Incremental, Orthorectified and Loop-independent Mosaicking of Aerial Images Taken by micro UAVs

Saeed Yahyanejad Markus Quaritsch Bernhard Rinner
Institute of Networked and Embedded Systems
Klagenfurt University, AUSTRIA
<firstname.lastname>@uni-klu.ac.at

***Abstract*—In this paper we survey thoroughly the problem of orthorectified and incremental image mosaicking of a sequence of aerial images taken from low-altitude micro aerial vehicles. Most of existing approaches have been exploiting the global optimization (in presence of a loop in the image sequences) to distribute and/or metadata to mitigate the accumulating stitching error. However, the resulting mosaic can be improved if the errors are diminished by studying their sources. Mostly the UAV aerial image mosaicking is affected by the following three important sources of error:** i) **a weak homography as a result of using unleveled ground control points (GCPs) for image registration,** ii) **a poor camera calibration and image rectification, and** iii) **deficiency of a well-defined projection model (cylindrical, planar, etc) and consequently an inappropriate transformation model. We investigate the influences of using a depth map to find the features from the same plane, geometric distortion correction and combining the appropriate choice of projection and transformation model for the mosaicking. We further quantify the improvement of orthorectification in mosaics by mitigating those errors and demonstrate the improvement on real-world mosaics.**

I. Introduction

Using unmanned aerial vehicles (UAVs) is growing rapidly for surveillance purposes. These UAVs are equipped with imaging sensors and they can easily provide aerial images taken from the target scene (see Figure 1). Image mosaicking is a noteworthy application of aerial imaging which could be used for further information retrieval from the target area. Note that in sensitive cases of surveillance each image might have critical details which need to be retained even after the image is placed in a mosaic. In cases where UAVs are supposed to fly and take images without any loop in their route (e.g., boarder control, road construction and object following) the problem of mosaicking and orthorectification gets more challenging.

In this paper we evaluate quantitatively the different parameters that affect these types of image mosaicking and we then find their potential to improve the outcome. In other words, we try to reduce the sources of errors which cause to lose the relative distances. To fulfil that we need to step back from visually appealing and non-rigid image blending methods and concentrate more on the origin of errors which accumulate over time.

The remainder of this paper is organized as follows. Section II gives an overview of related work in the domain of image mosaicking. In Section III we summarize the typical mosaicking procedure and define our problem. In Section IV we show the major sources of errors in mosaicking. Section V presents some practical results of mosaicked images taken by a UAV. Section VI finally concludes the paper.

(a) AscTec Pelican drone with an FLIR Photon 640 thermal camera. (b) MD4-200 drone with an RGB camera.

Figure 1. Two drones used for acquiring thermal and visible-light images.

II. Related work

UAVs are being used ubiquitously in many fields of aerial imaging. A huge number of aerial image mosaicking approaches rely on medium to large UAVs. These UAVs have more capabilities in aspects of their computational power, data transmission rate, payload capacity, accuracy of measurement devices and flight time. Based on these parameters a variety of approaches are proposed for mosaicking of images taken from UAVs. Automatic mosaicking by 3D-reconstruction and epipolar geometry [8], [14], combining global positioning system (GPS), inertial measurement unit (IMU) and video sensors for external distortion correction and geo-referencing [3], wavelet-based stitching [17], triangulated irregular network registration and perspective correction [13] and high altitude imaging and mosaicking [19], [5], [10], [13] are some of those examples. Schultz et al. [11] use a digital elevation model to mosaic images taken from an airplane. Hruska et al. [7] introduce an appropriate platform for small UAVs to be able to provide high resolution and georeferenced images by exploiting GPS and IMU. Afterwards they perform change detection by comparing different temporal images of a target area. In their work they remark the importance of internal geometric distortion correction but do not explain how it is used in mosaicking. Zhou [18] uses the video stream from a UAV (weight $10\,kg$) equipped with differential GPS, with an error range of a few centimeters, and real-time transmitter of video for further

