A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs

Saeed Yahyanejad, Bernhard Rinner

Institute of Networked and Embedded Systems,
Alpen-Adria Universität Klagenfurt, AUSTRIA
<firstname.lastname>@aau.at

Abstract

The use of multiple small-scale UAVs to support first responders in disaster management has become popular because of their speed and low deployment costs. We exploit such UAVs to perform real-time monitoring of target areas by fusing individual images captured from heterogeneous aerial sensors. Many approaches have already been presented to register images from homogeneous sensors. These methods have demonstrated robustness against scale, rotation and illumination variations and can also cope with limited overlap among individual images.

In this paper we focus on thermal and visual image registration and propose different methods to improve the quality of interspectral registration for the purpose of real-time monitoring and mobile mapping. Images captured by low-altitude UAVs represent a very challenging scenario for interspectral registration due to the strong variations in overlap, scale, rotation, point of view and structure of such scenes. Furthermore, these small-scale UAVs have limited processing and communication power. The contributions of this paper include (i) the introduction of a feature descriptor for robustly identifying corresponding regions of images in different spectrums, (ii) the registration of image mosaics, and (iii) the registration of depth maps. We evaluated the first method using a test data set consisting of 84 image pairs. In all instances our approach combined with SIFT or SURF feature-based registration was superior to the standard versions. Although we focus mainly on aerial imagery, our evaluation shows that the presented approach would also be beneficial in other scenarios such as surveillance and human detection.

Furthermore, we demonstrated the advantages of the other two methods in case of multiple image pairs.

Keywords: real-time UAV mapping, visual and thermal images, image registration, depth map, 3D registration

1. Introduction

Unmanned aerial vehicles (UAVs) are used extensively in the military domain and are becoming increasingly popular for other applications. Recent advances in technology, material science, and control engineering have made the development of small-scale UAVs possible and affordable. In particular, small-scale UAVs with a total weight of approximately $1\,kg$ and a diameter of less than $1\,m$ are now highly accessible for civilian use and pose new research questions. These UAVs are commonly equipped with sensors such as accelerometers, gyroscopes, and barometers that provide automatic stabilization as well GPS receivers that spatially report their location. Additionally, some UAVs can support the weight of additional sensors such as visual and infrared cameras. Figure 1 shows two such UAVs equipped with different sensors.

UAVs fitted with cameras can provide aerial views of regions that would otherwise be inaccessible. This type of UAV is useful in many applications such as environmental monitoring, surveillance and law enforcement, border control, farmland and crop monitoring, object detection, construction site assessment, and disaster management. Most importantly, such scenarios often involve highly dynamic target regions (e.g., a forest fire) and therefore existing information (e.g., maps) may no longer be valid. Hence, the overall goal is to provide a quick and accurate overview of the affected area, typically spanning hundreds of thousands of square meters. Composite aerial images can be refined and updated over time and can be augmented with additional information, such as detected objects or the changes in the scene. When covering large areas at conventional imaging resolutions using such small-scale UAVs, the overview image typically needs to be generated from dozens of individual images. Often, several UAVs are employed to cope with the stringent time constraints of dynamic environments and the flight time limitations of a single UAV. Commercially available UAVs are often required to fly at altitudes of less than $100\,m$ and use specialized software to convert the acquired images into a single mosaic image. This process, referred to as

(a) MD4-200 with a visual camera.

(b) AscTec Pelican with an FLIR Tau-2 thermal camera.

Figure 1: Two UAVs with different sensors.

image mosaicking, generates a broad overview image of the target area from which users can extract the desired information.

Fused images of a target scene from different sensors (e.g., visual and thermal) provide additional information for monitoring purposes. Nowadays exploiting imagery captured at different spectral bands is commonplace in various applications such as inspection, surveillance or recognition. In the case of fixed sensors or camera settings, the registration of images is relatively easy because the intrinsic and extrinsic parameters of the cameras can be determined, i.e., the relative orientation of stereo cameras is fixed and as a result the transformation for the image alignment can be computed based on purely geometric information. However, when this piece of information is missing or corrupt, the registration is more challenging and is typically accomplished using shared image feature recognition.

This paper deals with the robust registration of visual and thermal images captured by different sensors. The alignment process typically follows the following steps: (i) extraction of features in the individual images, (ii) matching corresponding feature points and identifying inliers between them, and (iii) computing the transformations for aligning individual images. Importantly, images of different spectrums may include rather distinct information. In general, the larger the wavelength difference between captured images, the greater the likelihood of feature dissimilarity. In this paper we

focus on extracting robust features which can be used to identify correspondences. Following a general discussion of feature point matching in thermal and visual images and we focus on the registration of low-altitude aerial images captured by small-scale UAVs. For such UAVs, the number of images and the positions from which they are captured are usually predefined by constraints on flight time, communication bandwidth and local processing power (Quaritsch et al., 2010; Wischounig-Strucl et al., 2011). In our experience, these images put heavy requirements on the registration algorithms because of the strong variations in overlap, scale, rotation, point of view and structure of the scene (Yahyanejad et al., 2010).

This paper introduces feature descriptors for the robust identification of correspondences between images of different spectrums, the registration of image mosaics and registration based on depth maps. We introduce a robust feature, derived from the edges, and demonstrate a quantifiable improvement in the quality of the points identified in the general case. Furthermore, we propose two methods to improve the registration of low-altitude aerial images. The first method exploits visual and thermal image mosaics, whereas the second utilizes a scene's depth map to perform feature extraction and registration.

In Section 2 related work is presented and compared with the presented approach and in Section 3 we analyse the performance of existing feature extraction methods for interspectral registration. Section 4 and 5 respectively introduce robust features along the edge and two registration methods for UAV imagery. Experimental results and analysis are presented in Section 6.

2. State of the art

The challenges of interspectral registration are largely determined by the type of the camera used. Important aspects include:

Altitude: Are the images taken from satellites or from lower altitudes (e.g., aircrafts)?

Quality: What is the noise level or resolution the images?

Timing: Are the images taken at the same time?

Spectrum: What is the difference in wavelengths of the different electromagnetic bands?

Overlap: How large is the overlap and what is the transformation function (relative translation, scale, rotation) between images? Are the images taken from different points of view?

Many interspectral registration techniques have been employed to register the different bands of satellite images. Fonseca and Costa (1997) proposed an automatic satellite image registration based on wavelets. They mainly focused on the registration of images taken from the same sensor. Mahdi and A. Farag (2002) proposed a cooperative parallel optimization based on a genetic algorithm to match the different bands of multispectral satellite images. For time efficiency, their method requires a parallel implementation of the genetic algorithms and a supervisor process. Hong and Zhang (2005) described an automated registration technique by combining feature-based and area-based matching for high resolution satellite images. They employed wavelet-based feature extraction and a relaxation-based image matching technique to reduce the local distortions caused by terrain relief. Although they managed to speed up the registration process, they only considered the registration of panchromatic with multispectral images which are almost in the same spectral range. Kim and Kim (2012) focused on the problem of parallax removal caused by different viewpoints. They improved the registration process by correcting the terrain relief using a rigorous sensor model with precise sensor parameters and ellipsoidal height information extracted from Digital Elevation Model (DEM) data. Kern and Pattichis (2007) proposed robust interspectral registration using mutual-information models. As the shape of a mutual-information surface is related to the frequency-domain characteristics of the imagery, this information may be used to iteratively optimize a target function and find the appropriate registration parameters. Lee (2010) performed coarse-to-fine multispectral satellite image registration, using the Speeded Up Robust Features (SURF) (Bay et al., 2008) for fast initial feature extraction and handling the possible differences in scale and orientation. He then applied the Harris operator to extract more features. In contrast, Teke and Temizel (2010) suggested the use of the SURF feature extraction method to perform the whole registration process. Teke's method takes advantage of the SURF parameters in cases when there is no rotation (Upright-SURF) or no scale differences (Scale-Restricted SURF) between two images.

Satellite remote sensing is not the only field where interspectral registration plays a critical role. Medical imaging, object or face detection, surveillance and UAV remote sensing are other fields that are growing rapidly and

where image registration has become very important. Schaefer et al. (2008) performed multi-modal (thermal and visual) medical image registration and overlay. By exploiting *a priori* knowledge of the human body, they segmented both types of image to find the matching body area. Kong et al. (2006) and Vaidehi et al. (2011) also used *a priori* knowledge about the human face to register and fuse the thermal and visual face images. Istenic et al. (2007) registered thermal and visual images of the facades of buildings. Since most buildings present straight lines, they performed a Hough transform over the images to extract the latter as mutual features. Likewise, the methods developed by Coiras et al. (2000) and Šegvic (2005) rely on having sufficient straight lines or structured polygons as a prerequisite for registration. Du and Raksuntorn (2008) presented an algorithm for the automatic registration of near-infrared and visual image sequences taken by a UAV. Since both cameras are mounted on a single UAV (i.e., the relative orientation is fixed), the computation of the extrinsic parameters of the cameras prior to flight is less expensive. Unlike all the mentioned works, which are based on individual images for registration, Joo et al. (2003) performed registration by using sequences of frames with moving objects. They first extract the moving region of each image as the target area, then they perform matching and registration over that region. While this method does have some advantages, it fails completely in the absence of moving objects.

