AUTOMATIC FUSION OF PARTIAL RECONSTRUCTIONS

Andreas Wendel, Christof Hoppe, Horst Bischof, and Franz Leberl

Institute for Computer Graphics and Vision
Graz University of Technology, Austria
{wendel, hoppe, bischof, leberl}@icg.tugraz.at
http://aerial.icg.tugraz.at

Commission III

KEY WORDS: Fusion, Reconstruction, Registration, Close Range, Aerial, Robotics, Vision

ABSTRACT:

Novel image acquisition tools such as micro aerial vehicles (MAVs) in form of quad- or octo-rotor helicopters support the creation of 3D reconstructions with ground sampling distances below $1\,cm$. The limitation of aerial photogrammetry to nadir and oblique views in heights of several hundred meters is bypassed, allowing close-up photos of facades and ground features. However, the new acquisition modality also introduces challenges: First, flight space might be restricted in urban areas, which leads to missing views for accurate 3D reconstruction and causes fracturing of large models. This could also happen due to vegetation or simply a change of illumination during image acquisition. Second, accurate geo-referencing of reconstructions is difficult because of shadowed GPS signals in urban areas, so alignment based on GPS information is often not possible.
In this paper, we address the automatic fusion of such partial reconstructions. Our approach is largely based on the work of (Wendel et al., 2011a), but does not require an overhead digital surface model for fusion. Instead, we exploit that patch-based semi-dense reconstruction of the fractured model typically results in several point clouds covering overlapping areas, even if sparse feature correspondences cannot be established. We approximate orthographic depth maps for the individual parts and iteratively align them in a global coordinate system. As a result, we are able to generate point clouds which are visually more appealing and serve as an ideal basis for further processing. Mismatches between parts of the fused models depend only on the individual point density, which allows us to achieve a fusion accuracy in the range of $\pm 1\,cm$ on our evaluation dataset.

1 INTRODUCTION

Novel image acquisition tools such as Micro Aerial Vehicles (MAVs) in form of quad- or octo-rotor helicopters support the creation of 3D reconstructions with ground sampling distances below $1cm$ and gain importance in photogrammetry (Eisenbeiss, 2004) (Eisenbeiss, 2009). The limitation of aerial photogrammetry to nadir and oblique views in heights of several hundred meters is bypassed, allowing close-up photos of facades and ground features.

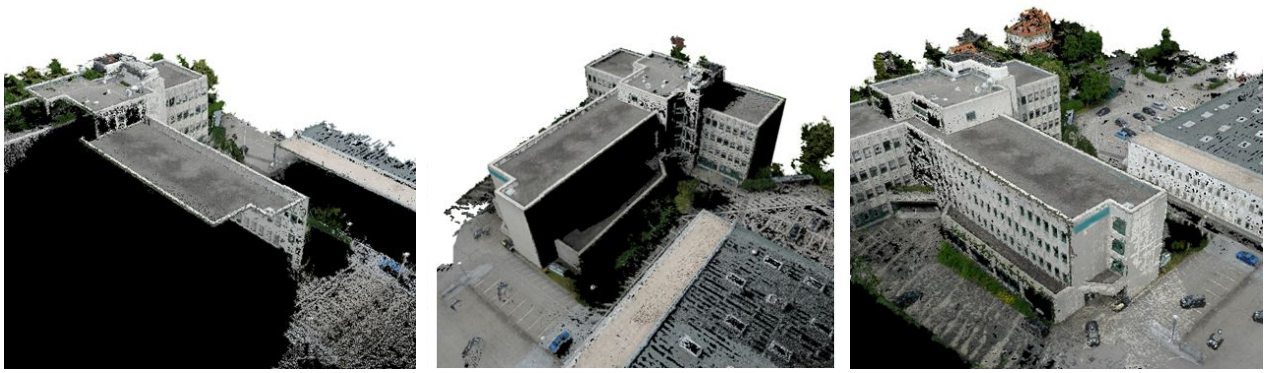
Next to several benefits, the new acquisition modality also introduces challenges: First, geo-referencing of 3D reconstructions is often difficult because of shadowed GPS signals in urban areas, so accurate alignment purely based on GPS information is not possible. However, proper alignment to a world coordinate system is not only beneficial to applications where the model should be set into context, for instance in industrial applications such as construction site monitoring (Kluckner et al., 2011), but crucial if further algorithmic steps depend on it as in automatic view planning (Schmid et al., 2012). Second, flight space might be restricted in urban areas, which leads to missing views for accurate 3D reconstruction and causes fracturing of large models. This could also happen due to vegetation or simply a change of illumination during image acquisition.

