A NEW PARADIGM FOR MATCHING UAV- AND AERIAL IMAGES

T. Koch[a, *], X. Zhuo[b], P. Reinartz[b], F. Fraundorfer[b,c]

[a]Remote Sensing Technology, Technische Universität München, 80333 München, Germany - tobias.koch@tum.de
[b] The Remote Sensing Technology Institute, German Aerospace Center, 82234 Wessling, Germany -
(xiangyu.zhuo, peter.reinartz, friedrich.fraundorfer)@dlr.de
[c] Institute for Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria

Commission III, WG III/1

KEY WORDS: Image matching, Feature-based matching, Image registration, Geo-registration, 3D Reconstruction, Navigation, SIFT, A-SIFT

ABSTRACT:

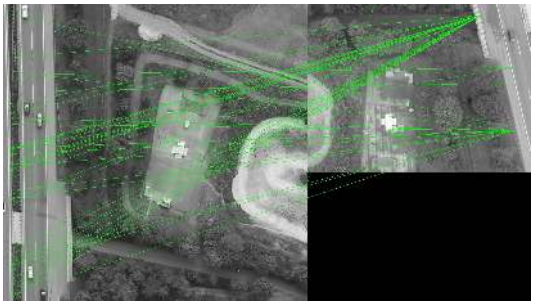
This paper investigates the performance of SIFT-based image matching regarding large differences in image scaling and rotation, as this is usually the case when trying to match images captured from UAVs and airplanes. This task represents an essential step for image registration and 3d-reconstruction applications. Various real world examples presented in this paper show that SIFT, as well as A-SIFT perform poorly or even fail in this matching scenario. Even if the scale difference in the images is known and eliminated beforehand, the matching performance suffers from too few feature point detections, ambiguous feature point orientations and rejection of many correct matches when applying the ratio-test afterwards. Therefore, a new feature matching method is provided that overcomes these problems and offers thousands of matches by a novel feature point detection strategy, applying a one-to-many matching scheme and substitute the ratio-test by adding geometric constraints to achieve geometric correct matches at repetitive image regions. This method is designed for matching almost nadir-directed images with low scene depth, as this is typical in UAV and aerial image matching scenarios. We tested the proposed method on different real world image pairs. While standard SIFT failed for most of the datasets, plenty of geometrical correct matches could be found using our approach. Comparing the estimated fundamental matrices and homographies with ground-truth solutions, mean errors of few pixels can be achieved.

1. INTRODUCTION

Image matching is a longstanding problem and widely used in many applications in the fields of Photogrammetry and Computer Vision. As a prior step of image registration, it is indispensable for many tasks like image stitching, mosaicing, 3d reconstruction, navigation, structure-from-motion, etc. Over the last decades, numerous different approaches have been developed, whereas feature-based methods with local descriptors are most commonly used when the image pairs have different viewpoints, resolutions and orientations (Zitova and Flusser, 2003). Although very good results can be achieved, even under very difficult conditions, it is interesting that there are still examples, where image matching is very problematic or even fails, surprisingly even for cases of image pairs that look very similar. In our context, this problem can be seen when trying to match UAV images and aerial images, which differ in geometric and temporal changes. With the increasing popularity of image acquisition using unmanned aerial vehicles, it seems reasonable to combine these datasets for joint 3d reconstructions. However, Figure 1 shows two of many typical examples, where feature-based matching of an down sampled UAV image and a cropped part of an aerial image fails.

SIFT matching is considered as the gold standard method by big parts of the community for image feature matching. And in fact it is true that SIFT matching can successfully be used for automatic matching for many different classes of image pairs. However, it does not seem to be apparent to the community that there still exists a large class of image pairs, which at first glance would look like easy matching candidates, but which nevertheless cannot successfully get matched using SIFT matching.

(a)



(b)

Figure 1: Image matching results of a downsampled UAV image (left) and a cropped aerial image (right) using SIFT-matching for the datasets ”Container” (a) and ”Highway” (b). Green lines indicate the apparent matches.

*Corresponding author

The main contributions of this paper are:

- a detailed investigation of SIFT-based matching of UAV images to aerial images with real world examples, in particularly regarding the influence of image rotation and the ratio-test.
- a new feature-matching approach, including a dense feature detection, a one-to-many matching scheme and a verification of the matches using pixel-distance histograms.

The paper is organized as follows: first, an overview about SIFT-related image matching is given in Section 2., followed by the analysis of the difficulties of matching UAV images to aerial images on real world examples. Section 4. presents the proposed feature matching approach. Experimental results are presented in Section 5., and finally, Section 6. concludes with discussions.

2. RELATED WORK

Image matching problems can be tackled with a variety of different approaches. Each of them has individual properties and is suited for different applications. Area-based methods, like Normalized-Cross-Correlation (NCC), can be applied if the images are perfectly aligned and the scene does not alter in appearance and temporal changes. However, a precise initialization of the unknown 2d translation vector must be given, otherwise they suffer from multi-modal matching results (Zitova and Flusser, 2003).