mosaicking purposes on the base-station. Xiang and Tian [15] also mention the role of high precision internal geometric distortion correction in georeferenced mosaic construction in addition to exploiting GPS and IMU. When flying with UAVs at a relatively low altitude (below $50\,m$), non-planar objects on the ground make the feature matching and image registration more difficult. As explained by Yahyanejad et al. [16], some limitations force UAVs to just take images at individual predefined picture points. This causes different angles of view looking to the same scene and this intensifies the problem of non-planar objects. In this paper we closely refer to the same scenario of surveillance in which we use UAVs that fly at a low altitude where images are taken at predefined picture points and the goal is to provide an orthorectified overview mosaic of a target area. Agarwala et al. [1] cope with a similar problem of producing multi-viewpoint panoramas of long, roughly planar scenes but on the ground (e.g. the facades of buildings along a city street). They use Markov Random Field optimization to construct a composite from arbitrarily shaped regions of the source images, rather than building the panorama from strips of the source images. They also consider a higher pairwise overlap (with approximately $1\,m$ distance between two picture-points) and the dominant plane of the photographed scene is defined by the user input.

In this paper we focus on the sources of error in mosaicking from a sparse set of aerial images. Our approach does not exploit GPS and IMU data, because these meta-data are typically unreliable for micro UAVs.

III. Typical pairwise mosaicking and problem definition

Pairwise image mosaicking is typically performed with the following steps:

1) **Correcting the internal geometric distortion.** Brown's distortion model [4] can tackle the radial and tangential distortion including the principal point estimation. Let $P = (x, y)$ be a normalized point in image reference system, the undistorted point P_u, using a 6th order radial and 2nd order tangential model can be acquired by:

$$P_u = \begin{bmatrix} x_u \\ y_u \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + (k_1 r^2 + k_2 r^4 + k_3 r^6) \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \begin{bmatrix} 2k_4 x_n y_n + k_5(r^2 + 2x_n^2) \\ 2k_5 x_n y_n + k_4(r^2 + 2y_n^2) \end{bmatrix} \tag{1}$$

where point (x_n, y_n) is in a coordinate considering the principal point (PP) as its origin ($x_n = x - PP_x, y_n = y - PP_y$), $r = \sqrt{x_n^2 + y_n^2}$ represents the distance from principal point, k_1, k_2 and k_3 are the radial distortion coefficients and k_4 and k_5 are the tangential distortion coefficients.

2) **Feature extraction and matching.** Different methods can be used to extract features which later will be used for image registration (e.g., by using SIFT [9], SURF[2], or Harris Corner[6]). Features extracted from the new unregistered image are matched with the previously registered image. Traditionally correspondences are determined by computing the similarity between descriptor vectors associated to each point.
3) **Defining the projection model.** For simplicity we assume a planar model, since we mostly fly over areas with a dominating ground plane. Of course there might be non-planar objects on the ground which we will discuss in Section IV-A.
4) **Defining the transformation for homography.** Based on different scenarios we can choose between different existing models such as translation, similarity or projective.
5) **Removing outliers and calculating the transformation function.** Performing RANdom SAmple Consensus (RANSAC) and least median of squares (LMS) are usually used to remove the outliers. Based on inliers, each iteration of RANSAC uses its own estimated transformation while trying to find the maximum set of matched pair-points. This set of matched pair-points should minimize a sum of squared distances among all estimated transformations:

$$\text{Minimize} \quad \lambda = \sum_i \|\mathbf{T} x_i - x_i'\| \tag{2}$$

where x_i are the points from unregistered image, x_i' are the points from the previously registered image and $\mathbf{T}$ is the optimized transformation matrix to change the coordinate between these two sets of points (all points and transformation matrices are in homogenous coordinates).
6) **Registration.** Transform the new unregistered image to the coordinate of previously registered image and perform the resampling (use an interpolation method).
7) **Mosaic construction.** Merging the transformed image with the mosaic constructed so far, in order to build the incremental mosaic.

Now imagine a case where we want to generate an incremental mosaic of consecutive images taken by UAVs without any loop. The challenge is how to preserve the orthorectification as much as possible without exploiting any metadata (e.g., GPS or IMU). Consider $\mathbf{O}_n$ as the overall image of the target area given a set of n consecutive images $\{\mathbf{I}_i | i = 1 \dots n\}$. The overall image can be iteratively constructed as follows:

$$\mathbf{O}_i = \mathit{Merge}(\mathbf{O}_{i-1}, T_i(\mathbf{I}_i)) \tag{3}$$

where $\mathbf{O}_0$ is an empty image, T_i are transformation functions (constructed from pairwise transformation matrices in Equation 2) and the *Merge* function combines each transformed image with the overall image. This mosaicking can be described as an optimization problem, in which we need to set the parameters in a way that it maximizes our orthorectification quality function μ. One way of constructing such a quality function is using a metric which evaluates the deformation of an image in different directions (horizontal, vertical and diagonal) compared to a reference image:

$$\mu = \frac{4}{\sum_{i=1}^{4} \frac{\max(l_i, l_i')}{\min(l_i, l_i')}} \tag{4}$$

where l_i are the length of width, height and two diagonals of each target image, and l'_i are the length of width, height and two diagonals of the reference image. In fact, by this metric we calculate the harmonic mean of horizontal, vertical and two diagonal deviation ratios.

