remote sensing

MDPI

Article

Automatic UAV Image Geo-Registration by Matching UAV Images to Georeferenced Image Data

Xiangyu Zhuo [1,*], Tobias Koch [2], Franz Kurz [1], Friedrich Fraundorfer [1,3] and Peter Reinartz [1]

1 Remote Sensing Technology Institute, German Aerospace Center, 82234 Wessling, Germany; franz.kurz@dlr.de (F.K.); fraundorfer@icg.tugraz.at (F.F.); peter.reinartz@dlr.de (P.R.)
2 Remote Sensing Technology, Technische Universität München, 80333 Munich, Germany; tobias.koch@tum.de
3 Institute for Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria
* Correspondence: xiangyu.zhuo@dlr.de; Tel.: +49-8153-28-4235

Academic Editors: Norman Kerle and Prasad S. Thenkabail
Received: 14 February 2017; Accepted: 9 April 2017; Published: 17 April 2017

Abstract: Recent years have witnessed the fast development of UAVs (unmanned aerial vehicles). As an alternative to traditional image acquisition methods, UAVs bridge the gap between terrestrial and airborne photogrammetry and enable flexible acquisition of high resolution images. However, the georeferencing accuracy of UAVs is still limited by the low-performance on-board GNSS and INS. This paper investigates automatic geo-registration of an individual UAV image or UAV image blocks by matching the UAV image(s) with a previously taken georeferenced image, such as an individual aerial or satellite image with a height map attached or an aerial orthophoto with a DSM (digital surface model) attached. As the biggest challenge for matching UAV and aerial images is in the large differences in scale and rotation, we propose a novel feature matching method for nadir or slightly tilted images. The method is comprised of a dense feature detection scheme, a one-to-many matching strategy and a global geometric verification scheme. The proposed method is able to find thousands of valid matches in cases where SIFT and ASIFT fail. Those matches can be used to geo-register the whole UAV image block towards the reference image data. When the reference images offer high georeferencing accuracy, the UAV images can also be geolocalized in a global coordinate system. A series of experiments involving different scenarios was conducted to validate the proposed method. The results demonstrate that our approach achieves not only decimeter-level registration accuracy, but also comparable global accuracy as the reference images.

Keywords: unmanned aerial vehicle; image registration; geo-registration; point cloud

1. Introduction

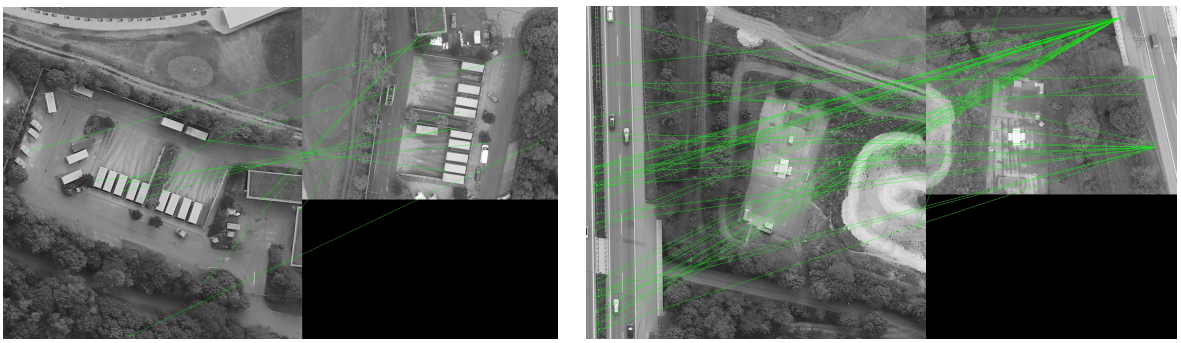
Emerging as novel image acquisition platforms, unmanned aerial vehicles (UAVs) bridge the gap between aerial and terrestrial photogrammetry and offer an alternative to conventional airborne image acquisition systems. In comparison to airborne or satellite remote sensing, UAVs stand out for low cost, the utility to be used in hazardous or inaccessible areas and the ability to achieve high spatial and temporal resolutions. Table 1 compares the main features of UAVs and manned aircraft based on the surveys of [1,2]. In contrast with manned aircraft, UAVs have smaller coverage due to lower flight altitude, but they are able to achieve high ground sampling distance (GSD) with lower cost and better flexibility. While manned aircraft require big landing fields and pilots, UAVs only need small landing sites and can be remotely controlled; therefore, they can work even in hazardous areas and severe weather conditions. Hence, UAVs have been widely involved in remote sensing applications, such as disaster management, urban development, documentation of cultural heritage or agriculture management [3].

Table 1. Comparison between UAV and manned aircraft photogrammetry.

	UAV Photogrammetry	**Manned Aircraft Photogrammetry**
Coverage	m^2–km^2	km^2
Image resolution/GSD	mm–cm	cm–dm
Geo-registration possibility	low quality GNSS/IMU meter-level accuracy	high quality GNSS/IMU centimeter-level accuracy
Price and operating cost	low-moderate	high
Flexibility	applicable in hazardous areas works in cloudy/drizzly weather remotely controlled	less mobile weather-dependent pilot needed

Accurate geo-registration of UAV imagery is a prerequisite for UAV geolocalization and many photogrammetric applications, such as generating georeferenced orthophotos, 3D point clouds or digital surface models (DSMs). However, accurate geo-registration of UAV imagery is still an open problem. Limited by on-board payload restrictions, UAVs are equipped with lightweight GNSS/IMU systems, whose georeferencing accuracies are in the range of meters [4] and far from the centimeter-level accuracy of airborne photogrammetry [5,6]. In order to achieve higher geo-registration accuracy beyond hardware limits, we use a pre-georeferenced aerial or satellite image as a reference and register the UAV image to the reference image with a novel feature-based image matching method.

In the field of image matching, numerous algorithms for different matching scenarios have been proposed in the last few decades. The biggest challenge for UAV and aerial image matching lies in the substantial differences in their scales, viewing directions and temporal changes. For instance, the flight altitude of UAV platforms is about 50 m–120 m above the Earth, whereas aerial images are usually captured at 800 m–1500 m from different viewing directions. Although state-of-the-art feature-based image matching methods are generally working fine for many different image pairs and are said to be invariant to changes in viewpoints, wider baselines and local changes of the scene, they surprisingly failed in many of our test cases. Figure 1 illustrates two typical cases of UAV and aerial image matching using SIFT [7].



(**a**) `Container` (**b**) `Highway`

Figure 1. Typical cases from the datasets (**a**) `Container` and (**b**) `Highway` showing the results of matching UAV and aerial images using SIFT, where the left of the subfigure is a downsampled UAV image and the right is a cropped aerial image. Green lines indicate the matches detected by SIFT; almost all of them are wrong.

Even though the scale difference has been eliminated by down-sampling the UAV image towards the aerial image and the aerial image has been cropped to the same region as the UAV image, no reliable set of correct matches could be found in the similar looking image pairs. This finding motivated us to analyze the reasons for the failure and to develop a new image matching strategy facilitating

a successful and robust matching of imagery with wide baselines and substantial geometrical and temporal changes. The obtained 2D matches are used for geo-registration of the UAV image with reference to the aerial image. The results demonstrate that our approach achieves decimeter-level co-registration accuracy and comparable absolute geo-registration accuracy as the reference image.

In summary, the main innovations of this paper cover the following aspects:

- An exhaustive analysis of limiting cases of SIFT-based image matching for UAV and aerial image pairs. The reasons for the matching failure are identified by investigating the influence of different SIFT and Affine-SIFT (ASIFT) parameters, image rotations and the ratio test.
- A novel feature matching pipeline constituted of a dense feature detection scheme, a one-to-many matching strategy and a global geometric verification scheme.
- A comprehensive analysis of the matching quality with ground-truth correspondences and a demonstration of various experiments for evaluating absolute and relative accuracies of generated photogrammetric 3D products.

The paper is organized as follows: Section 2 gives a review of related works; Section 3 introduces limiting cases for SIFT matching and outlines the key factors accounting for the failure of the matching. Section 4 proposes the novel feature matching method for a robust and reliable matching result for wide-baseline image pairs. In Section 5, various experiments are carried out to validate the accuracy of the proposed matching method. Besides a qualitative and quantitative analysis of the obtained matches of UAV and aerial images, 3D errors of triangulated matches from geo-registered UAV images are compared to 3D points from aerial imagery and to terrestrial measured ground control points (GCPs). Additionally, DSMs generated from geo-registered UAV images and from aerial images are compared, and a joint 3D point cloud is presented. Finally, Section 6 discusses the applicability and limitations of the proposed method, and Section 7 concludes the paper and describes further applications.

2. Related Work

The availability of georeferenced imagery is a prerequisite for many photogrammetric tasks, such as the generation of registered 3D point clouds, DSMs, orthorectification, mosaicking or 3D reconstructions of buildings. The key for precise georeferencing of the mentioned products lies in an accurate geo-registration of the captured images, which can be tackled in different ways. In the field of aerial photogrammetry, high-end GNSS/IMU localization sensors are used, which allow direct georeferencing of the images without the need for external GCPs or photogrammetric adjustments in a post-processing step. Many established systems in aerial photogrammetry have access to such accurate sensors and achieve centimeter-level registration accuracy. The relatively low-cost DLR 3K sensor system [8] presents a camera frame carried by either an airplane or helicopter and consists of three Canon EOS 1Ds Mark II cameras looking in nadir, forward and backward direction developed for real-time disaster monitoring. The synchronized image acquisition and localization information provided by the expensive and heavy GNSS/IMU system (4 kg in total) allows for direct georeferencing accuracies of 10 cm [9]. The Vexcel UltraCam [10] offers a high level optical sensor for high resolution aerial photogrammetry with more than 100 megapixels. Combined with the high-end UltraNav GNSS/IMU system [11], 5 cm accuracy for direct georeferencing can be achieved. Due to payload limitations, many commercial UAVs are usually equipped with lightweight sensors providing localization accuracies in the range of meters [12], which is not sufficient enough for photogrammetric applications using direct georeferencing. An investigation regarding the ability of direct georeferencing with UAV systems shows that the geolocalization accuracy of current UAV systems is still too low to perform direct applications of photogrammetry at very large scale [4].