To overcome these problems, local features are commonly used and work successfully in many different applications. In terms of the repeatability of the feature-detector and the distinctiveness of the descriptor, SIFT (Lowe, 2004) has proven remarkably successful. Numerous extensions and variants of SIFT have been developed, most of them dealing with the reduction of computational time by approximations in specific implementations details, but rarely outperformed SIFT. Detailed evaluations of SIFT and its variants can be found in (Miksik and Mikolajczyk, 2012) and (Juan and Gwun, 2009).

(Morel and Yu, 2009) shows that the performance of SIFT decreases in the case of larger geometric transformations and proposed A-SIFT to overcome these problems. It turned out that A-SIFT has a better performance than SIFT, due to multiple prior image-transformations before applying standard SIFT. A comprehensive investigation is presented in (Apollonio et al., 2014).

Although SIFT and A-SIFT are supposed to be scale-invariant, this is naturally limited to few scale spaces. In case of large scale differences, the images are usually sampled on the same scale level to improve the matching performance.

Using feature-based matching methods to images with repetitive structures often cause problems with similar descriptors of the local properties of the scene. In SIFT, the ratio-test is applied to discard mismatches by rejecting all potential matches with similar descriptors. If only few matches are available, this step is critical as presented in (Sur et al., 2011) and (Mok et al., 2011).

In the context of matching UAV images to aerial images, several works rely on prior scene knowledge and apply area-based matching methods. (Fan et al., 2010) uses GPS and IMU information for pre-aligning the UAV image and performs template matching, combining edges and entropy features. (Conte and Doherty, 2009) applies NCC on pre-aligned images for real-time navigation. (Huang et al., 2004) uses a hybrid method, combining feature-based and area-based matching by presenting salient region features and (Yuping and Medioni, 2007) make use of Mutual Information.

Considering the problems coming with area-based matching mentioned above, we would like to stay close to feature-based matching using the promising SIFT features.

3. ANALYSIS OF SIFT-BASED MATCHING

In this section, we present test cases where the standard SIFT matching workflow fails, very much to our surprise. We also analyse the reason of this failure by investigating into the influence of ratio-test and rotation. Experiment results show that the rotation invariance of SIFT is not as good as it has been considered to be and the deficiancy in the rotation estimation of SIFT leads to non-optimal matching results. Based on this conclusion, we propose a new strategy for matching aerial images and UAV images, and compare the performance of the proposed method with conventional SIFT and A-SIFT approaches on different datasets. For the experiments we use the SIFT implementation in OpenCV 3.0. Specifically, the contrast threshold is 0.04, the edge threshold is 10, the sigma of the Gaussian is 1.6 and each octave contains 3 layers.

3.1 SIFT

Unlike the matching between two aerial images, the matching between UAV and aerial images (typical altitudes are 80 - 120 m and 800 - 1500 m respectively) is more complicated and challenging due to the substantial scale difference in the range of 5 to 15 times. With the help of preliminary knowledge from GPS data, the scale difference can be estimated and eliminated by down sampling the UAV image. How to extract sufficient and reliable image correspondences is crucial for subsequent image matching and orientation. A popular workflow regarding this issue goes as follows (Sur et al., 2011):

1. Detect interest points of both images and compute the descriptors based on local photometry.
2. Match the interest points according to the similarity of descriptors.
3. Remove the mis-matches (outliers) by finding a subset of correspondences which are in accordance with the underlying epipolar geometry, e.g. fundamental matrix.
4. Determine further correspondences using guided matching, i.e., find more putative correspondences by relaxing step 2, and then prune the correspondences based on the estimated geometric constraint in step 3.
5. Prune the final set of image correspondences like in step 3.

In step 1, among the state-of-the-art matching algorithms, SIFT enjoys a high popularity for its scale and rotation invariance. Besides, the SIFT descriptor has been proved to surpass the other local descriptors (Mikolajczyk and Schmid, 2005). Considering the substantial scale and rotation difference between the UAV and aerial images, it makes sense to implement traditional SIFT detector and descriptor for interest point extraction.