In this paper we present three robust methods of feature extraction for interspectral image registration. Our methods do not rely on any prior knowledge about the scene and are able to deal with images with completely different spatial and temporal resolution, scale and orientation, as well as a low amount of overlap. Our first method extends existing robust feature extraction methods to the domain of interspectral image registration. The other two methods extract additional features in the presence of multiple pairs of thermal and visual images. Both methods are useful for the registration of low-altitude aerial images.

3. System overview

This work was performed as part of the project *Collaborative Microdrones (cDrones).* [1] The core idea of the project is to deploy multiple small-scale

[1] `http://uav.aau.at/`

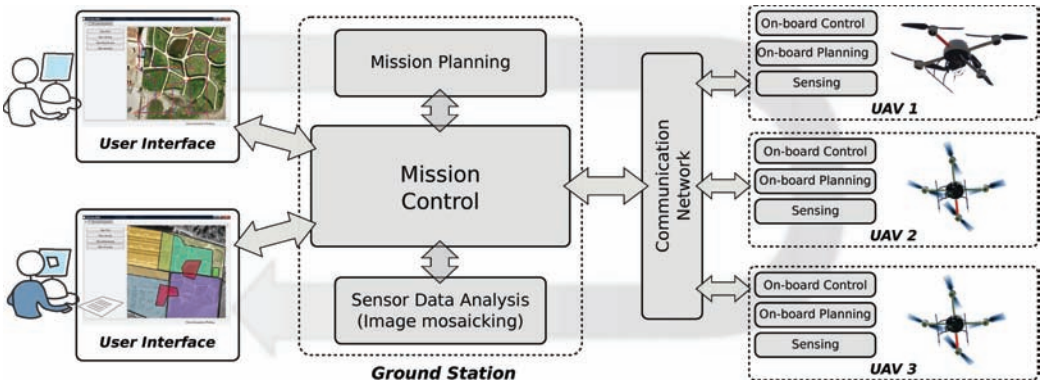


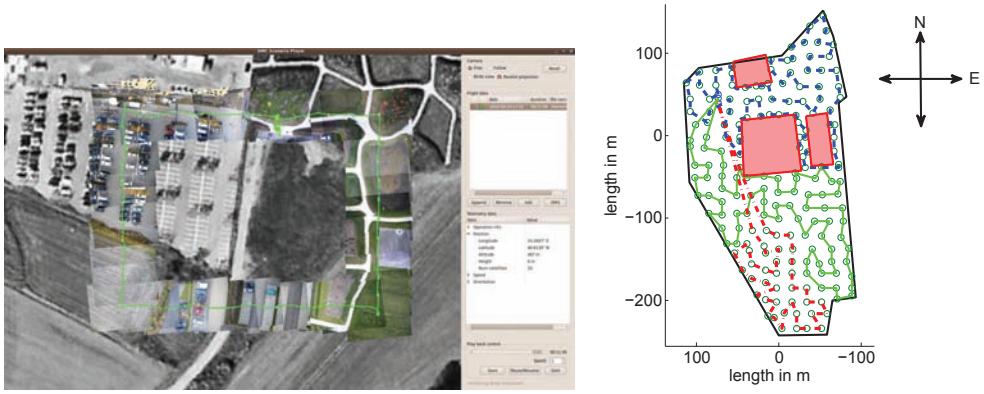
Figure 2: System architecture of our multi-UAV project.

UAVs to support first responders in disaster assessment and management. In particular, we use commercially available quadrocopters since they are agile, easy to fly and stable in the air due to their on-board sensors such as GPS and IMU. Each UAV is also equipped with a camera. Figure 2 describes the system architecture of the project.

The intended application can be sketched as follows: The operator first specifies the areas to be observed on a digital map and defines quality parameters for each area (Quaritsch et al., 2008). Quality parameters include, among others, the spatial and temporal resolution of the generated overview image and the minimum and maximum flight altitude (Quaritsch et al., 2010).

Based on the user's input, the system generates plans for the individual UAVs to cover the observation areas (Quaritsch et al., 2011). These are then partitioned into smaller areas corresponding to individual pictures to be taken by a particular UAV flying at a certain height. The partitioning must overlap neighboring images as required by the stitching process. Given a partitioning, we can discretize the continuous areas to be covered to a set of so-called picture-points. The picture-points are placed in the center of each partition at the chosen height. The pictures are taken with the camera pointing downwards (nadir-view).

The mission planner component generates routes for individual UAVs such that each picture-point is visited within the constraints of the UAV's resource limitations. The images together with metadata (i.e., the position and orientation of the camera) are transferred to the base-station during



(a) GUI showing new images over outdated background.

(b) Sample mission showing restricted areas and planned routes.

Figure 3: User interface and mission planning.

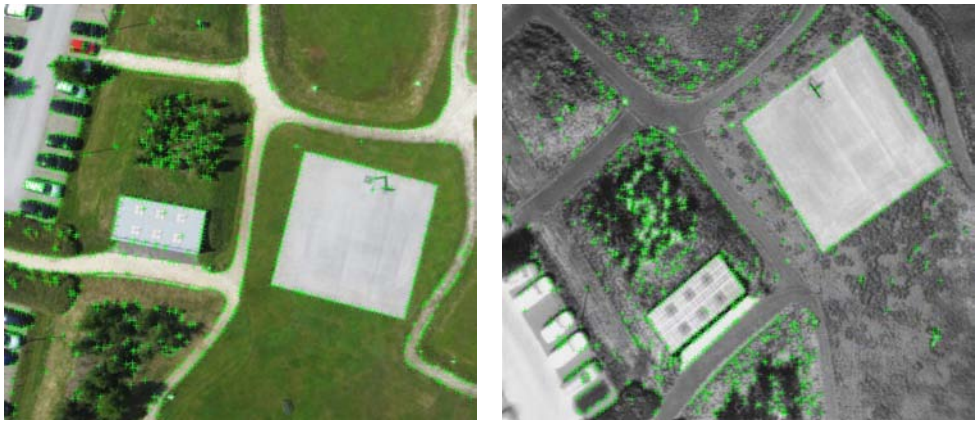
flights where the individual images are stitched into an overview image. Figure 3 illustrates samples of mission planning and aerial image acquisition within the scope of the project. Restricted areas (e.g., buildings and dangerous areas) are marked as obstacles. After planning is finished, the mission is executed. The UAVs take off fully autonomously, fly the specified routes and send the pictures to the ground station. Mosaicking the individual images and constructing an overview image is the final step of the procedure.

Additionally, various other applications have been considered within the scope of the project, such as object detection and tracking, multi-UAV area coverage to help the first responders with disaster management, construction site monitoring, and advertising. Sample images, videos and demonstrations are available on the project web-page. [2]

4. Analysis of existing feature extraction methods

As previously described, feature extraction is a fundamental step for registration. Although most of the conventional feature extraction methods (such as edge detection or corner detection) can be used to identify the mutual information between visual and thermal images, constructing an appropriate descriptor for finding matching pairs is not so simple. As an example,

[2] `http://uav.aau.at/`



(a) Harris corners in the visual image. (b) Harris corners in the thermal image.

Figure 4: The result of Harris operator over a pair of visual and thermal images.

Figure 4 shows the utilization of the Harris operator (Harris and Stephens, 1988) over a pair of visual and thermal images. This figure demonstrates the differences in the corners extracted in the visual and thermal images. Moreover, the correlation-based matching of the corners fails because of the different intensity pattern and the rotation difference between the two images. In general, the task of matching and removing the outliers and finding the transformation becomes challenging in presence of relative rotation and scale between images.

A multi-scale Harris operator and some other scale-invariant, rotation-invariant, illumination-invariant and affine-invariant feature extraction and matching methods have been proposed (Hansen and Morse, 1999; Dufournaud et al., 2000; Mikolajczyk and Schmid, 2004; Lin et al., 2010) to cope with this limitation. However, the methods with a well-defined robust local descriptor, such as SIFT (Lowe, 2004) and SURF (Bay et al., 2008), are gaining more attention. Equation 1 (Lowe, 2004) describes how the SIFT method detects the keypoint locations.

$$\begin{aligned} &D(x,y,\sigma) = (G_(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y), \\ &\text{where} \qquad G(x,y,\sigma) = \tfrac{1}{2\Pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \end{aligned} \tag{1}$$

The difference of the Gaussian is used as an approximation for the scale-normalized Laplacian of Gaussian, $\sigma^2\nabla^2 G$. The target keypoints are obtained

(a) SIFT features with 43 scale levels and Euclidean distance threshold 1.2 for matching.

(b) SURF features with sampling step 1 and initial lobe 1.

(c) Alignment and fusion.