In this paper, we address the automatic fusion of such partial Structure-from-Motion (SfM) 3D reconstructions and aim on generating a single, larger and denser 3D model of a scene. Our approach is largely based on the work of (Wendel et al., 2011a) who addressed the issue of geo–referencing partial reconstructions. In contrast, our approach does not require an overhead digital surface model for model fusion. Instead, we exploit that patch-based semi-dense reconstruction of the fractured model typically results in several point clouds covering overlapping areas, even if sparse feature correspondences cannot be established. We approximate orthographic depth maps for the individual parts and iteratively align them in a global coordinate system. As a result we are able to generate point clouds which are visually more appealing and serve as an ideal basis for further processing.

We evaluate our approach using two outdoor scenes, consisting of several partial reconstructions with more than a million points each. We demonstrate that our approach can not only fuse reconstructions from airborne imagery, but also closes the gap between aerial and terrestrial photos. Mismatches between parts of the fused models depend only on the individual point density, which allows us to achieve a fusion accuracy in the range of $\pm 1\,cm$ on our evaluation dataset. Figure 1 shows a typical fusion of two partial reconstructions.

2 RELATED WORK

The problem of aligning 2D images or 3D models to a 3D structure is well studied, especially in the context of large-scale city modeling. (Frueh and Zakhor, 2003) present an algorithm to fuse close–range facade models acquired at ground level with a far–range DSM recorded by a plane. The models are created using both ground-based and airborne laser scanners, as well as digital cameras for texturing. Their approach is based on registering the edges of the DSM image to the horizontal scans of a ground model using Monte-Carlo-Localization. Similarly, (Strecha et al., 2010) register facades segmented from a 3D point cloud to building footprints. Their approach combines various visual and geographical cues in a generative model, which allows robust treatment of outliers. However, both approaches are focused on large-scale city models with flat facades to both sides, resulting in fairly clean edges. In contrast, our approach takes the height over



(a) Western facades missing (b) Eastern facades missing (c) Fusion result

Figure 1: Typical results for the automatic fusion of two partial reconstructions, based on MAV imagery. Best viewed in color.

ground into account and therefore even benefits from complex structures.

(Wang and You, 2009) and (Wang and You, 2010) tackle the problem of registering images of 2D optical sensors and 3D range sensors, without any assumption about initial alignment. Their approach is based on region matching between optical images and depth images using Shape Context (Belongie et al., 2002). They extract regions from an ortho projection of the scene using an adjusted segmentation step, and connected regions of the same heights from a DSM. Again, this works well for large-scale city models, but would not work for partial SfM models. Additionally, 3D models created from ground level hardly show large regions which could be matched to a nadir view.

A popular approach to aligning and fusing SfM point clouds is to use random sample consensus (RANSAC)-based geometric verification (Fischler and Bolles, 1981). A typical issue is the estimation of a reasonable inlier threshold, however this has been resolved in recent work (Raguram and Frahm, 2011). Still, such an approach is not feasible for our purpose as on the one hand feature correspondences cannot be established and the algorithm would have to solve a huge combinatoric problem. On the other hand, we want to align data with significant variations of the ground sampling distance which would not be possible either.

Another well known method of aligning two point clouds is the Iterative Closest Points (ICP) algorithm (Zhang, 1994). ICP estimates a transform to minimize the overall distance between points by iteratively assigning closest points as correspondences and solving for the best rigid transform. While ICP is mainly used for registering 3D laser scans, (Zhao et al., 2005) use it to align dense motion stereo from videos to laser scan data. However, 3D ICP can take very long and suffers from getting stuck in local minima due to its typically small convergence radius. In other words, a good initialization is necessary for ICP to converge. Our evaluation in Section 5 shows that simply applying ICP is not sufficient for our challenging datasets; however, 3D ICP can still be exploited on top of our method to improve the results.