Step 2 is actually a critical part. Since the global threshold on the Euclidean distance between descriptors does not perform well (Lowe, 2004), the ”ratio-test” proposed by D. Lowe is widely

Dataset	Keypoints		Correct matches		
	aerial	UAV	nearest	ratio-test	nearest 100
Container	8682	2460	32	9	498
Highway	8926	2536	108	16	273
Urban	203026	8126	595	225	872
Urban2	32324	9766	252	30	880
Pool	3682	1835	224	79	344
Pool2	2866	2861	117	29	277
Building	4229	3883	185	34	558
Googlemaps	4374	3204	66	37	176

Table 1: Analysis of standard SIFT matching on the proposed datasets in Figure 7. Number of feature-points detected by the SIFT-detector, correct matches before and after applying the ratio-test and possible matches according to 100 nearest neighbors.

used to define correspondences, i.e., to impose that the ratio of the distances to the first and the second nearest neighbor is smaller than a certain threshold. This method works well in most cases, however, when the scene contains objects that have similar local properties, the distance ratio can be so high that these features with similar descriptors are defined as outliers. To investigate how many correct matches are actually discarded by the ratio-test, we implemented the standard SIFT method for matching and counted the correct matches before and after the ratio-test. Particularly, the distance of first two nearest neighbors are computed and compared with the threshold. After testing with a range of different values, the threshold was finally defined as 0.75, with which we can achieve the best matching result. Considering the number of matches can be numerous and it is unrealistic to check every single match manually, we therefore computed the fundamental matrix between the two images with dozens of manually selected image correspondences, and then apply the epipolar constraint using the derived fundamental matrix to filter the raw matches. Afterwards, the filtered matches are again checked by manual inspection to ensure the purity of correct matches. The final results of the eight datasets presented in Figure 7 are listed in Table 1. It needs to be pointed out that only a cropped part of the aerial image with almost the same image content of the UAV image was used for interest point detection, otherwise SIFT would fail to find a single correct match in datasets ”Container” and ”Highway”.

According to the experimental result, most of the correct matches are lost due to the ratio-test. As a consequence, the number of correct matches is too low for a reliable matching result in almost all datasets. However, the number of correct matches (using the same feature points and descriptors) can be significantly increased, if multiple nearest neighbors are considered as matching candidates. Figure 2 shows the cumulative number of correct matches for the first 100 nearest neighbors for the ”Container” dataset. The last column of Table 1 lists the number of possible matches for the other datasets.

After matching, a hypothesized correspondence set has been generated. As the features are based on local property only and not constrained by the global geometry, there may be many mis-matches in the putative correspondences, especially when the image contains many repeated patterns or similar objects. The popular way to deal with problem is to estimate the epipolar geometry using RANSAC based methods, but it fails to find the optimal set in the experiment. Figure 1 illustrates two examples where RANSAC failed to work.

Dataset	Inliers / Matches		
	Std. SIFT	Std. SIFT Rotation aligned	SIFT Rotation aligned Fixed-orientation
Container	4 / 108	8 / 99	25 / 126
Highway	2 / 178	24 / 118	65 / 122
Urban	194 / 543	185 / 551	342 / 654
Urban2	14 / 266	10 / 292	31 / 315
Pool	79 / 157	72 / 145	102 / 176
Pool2	25 / 93	21 / 96	55 / 136
Building	32 / 125	25 / 112	56 / 166
Googlemaps	31 / 137	32 / 119	56 / 164

Table 2: Analysis of the influence of image-rotation. Inliers and matches in case of original images, pre-aligned images and pre-aligned images with fixed orientation in the SIFT-detector.

3.2 Influence of rotation

As is shown in the above matching results, the SIFT method has unsatisfactory performance for the matching between UAV image and aerial image, and even fails to find any correct match in some cases. Considering the fact that the UAV and aerial images are both almost nadir view and the difference in scale has already been eliminated, the only observable difference is that the two images are not aligned in rotation. Therefore, the rotation invariant property of SIFT needs to be reconsidered and evaluated. To investigate into the problem, a series of experiments were carried out to test the influence of rotation. As listed in Table 2, firstly we implemented standard SIFT matching on the original unaligned images (denoted by ’Std. SIFT’) and on the aligned image (denoted by ’Std. SIFT Rotation aligned’); Besides, instead of letting SIFT assign the orientation for each keypoint, we forced the orientation of all the detected key points in the aligned images manually to be a fixed value, here it was 0 for aligned images (denoted by ’Fix-orientation’). The matching result was represented by the number of putative correspondences after ratio-test (denoted by ’Matches’) and the final correct matches pruned by RANSAC together with manual inspection (denoted by ’Inliers’). However, RANSAC does not work well for the matching using standard SIFT, and results in many remaining outliers. So the actual number of inliers can be fewer than listed in the table. It is worth noting that the performance of matching between rotation-aligned images using standard SIFT does not get improved; however, the number of inliers increased substantially after we fixed the orientation of the keypoints. The experiment result shows that the rotation invariance of SIFT does not always work well, at least for the scenes in our datasets.

For further investigation into the influence of rotation, we also made a comparison with the A-SIFT method, as Table 3 shows.

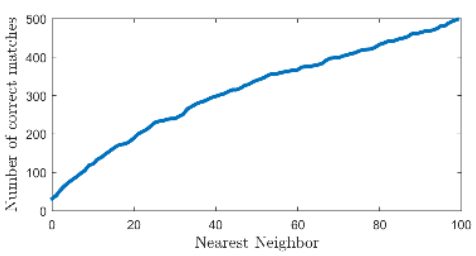


Figure 2: Cumulative number of possible correct matches after considering multiple nearest neighbors during feature matching for the ”Container” dataset.