In our work we combine different existing pairwise stitching methods and compare the resulting mosaics in terms of relative distances. Hence, we decide how to set the parameters to obtain the optimal result. Note that although we narrowed down the scope of our scenario, it is possible to simply merge the result with other approaches such as using metadata or bundle adjustment. Though the bundle adjustment methods are more efficient when either there are more than two viewpoints for most of the regions or in existence of a loop in the image sequences [12].

IV. Major sources of error in pairwise mosaicking

In order to use our metric and compare different mosaics we need a known and well-defined ground truth. For illustration we lined up a set of printed chess-board patterns plus some non-planar objects we put over and around those patterns along the scene. Then we use a camera with fixed custom settings (e.g., in our case focal length$= 28\,mm$, exposure time$= 1/500\,s$) and take consecutive images manually from top view with approximately 70% of overlap. By setting a low focal length and consequently a wider angle of view we increase the overlap ratio which leads to more matched feature and inliers. But note that in this case we also encounter a higher geometric distortion. In this way we can simulate the imaging from UAVs to a good extent.

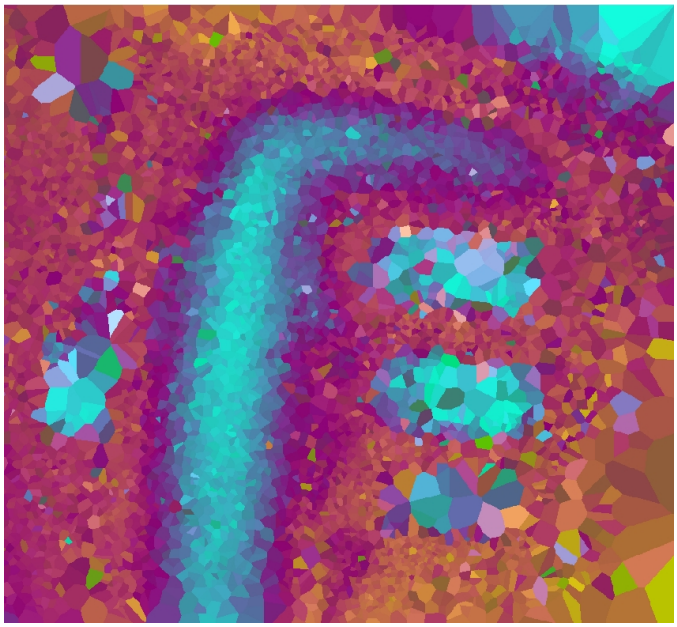
We tested different existing algorithms and parameters that are used for image mosaicking such as internal geometric distortion correction algorithms, feature extraction methods (SIFT, SURF and Harris corner) with different parameters, projection and transformations models and manual GCP selection. Among all, there are three main parameters that will affect the pairwise aerial image mosaicking more than the others namely i) using unleveled GCPs for image registration, ii) internal geometric distortion, and iii) choice of projection and transformation model.

In the following we discuss these parameters and quantify them based on our simulation data-set. In an ideal mosaic all chess-boards should have the same size and shape ($\mu = 1$).

A. Using unleveled features (GCPs) for image registration

Most of the existing mosaicking algorithms are built for panoramic imaging which consider all images are taken almost from the same spot. In this case the depth variation of the scene is not a problem (except small motions of camera or failure to rotate the camera around its optical center, which is usually handled by a parallax removal algorithm [12]). But in our aerial imaging scenario we take the images from significantly different points of view. As a result, non-planar features will produce a disparity when matching features from corresponding images. The disparity vector $\mathbf{d}$ of each transformed feature

(a) Disparity vectors shows the displacement of transformed feature points from their expected positions.