For this reason, image-based methods are usually utilized to facilitate geo-registration of UAV imagery in centimeter-level accuracy. One way to augment geo-registration results is to deploy GCPs, which is even recommended for high-end devices due to the existence of systematic errors [13]. Nevertheless, the deployment of GCPs is often expensive, requires fieldwork operations and is

unpractical or even impossible for hazardous or inaccessible regions. Due to the growing accessibility of high resolution aerial and satellite imagery, image matching approaches present a promising alternative for geo-registration. Here, geo-registration of UAV imagery is done by matching UAV images with georeferenced databases, such as 3D models, aerial images, orthophotos or satellite images. An accurate geo-registration of UAV images depends on the accuracy and reliability of the image matching result. Although image matching is a long-standing problem and much research has been performed in this area, still many cases exist where established methods fail or perform poorly. The task of matching UAV and aerial images can be characterized by wide baselines, large differences in viewpoints and geometrical, as well as temporal changes. Among intensity-based and frequency-based matching methods, local feature-based matching methods perform best with regard to these matching conditions [14]. Among various feature-based matching algorithms, SIFT [7] stands out for its robust scale- and rotation-invariant property. Although many variants and alternatives have been developed, such as its approximation SURF [15] and the binary descriptor BRIEF [16], investigations demonstrate that SIFT is still more robust to viewpoint changes and common image disturbances than both BRIEF and SURF [17]. ORB [18], which is a combination of the FAST detector [19] and the BRIEF binary descriptor, is a good choice for real-time applications, but several evaluations state that it cannot reach the repeatability and discriminative properties of SIFT [20–23]. KAZE [24] is a new development and succeeds especially in the presence of deformable objects. As a variant of SIFT, a full affine invariant matching framework ASIFT [25] was proposed to handle big differences in viewpoints by simulating a series of transformed images to cover the whole affine space. In the case of matching images with large differences in viewpoints, ASIFT has more robust performance than SIFT, which was also confirmed in the evaluation presented in [26].

Apart from feature-based wide-baseline matching, other concepts also investigate different methods for geo-registration of UAV imagery. Intensity-based methods, like an on-board correlation-based method to register UAV images towards aerial images in case of GNSS outages [27] or deformable template matching with image edges and entropy as feature representation [28], usually do not perform well in the case of temporal and geometrical changes. More recent work also focuses on matching terrestrial and aerial images showing extremely large viewpoint changes. A new feature representation using a convolutional neural network (CNN) is learned for geolocalizing ground-level images with an aerial reference database [29]. However, manual interventions are needed to estimate the scale for ground-level queries, and the absolute orientation of the query image can hardly be estimated. Shan et al. [30] synthesizes aerial views from pre-aligned Google Street View images using depth maps and corresponding camera poses, which are then matched with aerial images using SIFT. A similar approach is presented by Majdik et al. [31], where UAV images are matched with geo-tagged street view images using ASIFT, achieving meter-level global accuracy. However, only low altitudes and oblique images facing building facades are considered for the geo-registration. Aicardi et al. [32] adopts an image-based approach for co-registering multi-temporal UAV image datasets; however, it only estimates the relative transformation between the epochs, while the absolute transformation of the epoch is not solved. Finally, Xu et al. [33] presents an fast and efficient way for UAV image mosaicking without the explicit computation of camera poses; however, the image mosaics are not geo-registered.

Although considerable attempts and progress have been made regarding this topic, many of them rely on intensity-based matching methods, which are proven to be unstable in the case of geometric or temporal changes. In this sense, robust image matching against large-scale and viewpoint differences is the key to solve the problem for which feature-based approaches are still the methods of choice. Some mentioned approaches focus on improving the matching result for extremely large viewpoint changes, but still do not reach the desired global georeferencing accuracy.

Our approach is based on previous works [34,35], which have been proven to work for complex matching scenarios with multi-scale images. Compared with the state-of-the-art works mentioned above, our method is an advancement in the following aspects:

- To handle the large differences in scale and rotation between image pairs, we use a novel feature matching approach, which can overcome the challenge and robustly deliver abundant matches.
- Our method works for data of different scales, e.g., aerial images, aerial orthophotos and satellite images.
- Our method achieves not only decimeter-level co-registration accuracy, but also comparable absolute accuracy as that of the reference image, which is georeferenced in the conventional photogrammetric way.

3. Matching Performance Evaluation Using SIFT Features

This section introduces different UAV and aerial image pairs and a comprehensive analysis of the matching performance using SIFT and ASIFT. Although one would expect that SIFT matching can successfully match the presented images, a robust and successful matching is not possible. In order to figure out why the popular SIFT matching method surprisingly fails, we analyze the influence of different SIFT parameters, such as octaves and levels, the ratio test, but also image rotations. Experimental results demonstrate that the rotation invariance of SIFT is not as good as it has been considered to be, and the deficiency in the rotation estimation of SIFT leads to non-optimal matching results. In addition to that, many correct matches are either not nearest neighbors in the feature space or are rejected after applying the ratio test.

3.1. SIFT

Among the state-of-the-art matching algorithms, SIFT has been proven to be scale and rotation invariant and to outperform other local descriptors in various evaluations [20–23]. Besides, the ratio test proposed by Lowe [7] is widely applied to discard mismatches. In view of the substantial differences in scale and rotation of the UAV image and the aerial image, it makes sense to implement the SIFT matching algorithm (we use the OpenCV 3.0 implementation). This matching method is noted as "standard SIFT" in the following text.

The ratio test discards mismatches by rejecting all potential matches with similar descriptors. It works well in most cases; however, applying the ratio test in feature-based matching methods for images with repetitive structures often causes problems with similar descriptors. In this case, the distance ratio can be so high that these features would probably be defined as outliers. This can be critical especially when only a few correspondences remain after matching. To investigate how many correct matches are actually discarded by the ratio test, we implemented SIFT matching and counted the correct matches before and after the ratio test. Particularly, the distances of the first two nearest neighbors are computed and compared with the threshold. Considering that 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.

It needs to be pointed out that only a manually-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 correct matches for any dataset. This simplification of the matching problem is not feasible in practice and is only used for this analysis. The proposed method is able to match the original uncropped image pairs, as will be discussed in Section 5.

To ensure the best matching result using the SIFT detector and descriptor, we comprehensively tested different parameters. Specifically, we analyzed the effect of different ratio test thresholds and different parameters of the SIFT detection, like the number of octaves and levels per octave. Other parameters were kept constant as they have only a minor effect on the matching result. Concretely, we set the contrast threshold to 0.04, the edge threshold to 10 and the sigma of the Gaussian to 1.6. An extensive analysis was carried out for all of the datasets in Figure 2, while only

the results of the `Container` dataset is depicted. Nevertheless, we found similar results for all of our image pairs.

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

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

Figure 2. Datasets used in this paper: each column represents one (pre-processed) aerial reference image and two UAV target images. The UAV image in (**d**) should be matched to the aerial image (top right) and to a cropped part of a Google Maps image (**h**). (**a**) `Container`; (**b**) `Urban1`; (**c**) `Pool1`; (**d**) `Building`; (**e**) `Highway`; (**f**) `Urban2`; (**g**) `Pool2`; (**h**) `Googlemaps`.

In the first step of our analysis, we study the effect of different numbers of octaves and levels in the SIFT detection step, while fixing the ratio test threshold to a commonly-used value of 0.75. The number of octaves is related to different image samplings, while the number of levels represents the number of scale spaces per octave and is therefore related to the amount of image blurring. Table 2 lists the number of feasible correct matches from the set of remaining matches after applying the ratio test for different values of octaves and levels. Due to the low image sizes of the downsampled UAV image (664×885 pix) and cropped aerial image (971×665 pix), the number of keypoint detections saturates after two octaves. While increasing the number of levels per octave results in more matches surviving the ratio test, the number of inliers stays constant at a very low number of around 20 matches.

According to this experimental result, we analyze different thresholds of the ratio test in the next step while limiting the SIFT detector to three octaves and five levels. Like in the analysis above, we again count the number of remaining matches after the ratio test and the number of inliers among them, as illustrated in Figure 3a. A maximum number of around 100 correct matches can be found when only the first nearest neighbor is considered (equivalent to a threshold of one). Comparing this number to the total number of around 4000 matches, this is a very low ratio of inliers as can also be seen in Figure 3b. Increasing the impact of the ratio test (equivalent to lower values of the threshold), many correct matches are rejected due to a high similarity to other keypoint descriptors, while the ratio of outliers is decreasing at the same time.

According to the results in Figure 3b, the best ratio of inliers is suggested for threshold values between 0.3 and 0.5, but the absolute numbers of correct matches for these values is below ten and therefore not a reliable matching result.

Table 2. Analysis of SIFT performance with different octaves and levels for the `Container` dataset. Cells contain the number of correct matches (first number) from the set of remaining matches (second number) after applying the ratio test with a fixed threshold of 0.75. Due to the scale adaption of the UAV image, the number of keypoint detections saturates after two octaves. By increasing the levels, more keypoints can be detected, but the ratio of inliers decreases.

		Levels							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Octaves	**1**	12/50	15/61	15/64	17/74	17/84	11/91	17/78	14/91
	2	13/61	17/71	12/89	20/103	25/124	16/134	26/137	21/148
	3	13/63	17/76	13/93	22/108	26/131	17/142	27/148	22/153
	4	13/62	17/77	13/94	22/109	26/134	17/146	27/155	22/158
	5	13/62	17/77	13/93	22/110	26/136	17/148	27/157	22/159

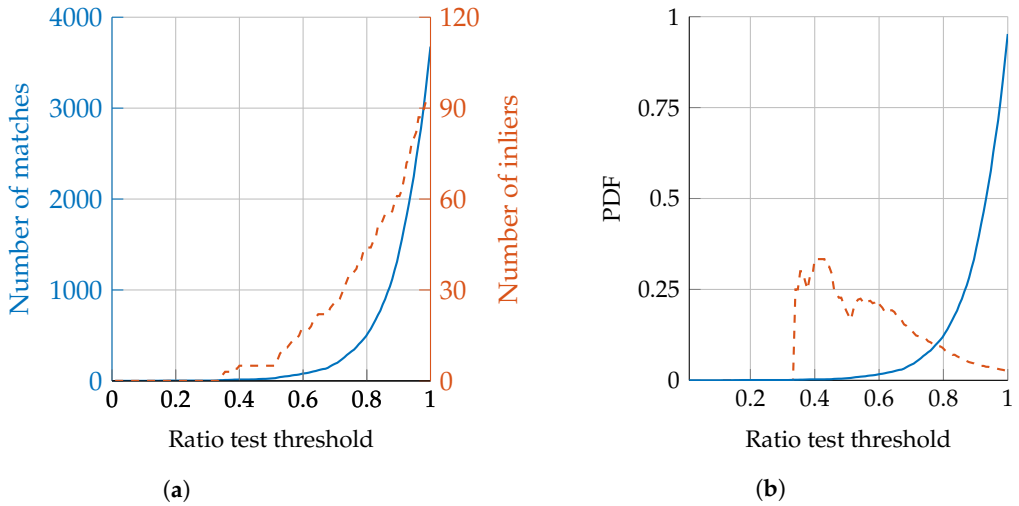


Figure 3. Influence of different ratio test thresholds for the `Container` dataset. (**a**) Number of remaining matches after applying the ratio test (solid) and the number of correct matches among them (dashed); (**b**) ratio of correct (dashed) and incorrect (solid) matches.

For our further analysis, we choose a ratio test threshold of 0.75, which is a good trade-off between rejecting most of the wrong matches and keeping a relatively high ratio of inliers.

Experimental results for the other datasets with these parameters are listed in Table 3, which confirmed the difficulty of matching this kind of image pair. Particularly in automatic registration systems for online geolocalization, it is crucial that the system is able to decide whether an image pair could be registered successfully or not. A high and reliable number of matches between 500 and 1000 is therefore indispensable for a trustable decision, compared to a rather low number below 50, as in our experiments, which could also satisfy random geometric transformations by chance.