Dataset	Inliers / Matches		
	SIFT Multi-rotation Fix-orientation	Std. A-SIFT	A-SIFT Fix-orientation
Container	52 / 217	25 / 281	46 / 283
Highway	79 / 219	56 / 249	70 / 237
Urban	673 / 1392	684 / 1800	1091 / 2132
Urban2	100 / 693	54 / 1069	109 / 994
Pool	346 / 509	355 / 600	375 / 620
Pool2	127 / 308	73 / 346	199 / 404
Building	132 / 356	101 / 382	180 / 424
Googlemaps	162 / 343	91 / 330	188 / 430

Table 3: Comparison with A-SIFT. Inliers and matches for pre-aligned images using standard SIFT with fixed orientation, A-SIFT and pre-aligned images on A-SIFT with fixed orientation.

First, we rotate the original images to multiple angles, here it was from -90 to 90 with a step of 20, and then fixed the orientation of each keypoint to 0 so that the same keypoint has different descriptors in different rotated images (denoted by 'SIFT Multi-rotation Fix-Orientaion'); In comparison, the standard A-SIFT (denoted by 'Std. A-SIFT') with a tilt value of 4 achieved fewer correct matches at even higher computation cost; Inspired by this finding, we also fix the orientation in A-SIFT (denoted by 'Fix-orientation') in the same way, and the matching performance get improved significantly. Comparing the results in column 2 and column 4, it can be seen that when the orientation is fixed, multi-rotation SIFT results in equivalent or even more inliers than A-SIFT for 2d-like scenes, e.g., Container, Highway and Urban2 dataset. However, when the scene contains objects with different scene depths, the simulation of tilt in A-SIFT can play a role and results in more inliers.

Step 4 and 5 are optional and aim at finding further correct correspondences and refine the estimation of epipolar geometry. This process can be iterated until the number of correspondences is stable. However, the guided matching works only under good initial estimation of epipolar geometry. If step 3 fails to achieve reliable correspondences and robust fundamental matrix, it does not make sense to apply guided matching to find more correspondences.

4. PROPOSED FEATURE MATCHING APPROACH

4.1 Concept

In this section, we propose a new method for finding correct feature correspondences using the SIFT-descriptor under the assumption of equally scaled and oriented images. In this case, corresponding feature points will only differ in a 2d translation. This translation vector can be extracted generating pixel-distance histograms of all putative matches (a similar idea has already successfully been used by (Baatz et al., 2010) for visual location recognition using mobile phones). If the set of matching hypotheses contains many correct matches, these will generate a distinct peak in the distance histograms, while wrong matches will be distributed randomly. To ensure a large number of correct matches in the set of matching hypotheses, we replace the original SIFT-feature-detector with a denser detection scheme by using the boundaries of superpixels as feature points on the one hand, and performing one-to-many matching of the feature descriptors on the other. The resulting large set of matching hypotheses is then verified by extracting geometrical correct matches for which the pixel-distances are close to the extracted peaks of the distance histograms.

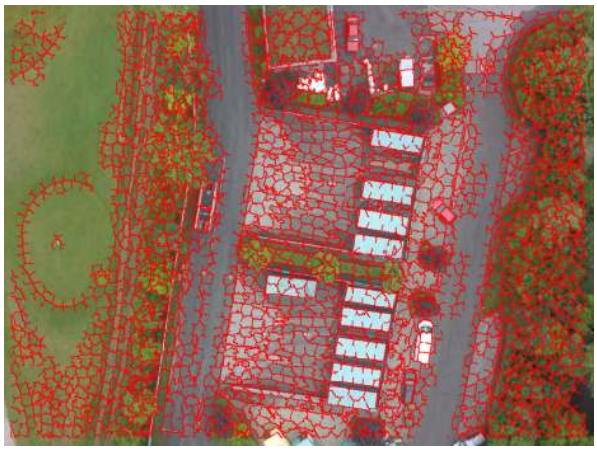


Figure 3: Feature points (red) defined as the border pixels of a superpixel segmentation step.