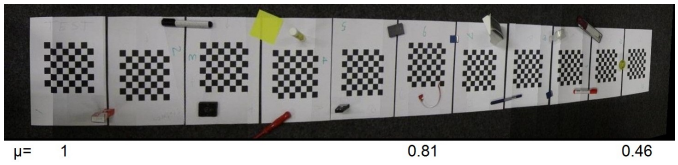


(b) A rough depth map is depicted by interpolating the disparity vectors.

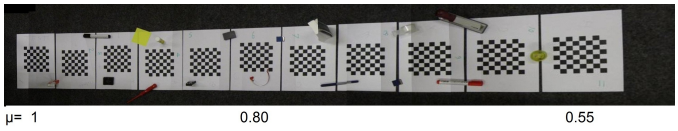
Figure 2. Depth information from stereo vision.

implies the vector from the expected feature point toward the estimated feature point ($\mathbf{d} = \overrightarrow{\mathbf{T}x_i - x'_i}$).

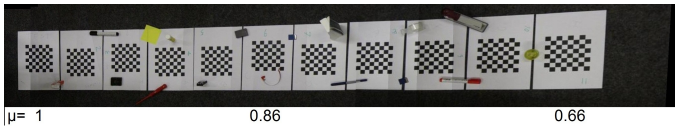
These disparities will impact the transformation estimation procedure as explained in Equation 2. To reduce this effect we need to extract the depth information to extract only the features from the same elevation level which later will be used for image homography. Some depth map construction algorithms use the whole image information (pixels), but we just use the displacement of feature points to speed up



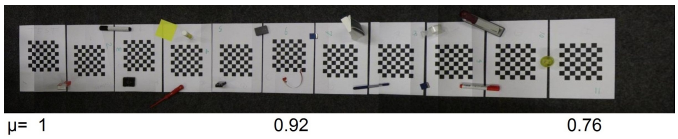
(a) Mosaic of raw (distorted) images.



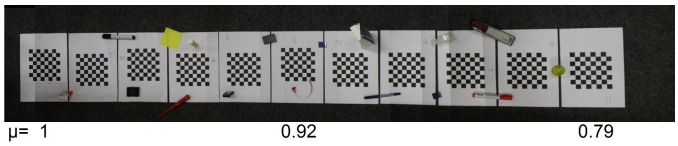
(b) Mosaic after 2nd order radial distortion correction.



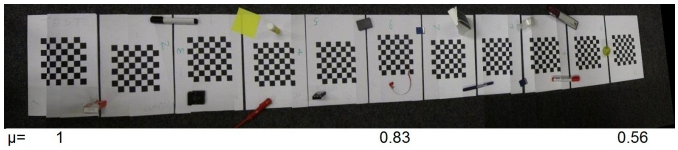
(c) Mosaic after 4th order radial distortion correction.



(d) Mosaic after 4th order radial plus tangential distortion correction.



(e) Mosaic after 6th order radial plus tangential distortion correction.



(f) Mosaic after 6th order radial plus tangential distortion correction but no depth consideration.

Figure 3. Resulting mosaic of 21 sequential images with different distortion correction and depth consideration parameters. Note that the μ values under the first, middle and last chess-board show the corresponding rectification quality.

the process. Sample disparity vectors from a set of stereo images taken by a UAV are shown in Figure 2(a). In order to visualize the corresponding information from these disparity vectors we depict a rough depth map in Figure 2(b). This false color depth map is constructed as follows:

Red component = x component of the displacement vector
Green component = magnitude of the displacement vector
Blue component = y component of the displacement vector

We remove features with magnitude of disparity vector larger than a threshold (varies based on height variation of the objects on the ground and flying altitude). Note that at the first glance it might look similar to setting the RANSAC threshold small, but in that case we might also reject some inliers just because of their small displacement which will slow down or even fail the convergence of RANSAC, especially in cases with low amount of overlap. Figure 3(f) shows a resulting mosaic of our test model without considering the depth information while in Figure 3(e) we see the result with taking the depth into account.

B. Internal geometric distortion

In this section we present the influence of different orders of geometric distortion correction (c.f. Equation 1) over the resulting mosaic of 21 consecutive images obtained as described for our test scenario. Figures 3(a) to 3(e) depicts the results under various distortion correction parameters (the depth information mentioned in IV-A is already considered). The pairwise stitching is performed from left to right. This will give us a visual understanding how much the polynomial orders in distortion correction procedure will affect the mosaicking.