Figure 5: Registration of a thermal and visual pair by using SIFT and SURF.

by calculating the differences between different scales of Gaussian blur over each octave and then by finding the local extrema based on comparing each sample point with its eight neighbors in the current image and its nine neighbors in the scale above and below. Since this method was initially designed for the registration of the images taken from homogeneous sensors, it fails if the parameters are not adjusted in the appropriate way. In other words, when comparing a pair of thermal and visual images taken from the same scene, we may receive matching keypoints at different scale levels even if the images have exactly the same scale ratio. By experiment, we realized that a larger number of scales in the SIFT method improves the registration quality but requires greater computational power because more features needs to be extracted. A detailed study regarding the scales of the SIFT method has been performed by Morel and Yu (2011). If the initial octave is set to -1, feature extraction starts with a double-sized image and consequently obtains more keypoints. In practice, however, the increased image size does not affect the registration quality because there are very few small features in most of the low-resolution thermal images. Nevertheless, setting a lower threshold for the multiplier coefficient of the Euclidean distance of the feature vectors is advantageous for calculating the matching pairs because the matching requirements are not so strict in case of different sensors. Figure 5(a) shows an example registration result using the SIFT feature extraction with 43 scale levels and an Euclidean distance threshold of $s = 1.2$ for matching. The default parameter setting does not lead to a successful registration.

Speeded-Up Robust Features (SURF) (Bay et al., 2008) analyzes the

different scale levels by up-scaling the box filter size rather than iteratively reducing the image size. In this way the performance is highly improved. The keypoint identification is based on an approximation of the determinant of Hessian—instead of the Laplacian of Gaussian in the SIFT descriptor. Figure 5(b) shows the same pair of images registered by SURF.

Both methods achieve approximately a 50% successful registration rate by adjusting their parameters based on each scenario. In our dataset we have tested different pairs of satellite images, images of human, images of the nature and surveillance, images taken from UAVs and images from facade of buildings. We show detailed results of this dataset in Section 7. Also with the fixed parameters (described in Figure 5) both methods achieve a similar performance. Figure 5(c) shows a sample aligned and fused result. In all our experiments, we have used RANdom SAmple Consensus (RANSAC) (Fischler and Bolles, 1981) and least median of squares (LMS) to remove the outliers (among all matched pair-points) and calculate the appropriate similarity transformation between images.

5. Robust features along the edge (RFAE)

Despite the acceptable performance of SIFT and SURF for interspectral image registration, they have failed in some scenarios in which mutual patterns are clearly available. Apparently both descriptors are inherently designed to emphasize the patterns of the gradient changes around a specific keypoint. Note that in different sensors, and more specifically considering thermal and visual sensors, we often record a completely different intensity value for each specific target region. This characteristic affects the matching between the descriptors.

To overcome this problem some authors extract line structures from the images to identify matching points (Coiras et al., 2000; Šegvic, 2005; Istenic et al., 2007). The main limitation of these methods is that they require a sufficient number of straight lines. Our approach extends this idea and extracts the edge structures in the images. It uses then SIFT or SURF to identify feature descriptors in the extracted binary edge image. This preserves the scale, rotation and illumination invariant characteristics. In our experiments we have used the Sobel operator as an approximation for the intensity gradient in the images (Gonzalez and Woods, 1992), i.e.,

(a) Successful registration by extracting the features along the edge.

(b) Overlaying the registration on the original image pair.

(c) Alignment and fusion.

Figure 6: Registration of a thermal and visual pair by using scale invariant features along the edge.

$$\mathbf{S}_x = \frac{1}{4}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \qquad \mathbf{S}_y = \frac{1}{4}\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix},$$

$$\begin{aligned} \text{Gradient:} \quad & \nabla\mathbf{I} \simeq (\mathbf{S}_x * \mathbf{I}, \mathbf{S}_y * \mathbf{I}), \\ \text{Gradient magnitude:} \quad & \|\nabla\mathbf{I}\| \simeq S(\mathbf{I}) = \sqrt{(\mathbf{S}_x * \mathbf{I})^2 + (\mathbf{S}_y * \mathbf{I})^2}, \\ \text{Gradient direction:} \quad & \Theta(\nabla\mathbf{I}) \simeq \arctan(\tfrac{\mathbf{S}_x * \mathbf{I}}{\mathbf{S}_y * \mathbf{I}}). \end{aligned} \tag{2}$$

To extract the edges, the resulting image is converted to binary by a cutoff threshold. Since the approximation of the gradient becomes bimodal, we reduce the sensitivity of the descriptors to the change of gradient in a neighborhood. Finally, the SIFT and SURF operate the difference of Gaussian or determinant of Hessian over this binary image. In case of SIFT we adopt the Equation 1 so that the difference of Gaussian along the edges is defined as

$$D_E(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * B \circ S(I(x, y), \theta), \tag{3}$$

where B is the binary operator based on the threshold θ and S is the Sobel operator constructed as explained in Equation 2. Figure 6 shows the registration of a pair of thermal and visual images taken from a building facade using our method. Note that the registration using merely the SIFT or SURF features failed. A remaining question is how to define the appropriate threshold, θ, for the conversion to the binary edge image. This threshold can be estimated by a statistical analysis of different sensors or different

image types. In our experiments, we extract the edges with three different threshold values ($\theta \in \{0.2, 0.4, 0.6\}$). We choose a pair (one threshold for the thermal image and one threshold for the visual image) for registration which maximizes our quality metric (cp. Section 7). Other edge detectors such as Laplacian of Gaussian, Canny (Canny, 1986), line segment detector (Grompone von Gioi et al., 2012) are applicable to this method, however they did not show considerable improvement in most of our scenarios.

Using this method we do not merely rely on long edges, but we also consider edges and corners comprised of only few pixels. One important advantage of RFAE over state of the art of registration using edge descriptors, such as the method proposed by Meltzer and Soatto (2008), is that RFAE does not necessarily consider a single edge. Since SIFT and SURF descriptors are calculated over blobs, therefore they include the gradient distribution of a region which may include multiple edges and their relative shape and orientation. RFAE does not cope with local deformations that some shape-based matching or medical image registration methods are handling (Klein et al., 2011; Müller et al., 2005).

6. Interspectral registration by multiple thermal-visual image pairs

So far we have considered the general case of registering thermal and visual images. In this section, we focus on registering low-altitude aerial images captured by small-scale UAVs. Because of the weight restrictions, these aerial robots are typically not able to carry both types of cameras. For these scenarios we improve the robustness of the registration by two methods. The first method exploits entire visual and thermal mosaics. The second method uses depth information to extract additional features for the registration. In all proposed methods we consider the limited on-board processing power and communication bandwidth to achieve a real-time monitoring.

6.1. Registration of mosaics.

As previously discussed, the RFAE method does not always improve interspectral registration. Figure 7 shows an example where RFAE did not improve the registration of two image pairs (i.e., I_{V_1} with I_{T_1} and I_{V_2} with I_{T_2}). The reason is that there are insufficient salient border lines and edges which are visible in both image types. As indicated with green lines in Figure 7, one pair (I_{V_1} with I_{T_1}) can be weakly registered with 7 SURF feature matches.

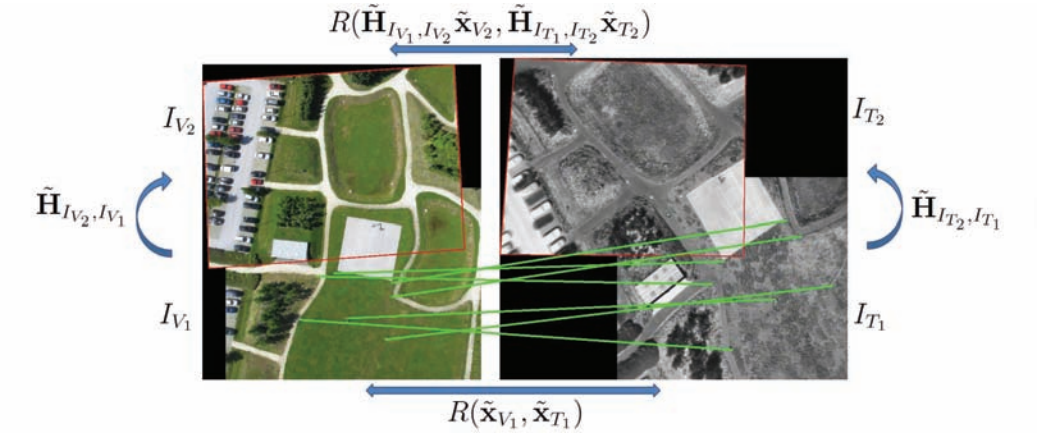


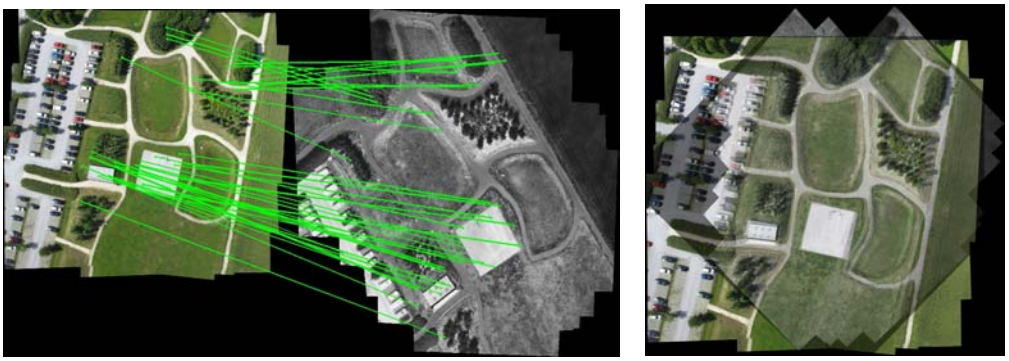
Figure 7: Interspectral registration by using multiple pairs.