The following sections provide a more detailed description of the proposed method, whose main contents are:

- Presentation of a new and dense feature-point detection scheme
- Motivation and description of the one-to-many feature matching
- Geometric verification of the matching hypotheses using pixel-distance histograms
- Extension for feature matching in situations of unknown image-rotations

4.2 Feature extraction

Firstly, we assume a pre-alignment of the UAV image towards the aerial image is possible by the information provided from the GPS/IMU sensors. To achieve a sufficient number of point correspondences with a homogeneous distribution in the image, we aim to extract feature points in all highly textured areas of the images. Since this is not always the case using SIFT-detector, we recommend using all pixels in both images as putative feature points. To reduce the computational overhead and reject weak SIFT-descriptors in homogeneous areas, we apply an image segmentation in advance using a superpixel method (SLIC, (Achanta et al., 2012)). The resulting boundaries of the superpixels are mostly located at highly textured areas of the images and therefore all pixels along these boundaries are considered as feature points. Figure 3 shows the resulting feature point detection of an UAV image after removing feature points at homogeneous areas at the superpixel boundaries. The number of superpixels, which can be set in the SLIC implementation, controls the level of detail and moreover the number of feature points. For each feature point given along the superpixel boundaries, a SIFT-descriptor is computed. As the scale of the UAV image is adapted, the scale space in the SIFT-descriptor should be equal for both images. The same should be done with the orientation of the feature points, when the image rotation is known beforehand. Section 4.5 will show that the image rotation can also be recovered if no information about the image heading is available. In this case, the orientation of the feature points should be computed using the SIFT-detector.

(a) (b)

Figure 4: Feature points in the aerial image with the closest descriptor distances (b) to a feature point at the corner of a container in the UAV image (a).

4.3 Feature matching

After computing SIFT-descriptors for all feature points, feature matching is performed by comparing these feature vectors by computing the euclidean distance for all pairs of feature points. We expect very similar matching distances for feature points describing comparable scene objects in the images. If only the nearest neighbor is considered as a match, most of the possible matches will be missed as shown in Table 1. To overcome this problem, a one-to-many matching scheme is applied, by taking the k-nearest neighbors as putative matches.

Figure 4 shows an example of the nearest matches for an ambiguous feature point. Using standard methods, the correct match would most probably be missed for two reasons: first, it is very unlikely that the correct match is the nearest neighbor and second, it would be rejected afterwards when applying the ratio-test, because the descriptors of the putative matches are very close to each other.
To speed-up the matching process, approximate nearest neighbor (ANN) can be used instead of an exhaustive search.

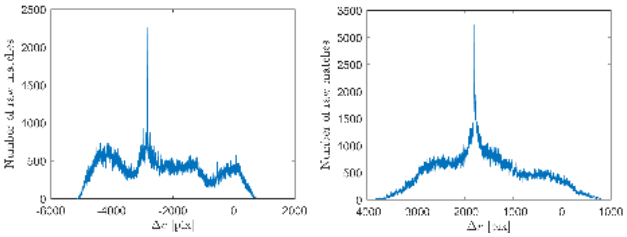
To reduce the number of wrong matching hypotheses, a threshold according to the feature matching-distance is applied to discard clear mismatches, as this is also proposed in the original SIFT-matching (Lowe, 2004). Our experiments showed, that 0.2 is a good trade-off between rejecting strong outliers and retaining enough correct matches.

4.4 Geometric match verification

The following section describes the strategy of extracting geometrical correct matches from the large set of putative matches with the help of pixel-distance histograms.

As we use a much denser feature point extraction and consider multiple nearest neighbors, we suppose that a lot of correct matches are inside this set of matches, although the ratio of outliers is expected to be very high. We use the fact, that geometrical correct matches only differ in an unknown global 2d translation vector. This translation can be recovered by simply computing coordinate differences of all putative one-to-many matches. The pixel differences $\Delta r^{i,j}$ and $\Delta c^{i,j}$ of a feature point i in the UAV image with row- and column coordinates r^i_{UAV} and c^i_{UAV} and all of its putative matches j ($j = 1 : k$) in the aerial image $r^{i,j}_{AIR}$ and $c^{i,j}_{AIR}$ are expressed as: $\Delta r^{i,j} := r^i_{UAV} - r^{i,j}_{AIR}$ and $\Delta c^{i,j} := c^i_{UAV} - c^{i,j}_{AIR}$.

Figure 5 shows an example of two corresponding distance histograms for the "Container" dataset. While distances of wrong



(a) (b)

Figure 5: Geometric verification of the matches with pixel-distance histograms. Distribution of pixel-distances for all putative matches according to the one-to-many matching in column- (a) and row- (b) direction. Distinct peaks represent unknown 2d translation.

matches are distributed randomly, a distinct peak is expected for geometrical correct matches. The coordinates of this peak p_r and p_c can be extracted and inliers are defined as matches which are close to both of the peaks. While this assumption only holds if the image planes are perfectly parallel, the scale and rotation of the images are identical and corresponding feature points have the same scene depth, a distance threshold T for extracting the correct matches is used to compensate any of these problems. That means, correct matches should satisfy $|p_r - \Delta r^{i,j}| \leq T \quad \cap \quad |p_c - \Delta c^{i,j}| \leq T$. The value of the distance threshold T is dependent on the scene depth and the quality of the pre-alignment of the images. Increasing this threshold allows to compensate larger uncertainties and more matches in general, but also accepting more outliers into the set of geometrical matching hypotheses.