C. Projection and transformation model

As we mentioned earlier, the planar projection model is an appropriate model for UAV imaging over a plane ground. Choosing the planar model demands a projective transformation to correct the perspective distortion of images taken while the camera was tilted. On the other hand, the projective transformation is quite susceptible to errors and a small deviation will spread after a number of images. Substituting the projective transformation with similarity transformation might help significantly to produce a more orthorectified mosaic, especially in cases in which the steps in Sections IV-A and IV-B did not manage to restrain the error propagation. The only drawback of using similarity transformation is that it might lead to small seams in pairwise mosaicking which can be ignored if UAV has almost a nadir view. In Figure 4(b) every other image is considered for mosaicking which reduces the overlap ratio. As μ values in Figures 4(a) and 4(c) show, using similarity transformation leads to a less deformation.

D. Summary

In calculation of μ values for all mosaics, l_i are the length of width, height and two diagonals of each chess-board which is measured within each mosaic image, and l'_i are the length of width, height and two diagonals of the first (leftmost) chess-board which is considered as our reference. In Figure 5 we used this metric to show how much each previously discussed approach or parameter will affect the mosaic integrity. As you see, using higher orders for radial distortion correction, tangential distortion correction, considering the depth information and using similarity transformation, all are the factors which can hep us to persist the correct size and preserve the relative distances along the incremental mosaicking process. This affect might not be sensed while using just a couple of images. As shown in this chart, the difference between the 4th order and the 6th order radial distortion correction is not so obvious until the middle of the mosaic, but at the end we

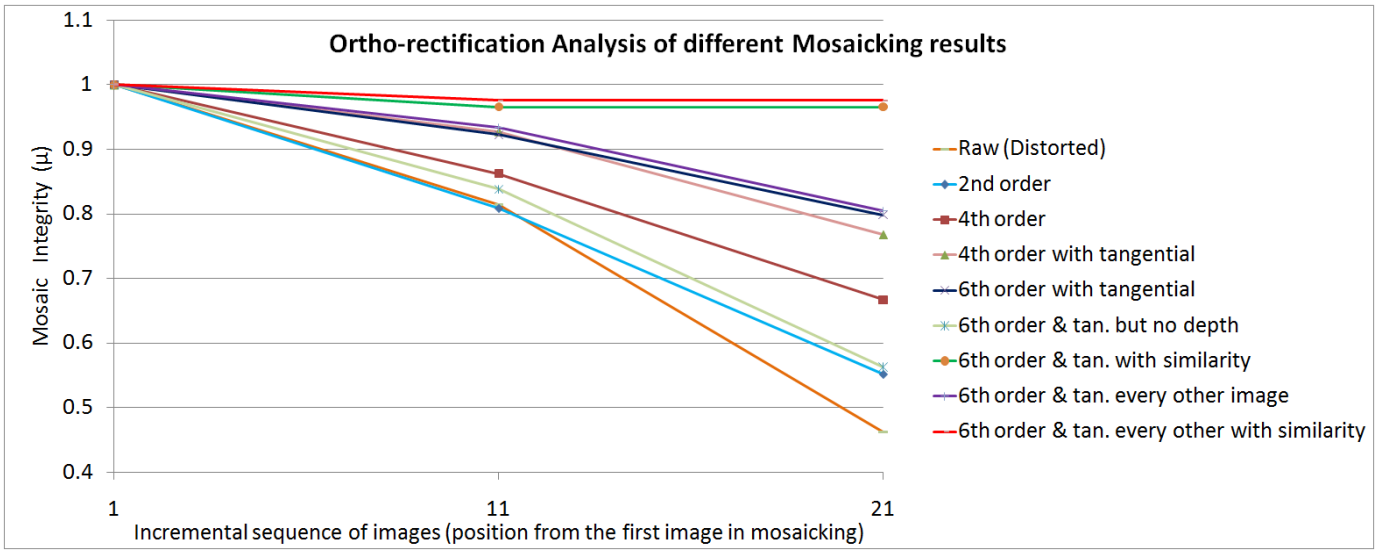
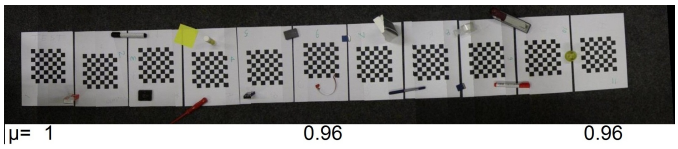
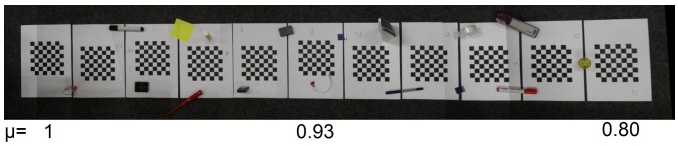