Here we present a new method to exploit the image mosaics to strengthen the interspectral registration. The mosaicking of aerial images taken from an identical sensor is based on the homography $\mathbf{H}$ corresponding to the perspective transformation between each pair of images. Although the target scene may not be completely planar, the choice of homography can be valid. In other words, the error is negligible as long as we choose the features from a dominant ground plane or the value of ϵ_d in Equation 6 is small (Yahyanejad et al. (2011b) and Szeliski (2010) explain these concepts in more details). Thus, pairwise registration can be seen as an initial step for mosaicking. Registration within a specific spectrum (identical sensor) is typically more robust and can be achieved even with a limited pairwise overlap. As shown in Figure 7, the visual image I_{V_2} is transformed to the coordinates of the visual image I_{V_1} by homography $\tilde{\mathbf{H}}_{I_{V_2},I_{V_1}}$. Similarly, the thermal image I_{T_2} is transformed to the coordinates of the thermal image I_{T_1} by homography $\tilde{\mathbf{H}}_{I_{T_2},I_{T_1}}$. By knowing one of the interspectral registration parameters, for example the corresponding pair points between images I_{V_1}, I_{T_1} shown as $R(\tilde{\mathbf{x}}_{V_1}, \tilde{\mathbf{x}}_{T_1})$, we can calculate the corresponding pair points between images I_{V_2}, I_{T_2} by

$$R(\tilde{\mathbf{x}}_{V_2}, \tilde{\mathbf{x}}_{T_2}) = R(\tilde{\mathbf{H}}_{I_{V_1},I_{V_2}}\tilde{\mathbf{x}}_{V_2}, \tilde{\mathbf{H}}_{I_{T_1},I_{T_2}}\tilde{\mathbf{x}}_{T_2}), \tag{4}$$

where $\tilde{\mathbf{x}}, \tilde{\mathbf{H}}$ are the points and the homography in homogeneous coordinates, respectively. This formulation helps us to generalize the pairwise registration.

The interspectral registration between large mosaics can be done in two



(a) Registered correspondences. (b) Overlayed and fused images.

Figure 8: Interspectral registration of two mosaics each constructed from 25 individual images.

different ways. The first way is to generalize the approach shown in Figure 7 over multiple pairs. No matter if the registration fails in some pairs, the mosaics can be registered as long as some of the thermal and visual images are registered. However, this method needs to consider all corresponding points, both within the homogeneous and the heterogeneous image types, for the global optimization. In other words, we need to find out the homographies which minimize the least mean squares (LMS) of the disparity error between all pair points. This increases the complexity of the homography estimation and the mosaic construction. Points that are considered in multiple image pairs (when more than two images overlap) will be over-weighted in this optimization. The knowledge of corresponding images, i.e., images which are supposed to have some overlap, is also required for this method. The second way is to first mosaic all the images from the same sensor and then register the two final mosaics together. The thermal and visual mosaics shown in Figure 8(a) are registered with this approach. The drawback here is that handling large image mosaics and large number of corresponding points is computationally expensive. In addition, errors in mosaicking homogeneous images affect the interspectral registration accuracy.

6.2. Exploiting the 3D structure

In scenarios where UAVs provide sparse pictures from different points of view, we can exploit methods from stereo vision to extract depth information. In this section we describe how depth information helps for the registration

of thermal and visual image mosaics. As explained by Yahyanejad et al. (2011b) we reveal depth information of a scene by using stereo images of the same scene taken from different points. We first calculate the disparity vectors from the displacement of all feature pairs in the two stereo images. Figure 9(b) depicts these disparity vectors as the displacement between two feature points after the alignment. Since this requires stereo images, we can only obtain the disparity vectors over the overlapping area of images taken from different positions. The magnitude of a disparity vector corresponds to the relative height difference of the corresponding feature point. The direction of the vector determines whether the feature point is below or above the average altitude. This helps us to construct a rough depth map as shown in Figure 9(c).

We define our false-color depth map image, **DM**, which is constructed using

$$\begin{aligned}
&\text{Red component} &&= \mathbf{DM}(x_{n_i}, y_{n_i}, 1) = \tfrac{d_{x_i} - \min_i d_{x_i}}{\max_i d_{x_i} - \min_i d_{x_i}},\\
&\text{Green component} &&= \mathbf{DM}(x_{n_i}, y_{n_i}, 2) = \tfrac{\|\mathbf{d}_i\| - \min_i \|\mathbf{d}_i\|}{\max_i \|\mathbf{d}_i\| - \min_i \|\mathbf{d}_i\|},\\
&\text{Blue component} &&= \mathbf{DM}(x_{n_i}, y_{n_i}, 3) = \tfrac{d_{y_i} - \min_i d_{y_i}}{\max_i d_{y_i} - \min_i d_{y_i}},\\
&\mathbf{d}_i = (d_{x_i}, d_{y_i}) = (x_{n_i} - \hat{x}_{n_i}, y_{n_i} - \hat{y}_{n_i}),\\
&\overline{\mathbf{x}}_{n_i} = (x_{n_i}, y_{n_i}, 1),\\
&\overline{\mathbf{H}(I_n, I_m)\tilde{\mathbf{x}}_{m_i}} = (\hat{x}_{n_i}, \hat{y}_{n_i}, 1),
\end{aligned} \tag{5}$$

where the last two equations represent the augmented vectors to convert the homogeneous coordinates back to the Cartesian coordinates; $\mathbf{d}$ represents the disparity vector as displacement between point $\mathbf{x}_n$ and its estimated position $\hat{\mathbf{x}}_n$; i is the index of the inliers among all corresponding feature points. All color components are normalized to fit in the image intensity range. In other words, the normalized x and y components of the disparity vector are used as the red and blue components of the depth map, and the green component of the depth map is the normalized magnitude of the disparity vector. The missing pixels of the depth map image are estimated by interpolation.

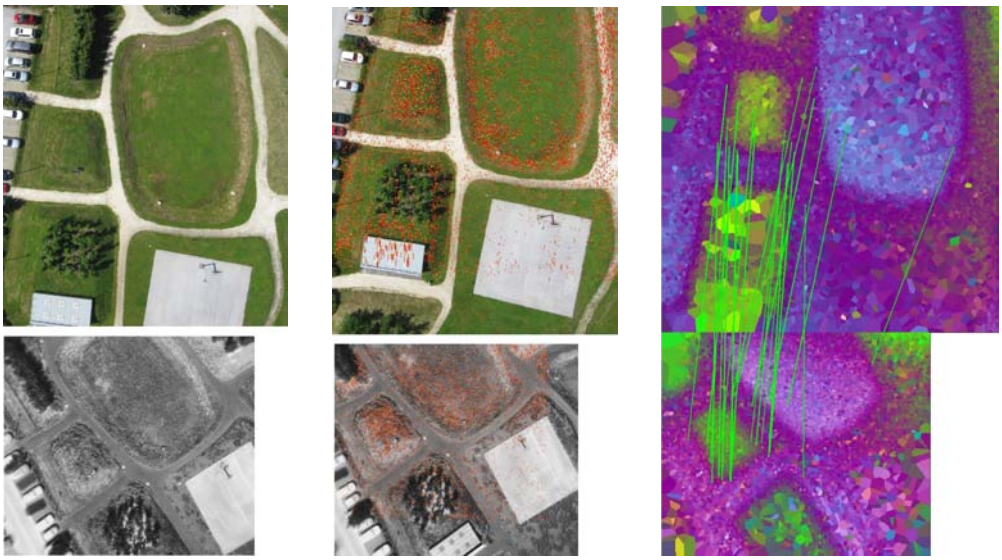
We remove features with magnitude of disparity vector $\|\mathbf{d}_i\|$ larger than a threshold ϵ_d. This threshold varies based on height variation of the objects on the ground and flying altitude. In our scenario, we calculate this threshold (in pixels) as follows:

$$\epsilon_d = \left\lceil \frac{\text{maximum height variation}}{\text{minimum flight altitude}} \times 50 \right\rceil, \tag{6}$$

where the height variation and the altitude are relative to the dominant ground plane and $\epsilon_d \geq 1$. For example, we set the threshold to 5 pixels if we have a maximum height variation of $4\,m$ and the flight height of $40\,m$. Note that at the first glance it might look similar to setting the RANSAC threshold small, but in that case we might also reject some inliers just because of their large displacement which slows down or even fails the convergence of RANSAC, especially in cases with low amount of overlap. On the other hand, as mentioned in Section 6.1, we use the features with less displacement which are considered to be almost on the same ground plane for the calculation of the homography. The features with larger displacement are exploited for constructing the depth map.