These raw matches can now be used for estimating a fundamental matrix or homography in combination with RANSAC methods in order to reject remaining outliers satisfying the geometric constraint. After computing the fundamental matrix, a guided matching method, as presented in Section 2, can be applied to find more matches if the threshold was chosen too small.

4.5 Handling bad initial rotation

While scale-adaption of UAV images can be done robustly using accurate pressure heights and accurate localization sensors on the plane, precise orientation-adaption fails for many UAVs, due to inaccurate heading information provided by low quality IMUs. Our assumption of reducing the matching problem to a 2d translation vector fails in case of different image rotations. However, we can estimate the image rotation by computing SIFT-orientations for all feature points and use them for matching. Although we showed in Section 3.2, that fixing the orientation of the feature points in the SIFT-descriptors performs better, a sufficient number of correct matches still can be found.

To obtain the unknown image rotation, it is equally divided into discrete rotation values between $[-180°, 180°[$. For each rotation, the coordinates of the feature-points in the UAV image are rotated around the image center and pixel-distance histograms are calculated according to Section 4.4. The maximum number of raw-matches (within the threshold T) is kept for all rotation values. Figure 6 shows the number of raw matches for different image rotations according to the "Container" dataset. The distinct peak at $-104°$ represents the unknown image rotation. Ideally, this method works for inaccurate image rotations provided by the IMU, but it may be used for a full 360° search as well.

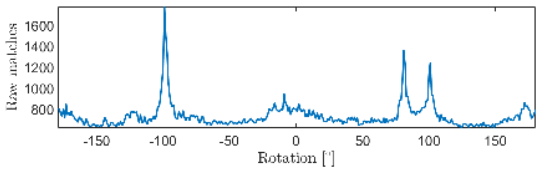


Figure 6: Recovering the unknown image rotation by rotating feature-points around image center before calculating distance-histograms. Maximum number of raw matches represents image rotation.

5. EXPERIMENTS

5.1 Data

In this section we present the quantitative and qualitative results of applying our method on several real images. In order to validate the robustness and accuracy of the proposed method, we use the same image pairs presented in Section 3., where standard SIFT performed poorly in most of the cases. In contrast to the results in Table 1, the matching should now be performed on the original, uncropped aerial images. As can be seen in Figure 7, only a small portion of the aerial images is pictured in the UAV images. Thus, it is also tested if the matching benefits from our geometric constraints in the presence of large search areas.

All image pairs are provided with information on positions and orientations from GPS and IMU, such that a pre-alignment of the images could be done beforehand.
Aerial images are captured at approximately 1000 meters altitude from a 18 MP Canon EOS-1D camera with a slightly oblique view. UAV images are captured from different cameras (GoPro Hero 3+ Black and Sony Nex7) in various altitudes (80 - 120 m) in almost nadir-directions. A scale difference of at least five times is expected in all datasets.

5.2 Processing details

Since the aerial images are captured at 12° boresight angle, a projective transformation with the known angle is applied in advance to get nadir-directed images.
Knowing internal camera parameters and altitude information, we can scale the UAV image towards the aerial image. Since the heading of the UAV (Asctec Falcon 8) can not be assumed to be very precise, the images are first rotated according to the Yaw-angle provided by the IMU, while a rotation search (as proposed in Section 4.5) is applied with an uncertainty of $\pm 7°$.
After pre-processing, the image pairs are processed accordingly to our method proposed in Section 4.. As parameters, we use 3500 superpixels for the UAV image, T = 12 [pix] as distance threshold and 50 nearest neighbors for the one-to-many feature matching.

5.3 Results

For evaluation, the raw-matches created with our method are used to estimate a homography H and fundamental matrix F together with RANSAC. Additionally, we use ground-truth homographies and fundamental matrices for each image pair. A mean transfer error (denoted by: Error_gt (H)) can be used to evaluate the estimated homographies by transforming inlier feature points in the UAV image with the estimated homography and calculate the euclidean distance to the projected feature point with the ground-truth homography. For evaluating the matches according to the epipolar geometry, the mean distance of all matches to their corresponding epipolar lines are computed (denoted by: Error_gt (F)).

	Raw matches (SIFT)	Inliers (F) / Error_gt (F)	Inliers (H) / Error_gt (H)
Container	58	14 / 666.26	9 / 1767.55
Highway	49	15 / 1996.30	9 / 2210.20
Urban	229	55 / 12.62	33 / 4.71
Urban2	151	15 / 1154.07	8 / 1616.19
Pool	162	52 / 0.83	33 / 1.63
Pool2	107	18 / 618.54	10 / 1308
Building	161	20 / 710.32	11 / 1273.87
Googlemaps	99	25 / 30.08	11 / 73.1

Table 4: Results using Standard-SIFT: Number of raw matches after applying our method for all datasets. Inliers after estimating fundamental matrix (F) and homography (H) using RANSAC. Mean errors [pix] according to ground-truth F and H.