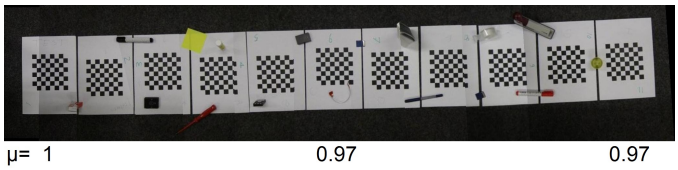
Figure 5. Comparison of orthorectification in different mosaics, built with different methods.



(a) Using similarity in Figure 3(e).



(b) Taking every other image in Figure 3(e) which reduces the overlap ratio.



(c) Taking every other image in Figure 3(e) and using similarity.

Figure 4. Resulting mosaic of 21 sequential images with different transformation model. μ is the rectification quality.

can see that 6th order leads to a slightly better quality. It also implies that similarity transformation significantly helps to mitigate the deformation error, since it does not produce and propagate any projective deformation.

V. Practical Result

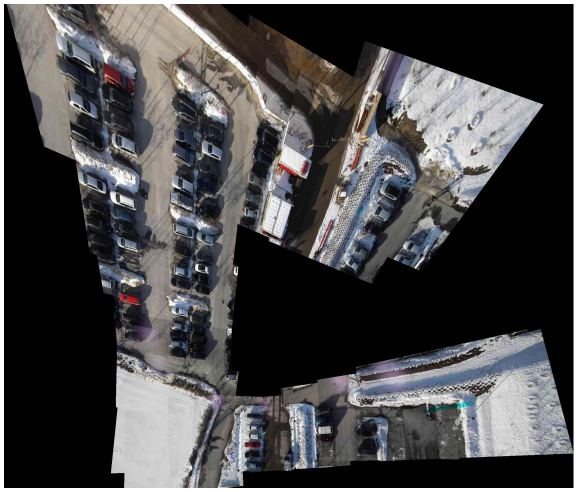
Here we show resulting mosaics of images taken by a UAV. In this scenario we took 27 images with approximately 60% of pairwise overlap. Figure 6(a) depicts the mosaicking result after 2nd order radial distortion correction without considering the depth information. Figures 6(b) and 6(c) show the corresponding mosaic considering the optimizations from the previous section with projective and similarity transformations, respectively. The existing mosaicking approaches mainly aim for visual appealing rather than preserving the integrity and relative distances. As we expected and as shown in Figure 6, mitigating the mentioned errors will noticeably improve the orthorectification.

VI. Conclusion

Note that in this paper we are not demonstrating a new mosaicking algorithm, alternately we quantify the influence of different parameters such as sensor distortion model, depth information of the scene and the choice of projection and transformation models over sequential, pairwise and loop-free image mosaicking. Understanding and comparing the sources of errors enables us to minimize those errors in a way that increases the orthorectification in aerial image mosaicking. Using higher polynomial orders in geometric distortion correction might not be noticeable in a pair of images, but at some point in incremental image mosaicking it will show its affect. To retain the relative distances, similarity transformation, despite its lower degree of freedom, is a good substitution for projective transformation if we have almost a nadir view of the camera. It is also shown how a simple depth map help us to choose the appropriate ground control points for an accurate mosaicking.

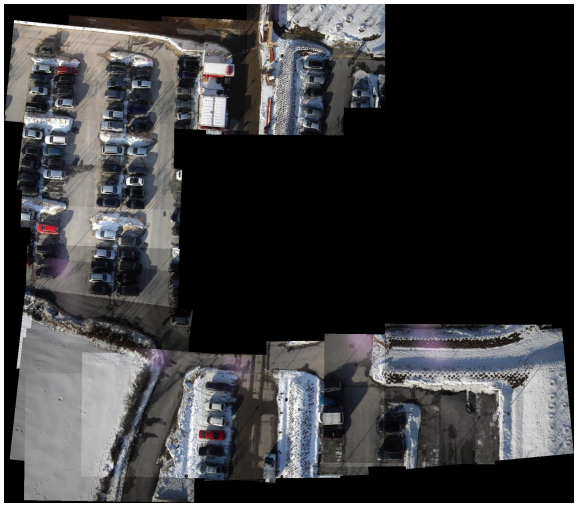
Acknowledgment

This work was performed in the project *Collaborative Microdrones (cDrones)* of the research cluster Lakeside Labs and was partly funded by the European Regional Development Fund, the Carinthian Economic Promotion Fund (KWF), and the state of Austria under grant KWF-20214/17095/24772.



