FUSION AND CLASSIFICATION OF AERIAL IMAGES FROM MAVS AND AIRPLANES FOR LOCAL INFORMATION ENRICHMENT

X. Zhuo, S. Cui, F. Kurz, P. Reinartz

German Aerospace Center, 82234 Wessling, Germany

ABSTRACT

Despite the existence of various matching algorithms, matching of images from Micro Aerial Vehicles (MAVs) and airplanes is still a tough problem due to the substantial differences in scale and rotation. This paper investigates the fusion of MAV imagery and airplane imagery and proposes a new robust image matching method with self-adaption to differences in scale and viewing direction. This method is further applied to register a MAV image block with reference to the orthophoto and DSM of a previously-geolocalized aerial image dataset. After registration, a fused 3D point cloud is generated and then combined with images as inputs for land cover (here roofs) classification. Experiments show that the proposed matching method outperforms SIFT/ASIFT methods in both quantity and reliability of matching results, while the registration of MAV imagery achieves decimeter-level accuracy without using any onboard GPS/IMU data. Besides, the pixel-level classification that integrates information of point clouds and images achieves significantly higher accuracy than simply image-based classification.

***Index Terms*—** Micro Aerial Vehicle, Multi-scale, Image Registration, Fusion, Point Cloud, Roof Classification

1. INTRODUCTION

Emerging as a novel tool for image acquisition, MAVs close the gap between aerial and terrestrial photogrammetry. Its maneuverability and flexibility enable rapid acquisition of images, which exhibit higher resolution and contain richer details in comparison with images captured from manned aircraft. On the other hand, MAV photogrammetry still suffers from the instability of the MAV platform and low accuracy of the on-board GPS/IMU. As is shown in Table 1, photogrammetry based on MAVs and manned aircrafts have complimentary characteristics, therefore it is promising to fuse MAV imagery and airborne imagery to achieve both high resolution and geo-localization accuracy.

	MAV photogrammtery	Manned aircraft photogrammtery
Advantage	flexible big overlap rich details	large coverage stable high positioning accuracy
Disadvantage	safety unguaranteed unstable inaccurate GPS/IMU	time/money expensive weather-dependent inadequate details

Table 1. Comparison between MAV and manned aircraft photogrammetry

As the fundamental step of fusion, image matching has a great impact on subsequent steps like registration and 3D reconstruction. A variety of matching algorithms have emerged and can be generally divided into three categories: pixel value-based methods, e.g., Normalized-Cross-Correlation(NCC), which are vulnerable to image intensity changes and geometric deformations; frequency domain-based methods, e.g., wavelet transform-based methods, which are sensitive to local distortions [1]; local feature-based methods, which outperform the other two methods in the mentioned aspects and are widely used for matching images with different viewpoints, resolutions and orientations [2]. Among various feature-based matching algorithms, SIFT stands out for its robust scale and rotation invariant property, however, it still fails in matching MAV and airborne images even after eliminating the differences in scale and rotation, which is caused by ambiguous keypoint orientations and the misuse of ratio test. This paper presents a novel robust matching method, which outperforms classic SIFT in terms of quantity and reliability of the matching result. Based on this method, we georeference a MAV imagery block without GPS/IMU information using the orthophoto and DSM of a previously-geolocalized airborne image dataset, and then achieve fusion in 3D-space using the DSM.

Traditional 2D image classification methods utilize spectral information contained in different channels [3] and are therefore vulnerable to occlusion, illumination etc. Nevertheless, the geometric information contained in point clouds provide valuable features for classification. For example, points on the roof can be distinguished by the relative height to the ground level, which serves as a valuable feature. In this paper, we make use of 3D point clouds together with images for an extended classification (here roofs) and prove that the classification accuracy get significantly improved compared to the classification using only 2D images.

2. METHODOLOGY

The fusion workflow starts with the robust matching between MAV and the orthophoto of airborne images, continues with a common bundle adjustment for the geo-localization of MAV images, creates dense surface models with outlier detection, and performs a classification based on the fused datasets.