By extracting the depth map of the overlapping area in both thermal and visual image pairs, we are able to register those images by registering their depth map. Regardless of existence of any mutual pattern or similarity between visual and thermal images, the depth information of a target scene provides consistent mutual information between two image types. In other words, it is not necessary that the features from different spectrum to be close to each other as long as we have enough sparse features. An automatic registration based on SURF features is shown by the green lines in Figure 9(c). Note that since the standard SURF is performed over gray-scale images, and since the orientation of the displacement is as important as its magnitude, we perform the SURF over all dimensions of the false-color depth map image. It can be done either over the average image of all the bands or by using the union of the descriptors of all 3 bands together, which in practice the results did not show much difference. We can generalize the depth map construction from a pairwise depth map to a mosaic depth map. The disparity vectors are constructed as explained in Equation 5. The process can be optimized by a 2D bundle adjustment (global optimization).

An alternative is a 3D optimization by a full bundle adjustment and estimating and reconstructing the 3D point positions (Snavely et al., 2006; Furukawa and Ponce, 2010; Tingdahl and Van Gool, 2011). The 3D models shown in Figure 10 are generated by such 3D reconstruction from the same 25 thermal and visual images as used in Figure 8. In general, a full 3D model reconstruction achieves higher accuracy with more images. Despite this fact, we often face more challenging scenarios such as sparse images with limited overlap. Furthermore, most of the thermal cameras mounted on small-scale UAVs have a lower resolution as compared to visual cameras and provide an analog image. These cameras often do not have a global shutter and



(a) Visual and thermal images taken at initial UAV position.

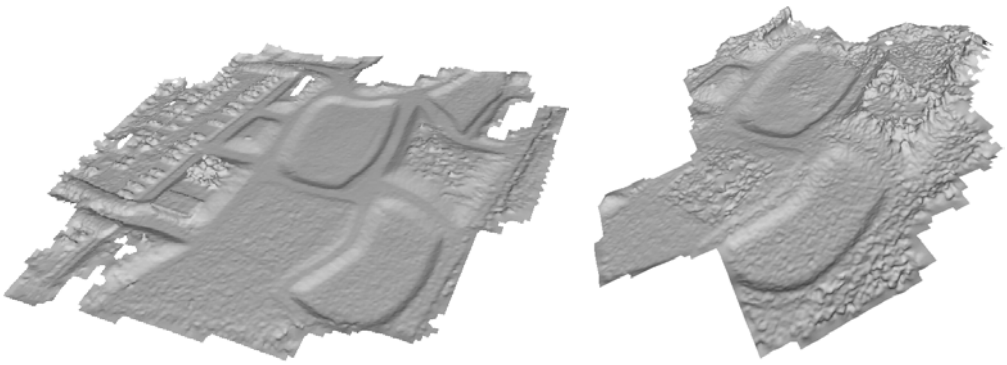
(b) Disparity vectors depicted in the overlapping area with images taken at a different UAV position.

(c) Computed depth maps of the overlap. The green lines indicate the registration based on SURF.

Figure 9: Construction of the depth map from two image pairs by calculating the disparity of the feature points. The upper row corresponds to the visual and the lower row to the thermal images.

correcting the lens distortion is not straightforward (cp. Yahyanejad et al. (2011a)). Hence, we typically see more noise in the 3D models constructed from thermal images than from visual images. It is therefore difficult to register these 3D models which also may be different in scale and orientation. Most of the existing point cloud or 3D-mesh registration methods consider a high accuracy of the models (e.g., laser scanned 3D models) or considering a rigid transformation between models (Flöry and Hofer, 2010; Sehgal et al., 2010; Song and Cheng, 2010).

Since we are more interested in 2D images and 2D registration, we perform a mapping transformation to convert our 3D models to an equivalent depth map. We first map all 3D points to a 2D plane Π, which is almost parallel to the ground and the camera planes. This is performed by finding $P_{i\Pi}$, the base point of the perpendicular of each 3D point $\mathbf{P}_i = (x_i, y_i, z_i)$, to the plane $\Pi : ax + by + cz + d = 0$ with a normal vector $\mathbf{n} = (a, b, c)$. Second, the



(a) 3D model from 25 visual images.

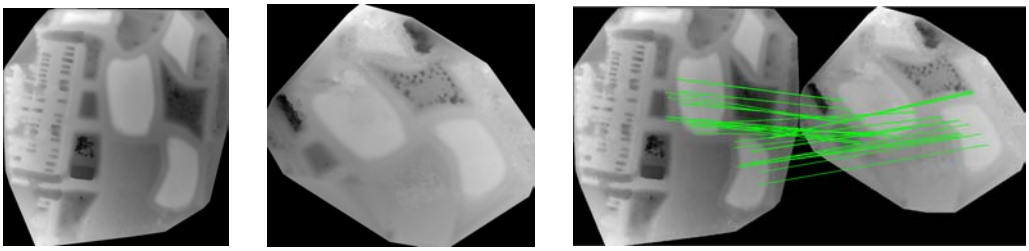
(b) 3D model from 25 thermal images.

Figure 10: 3D model reconstruction of the target area by using a full 3D bundle adjustment.

distance $d(\mathbf{P}_i, \Pi)$ of the point $\mathbf{P}_i$ to the plane Π is translated to the intensity value of the depth map image,

$$\mathbf{P}_{i\Pi} = \mathbf{P}_i - \frac{ax_i + by_i + cz_i + d}{a^2 + b^2 + c^2}\mathbf{n}, \quad d(\mathbf{P}_i, \Pi) = \frac{ax_i + by_i + cz_i + d}{\sqrt{a^2 + b^2 + c^2}}. \qquad (7)$$

Figures 11(a) and 11(b) show such depth maps constructed from the visual and thermal 3D models in Figure 10. Although the 3D reconstruction is generally slow, the depth map results show a smoother image compared to the fast depth map construction based on disparity. Figure 11(c) depicts the automatic registration of the resulting depth maps by using the SURF features. The registration in 2D can be done in few seconds, depending on image resolution and processing power, which is much faster compared to point cloud registration methods (for detailed computational complexity see Section 7.4). One obvious advantage of this registration method for thermal and visual images is the robustness against the image differences and spatial changes, since we are registering the depth information and not the image details. This is especially useful in cases with a high time difference between two remote sensing activities, such as registering images captured in different seasons of the year.



(a) Corr. depth map of the Figure 10(a).

(b) Corr. depth map of the Figure 10(b).

(c) Automatic registration done using the SURF features.

Figure 11: Extraction of the depth maps from the 3D models of Figure 10.

7. Results and discussion

This section presents further experimental results and discussions. First, we evaluate the performance of the RFAE wrt. other feature extraction methods. Second, we extend the discussion on our registration by exploiting images mosaics and depth maps.

7.1. Evaluating RFAE for interspectral registration

The evaluation of the RFAE methods is based on a heterogeneous dataset of 84 image pairs of different spectrums. This dataset consists of different types of satellite images, images of human bodies, general surveillance images and aerial images from low-altitude UAVs. The UAV-image dataset has higher variation in scale (due to the varying flight altitude) and point of view (due to the sparse picture points). On the other hand, in other scenarios, the point of view and the orientations are almost constant. This is done to emphasize the failure ratio of the other methods compared to RFAE, even in less challenging scenarios. The resolution varies for visual images between 320×240 and 1047×1061 pixels and for thermal images between 320×240 and 584×512 pixels. The overlap ratio between the image pairs varies between 50% and 100%. Table 1 summarizes our dataset.

We performed the interspectral image registration over this dataset by using the SIFT, SURF, upright SURF, RFAE and combination of SURF with RFAE. We use a quality metric to evaluate the extracted features for the purpose of interspectral registration. Although the number of corresponding matched features (inliers) is an important factor for registration, it does not provide any information regarding the distribution of the features. Irschara

Description	# of images	Overlap	Angle between principal axes	Resolution	Max. scale ratio
Satellite images from different bands	34	100%	$0\,^\circ$	$0.3 - 1.1\,MP$	1
Images including human body	16	$> 90\%$	$< 3\,^\circ$	$0.22 - 0.3\,MP$	1.1
Images taken from surveillance cameras	14	$> 85\%$	$< 5\,^\circ$	$0.08 - 0.5\,MP$	2.1
Aerial images from UAV	20	$> 50\%$	$< 17\,^\circ$	$0.3 - 1.1\,MP$	1.4

Table 1: Detailed description of 84 pair of images as our dataset.

et al. (2009) mention the importance of this distribution, yet a clear definition of their metric is missing. They consider the effective number of inliers in terms of coverage multiplied by the raw number of inliers. This way if we have two inliers over each other the metric improves which apparently should not change. In our case, we only use the number of inliers for acceptance or rejection of a registration based on a threshold. If a registration is accepted, we can use our quality metric to evaluate it or compare it with other methods. The success level of a registration increases when there are sufficient inliers and they are distributed uniformly over the image. Nevertheless, a metric which is modeling the deviation from a uniform distribution (cp. Schilcher et al. (2008)) is not appropriate for our case. Such a metric is built to quantify the inhomogeneity, so that adding an additional point very close to (or almost over) an existing point reduces the magnitude of the metric. In our case the metric should not change if we add a point exactly over an existing point and should increase slightly if we add a point in the close neighborhood of an existing point.