	Raw matches (our)	Inliers (F) / Error_gt (F)	Inliers (H) / Error_gt (H)
Container	24010	11215 / 2.59	6235 / 7.01
Highway	7937	5893 / 3.10	2904 / 2.88
Urban	60936	30339 / 1.83	15928 / 2.65
Urban2	19106	10215 / 1.31	4441 / 3.68
Pool	63569	39295 / 1.87	15386 / 1.87
Pool2	33237	16921 / 2.01	6792 / 4.12
Building	38174	18045 / 3.42	6363 / 2.17
Googlemaps	82763	32705 / 3.02	11883 / 2.40

Table 5: Results using proposed method: Number of raw matches after applying our method for all datasets. Inliers after estimating fundamental matrix (F) and homography (H) using RANSAC. Mean errors [pix] according to ground-truth F and H.

The quantitative results using standard-SIFT and our proposed method are summarized in Table 4 and Table 5.

Using entire aerial images (in contrast to the cropped aerial images in Section 3.), standard-SIFT failed in almost all datasets. Only "Pool" performed better than our method regarding the error of the ground-truth homography and fundamental matrix, but the number of inliers is very low. Even small changes in the scene (see "Pool2") can lead to a failure of the matching. The datasets "Urban" and "Googlemaps" are the only image pairs that found some correct matches, but still contain large errors.

Depending on the scene content, tens of thousands raw matches could be found satisfying the geometric constraint, listed in the first column of Table 5. Only exception is the "Highway" image pair providing remarkably less raw matches. This is caused by a large scene depth, lots of hardly matchable vegetation and dynamic objects on the pictured skate park. The ratio of inliers according to the estimated fundamental matrix (listed in the second column) ranges between 40 and 75%. If we use these inlier matches and calculate the mean distance error according to the ground-truth epipolar lines (second part of the second column) all errors are less than 3.5 pixels. Higher values are expected for image pairs with larger scene depth which is hard to recover using our assumption and a fixed value for the threshold T.

The second and third rows in Figure 8 show the locations of the matched feature points according to the fundamental matrix in the UAV images. It becomes apparent here that the matches are mostly located at highly textured image regions as a result of the superpixel segmentation. Compared to the SIFT-detector, the feature points are highly locally densified. The substitution of the ratio test with the geometry constraints shows that correct matches can be even found at repetitive image regions (like the container, swimming pool borders or roadsides). In the most

cases, a homogeneous distribution of the matches is achieved which stabilizes the estimation of the fundamental matrix and homography. An example of the superiority of feature-based methods compared to area-based methods is the ability to robustly match images which are only partially overlapped (see "Pool2").

The raw matches can also be used for estimating projective transformations. As homographies only consider plane-to-plane transformations, lots of mismatches at structures with different scene depths are discarded. The mean transfer error ranges between 2 and 3 pixels for the most cases, which represents a good result (ca. 20-30 cm ground distance). In some cases (Container, Pool, and Urban) too many raw matches are located at different scene depths. As result, the estimation provides wrong tilts, which explains the errors up to 7 pixel. The first row in Figure 8 shows the aerial images together with the projected UAV images after applying the estimated homography.

It should be noted that in all of our experiments the method was able to recover the unknown image orientation from the inaccurate initial values.

6. DISCUSSION

This paper dealt with SIFT-based image matching on problematic image pairs, like low altitude UAV images and high altitude aerial images. We showed, that the state-of-the-art SIFT and A-SIFT-methods often fail in case of large differences in image scaling, rotation and temporal changes of the scene. Even if the images are pre-aligned, only a low number of correct matches can be achieved which restricts an automatic image matching without user control. For this reason, a method was proposed that uses SIFT-descriptors together with a new feature matching approach, including a novel feature point detector, a one-to-many matching scheme and a geometric verification of the putative matches using pixel-distance histograms. A huge number of correct matches can be found, even at image regions with repetitive patterns.
Main limitation of this approach is the assumption of available information on position and orientation for a pre-alignment of the UAV image. Although missing image rotations can be recovered, a scale-adaption of the UAV image is indispensable. Further, we assume flat surfaces and small scene depths, which impedes our method in urban environments. In presence of various scene depths (like a ground surface and several buildings with flat roofs), multiple peaks will appear in the distance histograms that should be traced independently to find correct matches according to different scene depths.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and Susstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)., Vol. 34number 11, IEEE, pp. 2274–2282.

Apollonio, F., Ballabeni, A., Gaiani, M. and Remondino, F., 2014. Evaluation of feature-based methods for automated network orientation. International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences 40(5), pp. 47–54.

Baatz, G., Köser, K., Chen, D., Grzeszczuk, R. and Pollefeys, M., 2010. Handling urban location recognition as a 2d homothetic problem. In: Computer Vision–ECCV, Springer, pp. 266–279.