2.1. Robust self-adaptive matching based on SIFT

The challenges of matching MAV imagery and airborne imagery stem primarily from the following aspects: differences of scale and viewing direction, ambiguous keypoint orientations and misuses of ratio test. Our previous research[4] forcused on theoretical analysis of matching multi-scale images and has proved that:

1. The rotation invariance of SIFT does not work well for the matching of MAV imagery and airborne imagery. If we fix orientations of SIFT keypoints instead of letting SIFT determine the orientations itself, significantly more correct matches can be achieved;
2. Ratio test, an important step in the standard matching pipeline, actually discards many correct matches and lets wrong matches off when the image contains repeated structures, as the local descriptors of the repeated structure are quite similar;
3. The nearest neighbor does not necessarily contain the most correct matches. The nearest two or even nearest 100 neighbors contain about the same percentage of correct matches as the nearest neighbor;
4. Assuming the MAV image and the airborne image are both nadir view and aligned, the transformation between corresponding feature points approximates to a 2D-translation. This can serve as a geometric constraint for removing outliers.

Based on the findings above, we propose a self-adaptive matching method for matching imagery with considerable differences in scale and viewing direction. The steps are as follows:

1. Simulate all possible scales and viewing directions by downsampling and rotating the MAV image with a series of scaling and rotation factors, and then match each simulated image with the airborne image using NCC-based template matching. The convolution output should be highest for the simulated image which exhibits the same scale and viewing direction with the airborne image. Consequently, the scale difference s, rotation angle θ and translation vector T_x, T_y can be obtained, which roughly describe the transformation between the MAV image and the airborne image. Accordingly, the MAV image get roughly registered to the airborne image;
2. Extract SIFT features from the airborne image and the registered MAV image and assign each feature multiple orientations, i.e., rotate the image by different angles, e.g., $\pm 10^\circ, \pm 20^\circ$; each time the image is rotated, set the orientations of features to 0°. As a result, every SIFT feature is described by several different descriptors;
3. Involve all the descriptors in matching and select the nearest k neighbors as presumed correspondences;
4. Use the 2D-translation instead of ratio test as the geometric constraint to filter the presumed matches, i.e., for each keypoint i, whose pixel coordinates are (x^i_M, y^i_M) in the MAV image and (x^i_A, y^i_A) in the airborne image, and for each of its k presumed correspondences j $(j = 1 : k)$, whose pixel coordinates are (x^j_M, y^j_M) in the MAV image and (x^j_A, y^j_A) in the airborne image, calculate their coordinate differences $\Delta x^{i,j}$ and $\Delta y^{i,j}$ by $\Delta x^{i,j} = x^i_M - x^{i,j}_A$ and $\Delta y^{i,j} = y^i_M - y^{i,j}_A$. The correct matches are expected to fulfill the conditions $|T_x - \Delta x^{i,j}| \leq r$ and $|T_y - \Delta y^{i,j}| \leq r$, where r is the threshold related to the depth of the scene.

2.2. Co-registration of MAV images to reference aerial imagery

Limited by payloads, MAVs are often equipped with light GPS/IMU systems with low positioning accuracy. By contrast, manned aircrafts are usually equipped with high-end GPS/IMU and stable calibrated cameras, which can achieve centimeter-level accuracy. For this reason, we propose to register MAV images using the orthophoto and DSM of geolocalized airborne images for reference. Our previous work [5] still relied on GPS/IMU information to filter the SIFT matches, which were highly contaminated by outliers. By contrast, the proposed method does not need any GPS/IMU information and can achieve reliable matches and accurate geolocalization. Further details of the method are given below.

1. Match MAV images with the airborne orthophoto using the proposed matching method;
2. For each matching pair, $P_m(x_m, y_m)$ in the MAV image and $P_a(x_a, y_a)$ in the orthophoto, calculate its x and y coordinates in the world coordinate system (e.g. WGS84) referring to the orthophoto, and then look up its height z in the DSM;
3. Compute the corresponding 3D point (x, y, z) for each matching pair. Points appearing on more than two MAV images are then selected as ground control points (GCPs) in bundle adjustment.In this way, the MAV images are registered to the airborne images.

2.3. DSM generation with outlier detection

DSM generated from dense matching inevitably contain noises and outliers, which can be solved by fusion. Considering the DSM generated from MAV imagery exhibits much higher resolution than that of the airborne imagery, the latter should be firstly densified to the same resolution as the MAV DSM. Then the fused DSM can be calculated by adding the weighted height values of two DSMs.