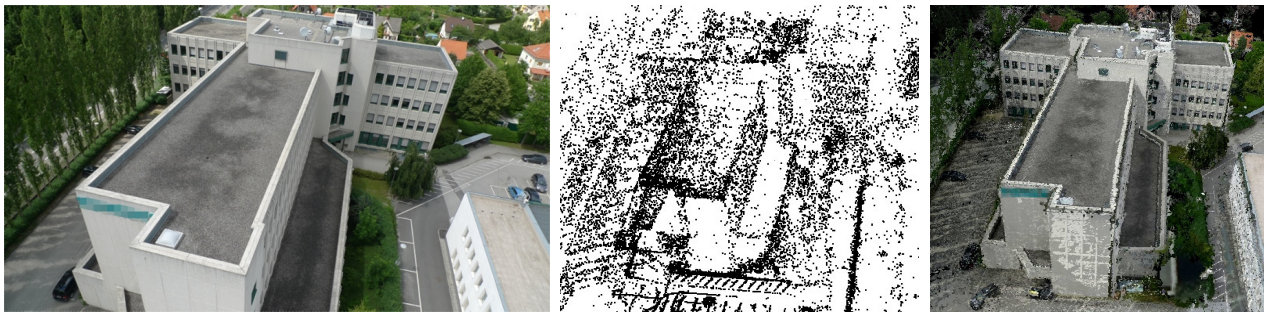
(Kaminsky et al., 2009) use 2D ICP to compute the optimal alignment of a sparse SfM point cloud to an overhead image using an objective function that matches 3D points to image edges. Additionally, the objective function contains free space constraints which avoid an alignment to extraneous edges in the overhead image. While their approach is suitable to align many 3D models obtained from ground level, it has problems with points on the ground and would therefore fail to align the models acquired using our micro aerial vehicle.

Our fusion approach builds on previous work of (Wendel et al., 2011a). This approach has shown to work in complex scenarios with models acquired in different seasons where sparse feature correspondences could not be established. Given a sufficient point density in the reconstruction, the approach is less prone to errors caused by objects on the ground than previous work, it implicitly follows a free-space constraint and it works with models covering a small area. In the following we demonstrate how the approach can be extend for automatic fusion of partial reconstructions without the need for an overhead digital surface model.

3 OBTAINING THE 3D MODELS

For 3D model reconstruction we rely on a Structure from Motion (SfM) approach that is able to reconstruct a scene from unorganized image sets. Structure from Motion deals with the problem of estimating the 3D structure of a scene and camera orientations from 2D image measurements only. Our solution to the 3D reconstruction problem is based on the work of (Irschara et al., 2010) and (Wendel et al., 2011b). It is widely applicable since no prior knowledge about the scene is necessary (i.e. no sequential ordering of the input images has to be provided) and can therefore be applied to terrestrial as well as aerial imagery. To accelerate the computations we take advantage of graphic processing units (GPUs) for efficient parallelized computing (Frahm et al., 2010).

In particular our framework consists of three processing steps, namely feature extraction, matching, and geometry estimation. First, we extract SIFT features (Lowe, 2004) from each frame. We then match the keypoint descriptors between each pair of images and perform geometric verification based on the Five-Point algorithm (Nistér, 2004). Since matches that arise from descriptor comparisons are often highly contaminated by outliers, we employ a RANSAC (Fischler and Bolles, 1981) algorithm for robust estimation. The matching output is a graph structure denoted as epipolar graph $\mathcal{EG}$, that consists of the set of vertices $\mathcal{V} = \{I_1 \dots I_N\}$ corresponding to the images and a set of edges $\mathcal{E} = \{e_{ij} | i, j \in \mathcal{V}\}$ that are pairwise reconstructions. Our SfM method follows an incremental approach (Snavely et al., 2006) based on the epipolar graph $\mathcal{EG}$. We initialize the geometry as proposed in (Klopschitz et al., 2010). Next, for every image I that is not reconstructed and has a potential overlap to the current 3D scene (estimated from the $\mathcal{EG}$ graph), 2D–to–3D correspondences are established. A three-point pose algorithm (Haralick et al., 1991) inside a RANSAC loop is used to insert the position of a new image. When a pose can be determined (i.e. a sufficient inlier confidence is achieved), the structure is updated with the new camera and all measurements visible therein. A subsequent procedure expands the current 3D structure by triangulation of



(a) (b) (c)

Figure 2: Scene reconstruction. (a) Original view of the scene. (b) Sparse model obtained by structure from motion reconstruction. (c) Semi-dense point model obtained by refining the sparse model with PMVS and fusing it using the proposed algorithm.

new correspondences. Bundle adjustment (Triggs et al., 2000) is used to globally minimize the reprojection error over all measurements. The triangulated points are then back-projected and searched for in every image. To this end we utilize a 2D kd-tree for efficient correspondence search in a local neighborhood of projections. This method ensures strong connections within the current reconstruction. Whenever a number of N images is added (we use $N = 10$), bundle adjustment is used to simultaneously optimize structure and camera pose. The sparse reconstruction result can be seen in Figure 2(b).