Conte, G. and Doherty, P., 2009. Vision-based unmanned aerial vehicle navigation using geo-referenced information. EURASIP Journal on Advances in Signal Processing 2009, pp. 10.

Fan, B., Du, Y., Zhu, L. and Tang, Y., 2010. The registration of uav down-looking aerial images to satellite images with image entropy and edges. In: Intelligent Robotics and Applications, Springer, pp. 609–617.

Huang, X., Sun, Y., Metaxas, D., Sauer, F. and Xu, C., 2004. Hybrid image registration based on configural matching of scale-invariant salient region features. In: Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)., IEEE, pp. 167–167.

Juan, L. and Gwun, O., 2009. A comparison of sift, pca-sift and surf. International Journal of Image Processing (IJIP) 3(4), pp. 143–152.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2), pp. 91–110.

Mikolajczyk, K. and Schmid, C., 2005. A performance evaluation of local descriptors. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)., Vol. 27number 10, IEEE, pp. 1615–1630.

Miksik, O. and Mikolajczyk, K., 2012. Evaluation of local detectors and descriptors for fast feature matching. In: 21st International Conference on Pattern Recognition (ICPR), IEEE, pp. 2681–2684.

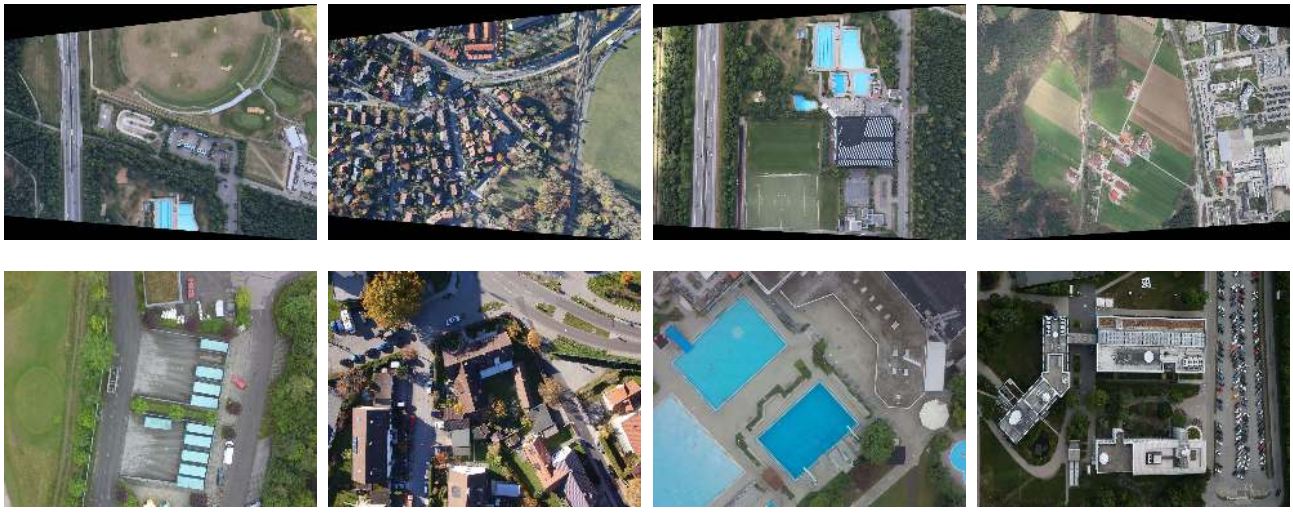
Mok, S. J., Jung, K., Ko, D. W., Lee, S. H. and Choi, B.-U., 2011. Serp: Surf enhancer for repeated pattern. In: Advances in Visual Computing, Springer, pp. 578–587.

Morel, J.-M. and Yu, G., 2009. Asift: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences 2(2), pp. 438–469.

Sur, F., Noury, N. and Berger, M.-O., 2011. Image point correspondences and repeated patterns. Research Report RR-7693, INRIA.

Yuping, L. and Medioni, G., 2007. Map-enhanced uav image sequence registration and synchronization of multiple image sequences. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Vol. 17.

Zitova, B. and Flusser, J., 2003. Image registration methods: a survey. Image and vision computing 21(11), pp. 977–1000.



(a) Container (b) Urban1 (c) Pool1 (d) Building

(e) Highway (f) Urban2 (g) Pool2 (h) Googlemaps

Figure 7: Datasets used in this paper: Each column represents one (pre-processed) aerial image (first row) and two UAV images ((a) - (d) and (e) - (g)) that should be matched to the aerial image. For the ”Building” dataset, the UAV image (d) should be matched to the aerial image (top right) and to a cropped part of a googlemaps image (h).

Figure 8: Qualitative results: UAV images are projected with estimated homographies to the aerial image (first row). Second and third row show the distribution of the matched feature points (yellow dots) remained after estimating the fundamental matrix with RANSAC.