Table 3. Analysis of standard SIFT matching on the proposed datasets in Figure 2. Matching was performed on downsampled UAV images and cropped aerial images on the same image content of the UAV image. Keypoint detection was limited to 3 octaves and 5 levels, and the ratio test threshold was set to 0.75. Results show the number of feature points detected by the SIFT-detector, correct matches considering only the first nearest neighbor, after applying the ratio test and possible matches according to 100 nearest neighbors.

Scenario	Image Fragment Size (pix)		Keypoints		Correct Matches		
	Aerial	UAV	Aerial	UAV	Nearest	Ratio Test	Nearest 100
`Container`	971 × 665	664 × 885	3763	3682	81	27	690
`Highway`	617 × 908	571 × 762	2768	2560	46	22	521
`Urban1`	1197 × 1643	871 × 1307	10,335	6266	47	27	304
`Urban2`	1199 × 1603	871 × 1307	9642	5757	293	176	1031
`Pool1`	838 × 1075	804 × 1071	5096	4202	87	47	451
`Pool2`	976 × 1074	799 × 1065	5788	4047	152	103	675
`Building`	1100 × 830	687 × 1030	4072	3270	76	39	498
`Googlemaps`	630 × 944	924 × 1668	3411	5963	45	21	565

However, the number of correct matches (using the same feature points and descriptors) can be significantly increased, if multiple nearest neighbors in feature space are considered as matching candidates. Figure 4 shows the cumulative number of correct matches for the first 100 nearest neighbors for the `Container` dataset. The last column of Table 3 lists the number of possible matches for the other datasets. This significant increase of correct matches for all datasets indicates that many corresponding keypoints in an image pair are not described perfectly by the SIFT descriptor, but can still be found among the first nearest neighbors in the feature space.

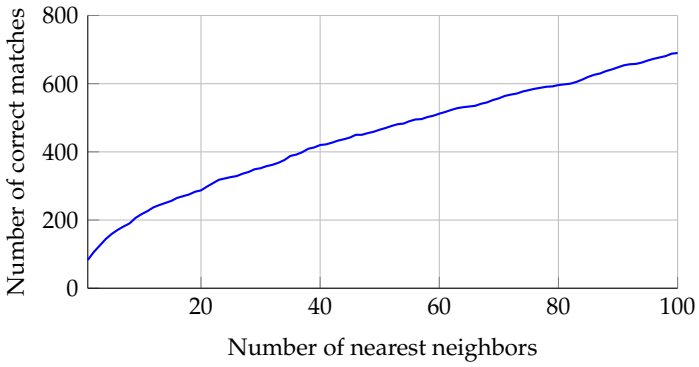


Figure 4. Cumulative number of possible correct matches considering multiple nearest neighbors in the feature matching for the `Container` dataset.

3.2. *Influence of Rotation*

As shown in the matching results above, SIFT has unsatisfactory performance for matching UAV and aerial images. 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 invariance property of SIFT needs to be reconsidered and evaluated. To investigate the problem, a series of experiments was carried out to test the influence of rotation. As listed in Table 4, we compare the standard SIFT matching on the original unaligned images (denoted by 'Std. SIFT') from Table 3 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 of the detected key points in the aligned images manually to be a fixed value; here, it was 0° for aligned images (denoted by 'Fixed-orientation'). The matching result was

represented by the number of putative correspondences after ratio test (denoted by 'Matches') and the correct matches among them (denoted by 'Inliers'). It is worth noting that the performance of matching between rotation-aligned images using standard SIFT does not improve; 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.

Table 4. Analysis of the influence of image rotation on matching performance. Inliers and matches for downsampled UAV images and cropped aerial images, rotation-aligned UAV images and rotation-aligned UAV images with fixed orientation in the SIFT-detector. Std., standard.

Scenario	Inliers/Matches		
	Std. SIFT	**Std. SIFT Rotation Aligned**	**SIFT Rotation Aligned Fixed-Orientation**
`Container`	27/320	22/349	30/306
`Highway`	22/204	26/263	52/277
`Urban1`	27/471	17/496	43/478
`Urban2`	103/635	179/677	267/734
`Pool1`	47/391	65/446	92/404
`Pool2`	103/635	179/677	267/734
`Building`	39/349	27/381	51/396
`Googlemaps`	21/535	21/509	35/394

For further investigation into the influence of rotation, we also made a comparison with the ASIFT method, as Table 5 shows. First, we compared the fixed-orientation SIFT with standard ASIFT on aligned images. As we achieved fewer correct matches for a tilt value of four at even higher computation cost, we, inspired by this finding, also fixed the orientation in ASIFT (denoted by 'Fixed-orientation') in the same way, and the matching performance improved significantly. Comparing the results in Column 2 and Column 4, it can be seen that when the orientation is fixed, SIFT results in almost equivalent inliers as ASIFT; however, for a robust matching, the number of inliers is still far from enough.

Table 5. Comparison with ASIFT. Inliers and matches for pre-aligned images using standard SIFT with fixed orientation, ASIFT and pre-aligned images on ASIFT with fixed orientation.

Scenario	Inliers/Matches		
	SIFT Rotation Aligned Fixed-Orientation	**Std. ASIFT**	**ASIFT Rotation Aligned Fixed-Orientation**
`Container`	30/306	25/281	46/283
`Highway`	52/227	56/249	70/237
`Urban1`	43/478	46/512	61/508
`Urban2`	267/734	254/1069	281/994
`Pool1`	92/404	73/346	109/404
`Pool2`	267/734	255/600	375/620
`Building`	51/396	45/382	78/424
`Googlemaps`	35/394	42/330	47/430

Based on the above findings, we summarize that the challenges of matching UAV imagery and airborne imagery stem mainly from the following aspects: inadequate matching candidates, ambiguous keypoint orientations and misuse of the ratio test. To be more specific:

- The rotation invariance of SIFT does not work well when the images have large differences in scales and viewpoints. In standard SIFT, the dominant orientation is detected automatically.

Instead, if we fix the orientations of SIFT keypoints, the number of correct matches increases significantly.

- When the image has repeated patterns, the local descriptors of the repeated structure can be so similar that the distance ratio between the nearest and second nearest neighbor is no more distinctive. As an important step in the standard matching pipeline, the ratio test actually discards many correct matches, and the remaining correspondences are not reliable. In contrast, considering multiple nearest neighbors as matching hypotheses can help to increase the matching performance enormously.

4. Proposed Image Matching Method

According to the reasons for the matching failure presented in Section 3, the new matching approach is designed to eliminate each of the exposed bottlenecks. A new feature detection scheme increases the number of matchable keypoints, which is necessary for a reliable matching result. To avoid loosing many correct matches that are not nearest neighbors in feature space or that are rejected by the ratio test, we introduce a one-to-many matching scheme. To extract correct matches among them, a direct method using histogram voting is performed instead of the commonly-used RANSAC scheme. An extension of this method can also handle unknown image rotations. In the end, the detected matches are used to estimate camera poses of the UAV images in the coordinate system of the reference images.

4.1. Prerequisites

The proposed method assumes that the scale difference between both images can be estimated and mostly eliminated in advance. This requirement can be generally fulfilled, as accurate positional information of aerial images is always available, and UAV images are tagged with both GNSS and barometric altitude information. One of both sensors should deliver reliable data in any case.

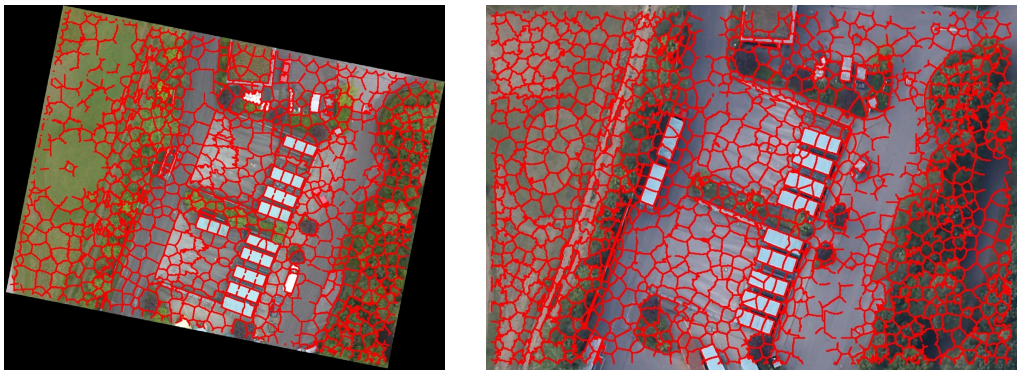
Secondly, a pre-alignment with respect to the image rotation can be achieved using the on-board compass of the UAV. The next sections assume that a rough pre-alignment of the image pairs is feasible, but in the case of no or only imprecise image heading information, Section 4.5 presents an extension of the proposed method that allows recovering an unknown image rotation.

4.2. Dense Feature Extraction

The essential prerequisites of robust matching are sufficient and uniformly distributed features whose density should reflect the information content of the image. According to the results in Tables 4 and 5, established keypoint detectors, such as in SIFT, do not always find a sufficient number of matchable features. To ensure a large number of inliers, a dense detection scheme is desired, but instead of using all pixels as potential feature points, only keypoints should be considered that are located along strong image gradients. This does not only reduce computational time, but also rejects hardly matchable feature points at homogeneous areas with weak descriptors.

In view of the fact that image segmentation using SLIC (simple linear iterative clustering) [36] can efficiently generate compact and highly uniform superpixels, whose boundaries mostly define strong variations in the intensities of the local neighborhood, like edges and corners, we therefore adopt all of the pixels at the boundaries of superpixels as feature points. In practice, the number of desired superpixels can be specified according to the need for feature density and compactness. Since the relative scale difference of both images is known beforehand, the number and compactness of superpixels in both images are similar and therefore ensure the extraction of identical object boundaries. Figure 5 highlights the feature points of a UAV and aerial image, namely all of the pixels at the boundaries of superpixels, after removing those feature points located at homogeneous areas.

Afterwards, a SIFT-descriptor for each detected feature point is computed. Since the UAV image is already aligned with the reference image, the scale space and feature orientation of SIFT-descriptors should be identically assigned for both images.



(**a**) UAV image (**b**) Aerial Image

Figure 5. Feature points highlighted in red, namely all of the pixels at the boundaries of superpixels, after removing those feature points located at homogeneous areas for (**a**) the pre-aligned UAV image and (**b**) the aerial image of the `Container` dataset with 1000 simple linear iterative clustering (SLIC) superpixels.

4.3. One-To-Many Feature Matching

In this phase, a feature descriptor in one image is matched with all other features in the other image using the euclidean distance calculation. In standard SIFT, only the first and second nearest neighbors are taken into account, so that many correct matches are actually discarded as presented in Table 3. An example of ambiguous feature matching is demonstrated in Figure 6. The correct feature point (left) would mainly be discarded for two reasons: first, the correct match may not be the first nearest neighbor in feature space; second, it may not pass the ratio test due to the high similarity of the local descriptors.