We further rely on the publicly available patch–based multiview stereo (PMVS) (Furukawa and Ponce, 2009) reconstruction software to densify aerial and terrestrial reconstructions. Since PMVS requires textured surfaces for densification, it does not guarantee a constant ground sampling distance and therefore the reconstruction of weakly textured surfaces may fail. The semi-dense PMVS reconstruction result, fused from two partial reconstructions using our approach, is depicted in Figure 2(c).

4 AUTOMATIC FUSION

Our approach to the automatic fusion of partial reconstructions is based on the alignment of 2D depth maps showing the scene in a nadir view, rather than 3D point clouds. We can thus handle geometric configurations where ICP fails due to local minima, as well as significant differences in appearance because there is no need to establish sparse feature correspondences. Instead, we exploit that patch-based semi-dense reconstruction of the fractured model typically results in several point clouds covering overlapping areas. We automatically rotate all available partial reconstructions into a common coordinate system and project them to a plane parallel to the ground. Additionally available GPS information might be used to roughly align the models within that plane; however, the user's knowledge about acquiring the models at roughly the same place is also good enough. Finally, the partial reconstructions are iteratively aligned to each other by correlating the successively improving depth map stored in the plane with the individual depth maps generated from the models. This corrects for the initial alignment uncertainties and results in precisely fused models. In the following paragraphs a detailed description of our approach is given.

4.1 Iterative Processing Scheme

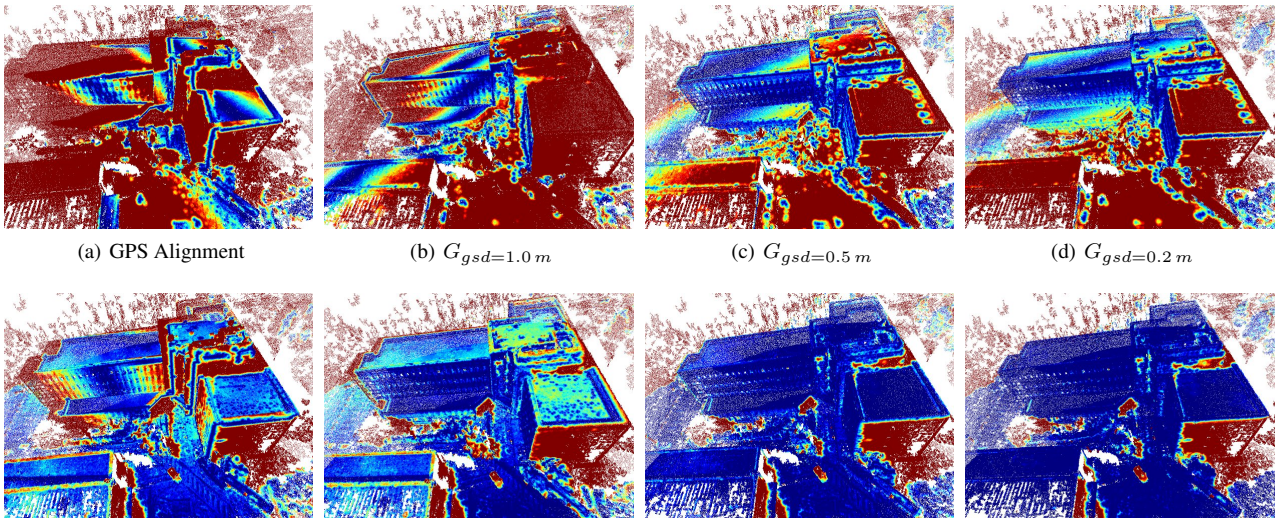
We employ an iterative processing scheme for fusing partial reconstructions. Similar to the original approach by (Wendel et al., 2011a) we store information about surface heights in a plane parallel to the ground. However, in contrast to using the digital surface model (DSM) we initialize these heights to be *undefined* and successively improve the depth map with every alignment of a partial reconstruction. In other words, the first partial model is only roughly projected to the plane, but all further iterations can resort to the evolving surface model. While we do not require the initial DSM information for fusion anymore, it can still be used to improve geo–referencing as in the original approach.

