired from UAVs and with GPS Cameras. – Surveying and Land Information Science **70** (3): 151–159.

Bodenmüller, T., 2009: Streaming Surface Reconstruction from Real Time 3D Measurements. – Dissertation, Technische Universität München.

Fischler, M. & Bolles, R., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. – Communications of the ACM **24** (6): 381–395.

Frahm, J.-M., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S. & Pollefeys, M., 2010: Building Rome on a Cloudless Day. – Eleventh European Conference on Computer Vision **IV:** 368–381.

Grün, A., 1985: Adaptive Least Squares Correlation: A Powerful Image Matching Technique. – South African Journal of Photogrammetry, Remote Sensing and Cartography **14** (3): 175–187.

Havlena, M., Torii, A. & Pajdla, T., 2010: Efficient Structure from Motion by Graph Optimization. – Eleventh European Conference on Computer Vision **II:** 100–113.

Hirschmüller, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information. – IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2): 328–341.

Hirschmüller, H. & Scharstein, D., 2009: Evaluation of Stereo Matching Costs on Images with Radiometric Differences. – IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (9): 1582–1599.

Huber, P., 1981: Robust Statistics. – John Wiley & Sons, Inc. New York, USA.

Lindeberg, T., 1994: Scale-Space Theory in Computer Vision. – Kluwer Academic Publishers, Boston, USA.

Lowe, D., 2004: Distinctive Image Features from Scale-Invariant Keypoints. – International Journal of Computer Vision **60** (2): 91–110.

Morel, J. & Yu, G., 2009: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. – SIAM Journal on Imaging Sciences **2** (2): 438–469.

Nistér, D., 2004: An Efficient Solution to the Five-Point Relative Pose Problem. – IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (6): 756–770.

Pollefeys, M., Vergauwen, M. & Van Gool, L., 2000: Automatic 3D Modeling from Image Sequences. – International Archives of Photogrammetry and Remote Sensing **33** (B5/2): 619–626.

Pollefeys, M., Verbiest, F. & Van Gool, L., 2002: Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery. – Seventh European Conference on Computer Vision, Volume **II:** 837–851.

Pollefeys, M., Nistér, D., Frahm, J.-M., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S.-J., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewénius, H., Yang, R., Welch, G. & Towles, H., 2008: Detailed Real-Time Urban 3D Reconstruction from Video. – International Journal of Computer Vision **78** (2–3): 143–167.

Prim, R., 1957: Shortest Connection Networks and some Generalizations. – Bell Systems Technical Journal **36:** 1389–1401.

Schaffalitzky, F. & Zisserman, A., 2002: Multiview Matching for Unordered Images Sets, or "How Do I Organize My Holiday Snaps?". – Seventh European Conference on Computer Vision **I:** 414–431.

Sklansky, J., 1982: Finding the Convex Hull of a Simple Polygon. – Pattern Recognition Letters **1:** 79–83.

Snavely, N., 2010: Bundler: Structure from Motion for Unordered Image Collections. – phototour.cs.washington.edu/bundler (9.5.2012).

Strecha, C., von Hansen, W., Van Gool, L., Fua, P. & Thoennessen, U., 2008: On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. – Computer Vision and Pattern Recognition: 1–8.

Strecha, C., Pylvänäinen, T. & Fua, P., 2010: Dynamic and Scalable Large Scale Image Reconstruction. – Computer Vision and Pattern Recognition: 358–365.

Torr, P. & Zisserman, A., 1997: Robust Parametrization and Computation of the Trifocal Tensor. – Image and Vision Computing **15:** 591–605.

Wu, C., 2007: SiftGPU: A GPU Implementation of Scale Invariant Feature Transform (SIFT). – cs.unc.edu/~ccwu/siftgpu (9.5.2012).

Addresses of the Authors:

Dipl.-Inf. Jan Bartelsen, Prof. Dr.-Ing. Helmut Mayer, Dipl.-Inf. Andreas Kuhn, M.Sc. Mario Michelini, Bundeswehr University Munich, Institute of Applied Computer Science, D-85577 Neubiberg, Tel.: +49-89-6004-3429, Fax: +49-89-6004-

3906, e-mail: {jan.bartelsen}{helmut.mayer}{andreas.kuhn}{mario.michelini}@unibw.de

Dr. Heiko Hirschmüller, Dipl.-Inf. Andreas Kuhn, Department of Perception and Cognition, Institute of Robotics and Mechatronics, German Aerospace Center (DLR), e-mail: heiko.hirschmueller@dlr.de

This article is a modified version of a paper accepted for the ISPRS Congress 2012 in Melbourne, Australia.

Manuskript eingereicht: März 2012
Angenommen: Mai 2012