Suppose that an optimal feature distribution for the purpose of registration demands at least one feature point within a distance δ to any random point in the image, we construct our metric as follows:

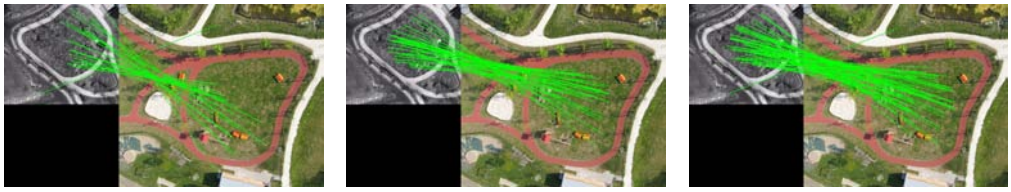
$$Q = \frac{|\bigcup_i \{\mathbf{x} : \|\mathbf{x} - \mathbf{f}_i\| < \delta\}|}{A} \times \max_{i,j} \frac{\|\mathbf{f}_i - \mathbf{f}_j\|}{d}, \tag{8}$$

where $\mathbf{x}$ represents a point in the image, $\mathbf{f}$ an inlier feature point in the image, i, j indexes of the inliers, A the area of the image which is equivalent to the image resolution in pixels and d is the length of the diagonal of the image. The denominators of the fractions aim to normalize the metric to the range $[0, 1]$. The numerator of the first fraction represents the aggregated area of all circles with radius δ centered at feature points. The first fraction of the equation implies the coverage ratio of the feature points within the image. The remainder of the equation describes the normalized maximum distance between all possible feature point pairs. This favors the sparse features rather than dense features which is an important factor for a successful registration. The Q value takes its maximum, 100%, if and only if we have at least two features over the two far corners of the image and there exist no circle with radius δ in the image so that no feature falls inside this circle.

Defining the value for δ depends on different factors such as image quality and resolution. In our experiments we set δ to 10% of the image width. We set a threshold and we accept only the registrations with at least 9 corresponding feature pairs (inliers). In our empirical study of more than 500 pairwise image registration we realized that at least 9 inliers are required to achieve a reliable registration. We set the Q value to zero for unaccepted registrations. We use this metric primarily for comparing different methods of registration over the same pair of images. For instance we can use it to find the best parameters for an individual feature extraction method (cp. Figure 12) or compare the registration performance between different methods (cp. Figure 13). These samples also show that more inliers do not necessarily result in an increase of the Q value. Note that the high Q value in Figure 12(a) is caused by a single distant feature point, despite its small number of inliers. This is exactly what we tried to achieve with our metric. Dense and concentrated set of features are not reliable for registration, since small errors cause large deviation in homography. Yet, one single distant feature prevents the large deviation in the homography estimation. The sparse feature distribution in Figure 13(b) achieves also a high Q value with far less inliers than in Figure 13(a).

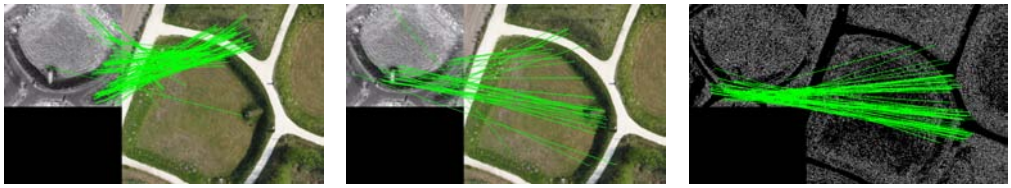
This metric is used to identify the best registration for each pair of images and classify our dataset as shown in Table 2. The fractions shown in this table represent the ratio of the number of acceptable registrations to all number of pairs.

Since satellite images have a relatively high overlap and are usually aligned quite well with a fixed rotation and scale, most of the feature extraction meth-



(a) SURF initial lobe=5, inliers=22, Q=17.2%. (b) SURF initial lobe=3, inliers=38, Q=8%. (c) SURF initial lobe=1, inliers=62, Q=16.6%.

Figure 12: Finding the initial lobe parameter of SURF which maximizes the Q value.



(a) SIFT method with inliers=102 and Q=22%. (b) SURF method with inliers=31 and Q=23%. (c) RFAE method with inliers=59 and Q=12%.

Figure 13: Finding the best method which maximizes the Q value. In this example SURF has a higher Q value.

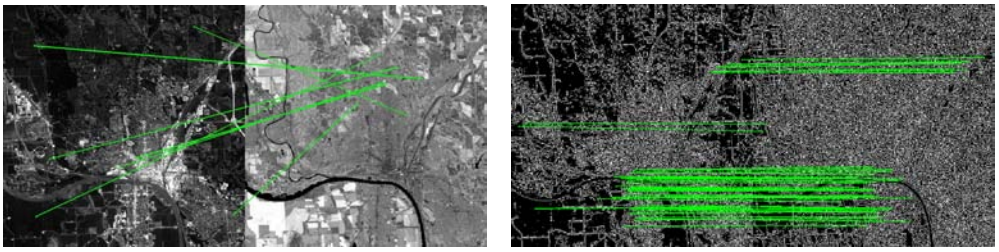
	SIFT	SURF	U-SURF	RFAE	SURF+ RFAE
Satellite (low deviation)	24/24	24/24	24/24	24/24	24/24
Satellite (high deviation)	7/10	6/10	9/10	10/10	10/10
Human	0/16	0/16	2/16	12/16	12/16
Surveillance	1/14	2/14	4/14	12/14	12/14
UAV	13/20	14/20	2/20	13/20	17/20

Table 2: Successful registration ratios based on different feature extraction methods and types of images.

ods succeed to register these images. In cases with high deviation between the spectral bands (in which SURF, SIFT and upright SURF failed with the registration), the RFAE method performs better for registration as shown in Figure 14. The RFAE method shows the highest improvement for images of human bodies. Figure 15 depicts a sample of such a thermal and visual image

	SIFT	SURF	U-SURF	RFAE	SURF+ RFAE
Satellite (low deviation)	72%	66%	71%	53%	68%
Satellite (high deviation)	45%	41%	49%	46%	52%
Human	0%	0%	1%	8%	8%
Surveillance	1%	3%	4%	18%	19%
UAV	9%	10%	1%	6%	11%

Table 3: Average Q values based on different feature extraction methods and types of images.

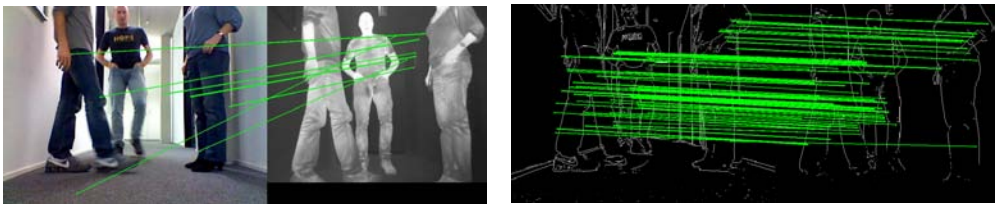


(a) Failed registration using SURF. (b) Successful registration using RFAE.

Figure 14: Registration between bands 1 and 4 of the Landsat satellite image of Iowa state (Source: NASA/USGS).

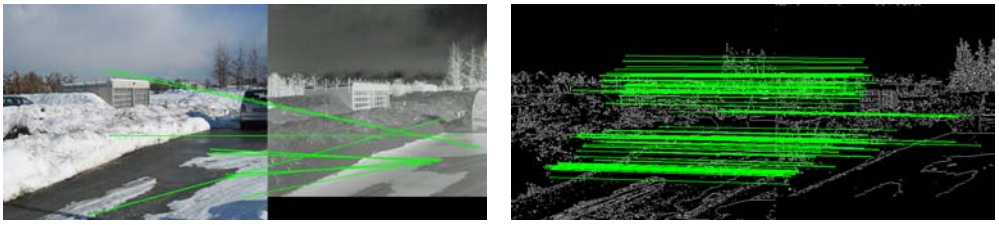
registration. A similar improvement can be seen for surveillance scenarios (cp. Figure 16). The interspectral registration of low-altitude aerial images has turned out to be more challenging. Whenever a pair of aerial images does not share enough mutual edge information, the RFAE method shows a weak performance. However, in cases with more mutual edge information (such as Figure 17) RFAE dominated the other methods. We therefore combine the RFAE and SURF methods and choose the feature extraction method with highest Q value for the registration of the images. As shown in the last column of Table 2, this combination chooses the best result among RFAE and SURF and achieves the best overall registration performance.