In our processing scheme we represent all coordinates in a local Earth-centered, Earth-fixed (local ECEF) coordinate system. While the global ECEF coordinate system has its origin at the center of the Earth, with the x axis passing through the equator at the prime meridian and the z axis passing through the north pole, local ECEF employs a tangent plane to the Earth's surface at a reference point. By definition, the x axis heads East, the y axis North, and the z axis up into the sky (Snyder, 1987). Storing data in local ECEF format has two advantages over the original method proposed by (Wendel et al., 2011a): First, the plane for creating depth maps over ground is inherently given and all coordinates in this plane have metrical values. This is useful for defining parameters in the alignment process. Second, the issue of storing huge numerical values as in ECEF format and the resulting inaccuracies are resolved by subtracting the local reference point.

4.2 Rough Alignment

Structure from Motion (SfM) pipelines typically store resulting models in the coordinate system of the first camera. As a result, the axes of partial reconstructions do not align at all and have to be rotated to a common ground plane. We employ a reasonable assumption to approximate this plane, namely that the horizontal axis in every image coordinate system is approximately parallel to the ground plane, which is the case when taking upright photographs from the ground, but also when taking nadir and oblique pictures on a micro aerial vehicle. The approach of (Szeliski, 2006) can then be used to compute the ground plane normal and the corresponding rotation.

If GPS coordinates corresponding to the centers of the cameras used for reconstruction are available (as for our aerial data), they can be incorporated to position the model in a world coordinate system. In this case our approach is to robustly solve for a 2D similarity transform between the camera positions in the SfM model and the GPS coordinates in local ECEF format using RANSAC (Fischler and Bolles, 1981). As GPS coordinates are noisy, this only results in a rough alignment. If GPS coordinates are not available but all partial reconstructions result from reconstructing the same structure, rough alignment can be achieved by placing all models in the origin and adjusting the parameters for the following precise alignment step. However, this only works if the models are of sufficient complexity (i.e., a model of a single corner is too ambiguous) so using GPS data is preferred.



(a) GPS Alignment (b) $G_{gsd=1.0\,m}$ (c) $G_{gsd=0.5\,m}$ (d) $G_{gsd=0.2\,m}$

(e) GPS Alignment & ICP (f) $G_{gsd=1.0\,m}$ & ICP (g) $G_{gsd=0.5\,m}$ & ICP (h) $G_{gsd=0.2\,m}$ & ICP

Figure 3: Alignment accuracy. Red points have a Hausdorff distance larger than 10 cm. Blue points are closer than 1 cm to the second model. (a)-(d) Accuracy after alignment using different ground sampling distances for G_{gsd} without ICP. (e)-(h) Remaining error after solving Equation 3 with subsequent ICP. Note that points present in only one part of the reconstruction by definition have a Hausdorff distance of more than 10 cm and are thus colored in red. Best viewed in color.

4.3 Precise Alignment

Given the rough alignment, we use correlation for precise alignment. We project the semi-dense 3D point cloud into the pixel grid of the evolving DSM image, storing only the maximum height value per pixel. Pixel clusters with a radius $r \leq 2px$ are removed using morphological operations to get rid of reconstruction outliers. The model template M_0 is finally created by cropping an axis-aligned box containing the defined values.

As the uncertainty of the rough alignment can introduce rotation and scale errors next to the translational uncertainty ΔT, we rotate the model by the angles $\Delta\phi$, $\Delta\theta$, and $\Delta\psi$ (roll, pitch, yaw) and scale it with a factor $s = 1.0 \pm \Delta s$ to generate the model templates M_t. We cover the search space using the coarse–to–fine approach by (Kaminsky et al., 2009) to speed up computation, and crop the ground template G_{gsd} from the evolving DSM image according to the pyramid level gsd.