(**a**) (**b**)

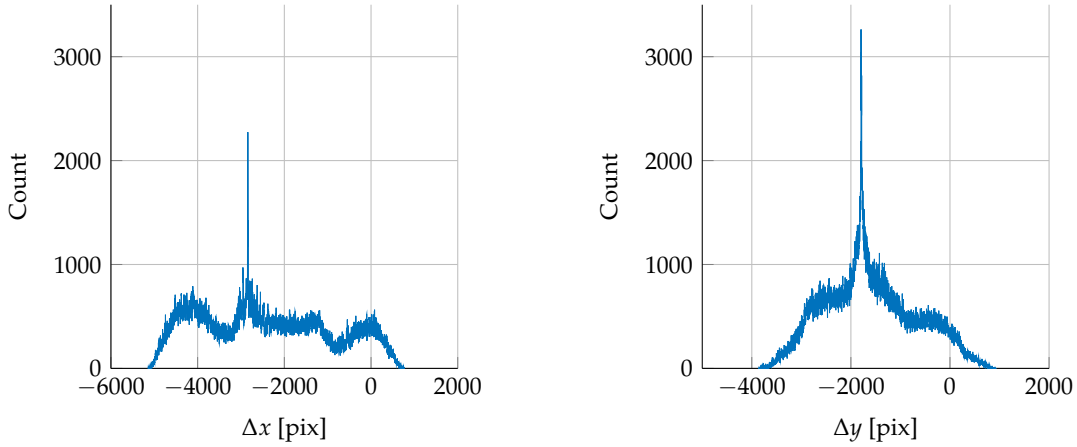
Figure 6. Challenge of ambiguous feature matching. One feature point at a corner of a container in the UAV image (**a**) corresponds to many feature points in the aerial image with similar descriptors (**b**). The correct match often can be found among a set of multiple nearest neighbors. These ambiguities need to be solved in order to extract the correct match.

To solve this problem, we propose a one-to-many matching scheme by taking the k-nearest neighbors as matching candidates to ensure that correct matches can be even found for corresponding keypoints that do not show nearest descriptors distances. Besides, the approximate nearest neighbor method (ANN) is applied to avoid the exhaustive search and to speed up the matching process. Although the idea of using a one-to-many matching scheme is not new, the next section proposes a new approach for how to extract the correct matches among them.

4.4. Geometric Match Verification with Histogram Voting

It is pointed out in Section 3 that the commonly-used ratio test in SIFT does not effectively determine whether a feature point is a correct match. As a substitute, we use pixel-distances as a global geometric constraint to verify the matching hypotheses. The superpixel-based feature point extraction and one-to-many matching strategy result in a plethora of putative matches, which ensures a sufficient number of correct matches, but also inevitably contains a massive number of mismatches. Postulating that the UAV and reference image both contain the same planar scene and the differences in their scales and rotations have already been eliminated, the transformation between the two aligned images can be simply approximated as a 2D-translation. Particularly, for each keypoint i, whose image coordinates are (x_u^i, y_u^i) in the UAV image and (x_r^i, y_r^i) in the reference image, and for each of its k matching hypotheses j ($j = 1 : k$), whose image coordinates are (x_r^j, y_r^j) in the reference image, we calculate their coordinate differences $\Delta x^{i,j}$ and $\Delta y^{i,j}$ by $\Delta x^{i,j} = x_u^i - x_r^{i,j}$ and $\Delta y^{i,j} = y_u^i - y_r^{i,j}$. Correct matches are expected to satisfy the conditions $|T_x - \Delta x^{i,j}| \leq R \wedge |T_y - \Delta y^{i,j}| \leq R$, where R is a threshold related to the scene depth and T_x and T_y are the parameters of the unknown 2D translation. We can recover this translation by a simple histogram voting scheme. After computing $\Delta x^{i,j}$ and $\Delta y^{i,j}$ for all putative matches, distinctive peaks T_x and T_y in the both histograms are extracted.

Figure 7 presents an example for this histogram voting regarding the `Container` scenario. While distances of wrong matches are randomly distributed, those of geometrically correct matches concentrate on or aggregate around a common value (T_x, T_y), thus shaping a distinct peak in the histogram. To allow for minor changes of image scene depth, we determine the matches located at close range to (T_x, T_y) as possibly correct matches; the distance threshold is denoted by R. The value of R is related to the change of scene depth, as well as the accuracy of pre-alignment. A larger threshold R can compensate for these impacts and result in more matches; on the other hand, more outliers would also be introduced into the raw matches.



(**a**) Row direction (**b**) Column direction

Figure 7. Geometric match verification of the `Container` scenario with histogram voting. Distribution of pixel distances for all putative matches according to the one-to-many matching in the (**a**) row and (**b**) column direction. Distinct peaks represent unknown 2D-translation.

4.5. Eliminating Differences in Image Rotation

The scale difference between the UAV image and the reference image can be derived using either the on-board GNSS information or the barometric altitude sensor. In contrast, precise orientation-adaption fails for many UAVs due to inaccurate heading information provided by the low quality IMUs. Our assumption that correct matches follow a simple 2D translation fails in the case of unaligned images. However, we can estimate the unknown image rotation by adapting the

proposed matching approach with a rotation search scheme. Although Section 3.2 shows that fixing the orientation of the feature points in the SIFT descriptors results in a better matching performance if the images are pre-aligned, a sufficient number of correct matches can still be found for unaligned images with the keypoint orientation estimation of SIFT when using a denser feature detection like the one presented in Section 4.2.

After generating a set of putative one-to-many matches for unaligned images, the unknown image rotation is obtained by first dividing the rotation ψ equally into discrete rotation values $\psi^a = [-180, 180]$ deg. For each rotation ψ^a, the feature points of the UAV images $pt_u^i = (x_u^i, y_u^i, 1)^T$ are rotated around the image center $pt_{u,rot}^{i,a} = M(p, \psi^a) \cdot pt_u^i$ with a transformation matrix $M(p, \psi^a) = [T(p)R(\psi^a)T(-p)]$, where $T(p)$ is a translation matrix with the coordinates of the image center p and R a rotation matrix with rotation angle ψ^a. Pixel distances are calculated according to $\Delta x_{rot}^{i,j,a} = x_{u,rot}^{i,a} - x_r^{i,j}$ and $\Delta y_{rot}^{i,j,a} = y_{u,rot}^{i,a} - y_r^{i,j}$, and histogram voting from Section 4.4 is performed for each rotation. The maximum number of raw matches satisfying the threshold T_x^a and T_y^a is kept for all rotation values ψ^a. Figure 8 shows the number of raw matches for different image rotations according to the `Container` dataset. The distinct peak at -104 deg represents the unknown image rotation.

This method may be used for a full 360 deg search; however, the search range can be reduced in the case of available inaccurate rotations from the on-board IMU. After recovering the unknown image rotation, further matches can be determined with fixed orientations according to the previous sections.

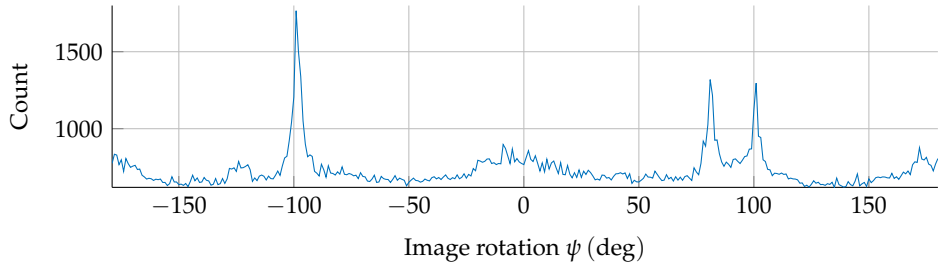
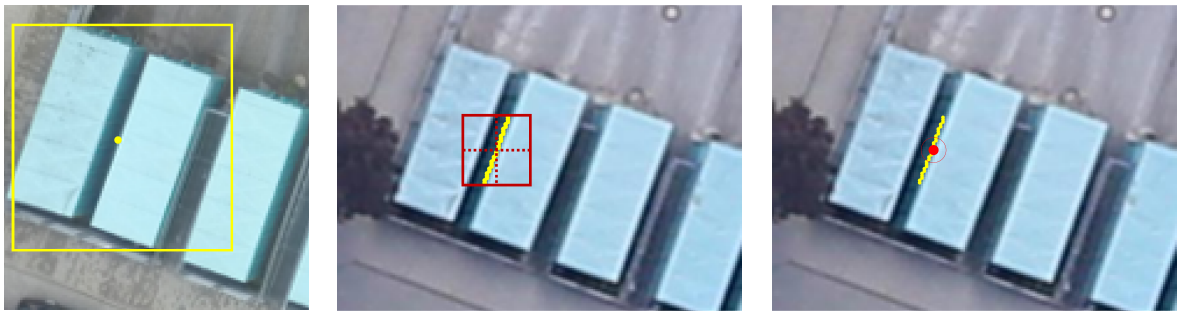


Figure 8. Recovering the unknown image rotation in the case of unavailable or inaccurate UAV IMU data. Extending the proposed method by transforming UAV feature points with multiple rotation values before the histogram voting step. The figure shows the rotation histogram for the `Container` dataset. The maximum number of raw matches represents unknown image rotation.

4.6. Match Refinement

After the geometric verification of the one-to-many matches, it is likely for some keypoints that they share multiple adjacent feature points in the other image as geometric correct matches. This is caused by the dense feature point extraction, which generates dense feature points especially along strong image edges. The distance threshold R allows multiple geometric correct matches for adjacent feature points for which the distance to T_x and T_y is below R. Figure 9 illustrates these local ambiguities of the feature matches. One feature point in the UAV image in Figure 9a corresponds to multiple geometrical correct matches in the aerial image in Figure 9b. Even a successive RANSAC filtering step according to geometrical transformations will not truly solve these ambiguities, if Sampson distances or transfer errors of neighboring matches are below the filtering threshold. In order to ensure geometrically-correct and unique one-to-one matches, a refinement step is applied for all geometrically-correct matches by eliminating the ambiguities and optimizing the location of the feature points. The superpixel segmentation cannot guarantee exact locations of corresponding pixels in both images. The refinement consists of a normalized cross correlation (NCC) of a template in the local neighborhood of the UAV feature point (yellow rectangle in Figure 9a). For all corresponding matching hypotheses (yellow dots in Figure 9b), the corresponding patch is searched in a local search window around the feature points (red rectangle in Figure 9b). The size of the search window for all aerial feature points can be set to the threshold R of the geometric verification. The NCC

optimizes all matching hypotheses to the correct location, illustrated by the red dot in Figure 9c. This method eliminates duplicate matches and refines feature point locations for inaccurate keypoints in a local neighborhood of the initial keypoints. These raw matches can now be used to estimate the fundamental matrix or homography in combination with RANSAC methods and to reject remaining outliers satisfying the geometric constraint. After computing the fundamental matrix, a guided matching method, as presented in Section 3, can be applied to find more matches if the threshold was chosen too small.



(**a**) (**b**) (**c**)

Figure 9. Refinement and duplicate elimination of geometric correct matches. (**a**) One feature point in the UAV image (yellow dot) and its template size (rectangle); (**b**) corresponding geometric inliers (yellow dots) in the aerial image and size of the search window for one match (red rectangle); (**c**) all geometric inliers will share the same optimized pixel location after refinement (red dot).