(a) Images are mosaicked with 2nd order radial distortion correction and without depth consideration.

(b) Our approach with projective transformation.



(c) Our approach with similarity transformation.

Figure 6. Resulting mosaic of 27 images taken from UAV.

REFERENCES

[1] A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski. Photographing long scenes with multi-viewpoint panoramas. *ACM Trans. Graph.*, 25:853–861, July 2006. 2

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, 110(3):346–359, 2008. 2

[3] A. Brown, C. Gilbert, H. Holland, and Y. Lu. Near Real-Time Dissemination of Geo-Referenced Imagery by an Enterprise Server. In *Proceedings of 2006 GeoTec Event*, Ottawa, Ontario, Canada, June 2006. 1

[4] D. C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, 1966. 2

[5] X. Han, H. Zhao, L. Yan, and S. Yang. An approach of fast mosaic for serial remote sensing images from uav. In *FSKD '07: Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, pages 11–15, Washington, DC, USA, 2007. IEEE Computer Society. 1

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988. 2

[7] R. Hruska, G. Lancaster, J. Harbour, and S. Cherry. Small uav-acquired, high-resolution, georeferenced still imagery. In *Proceedings of AUVSI Unmanned Systems North America*, pages 837–840, June 2005. 1

[8] L. Lou, F.-M. Zhang, C. Xu, F. Li, and M.-G. Xue. Automatic registration of aerial image series using geometric invariance. In *Proceedings of IEEE International Conference on Automation and Logistics*, pages 1198–1203, 2008. 1

[9] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 1150, Washington, DC, USA, 1999. IEEE Computer Society. 2

[10] P. Pesti, J. Elson, J. Howell, D. Steedly, and M. Uyttendaele. Low-cost orthographic imagery. In *GIS '08: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–8, New York, NY, USA, 2008. ACM. 1

[11] H. Schultz, A. R. Hanson, E. M. Riseman, F. Stolle, Z. Zhu, C. D. Hayward, and D. Slaymaker. A system for real-time generation of geo-referenced terrain models. In *In Proc. SPIE Symposium on Enabling Technolgies for Law Enforcement*, 2000. 1

[12] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., 1st edition, 2010. 3

[13] W. H. WANG Yue, WU Yun-dong. Free image registration and mosaicing based on tin and improved szeliski algorithm. In *Proceedings of ISPRS Congress*, volume XXXVII, Beijing, 2008. 1

[14] D. Wischounig-Strucl, M. Quaritsch, and B. Rinner. A structure based mosaicking approach for aerial images from low altitude of non-planar scenes. In A. Wendel, S. Sternig, and M. Godec, editors, *Proceedings of the 16th Computer Vision Winter Workshop*, pages 51–58. Graz University of Technology, 2 2011. 1

[15] H. Xiang and L. Tian. Method for automatic georeferencing aerial remote sensing (rs) images from an unmanned aerial vehicle (uav) platform. *Biosystems Engineering*, 108(2):104–113, 2011. 2

[16] S. Yahyanejad, D. Wischounig-Strucl, M. Quaritsch, and B. Rinner. Incremental mosaicking of images from autonomous, small-scale uavs. *Advanced Video and Signal Based Surveillance, IEEE Conference on*, 0:329–336, 2010. 2

[17] C. Yuanhang, H. Xiaowei, and X. Dingyu. A mosaic approach for remote sensing images based on wavelet transform. In *WiCOM '08: Proceedings of the Fourth International Conference on Wireless Communications, Networking and Mobile Computing*, pages 1–4, 2008. 1

[18] G. Zhou. Geo-referencing of video flow from small low-cost civilian uav. *IEEE T. Automation Science and Engineering*, 7(1):156–166, 2010. 1

[19] Z. Zhu, E. M. Riseman, A. R. Hanson, and H. J. Schultz. An efficient method for geo-referenced video mosaicing for environmental monitoring. *Mach. Vis. Appl.*, 16(4):203–216, 2005. 1