The score of template t is finally computed by normalized cross-correlation of ground and model templates,

$$d(t) = \frac{1}{n_t - 1} \sum_{x,y} \frac{(G_{gsd}(x,y) - \overline{G_{gsd}})(M_t(x,y) - \overline{M_t})}{\sigma_{G_{gsd}} \sigma_{M_t}}, \tag{1}$$

where n_t is the number of *defined* pixels for every template t, and $\overline{G_{gsd}}$, $\sigma_{G_{gsd}}$, $\overline{M_t}$ as well as σ_{M_t} are computed only for *defined* pixels. Additionally, we introduce a term for penalizing alignments which contain a large amount of *undefined* pixels,

$$r(t) = \frac{n_t}{N_t}, \tag{2}$$

where N_t is the number of all pixels in template t. The best height map alignment is then associated with the best model template

$$t_{best} = \arg\max_t \; d(t) + \lambda r(t). \tag{3}$$

In contrast to (Wendel et al., 2011a), we found the mode of the difference between the ground template and the best model template to be more robust for estimating the translation along the vertical axis.

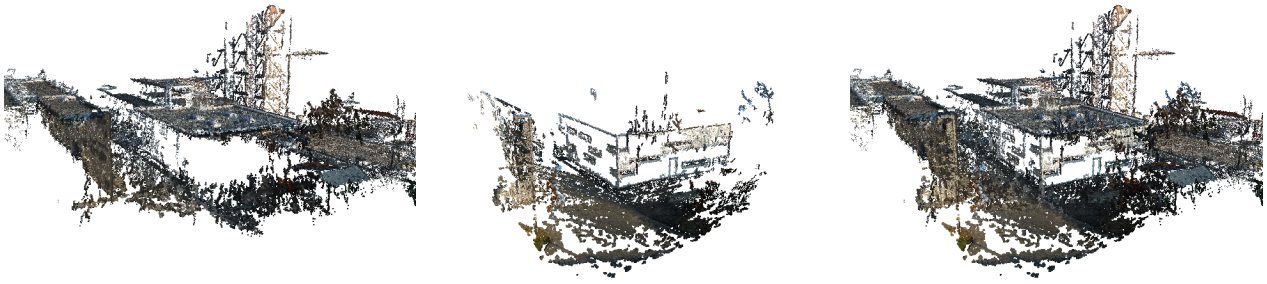
The previous step delivers a translation, rotation, and scaling which is used to transform the partial reconstruction to the iteratively growing point cloud. While the discrete correlation approach successfully avoids getting stuck in local minima, continuous optimization can nearly always improve the final result. We thus further improve the accuracy by applying 3D ICP (Zhang, 1994). Given the already very good alignment, a sparse set of N_{icp} points is selected and ICP typically converges within seconds. Finally, the point cloud is projected to the evolving DSM and serves as a basis for alignment of the next partial reconstruction.

5 RESULTS

In order to demonstrate the accuracy of our alignment and fusion approach, we perform experiments on two different datasets. We compare the accuracy of our fusion algorithm to a standard 3D ICP registration method and demonstrate that our approach is able to fuse reconstructions obtained from airborne images as well as from images acquired on the ground level.

The first dataset (Figure 1) shows a large office building which is reconstructed from 400 images acquired by a manually controlled MAV. Due to limited power supply, images were acquired in two different flights resulting in two partial reconstructions with 2 million and 1.3 million 3D points, respectively. The second dataset consists of two partial reconstruction where one part is reconstructed from 35 nadir images taken by our MAV at a height of 60 m above the building (see Figure 4(a)). The other partial model (Figure 4(b)) is reconstructed from 44 images acquired at ground level.

We obtained all images using a Panasonic DMC-LX3 camera at a resolution of $3968 \times 2232px$, both at ground level and airborne using an Ascending Technologies Falcon 8 octo–rotor MAV. Additionally, the MAV is equipped with a consumer–grade GPS, which allows rough geo–referencing of the reconstructions. The distance between the object and the camera position varies between 20 m and 60 m which results in a ground sampling distance (GSD) of 8 mm to 24 mm per pixel; however, as our reconstructions are only semi-dense the resulting point clouds are often sparser if texture is missing.



(a) Aerial MAV imagery (b) Terrestrial imagery (c) Fused model

Figure 4: Fusion of two models computed from airborne and ground–level images. (a) Reconstruction obtained by nadir images at 60 m height. (b) Model obtained from ground images. (c) Both models fused by our approach.

We reduce the search space for precise alignment according to the expected uncertainty of the rough GPS alignment, with a translational uncertainty $\Delta T = \pm 5m$, roll $\Delta\phi = \pm 10°$, pitch $\Delta\theta = \pm 10°$, yaw $\Delta\psi = \pm 20°$, and scale $s = 1.0 \pm 0.2$. A brute force approach for checking all combinations is not feasible due to the huge amount of required templates. Therefore, we exploit a coarse–to–fine scheme that iteratively increases the GSD of the evolving DSM for every pyramid level G_{gsd}. We weight the penalizing term with $\lambda = 0.1$ and set the number of sampled points for subsequent ICP to $N_{icp} = 1000$.