Table 3 shows the average Q values of the same dataset used in Table 2. Image pairs with high correlation (e.g., most of the satellite images) show a high Q value. For these images, the success rate of registration by using standard methods is higher rather than RFAE. The reason is that standard methods are able to extract more detailed features as compared to



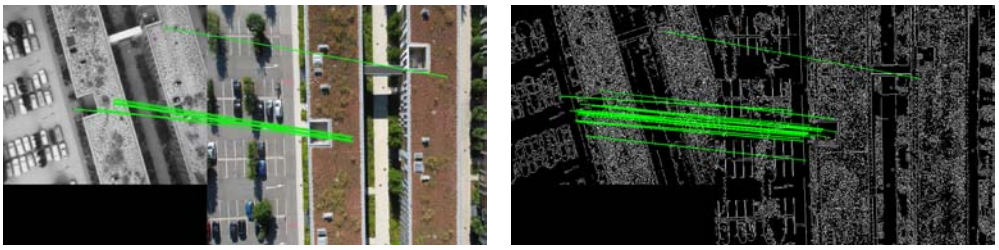
(a) Failed registration using SURF. (b) Successful registration using RFAE.

Figure 15: Registration of thermal and visual images of humans.



(a) Failed registration using SURF. (b) Successful registration using RFAE.

Figure 16: Registration of thermal and visual surveillance images.



(a) Failed registration using SURF. (b) Successful registration using RFAE.

Figure 17: Registration of thermal and visual aerial images taken from low-altitude UAV.

RFAE which extracts merely features along the edges. On the other hand, RFAE is dominant when images have a high deviation, yet with sufficient mutual edges. The average Q values corresponding to images of surveillance or human bodies imply the better performance of RFAE. Table 2 represents mainly the interspectral registration acceptance rate while Table 3 represents the quality of extracted features for the purpose of interspectral registration.

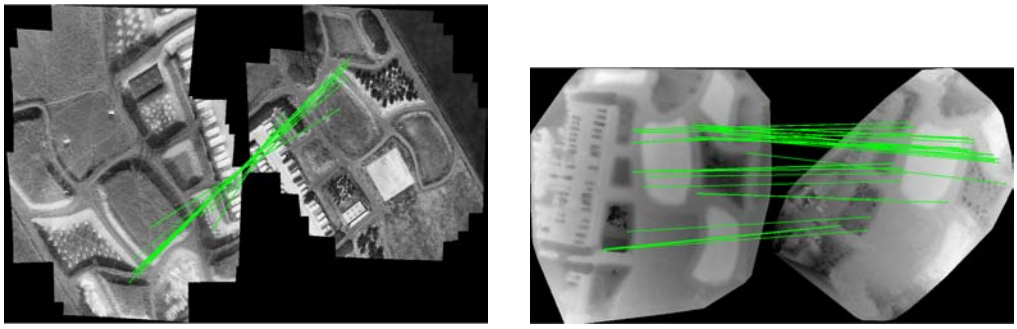
7.2. Mosaic registration

As explained in Section 6.1, we are able to register two mosaics together as long as there is one pair of images which can be registered among the entire mosaics. Although we can achieve a higher accuracy when more image pairs are registered, there is a general drawback in this type of registration. In most of the constructed mosaics there is some deformation. This can be due to the different angles of imaging from non-flat objects or some non-rigid transformations performed in mosaicking. This problem in addition to the accumulated error near the borders sometime cause some misalignment and ghosting effect in the fusion as you can see in Figure 8(b).

7.3. Depth map registration

The registration of the depth map instead of the mutual image information is often more complex and computationally expensive. However, there is an advantage in cases of aerial imagery with low temporal resolution. To demonstrate this advantage we tested the registration of aerial images taken in summer and winter (cp. Figure 8). The images taken in different seasons exhibit a very high variation. As shown in Figure 18(a), images of the same spectrum (i.e., thermal mosaic in winter and thermal mosaic in summer) can be only weakly registered by standard SURF. On the other hand, by our depth map method we were able to successfully register even the more complex scenario of interspectral registration with a high image variation over time. Figure 18(b) shows such successful registration between thermal winter mosaic and the visual summer mosaic.

In our proposed system, the visual computation is mainly done on the ground station (cp. Section7.4). Despite this fact, our sparse depth map construction and registration can be done in real-time onboard the UAVs. The more computationally expensive tasks cannot be performed onboard. The future work may include the implementation of advanced depth-map construction for real-time registration (Banz et al., 2010; Hirschmuller, 2005; McKinnon et al., 2012). The advantage of semi global matching is that the output depth-map is already aligned with the image plane. However so far these methods are focused on stereo or trinocular images (Heinrichs et al., 2007) and an orthorectified depth-map mosaic based on multiple images is challenging and time consuming.



(a) Weak registration between thermal mosaics in summer and winter (same spectrum).

(b) Registration between thermal depth map in winter and visual depth map in summer (interspectral).

Figure 18: Registration between two set of aerial images taken in different seasons.

7.4. Further discussion

Although our experiments have been carried out using small-scale UAVs, our results and methods may be extended to larger UAVs. As long as it is possible to plan a flight mission with identical picture-points, the acquired aerial images will be the same and the same methods are applicable. Additionally, larger UAVs with more accurate sensors have the advantage of providing more reliable metadata regarding flight behavior, in turn reducing the search space for finding appropriate mosaicking parameters. This may by crucial for mosaicking areas with few features or correspondences such as water, desert, or highly dynamic scenes.

For disaster management and aerial monitoring images must be transmitted to the ground station immediately. Hence, in our proposed system, most of the image processing is done on the ground. This also avoids the bottleneck of transmitting large mosaics with limited bandwidth. However, performing the visual computation onboard is useful in scenarios where the UAVs are required to follow a target or cooperate autonomously. Thus we tested the onboard registration of a pair of image (Intel Atom $1.6\,GHz$ processor on AscTec Pelican) and ascertained that on average it is twice as slow as registration on ground stations (standard PC running at $2.66\,GHz$). Both cases are feasible in real-time, since the UAVs require around 5 seconds to move between picture points. In our work, several tasks are executed onboard, such as picture-point navigation, image compression and prioritized

data transmission (Wischounig-Strucl et al., 2011). The complexity of performing these tasks in real-time is within the limits of onboard processing power, but a detailed description is beyond the focus of this paper.

On average, using Matlab on a standard PC running at $2.66\,GHz$, we were able to perform a single-pair registration of 1 megapixel images in $2\,s$. The computational complexity of the interspectral registration between two mosaics (considering each mosaic from n individual images) is approximately n times more than a single-pair registration. The registration of the mosaics in Figure 8 was performed in $57\,s$. The computation time for depth map registration based on disparity vectors (cp. Figure 9) is on average 4 times that of an equivalent single-pair registration. This is because we need to register two pairs of images in order to construct a pair of depth maps, resulting in 3 registrations and 2 depth map constructions. However, the computational complexity of depth map registration based on a full bundle adjustment is much higher. For example, the registration shown in Figure 11 required about 2 hours.

8. Conclusion

In this paper, we introduce methods for robust interspectral image registration for real-time monitoring. The methods are applicable for pairs of images acquired at different wavelengths, however we mainly focused on thermal and visual image registration. First we presented a general method (RFAE) which exploits the existing scale-invariant feature extraction methods such as SIFT and SURF in order to extract the robust features along the edges. Based on experimental results, our approach improves interspectral registration noticeably. Second, we proposed two methods for increasing the robustness of image registration and extracting additional features in cases when more than one pair of images is available. The latter scenario was studied with a focus on thermal and visual aerial images acquired by low-altitude UAVs. In scenarios involving multiple image pairs, image mosaics may be used for interspectral registration or depth maps of a target scene may be used for the feature extraction.

Acknowledgment

This work was performed in the project *Collaborative Microdrones (cDrones)* of the research cluster Lakeside Labs and was partly funded by the Euro-

pean Regional Development Fund, the Carinthian Economic Promotion Fund (KWF), and the state of Austria under grant KWF-20214/17095/24772.

We would like to thank Daniel Wischounig-Strucl and Vera Mersheeva for their help in the preparation of the thermal and visual dataset.

Banz, C., Hesselbarth, S., Flatt, H., Blume, H., Pirsch, P., 2010. Real-time stereo vision system using semi-global matching disparity estimation: Architecture and fpga-implementation, in: Proceedings of International Symposium on Systems, Architectures, Modeling, and Simulation (SAMOS), Samos, Greece. pp. 93–101.

Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. SURF: Speeded-up robust features. Computer Vision and Image Understanding (CVIU) 110, 346–359.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8, 679–698.