2.4. Classification with 2D image features and 3D point cloud features

Points on the roof can be distinguished by the relative height to the ground level, which serve as a valuable feature. The classification proceeds as follows: first,the relative heights to the ground level of all 3D points are computed and normalized, while the image features are also extracted from RGB images. By projecting the image features to 3D space, we can find the corresponding 3D point for each Gabor feature. In this way, we use a combination of the image features together with the corresponding height features in a classification, resulting in a preliminary classification result. In practice, point clouds generated by dense matching often contain holes due to occlusion, reflection etc. To compensate such loss in the coverage, we calculate a probability map of the preliminary classification result using a sigmoid function and then interpolate over the whole area. After that, a Markov Random Field (MRF)-based classification is carried out using the interpolated probability map, which results in a refined classification with fewer holes.

3. EXPERIMENTS

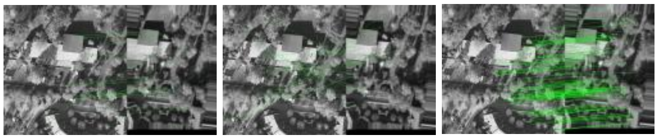
In this section experiments are carried out to test the proposed method. Qualitative and quantitative analysis of experimental results are presented.

3.1. matching

In order to verify the reliability and robustness of the proposed matching method, we tested it on several datasets which were captured at different time with different scales. Two aerial imagery datasets were acquired with the DLR 4k sensor system, "Eichenau" on November 2^{nd}, 2015 at an altitude of about 600m while "Germering" on June 17^{th}, 2015 at an altitude of about 1000m above ground level; With a slight time delay, MAV images were acquired from both test regions, "Eichenau" on 11^{nd}, 2015 with Sony Nex-7 and "Germering" on 11^{th}, 2014 with a GoPro Hero 3+ Black both at an altitude of around 100m. Besides, "Googlemaps" uses screenshot from Googlemaps for matching with MAV images, thus verifying that the method is also applicable for matching satellite imagery and MAV imagery. In "Ortho", the MAV image is registered to the airborne orthophoto with 9 times scale difference.

Dataset	Scaling	Rotation	Correct Matches		
			SIFT	ASIFT	Proposed method
Germering	5	101°	4	25	292
Eichenau	4	22°	14	54	511
Building	6	172°	32	101	132
Googlemaps	5	118°	31	91	679
Ortho	9	84°	9	15	283

Table 2. Number of correct matches generated by classic SIFT, ASIFT and proposed method.



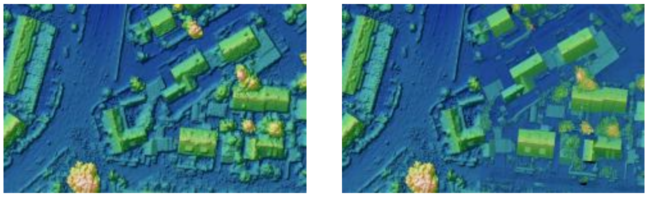
(a) SIFT (b) ASIFT (c) Proposed method

Fig. 1. Matching of MAV image and airplane orthophoto using SIFT, ASIFT and proposed method respectively

As described in the proposed method, the scaling, rotation and translation factors are computed automatically so that the MAV and airborne images are roughly aligned before matching. The threshold of ratio test in SIFT/ASIFT was set to 0.75 and the tilt value in ASIFT was set to 4, with which the the optimal result can be achieved. As for the proposed method, the nearest 3 neighbors were selected as match candidates as a trade off between speed and the number of matches, and the thresholds of translation in x and y directions were both set to 30 pixels. The matching results of MAV images and airplane orthophoto is illustrated in Figure 1, while the numbers of correct matches of each method are listed in Table 2, it can be seen that the proposed method results in remarkably more reliable matches than SIFT and ASIFT for flat scenes such as "Germering" and "Eichenau". However, when the scene depth of the scene is larger, like in dataset "Building", the simulation of tilt in ASIFT can play its role and therefore results in almost comparable matches.