In our first experiment, we align both reconstructions using the available GPS information as described in Section 4.2 and measure the alignment error by evaluating the Hausdorff distance from the smaller to the larger part (Huttenlocher et al., 1993). Figure 3(a) shows the resulting pseudo color visualization, where red points have an error larger than 10 cm and blue ones have an error smaller than 1 cm. After the previous rough alignment, we perform 3D ICP to reduce the error. However, ICP gets stuck in a local minimum and therefore the alignment is far away from the desired registration (Figure 3(e)). This experiment demonstrates that, due to the small convergence radius of ICP, a consumer–grade GPS alignment is not sufficient to obtain an accurate fusion using a standard 3D ICP algorithm. On a closer look, we also observe that the scale estimated by the rough GPS alignment differs considerably between the two parts. Therefore, a proper fusion method has to estimate the scaling between the parts.

In the second experiment, we apply our proposed algorithm as described in Section 4.3 to the first dataset. As shown in Figure 3(b)-3(d), the alignment error reduces with increasing resolution of the ground template G_{gsd}. On a scale level of $G_{gsd=0.5\,m}$, the fusion method already allows the subsequent ICP to converge to the global optimum. For most points, the remaining alignment error is smaller than 1 cm. Points that are present in only one part of the reconstruction have a Hausdorff distance of more than 10 cm which is obvious since they do not have a corresponding counterpart. Since we correctly estimate the scale difference of $\Delta s = 0.04$, the result does not show errors in scaling. This observation is confirmed by Figure 3(h) which shows a constant error over the entire surface. Subsequently applying ICP has shown to converge within seconds and mainly corrects for rotational errors introduced by the discrete search space of our approach. The entire fusion process requires less than 5 minutes on any of our evaluation datasets for finding the correct transformation with an accuracy in the range of $\pm 1\,cm$. This corresponds to the point density of the model.

Our method also bridges the gap between reconstructions computed from airborne and ground–level imagery. Figure 4(a) shows the partial reconstruction of a building that is obtained from nadir images taken by our MAV at a height of 60 m, and thus does not show facade details. The reconstruction of the same building from ground–level images is shown in Figure 4(b). Due to the weakly textured surfaces, the utilized densification algorithm does not perform very well and the GSD of the model varies. Even using such challenging data, our algorithm fuses both parts to a comprehensive model. The resulting Figure 4(c) demonstrates that we can merge reconstructions obtained from very different viewpoints and with highly varying GSD to an accurate 3D model.

Figure 5: Fusion and geo–referencing of partial reconstructions. The medieval clocktower in front has been fused and aligned to the geo–referenced, approximated DSM of the city in the background.

Of course, our approach can also be used in conjunction with the original approach of (Wendel et al., 2011a) to geo–reference several partial reconstructions in a world coordinate frame given an accurate (i.e. aerial triangulation based) or approximated (i.e. OpenStreetMap based) DSM. An example can be found in Figure 5.

6 CONCLUSION

We have presented a novel technique for the automatic fusion of partial 3D reconstructions based on the correlation of orthographic depth maps. We can handle complex cases where previous methods had problems, including models which do not share any appearance features and models with considerably different ground sampling distances. This allows not only to fuse data acquired by an MAV, but also combining aerial and terrestrial data sources. Our qualitative and quantitative evaluation using two outdoor scenes shows that we can achieve a fusion accuracy in the range of $\pm 1\,cm$, and that we are able to generate point clouds

which are visually more appealing and serve as an ideal basis for further processing.

In future work we plan to adjust our MAV image acquisition strategy according to the findings of this work. We will first acquire nadir imagery for creating an initial DSM, and then fuse further oblique aerial and terrestrial views into the model. As a result, the spatial area which can be represented in a single model is increased even further.

ACKNOWLEDGMENTS

This work has been supported by the Austrian Research Promotion Agency (FFG) FIT-IT projects Pegasus (825841), Construct (830035), and Holistic (830044).