Coiras, E., Santamaria, J., Miravet, C., 2000. Segment-based registration technique for visual-infrared images. Optical Engineering 39, 282–289.

Du, Q., Raksuntorn, N., 2008. Automatic registration and mosaicking for airborne multispectral image sequences. Photogrammetric Engineering & Remote Sensing 74, 169–181.

Dufournaud, Y., Schmid, C., Horaud, R., 2000. Matching images with different resolutions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hilton Head, SC, USA. pp. 612–618.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24, 381–395.

Flöry, S., Hofer, M., 2010. Surface fitting and registration of point clouds using approximations of the unsigned distance function. Computer Aided Geometric Design 27, 60–77.

Fonseca, L.M.G., Costa, M.H.M., 1997. Automatic registration of satellite images, in: Proceedings of Brazilian Symposium on Computer Graphics and Image Processing, Campos do Jordão, SP, Brazil. pp. 219–226.

Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1362–1376.

Grompone von Gioi, R., Jakubowicz, J., Morel, J.M., Randall, G., 2012. LSD: A line segment detector. Image Processing On Line 2012, 1–19.

Gonzalez, R.C., Woods, R.E., 1992. Digital image processing. Addison-Wesley Publishing Company, Reading, Mass., Menlo Park, Calif., New York.

Hansen, B.B., Morse, B.S., 1999. Multiscale image registration using scale trace correlation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Ft. Collins, CO, USA. pp. 202–208.

Harris, C., Stephens, M., 1988. A combined corner and edge detector, in: Proceedings of Fourth Alvey Vision Conference, Manchester, UK. pp. 147–151.

Heinrichs, M., Rodehorst, V., Hellwich, O., 2007. Efficient semi-global matching for trinocular stereo. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (3/W49A), 185–190.

Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA. pp. 807–814.

Hong, G., Zhang, Y., 2005. The image registration technique for high resolution remote sensing image in hilly area. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (8/W27), (on CDROM).

Irschara, A., Zach, C., michael Frahm, J., Bischof, H., 2009. From structure-from-motion point clouds to fast location recognition, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR, Miami, Florida, USA. pp. 2599–2606.

Istenic, R., Heric, D., Ribaric, S., Zazula, D., 2007. Thermal and visual image registration in Hough parameter space, in: Proceedings of 14th International Workshop on Systems, Signals and Image Processing (IWSSIP), Maribor, Slovenia. pp. 106–109.

Joo, J., Choi, J., Cho, D., 2003. Robust registration in two heterogeneous sequence images on moving objects, in: Proceedings of Sixth International Conference of Information Fusion, Cairns, Queensland, Australia. pp. 277–282.

Kern, J., Pattichis, M., 2007. Robust multispectral image registration using mutual-information models. IEEE Transactions on Geoscience and Remote Sensing 45, 1494–1505.

Kim, H., Kim, M.G., 2012. Image registration using terrain relief correction based on the rigorous sensor models. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39-B1, 235–238.

Klein, A., Kroon, D.J., Hoogeveen, Y., Schultze Kool, L.J., Renema, W.K.J., Slump, C.H., 2011. Multimodal image registration by edge attraction and regularization using a b-spline grid, in: Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE), Orlando, Florida, USA. pp. 7962:20–7962:28.

Kong, S.G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B.R., Koschan, A., Yi, M., Abidi, M.A., 2006. Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. International Journal of Computer Vision 71, 215–233.

Lee, S., 2010. A coarse-to-fine approach for remote-sensing image registration based on a local method. International Journal on Smart Sensing and Intelligent Systems 3, 690–702.

Lin, H., Du, P., Zhao, W., Zhang, L., Sun, H., 2010. Image registration based on corner detection and affine transformation, in: Proceedings of 3rd International Congress on Image and Signal Processing, Yantai, China. pp. 2184–2188.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110.

Mahdi, H., A. Farag, A., 2002. Image registration in multispectral data sets, in: Proceedings of International Conference on Image Processing (ICIP),, IEEE, Rochester, NY, USA. pp. 369–372.

McKinnon, D., Smith, R.N., Upcroft, B., 2012. A semi-local method for iterative depth-map refinement, in: Proceedings of International Conference on Robotics and Automation (ICRA), IEEE, St. Paul, MN, USA. pp. 758–763.

Meltzer, J., Soatto, S., 2008. Edge descriptors for robust wide-baseline correspondence, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, USA. pp. 1–8.

Mikolajczyk, K., Schmid, C., 2004. Scale & affine invariant interest point detectors. International Journal of Computer Vision 60, 63–86.

Morel, J.M., Yu, G., 2011. Is SIFT scale invariant? Inverse Problems and Imaging (IPI) 5, 115–136.

Müller, M., Heidelberger, B., Teschner, M., Gross, M., 2005. Meshless deformations based on shape matching. ACM Transactions on Graphics 24, 471–478.

Quaritsch, M., Kruggl, K., Wischounig-Strucl, D., Bhattacharya, S., Shah, M., Rinner, B., 2010. Networked UAVs as aerial sensor network for disaster management applications. Elektrotechnik und Informationstechnik (e&i) 127, 56–63. Springer.

Quaritsch, M., Kuschnig, R., Hellwagner, H., Rinner, B., 2011. Fast aerial image acquisition and mosaicking for emergency response operations by collaborative UAVs, in: Proceedings of 8th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2011), Lisbon, Portugal. pp. 1–5.

Quaritsch, M., Stojanovski, E., Bettstetter, C., Friedrich, G., Hellwagner, H., Rinner, B., Hofbauer, M., Shah, M., 2008. Collaborative microdrones: Applications and research challenges, in: Proceedings of the Second International Conference on Autonomic Computing and Communication Systems, ACM Press, Turin, Italy. pp. 38:1–38:7.

Schaefer, G., Tait, R., Howell, K., Hopgood, A., Woo, P., Harper, J., 2008. User Centered Design for Medical Visualization. IGI Global. chapter Automated overlay of infrared and visual medical images. pp. 174–183.

Schilcher, U., Gyarmati, M., Bettstetter, C., Chung, Y.W., Kim, Y.H., 2008. Measuring inhomogeneity in spatial distributions, in: Proceedings of Vehicular Technology Conference (VTC Spring), Marina Bay, Singapore. pp. 2690–2694.

Sehgal, A., Cernea, D., Makaveeva, M., 2010. Real-time scale invariant 3d range point cloud registration., in: Proceedings of International Conference on Image Analysis and Recognition (ICIAR), Springer, Povoa de Varzim, Portugal. pp. 220–229.

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3d. ACM Transactions on Graphics 25, 835–846.

Song, Z., Cheng, X., 2010. A new search engine filtering scheme based on improved neural network and ontology, in: Peoceedings of International Conference on Computational and Information Sciences, IEEE, Chengdu, Sichuan, China. pp. 178–181.

Szeliski, R., 2010. Computer Vision: Algorithms and Applications. 1st ed., Springer-Verlag New York, Inc., New York, NY, USA.

Teke, M., Temizel, A., 2010. Multi-spectral satellite image registration using scale-restricted SURF, in: Proceedings of 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey. pp. 2310–2313.

Tingdahl, D., Van Gool, L., 2011. A public system for image based 3d model generation, in: Proceedings of the 5th international conference on Computer vision/computer graphics collaboration techniques, Springer-Verlag, Rocquencourt, France. pp. 262–273.

Vaidehi, V., Ramya, R., Prasannadevi, M., Naresh Babu, N.T., Balamurali, P., Girish Chandra, M., 2011. Fusion of multi-scale visible and thermal images using EMD for improved face recognition, in: Proceedings of International MultiConference of Engineers and Computer Scientists, Hong Kong. pp. 543–548.

Šegvic, S., 2005. A multimodal image registration technique for structured polygonal scenes, in: Proceedings of Image and Signal Processing and Analysis (ISPA), Zagreb, Croatia. pp. 500– 505.

Wischounig-Strucl, D., Quartisch, M., Rinner, B., 2011. Prioritized data transmission in airborne camera networks for wide area surveillance and image mosaicking, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Colorado Springs, CO, USA. pp. 17 –24.

Yahyanejad, S., Misiorny, J., Rinner, B., 2011a. Lens distortion correction for thermal cameras to improve aerial imaging with small-scale UAVs, in: Proceedings of IEEE International Symposium on Robotic and Sensors Environments (ROSE), Montreal, QC, Canada. pp. 231–236.

Yahyanejad, S., Quaritsch, M., Rinner, B., 2011b. Incremental, orthorectified and loop-independent mosaicking of aerial images taken by micro UAVs, in: Proceedings of IEEE International Symposium on Robotic and Sensors Environments (ROSE), Montreal, QC, Canada. pp. 137–142.

Yahyanejad, S., Wischounig-Strucl, D., Quaritsch, M., Rinner, B., 2010. Incremental mosaicking of images from autonomous, small-scale UAVs, in: Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE Computer Society, Boston, MA, USA. pp. 329–336.