4.7. Geo-Registration of UAV Images

As the UAV image and the reference image have overlapping areas, one 3D point in the object space could be visible both in the reference image and the UAV image. Such 3D points can be used as reference 3D points for geo-registration of UAV images. The prerequisite of the geo-registration is the available georeferenced aerial image together with its height map or one orthorectified mosaic with a high resolution DSM.

- Match a UAV image U with the reference image R using the proposed matching method. Assume a feature point (x_r, y_r) in the reference image is matched to feature point (x_u, y_u) in the UAV images, this matching pair corresponds to a 3D point $P(X, Y, Z)$ in the object space.
- If image R is an individual georeferenced aerial or satellite image, we assume its height map is available, which can be generated in the process of dense matching with neighboring images [37]. The height Z can be looked up in the height map, and the planar coordinates X and Y can be calculated using the orientation parameters of R. If image R is an aerial orthophoto that is generated by an orthographic projection of the aerial image mosaic onto a high resolution DSM, the planar coordinates (X, Y) are namely the corresponding georeferenced coordinates of the pixel (x_r, y_r) in the orthophoto, and Z is namely the corresponding height at (X, Y) of the DSM.
- As the proposed matching method generates thousands of matches and each match results in a 3D point, those points can be used as reference 3D points to transform the UAV image to the same global coordinate system of the reference image. If there are UAV image sequences, a bundle adjustment can be performed to improve the global geo-registration accuracy.

5. Experiments

In order to verify the robustness and reliability of the proposed matching method, we compare the performance of our method with standard SIFT on different datasets. Furthermore, the generated matches are used for geo-registration and 3D reconstruction of the UAV images. Qualitative and

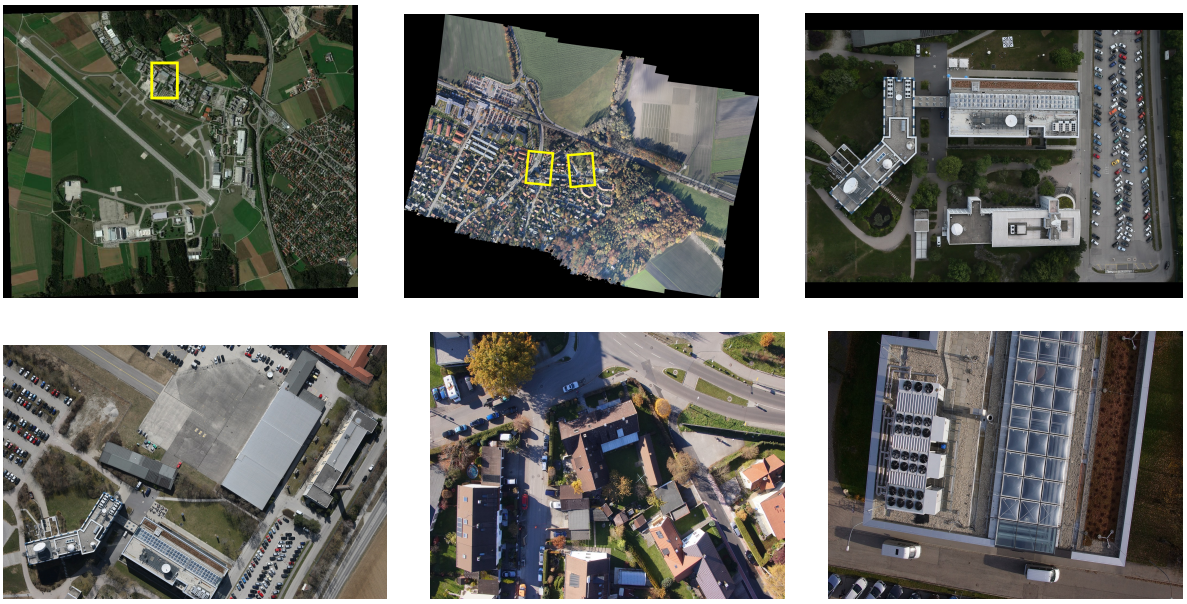
quantitative analyses are presented to validate the accuracy of geo-registration, and on this basis, photogrammetric 3D products, such as orthophotos, DSMs and merged points clouds are discussed.

5.1. Data Acquisition

Experiments were carried out based on offline flight data of four datasets: `Eichenau`, `Germering`, `EOC`(Earth Observation Center) and `WV2` (WorldView-2). It is worth noting that for datasets `Eichenau`, `Germering` and `EOC`, which contain 72, 58 and 11 UAV images, respectively, the whole of the UAV sequences were matched in an automatic manner. Showing the results for all image pairs is beyond the scope of this paper, so we focused on the same image pairs which were already introduced in Section 3. The `Eichenau` dataset contains two scenarios: `Urban1` and `Urban2`. The UAV images were acquired with a Sony Nex-7 camera simultaneously with the reference aerial images on 2 November 2015. For both scenarios, we matched UAV images not only to aerial images, but also to aerial orthophotos, which are generated by an orthographic projection onto a high resolution DSM [37,38]. The `Germering` dataset is comprised of four different scenarios: `Container`, `Highway`, `Pool1` and `Pool2`. The reference aerial images of this dataset were captured on 17 June 2014, whereas the UAV images were captured with a slight time delay on 11 July 2014 with a GoPro Hero 3+ Black camera. The aerial images in the `EOC` dataset were acquired on 16 June 2014, and the UAV images were captured on 12 November 2014 with a Sony Nex-7 camera. In the `EOC` dataset, all aerial images are almost nadir, whereas the UAV images have both nadir views of the building roof and oblique views of the building facades. Only the nadir-view UAV images are matched with the aerial images, and the generated 3D points are used to geo-register the whole of the UAV image block, including both nadir and oblique images. In addition, the nadir UAV images are also matched with a screenshot of Google Maps. In the `WV2` dataset [39], we match an aerial image from the `EOC` dataset with a WorldView-2 RGB satellite image of the year 2010 to validate the generalization ability of the proposed method and its robustness against large temporal changes. Besides, the datasets `Eichenau`, `Germering` and `EOC` are not significantly affected by temporal changes, as the vegetation periods are the same (except for `EOC`) and the appearances of buildings have not changed. All of the aerial images were captured by a Canon EOS-1DX camera mounted on the DLR 4K sensor system [9], which consists of two cameras with a 15° sidewards looking angle and an FOV of 75° across. In data pre-processing, an orthographic projection of the aerial imagery was performed to generate nadir-view images. Figures 2 and 10 illustrate all datasets used in the experiments, where the first row shows the reference images (pre-processed nadir-view aerial images and satellite image) and the other two rows are the corresponding target images (UAV and aerial images) to be matched. Detailed characteristics of the datasets are listed in Table 6.

Table 6. Characteristics of the datasets used in the experiment. Target images are pre-aligned towards the reference image using GNSS/IMU data. AI: aerial imagery; AO: aerial orthophoto; SI: satellite imagery; UI: UAV imagery.

Dataset	Reference Image				Target Image			
	Type/Date	Resolution (pix)	Height (m)	GSD (cm)	Type/Date	Resolution (pix)	Height (m)	GSD (cm)
`Eichenau`	AO 11/2015	9206 × 7357	600	20	UI 11/2015	573 × 794	100	1.8
`Germering`	AI 06/2014	5184 × 3902	700	9.4	UI 07/2014	823 × 996	100	2
`EOC`	AI 06/2014	5184 × 3902	340	4.6	UI 11/2014	1106 × 807	25–40	0.5–0.8
`WV2`	SI 2010	5292 × 6410	770,000	46	AI 2015	497 × 332	350	4.4



(**a**) WV2 (**b**) Eichenau (**c**) EOC

Figure 10. Additional datasets for the experiment. Top: reference images. Bottom: target images. Overlapping areas are highlighted by yellow rectangles in the reference images. (**a**) WV2; (**b**) Eichenau; (**c**) EOC.

5.2. Performance Test of Matching UAV Images with a Reference Image

In order to validate the robustness and accuracy of the proposed method, we use the same image pairs presented in Section 3, where the standard SIFT performed poorly in most of the cases. Different from the results in Table 3, the matching is now performed with original aerial images other than the cropped images. As can be seen in Figure 2, 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 searching areas.

All image pairs are provided with rough information of positions and orientations from GNSS and IMU, so that the images could be pre-aligned beforehand. Then, the target images and the reference images were matched with the proposed matching method and standard SIFT. Specifically, 750 superpixels were segmented from the UAV images; the threshold for the feature matching-distance was set to 0.2 as a trade-off between discarding apparent outliers and retaining enough matching hypotheses. Fifty nearest neighbors were selected as matching candidates for the one-to-many matching, and the distance threshold R for the geometric verification was set to 12 pixels. As for matching using SIFT, the threshold of ratio test was set to 0.75.

In order to evaluate the matching accuracy, we created ground-truths of feature point correspondences for each dataset using manually-selected and automatically-detected matching correspondences. The quantitative results using standard SIFT and our proposed method are summarized in Tables 7 and 8, where Error (homography (H)) denotes the mean transfer error (the euclidean distance between a point's true correspondence and the point mapped by the homography matrix H, which is estimated from matching correspondences) and Error (fundamental (F)) denotes the mean Sampson distance (the distance between a point to the corresponding epipolar line). Standard SIFT failed for almost all scenarios, while the proposed method found abundant matches with much smaller errors.

Regarding matching accuracy, standard SIFT outperformed our method only for the Pool1 scenario. As homography only considers transformation between two planes, those mismatches for areas with apparently different scene depths were discarded. The mean transfer errors were only 2–3 pixels in most cases, corresponding to a ground distance of about 20–30 cm.

Table 7. Results using standard SIFT: number of raw matches after applying SIFT for all scenarios. Inliers after estimating the fundamental matrix (F) and homography (H) using RANSAC. Mean errors (in pixel) according to ground-truths F and H.

Scenario	Raw Matches (SIFT)	Inliers F/Error (F)	Inliers H/Error (H)
Container	58	14/666.26	9/1767.55
Highway	49	15/1996.30	9/2210.20
Pool1	162	52/0.83	33/1.63
Pool2	107	18/618.54	10/1308.02
Eichenau1	287	45/19.11	48/3.63
Eichenau2	436	140/1.11	146/3.64
EOC	446	16/959.87	6/877.21
WV2	117	19/175.73	19/4.03
Building	553	16/595.06	11/317.59
Googlemaps	522	19/195.34	8/919.48

Table 8. Results using the proposed method: number of raw matches after applying our method for all scenarios. Inliers after estimating fundamental matrix (F) and homography (H) using RANSAC. Mean errors (in pixel) according to ground-truths F and H.