3.2. registration and fusion

Following the proposed pipeline, the 3D coordinates of each match were computed with reference to the aerial orthophto and DSM generated from dense matching. 157 points were finally selected as GCPs to geolocalize the MAV images in a bundle adjustment, resulting in an average reprojection error of 0.816 pixel, afterwards the orthophoto and DSM of MAV images were generated. To evaluate the accuracy of registration, we measured the coordinates of several checkpoints in the orthophoto and DSM of airborne images and MAV images respectively and then calculate the difference. Table



(a) Before fusion (b) After fusion

Fig. 2. Comparison of DSMs(shaded) before and after fusion

Checkpoint	Error (m)		
	Δx	Δy	Δz
1	-0.21	-0.51	-0.64
2	0.28	-0.38	-0.029
3	0.98	-0.07	0.139
4	0.71	0.41	-0.206
5	0.26	0.37	-0.177
6	-0.25	0.38	-0.501

Table 3. Geolocalization accuracy of the MAV image dataset compared to the airborne image dataset

3 lists the errors of check points.

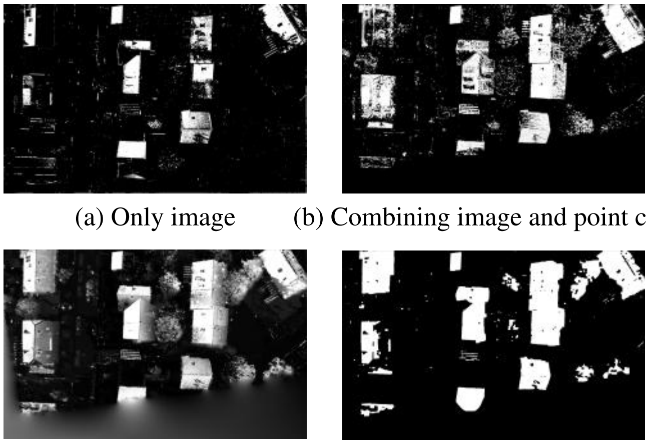
After registration, DSM fusion was then implemented. In Figure 2, (a) shows the original airborne DSM with blurred edge and inadequate details. (b) shows the fusion result of MAV DSM (covering residential area) and airborne DSM, which contains much richer textures and sharper edges.

3.3. Classification

To reduce the computation time, we carried out experiment on a piece of cropped data, chose 5% image pixels for training and used the linear SVM as the classifier. The results are illustrated in Figure 3. Specifically, (a) shows the pixel-level classification using only image features, whose accuracy was 84.75%; after involving additional height features from point cloud, the accuracy increased to 91.83%, but some parts of the roof were missing due to holes in the point cloud, as is shown in (b); (c) illustrates the interpolated probability map, which was then used as input in MRF-based classification. (d) shows the final result, which achieved an accuracy of 92.13%.

4. DISCUSSION AND CONCLUSION

This paper presents a new robust image matching approach with self-adaption to substantial differences in scale and viewing direction. The method is comprised of a scale and rotation-adapted scheme, a multi-orientation simulation scheme, a one-to-many matching strategy and a geometric constraint for outliers detection. Experiments show that our method achieved far more reliable matches than SIFT/ASIFT.



(a) Only image (b) Combining image and point cloud

(c) Interpolated probability map (d) MRF-based classification

Fig. 3. Comparison of classification results

In addition, the matching method was applied in georeferencing a MAV image block using the orthophoto and DSM of an airborne image dataset, which enables a decimeter-level geolocalization without using GPS/IMU data. In the end, we combined features from image and point cloud together in a classification and achievd significant improvement compared to traditional image-based classification. In summary, the fusion of multimodal data holds promising potential. Our future work will focus on the selection or design of point cloud features for multi-category classification.

5. REFERENCES

[1] M. Chen, Z. Shao, D. Li, and J. Liu, “Invariant matching method for different viewpoint angle images,” *Applied optics*, vol. 52, no. 1, pp. 96–104, 2013.

[2] B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[3] Xiaoying Jin and Curt H Davis, “Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information,” *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 14, pp. 1–11, 2005.

[4] T. Koch, X. Zhuo, F. Fraundorfer, and P. Reinartz, “A new paradigm for matching uav- and aerial images,” *submitted to XXIII ISPRS Congress*, 2016.

[5] X. Zhuo, F. Kurz, and P. Reinartz, “Fusion of multi-view and multi-scale aerial imagery for real-time situation awareness applications,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W4, pp. 201–206, 2015.