REFERENCES

Belongie, S., Malik, J., and Puzicha, J., 2002. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4), pp. 509–522. 2

Eisenbeiss, H., 2004. A mini unmanned aerial vehicle (UAV): System overview and image acquisition. In: International Workshop on Processing and Visualization using High-Resolution Imagery. 1

Eisenbeiss, H., 2009. UAV Photogrammetry. PhD Thesis, ETH Zuerich, Switzerland. 1

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communication Association and Computing Machine 24(6), pp. 381–395. 2, 3

Frahm, J.-M., Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S. and Pollefeys, M., 2010. Building rome on a cloudless day. In: European Conference on Computer Vision (ECCV). 2

Frueh, C. and Zakhor, A., 2003. Constructing 3D city models by merging ground-based and airborne views. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1

Furukawa, Y. and Ponce, J., 2009. Accurate, dense, and robust multi-view stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 3

Haralick, R. M., Lee, C., Ottenberg, K. and Nölle, M., 1991. Analysis and solutions of the three point perspective pose estimation problem. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 592–598. 2

Huttenlocher, D., Klanderman, G. and Rucklidge, W., 1993. Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 15(9), pp. 850–863. 5

Irschara, A., Kaufmann, V., Klopschitz, M., Bischof, H. and Leberl, F., 2010. Towards fully automatic photogrammetric reconstruction using digital images taken from UAVs. In: Proceedings of the ISPRS Symposium, 100 Years ISPRS - Advancing Remote Sensing Science. 2

Kaminsky, R. S., Snavely, N., Seitz, S. M. and Szeliski, R., 2009. Alignment of 3D point clouds to overhead images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, IEEE, pp. 63–70. 2, 4

Klopschitz, M., Irschara, A., Reitmayr, G. and Schmalstieg, D., 2010. Robust incremental structure from motion. In: International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT). 2

Kluckner, S., Birchbauer, J. A., Windisch, C., Hoppe, C., Irschara, A., Wendel, A., Zollmann, S., Reitmayr, G. and Bischof, H., 2011. Construction site monitoring from highly-overlapping mav images. In: Proceedings of the IEEE International Conference on Advanced Video- and Signal-based Surveillance (AVSS), Industrial Session. 1

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60(2), pp. 91–110. 2

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 26(6), pp. 756–770. 2

Raguram, R. and Frahm, J.-M., 2011. RECON: Scale-adaptive robust estimation via residual consensus. In: IEEE International Conference on Computer Vision (ICCV). 2

Schmid, K., Hirschmueller, H., Doemel, A., Grixa, I., Suppa, M. and Hirzinger, G., 2012. View planning for multi-view stereo 3d reconstruction using an autonomous multicopter. Journal of Intelligent Robotic Systems 65, pp. 309–323. 1

Snavely, N., Seitz, S. and Szeliski, R., 2006. Photo tourism: Exploring photo collections in 3D. In: Proceedings of SIGGRAPH 2006. 2

Snyder, J. P., 1987. Map Projections - A Working Manual. U.S. Geological Survey Professional Paper 1395. United States Government Printing Office, Washington, D.C. 3

Strecha, C., Pylvaenaeinen, T. and Fua, P., 2010. Dynamic and scalable large scale image reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1

Szeliski, R., 2006. Image alignment and stitching: a tutorial. Foundations and Trends in Computer Graphics and Vision 2(1), pp. 1–104. 3

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A., 2000. Bundle adjustment – A modern synthesis. In: Vision Algorithms: Theory and Practice, pp. 298–375. 3

Wang, Q. and You, S., 2009. A vision-based 2D-3D registration system. In: Workshop on Applications of Computer Vision (WACV), IEEE. 2

Wang, Q. and You, S., 2010. Automatic registration of large-scale multi-sensor datasets. In: European Conference on Computer Vision (ECCV) Workshops. 2

Wendel, A., Irschara, A. and Bischof, H., 2011a. Automatic alignment of 3D reconstructions using a digital surface model. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on Aerial Video Processing. 1, 2, 3, 4, 5

Wendel, A., Irschara, A. and Bischof, H., 2011b. Natural landmark-based monocular localization for MAVs. In: IEEE International Conference on Robotics and Vision (ICRA). 2

Zhang, Z., 1994. Iterative point matching for registration of free-form curves and surfaces. Int. Journal of Computer Vision. 2, 4

Zhao, W., Nister, D. and Hsu, S., 2005. Alignment of continuous video onto 3D point clouds. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) pp. 1305–1318. 2