Scenario	Raw Matches (Our)	Inliers F/Error (F)	Inliers H/Error (H)
Container	8264	4876/2.59	2835/7.01
Highway	1979	1184/2.79	1230/1.20
Pool1	6593	3599/1.87	2188/1.87
Pool2	14,091	7555/2.01	4199/2.03
Eichenau1	4018	1850/4.35	1165/3.53
Eichenau2	5846	3204/1.09	3077/4.65
EOC	6834	3949/2.92	2586/3.18
WV2	15,131	6290/2.22	6760/3.57
Building	9113	3526/3.15	1932/2.36
Googlemaps	15,437	5120/3.42	3217/2.82

The matched feature points are marked in the UAV images (the second and third rows in Figure 11). As a result of the superpixel segmentation, most matches are located at regions with rich textures and have apparently much higher density than SIFT-features. The projection transformations can then be estimated using these matches. The first row in Figure 11 depicts the projected UAV images on the aerial images by the estimated homography.

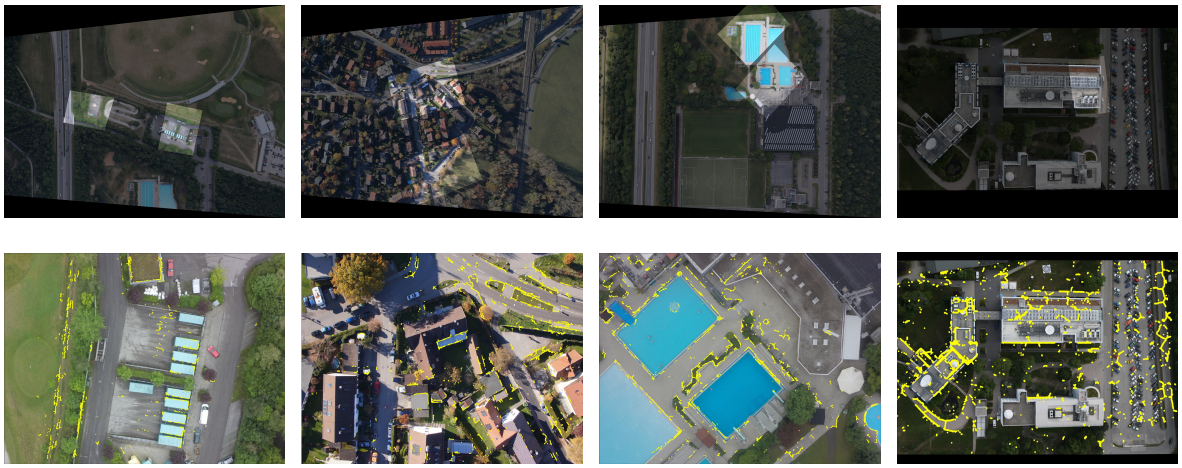


Figure 11. *Cont.*

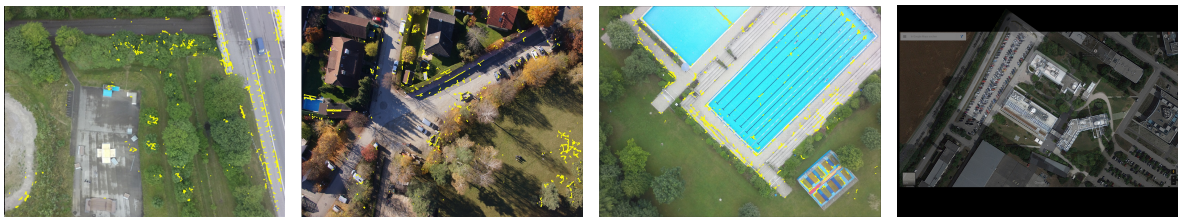


Figure 11. Qualitative results of the proposed matching method according to the image pairs in Figure 2. The first row shows the overlapped UAV and aerial image pairs after applying an estimated homography calculated from our matches (also for the figure on the bottom right). The second and third row show the distribution of the geometrically-correct matches in the UAV images (yellow dots).

5.3. Evaluation of the Geo-Registration of UAV Images

Following the proposed pipeline in Section 4.7, plenty of 3D reference points were computed and then used as GCPs in a bundle block adjustment to geo-register the UAV images to the global coordinate frame.

In order to verify the accuracy of the geo-registration of UAV images, several evenly-distributed ground check points were selected across the survey area, and their actual coordinates P_{rtk} were measured using an RTK GNSS receiver. Meanwhile, these ground check points were marked in all UAV images, and their theoretical 3D coordinates P_{uav} were computed by triangulating the geo-registered UAV images. The columns "$Error_{\text{rtk}}$" in Tables 9 and 10 list the errors $P_{\text{uav}} - P_{\text{rtk}}$ of `Eichenau` and `Germering` datasets, respectively. The height errors in "$Error_{\text{rtk}}$" are around 2 meters; this is mainly caused by the systematic errors of the global digital elevation model like SRTM [40], which was used as the height reference during the processing of the reference images.

In order to validate accuracy of co-registration, the coordinates triangulated by geo-registered UAV images, P_{uav}, were compared with the identical points on the reference image, as well. In the `Eichenau` dataset, the reference image was an aerial orthophoto (with a high resolution DSM), so the corresponding coordinates P_{ref} were manually looked up in the orthophoto and DSM, as explained in Section 4.7. In the `Germering` dataset, the reference image was an individual aerial image from a pre-georeferenced aerial images dataset, so the corresponding coordinates P_{ref} were triangulated using multiple pre-georeferenced aerial images from that dataset. The column "$Error_{ref}$" in Tables 9 and 10 lists the error $P_{\text{uav}} - P_{\text{ref}}$.

Afterwards, the orthophoto and DSM were reconstructed from the geo-registered UAV images using the software SURE [41]. Figure 12 illustrates the aerial orthophoto and the UAV orthophoto of `Eichenau` dataset. More specifically, Figure 12a depicts the aerial orthophoto of the `Eichenau` dataset, whose resolution is 20 cm; Figure 12b shows the UAV orthophoto of the `Eichenau` dataset, whose resolution is 2 cm. It is obvious that the UAV orthophoto has higher resolution and contains more details than the aerial orthophoto. Figure 12c displays the UAV orthophoto overlapping on the aerial orthophoto with 50% transparency; it can be seen that the two orthophotos are precisely aligned using the proposed geo-registration method. Figure 12d–g compare the appearance of corresponding objects on the aerial orthophoto and UAV orthophoto, demonstrating that the UAV orthophoto contains richer textures than the aerial orthophoto. Figure 13 illustrates the estimated camera poses, as well as the reconstructed point cloud of the geo-registered UAV image blocks and the aerial image blocks. Despite the considerable scale difference, our matching approach still succeeds in an accurate registration. Similarly, Figure 14 demonstrates the aerial orthophoto and UAV orthophoto of the `Germering` dataset, whose resolutions are 20 cm and 2 cm, respectively.

Table 9. Errors of the coordinates of check points compared to RTK GNSS measurements and the coordinates looked up in the aerial orthophoto and DSM: `Eichenau` dataset.

Check Point	$\text{Error}_{\text{ref}}$ (m)			$\text{Error}_{\text{rtk}}$ (m)		
	Δx	Δy	Δz	Δx	Δy	Δz
1	0.04	−0.51	−0.21	−0.04	−0.39	−1.74
2	−0.05	−0.07	−0.15	−0.11	−0.40	−1.90
3	0.04	−0.41	−0.36	−0.10	−0.83	−2.04
4	−0.14	0.80	0.70	−0.35	−0.33	−1.91
5	−0.04	0.49	−0.17	−0.05	−0.21	−1.81
6	−0.03	0.12	−0.10	0.12	−0.36	−1.63

Table 10. Errors of the coordinates of check points compared to RTK GNSS measurements and the coordinates triangulated using aerial images: `Germering` dataset.

Check Point	$\text{Error}_{\text{ref}}$ (m)			$\text{Error}_{\text{rtk}}$ (m)		
	Δx	Δy	Δz	Δx	Δy	Δz
1	−0.06	−0.14	−0.38	0.34	−0.01	1.49
2	0.16	−0.67	0.37	0.43	−0.54	1.68
3	0.14	−0.02	0.46	0.56	0.16	1.76
4	0.11	−0.76	0.26	0.44	−0.76	1.71
5	0.19	−0.10	0.50	0.55	−0.06	0.75
6	−0.05	0.18	0.18	0.39	0.36	1.30
7	−0.08	0.41	−0.06	0.41	0.50	1.42

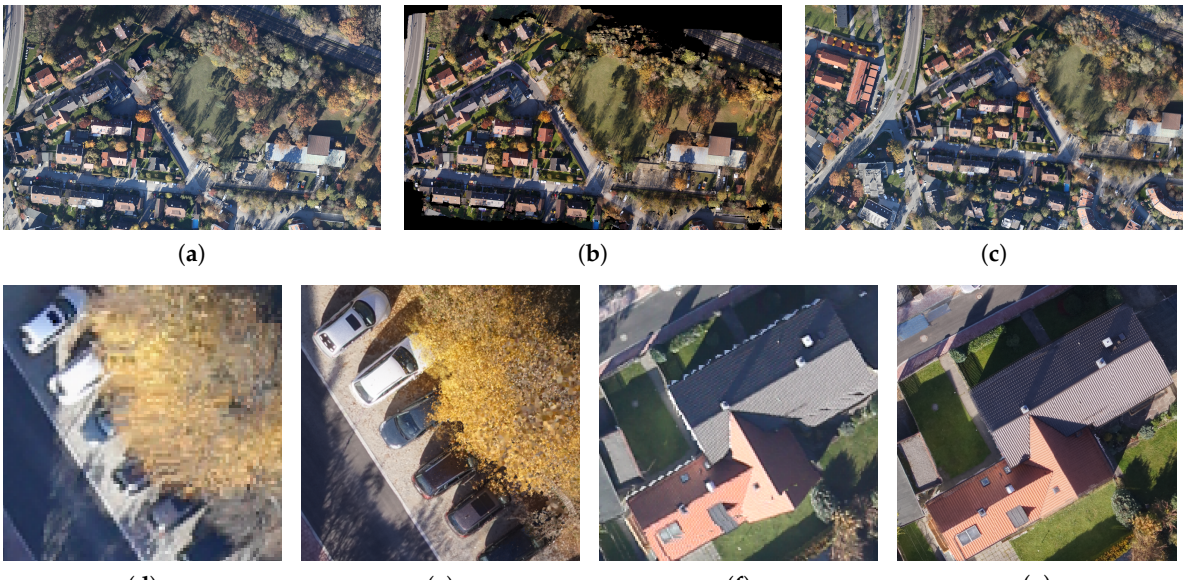


Figure 12. Comparison of (**a**) the aerial orthophoto with 20 cm GSD and (**b**) the UAV orthophoto with 2 cm GSD of the `Eichenau` dataset; (**c**) 50% transparent overlap of both orthophotos; (**d**,**e**) compare cars and (**f**,**g**) show a roof on the aerial and UAV orthophoto, respectively.

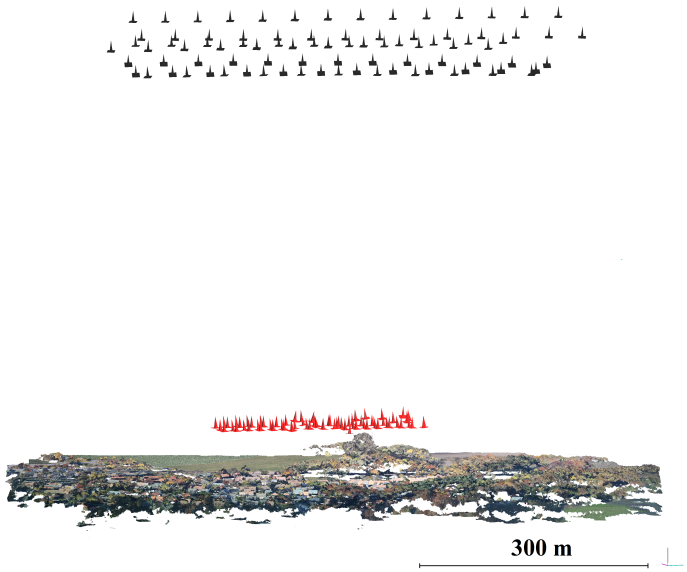


Figure 13. Camera pose visualization for the `Eichenau` dataset, showing camera poses of the geo-registered UAV image block at a 100-m altitude (red) and the aerial image block at a 600-m altitude (black).

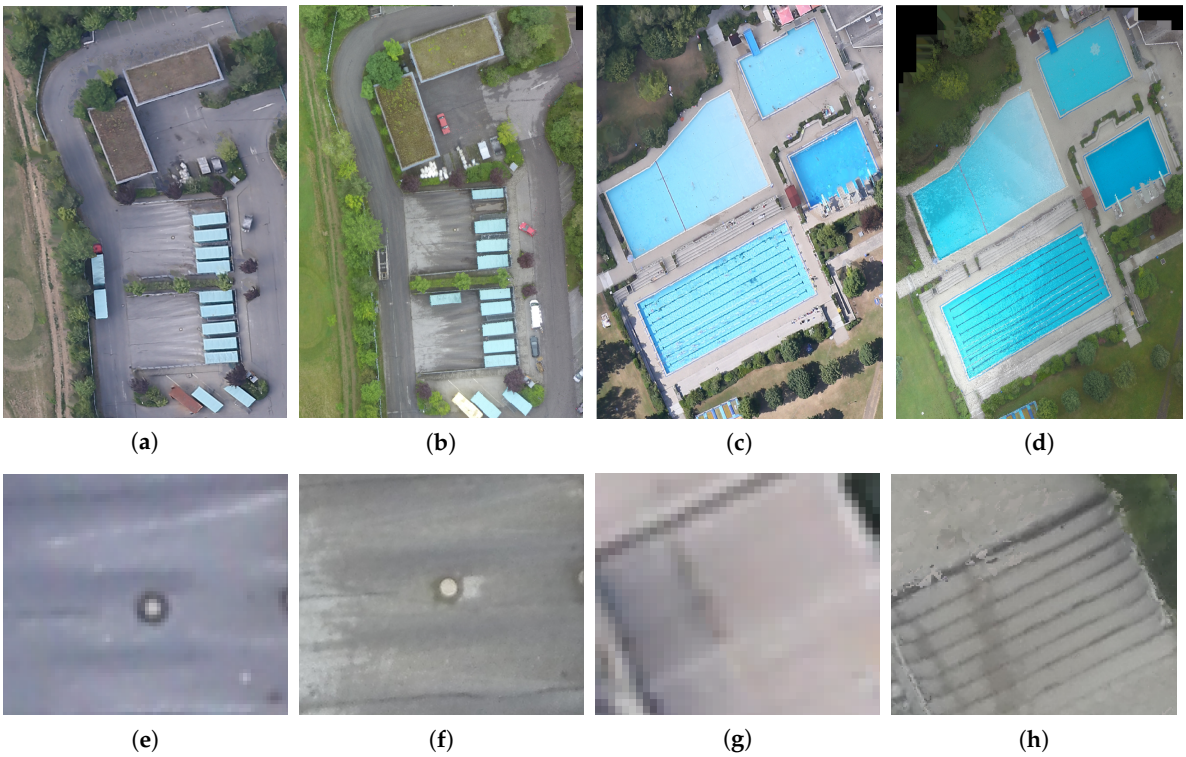
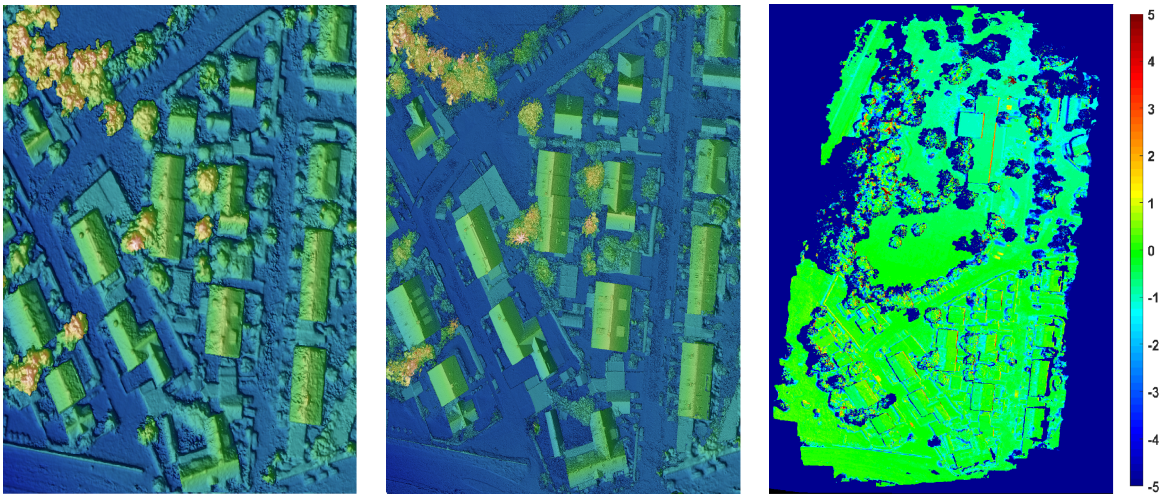


Figure 14. Comparison of (**a**,**c**) aerial orthophotos with 20-cm GSD and (**b**,**d**) UAV orthophotos with 2-cm GSD of the `Germering` dataset; (**e**,**f**) compare a manhole and (**g**,**h**) staircases on the aerial and UAV orthophoto, respectively.

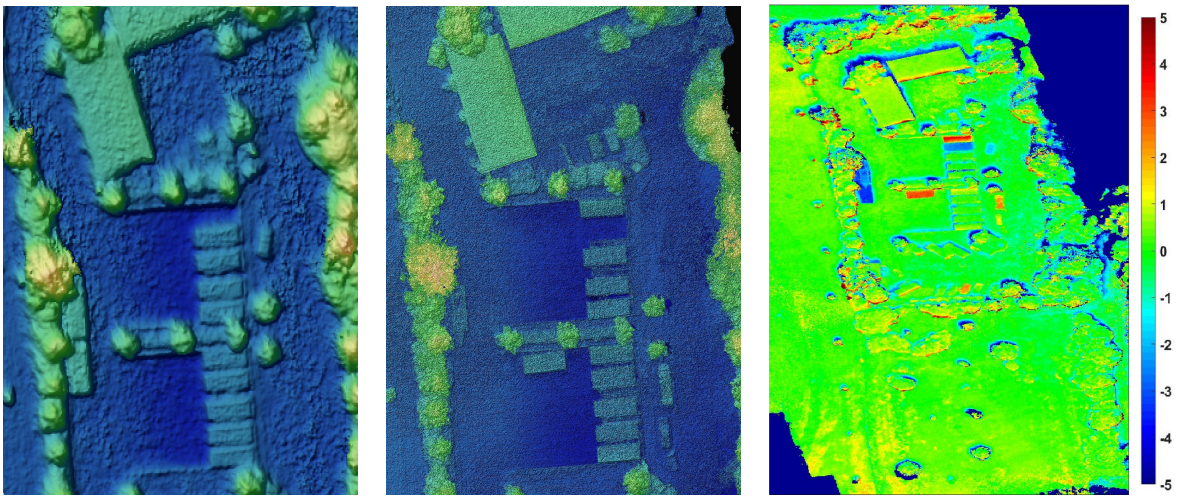
Figures 15 and 16 illustrate the aerial DSMs with 20 cm resolution and UAV DSMs with 2 cm resolution of the `Eichenau` and `Germering` dataset, respectively. The aerial DSM in (a) has a blurred edge and inadequate details, while the UAV DSM in (b) represents more refined details and sharper edges. Then, the UAV DSM was resampled by bilinear interpolation to the same resolution of the aerial DSM, and their height differences were calculated. (c) illustrates the colorized height differences

ranging from −5 m to 5 m, and it is apparent that the errors are mostly smaller than 1 m. Note that the two red and one blue spots on the container site in Figure 16c indicate movements of the containers due to different acquisition times of the captured images. In this sense, our matching method is able to cope with such temporal changes in the scene. Figure 17 shows the histograms of the height differences for both datasets.



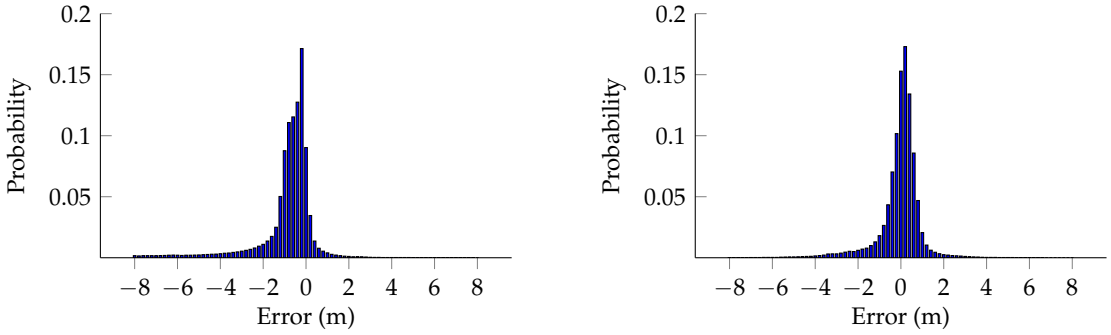
(**a**) Aerial DSM (**b**) UAV DSM (**c**) DSM differences (m)

Figure 15. Comparison of (**a**) aerial and (**b**) UAV DSM of the `Eichenau` dataset. 20-cm GSD for aerial and 2-cm GSD for UAV DSM. (**c**) Color map illustrating the height differences between the two DSMs in meters.



(**a**) Aerial DSM (**b**) UAV DSM (**c**) DSM differences (m)

Figure 16. Comparison of (**a**) aerial and (**b**) UAV DSM of the `Germering` dataset. 20-cm GSD for aerial and 2-cm GSD for UAV DSM. (**c**) Color map illustrating the height differences between the two DSMs in meters.



(**a**) `Eichenau` (**b**) `Germering`

Figure 17. Histograms of the height differences between the aligned DSMs generated from UAV and aerial images for the (**a**) `Eichenau` and (**b**) `Germing` datasets.

5.4. Application Scenario: Enriching 3D Building Models

The `EOC` dataset represents an urban scene, demonstrating the benefits of the joint use of aerial and UAV imagery. Figure 18a displays a dense georeferenced 3D point cloud generated solely from aerial images. Since the aerial images only contain nadir views of the scene, the reconstructed building facades are not complete, which is a typical problem for aerial photogrammetry.

We automatically geo-registered a sequence of nadir-view UAV images (see the result for one image pair in Table 8) to the aerial images. In addition, we also registered oblique UAV images facing the facades of the building to the already geo-registered UAV nadir views in a conventional photogrammetric way. Afterwards, a dense 3D point cloud was generated using all of the geo-registered UAV images, resulting in a complete reconstruction of the building with a much higher GSD than the aerial point cloud. The accurate geo-registration of the UAV images enables us to merge the UAV and aerial point cloud and leads to a comprehensive representation of the scene, as illustrated in Figure 18b. It can be seen that the UAV point cloud is precisely aligned with the aerial point cloud. While the aerial point cloud covers a large area of the scene, the UAV point cloud contributes to information of the building facades (particularly at positions indicated by yellow arrows) and enriched details of the reconstructed building.

(**a**) Aerial point cloud (**b**) Merged point cloud

Figure 18. Comparison of the dense point clouds for (**a**) only aerial images and (**b**) additional registered nadir and oblique UAV images of the `EOC` dataset. The combination of aerial and UAV images can enrich 3D models for more details and add facades to buildings.

6. Discussion

Our method achieves robust and accurate co-registration of images acquired from different acquisition platforms, thus opening up the possibility to integrate the information from multi-source

images and achieve a more comprehensive understanding of the scene. Besides, repetitive image acquisition with manned aircraft or satellites is quite expensive, whereas it is convenient to perform with UAVs. The robust registration enables the timely update of pre-existing remote sensing data using UAVs, which can also be applied in environment monitoring and change detection.

The main limitation of our method is that it only works for nadir or slightly tilted images. When a conspicuous height jump exists, the histogram may present multiple peaks, e.g., one representing matches on the ground-level and one matches on a higher level (like roofs). Therefore, manual inspection is needed in this case. Moreover, it is difficult to determine the translation threshold R if the scene depth changes continuously in the image. As listed in the first column of Table 8, there were remarkably fewer raw matches in the `Highway` scenario than in the other ones due to topographic changes. Furthermore, those scenarios containing various scene depths (e.g., `Container` and `Eichenau`) resulted in wrong tilts when estimating the homography, leading to higher mean transfer errors (up to seven pixels) compared to the scenarios with flat landscape.

7. Conclusions

This paper investigates into UAV geo-registration by matching UAV images with already georeferenced aerial imagery. On the basis of an extensive analysis of why SIFT performs poorly for this kind of image pair, a robust image matching approach is proposed to deliver a large number of reliable matching correspondences between the UAV and a reference image. The method is comprised of a novel feature detector, a one-to-many matching strategy and a global geometric constraint for outliers' detection. The prerequisite of our proposed method is the availability of rough GNSS/IMU data of the UAV images to eliminate scale differences in the images and if possible to pre-align the images with respect to the image rotation, although an extension of the method can handle unknown or imprecise image rotations.

Experimental results prove that our method outperforms SIFT/ASIFT in the aspects of quantity and accuracy of the detected matches. These matches are used to align UAV image blocks towards the reference images in a bundle block adjustment, which achieves a registration accuracy of 1–3 GSD. A global accuracy evaluation of 3D points from geo-registered UAV images and terrestrial measurements from RTK GNSS shows 0.5 m horizontal 1.5 m vertical deviations, which mainly stem from the inaccurate georeferencing of the reference image.

Acknowledgments: This research was funded by the German Research Foundation (DFG) for Tobias Koch and the German Academic Exchange Service (DAAD:DLR/DAAD Research Fellowship Nr. 50019750) for Xiangyu Zhuo.

Author Contributions: Xiangyu Zhuo and Tobias Koch contributed equally to the development of the methodology and the design and analysis of the conducted experiments. Besides, they shared the acquisition of the UAV imagery. Friedrich Fraundorfer and Peter Reinartz supervised the concept of the research and interpretation of the results. Xiangyu Zhuo and Tobias Koch wrote the manuscript supported by Friedrich Fraundorfer, Peter Reinartz and Franz Kurz.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Eisenbeiß, H. UAV Photogrammetry. Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich, Switzerland, 2009.
2. Nex, F.; Remondino, F. UAV for 3D mapping applications: A review. *Appl. Geomat.* **2014**, *6*, 1–15.
3. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97.
4. Chiabrando, F.; Lingua, A.; Piras, M. Direct photogrammetry using UAV: Tests and first results. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, pp. 81–86.
5. Jacobsen, K.; Cramer, M.; Ladstädter, R.; Ressl, C.; Spreckels, V. DGPF-Project: Evaluation of Digital Photogrammetric Camera Systems Geometric Performance. *Photogramm. Fernerkund. Geoinf.* **2010**, *2010*, 83–97.

6. Zhao, H.; Zhang, B.; Wu, C.; Zuo, Z.; Chen, Z.; Bi, J. Direct georeferencing of oblique and vertical imagery in different coordinate systems. *ISPRS J. Photogramm. Remote Sens.* **2014**, *95*, 122 – 133.
7. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.
8. Kurz, F.; Meynberg, O.; Rosenbaum, D.; Türmer, S.; Reinartz, P.; Schroeder, M. Low-cost optical camera system for disaster monitoring. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *39*, 33–37.
9. Kurz, F.; Rosenbaum, D.; Meynberg, O.; Mattyus, G.; Reinartz, P. Performance of a real-time sensor and processing system on a helicopter. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL-1*, 189–193.
10. Vexcel UltraCam. Available online: http://www.vexcel-imaging.com/ (accessed on 10 February 2017).
11. Vexcel UltraNav. Available online: http://www.vexcel-imaging.com/wp-content/uploads/2016/09/Brochure_UltraNav.pdf (accessed on 10 February 2017).
12. Verhoeven, G.; Wieser, M.; Briese, C.; Doneus, M. Positioning in time and space: Cost-effective exterior orientation for airborne archaeological photographs. In Proceedings of the 24th International CIPA Symposium (ISPRS), Strasbourg, France, 2–6 September 2013; pp. 313–318.
13. Gerke, M.; Przybilla, H.J. Accuracy analysis of photogrammetric UAV image blocks: Influence of onboard RTK-GNSS and cross flight patterns. *Photogramm. Fernerkund. Geoinf.* **2016**, doi:10.1127/pfg/2016/0284.
14. Zitová, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977 – 1000.
15. Bay, H.; Ess, A.; Tuytelaars, T.; van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359.
16. Calonder, M.; Lepetit, V.; Strecha, C.; Fua, P. Brief: Binary robust independent elementary features. In Proceedings of the European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010; Springer: Berlin, Heidelberg, 2010; pp. 778–792.
17. Calonder, M.; Lepetit, V.; Ozuysal, M.; Trzcinski, T.; Strecha, C.; Fua, P. BRIEF: Computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1281–1298.
18. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
19. Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119.
20. Heinly, J.; Dunn, E.; Frahm, J.M. Comparative evaluation of binary features. In *Computer Vision—ECCV 2012*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 759–773.
21. Bekele, D.; Teutsch, M.; Schuchert, T. Evaluation of binary keypoint descriptors. In Proceedings of the 2013 20th IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 15–18 September 2013; pp. 3652–3656.
22. Dwarakanath, D.; Eichhorn, A.; Halvorsen, P.; Griwodz, C. Evaluating performance of feature extraction methods for practical 3D imaging systems. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, Dunedin, New Zealand, 26–28 November 2012; pp. 250–255.
23. Juan, L.; Gwun, O. A comparison of sift, pca-sift and surf. *Int. J. Image Process.* **2009**, *3*, 143–152.
24. Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE features. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012, Springer: Berlin/Heidelberg, Germany, 2012; pp. 214–227.
25. Yu, G.; Morel, J.M. ASIFT: An algorithm for fully affine invariant comparison. *Image Process. On Line* **2011**, *1*, 11–38.
26. Apollonio, F.; Ballabeni, A.; Gaiani, M.; Remondino, F. Evaluation of feature-based methods for automated network orientation. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *40*, 47–54.
27. Conte, G.; Doherty, P. Vision-based unmanned aerial vehicle navigation using geo-referenced information. *EURASIP J. Adv. Signal Process.* **2009**, doi:10.1155/2009/387308.

28. Fan, B.; Du, Y.; Zhu, L.; Tang, Y. The registration of UAV down-looking aerial images to satellite images with image entropy and edges. In Proceedings of the International Conference on Intelligent Robotics and Applications, Shanghai, China, 10–12 November 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 609–617.
29. Lin, T.Y.; Cui, Y.; Belongie, S.; Hays, J. Learning deep representations for ground-to-aerial geolocalization. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5007–5015.
30. Shan, Q.; Wu, C.; Curless, B.; Furukawa, Y.; Hernandez, C.; Seitz, S.M. Accurate geo-registration by ground-to-aerial image matching. In Proceedings of the 2014 2nd International Conference on 3D Vision, Tokyo, Japan, 8–11 December 2014; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2014; Volume 1, pp. 525–532.
31. Majdik, A.L.; Verda, D.; Albers-Schoenberg, Y.; Scaramuzza, D. Air-ground matching: Appearance-based GPS-denied urban localization of micro aerial vehicles. *J. Field Robot.* **2015**, *32*, 1015–1039.
32. Aicardi, I.; Nex, F.; Gerke, M.; Lingua, A.M. An image-Based approach for the co-Registration of multi-Temporal UAV image datasets. *Remote Sens.* **2016**, *8*, 779–798.
33. Xu, Y.; Ou, J.; He, H.; Zhang, X.; Mills, J. Mosaicking of Unmanned Aerial Vehicle Imagery in the Absence of Camera Poses. *Remote Sens.* **2016**, *8*, 204.
34. Koch, T.; Zhuo, X.; Reinartz, P.; Fraundorfer, F. A new paradigm for matching UAV-and aerial images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 83–90.
35. Zhuo, X.; Cui, S.; Kurz, F.; Reinartz, P. Fusion and classification of aerial images from MAVS and airplanes for local information enrichment. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 20166; pp. 3567–3570.
36. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282.
37. D'Angelo, P.; Reinartz, P. Semiglobal matching results on the ISPRS stereo matching benchmark. In Proceedings of the ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, 14–17 June 2011.
38. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.
39. Koch, T.; d'Angelo, P.; Kurz, F.; Fraundorfer, F.; Reinartz, P.; Korner, M. The TUM-DLR multimodal earth observation evaluation benchmark. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 27–30 June 2016.
40. Rabus, B.; Eineder, M.; Roth, A.; Bamler, R. The shuttle radar topography mission-a new class of digital elevation models acquired by spaceborne radar. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 241–262.
41. Rothermel, M.; Wenzel, K.; Fritsch, D.; Haala, N. SURE: Photogrammetric surface reconstruction from imagery. In Proceedings of the LC3D Workshop, Berlin, Germany, 4–5 December 2012.