remote sensing

MDPI

Article

Online Correction of the Mutual Miscalibration of Multimodal VIS–IR Sensors and 3D Data on a UAV Platform for Surveillance Applications

Piotr Siekański *, Sławomir Paśko, Krzysztof Malowany and Marcin Malesa

Institute of Micromechanics and Photonics, Warsaw University of Technology, 8 Św. A. Boboli St., 02-525 Warsaw, Poland; s.pasko@mchtr.pw.edu.pl (S.P.); k.malowany@mchtr.pw.edu.pl (K.M.); m.malesa@mchtr.pw.edu.pl (M.M.)

* Correspondence: p.siekanski@mchtr.pw.edu.pl

Received: 9 September 2019; Accepted: 21 October 2019; Published: 23 October 2019

Abstract: Unmanned aerial vehicles (UAVs) are widely used to protect critical infrastructure objects, and they are most often equipped with one or more RGB cameras and, sometimes, with a thermal imaging camera as well. To obtain as much information as possible from them, they should be combined or fused. This article presents a situation in which data from RGB (visible, VIS) and thermovision (infrared, IR) cameras and 3D data have been combined in a common coordinate system. A specially designed calibration target was developed to enable the geometric calibration of IR and VIS cameras in the same coordinate system. 3D data are compatible with the VIS coordinate system when the structure from motion (SfM) algorithm is used. The main focus of this article is to provide the spatial coherence between these data in the case of relative camera movement, which usually results in a miscalibration of the system. Therefore, a new algorithm for the detection of sensor system miscalibration, based on phase correlation with automatic calibration correction in real time, is introduced.

Keywords: UAVs; calibration; multimodal sensors; thermovision; camera; surveillance

1. Introduction

Drones are used in inspection and security systems because they are able to quickly monitor large areas, work autonomously, and are low cost compared to traditional human inspection. Drone-mounted cameras operating in the visible band are commonly used in these systems because they provide information that can be easily interpreted, either by a human operator or automatically, using computer vision algorithms. However, this approach may be insufficient in the case of systems used for monitoring critical infrastructure. These systems must be able to detect threats under adverse conditions (e.g., at night when there is not enough light to detect objects) or hidden objects (e.g., persons hiding under the trees who cannot be viewed from above by an ordinary camera). However, because of the differences in thermal radiation, they may be distinguished in the infrared band [1]. 3D data may be helpful in the detection of a relative change in the geometric parameters of the objects. For instance, masked objects, such as sheds covered by sand or snow, may not be detected using an ordinary visible (VIS) camera; however, they are detectable using 3D data [1]. Therefore, we propose a fusion of data, including color imaging (VIS), thermal imaging (IR), and 3D data (Figure 1). To maintain spatial data coherence, the presented system must be as resistant as possible to system miscalibration, which is caused by device vibrations that occur during flight or result from possible impacts of the multimodal head with the ground during unmanned aerial vehicle (UAV) landing. To meet these requirements, we have developed a procedure for the mutual calibration of infrared (IR) and VIS sensors using a

single calibration artefact, and thanks to the use of the SfM method [2] to obtain 3D data offline, we can be sure that these data will also be consistent with the adopted coordinate system. The multimodal system miscalibration detection procedure and calibration correction algorithm work on the drone in real time.

(a) (b) (c)

Figure 1. Multimodal data: (**a**)—visible (VIS) image; (**b**)—infrared (IR) image; and (**c**) 3D data in the form of a point cloud.

2. Related Work

2.1. IR–VIS Image Registration

Combining multimodal IR and VIS images with 3D data is normally used in the analysis of data collected by drones. It is common practice to create a georeferenced orthophoto map from VIS images and to map IR data to VIS using GPS [3–5]. In some cases, the GPS position may not be accurate enough to properly register the images or, because of interference, may not be available at all [3]. The registration of IR and VIS data can be improved using well-defined markers (i.e., ground control points, GCPs) that are detectable in both modalities and whose positions must be accurately determined in space [4]. This approach, like the previous one, requires an orthophoto map to be generated beforehand, which makes it unsuitable for online use.

In the absence of GCP, image analysis algorithms for VIS and IR images may be used to determine the transformation between them. This is a demanding task because, in general, the cameras used for imaging have different focal lengths, aberrations, resolutions, and spectral ranges and observe the scene from a slightly different perspective. Multiple methods exist for matching multimodal images. They can be combined by analyzing the spectra of both images, after transforming them to the frequency domain, or by directly analyzing the features visible in both images. In the first case, we can distinguish between approaches based on the Fourier transform [6,7] and those based on the wavelet transform [8]. Image-based methods are used much more often to register images from IR and VIS cameras [9]. Approaches based on feature points [10], lines [9], or regions [11] can be distinguished. It is often necessary to combine more than one method for extracting the features in images. Yahyanejad and Rinner [12] proposed an analysis of feature points extracted from the edges of images in both modalities, and Li et al. [13] used feature points extracted from both images using the SIFT (Scale Invariant Feature Transform) algorithm [14] and phase correlation to search for correspondences. None of the above methods takes into account the image aberration caused by the low-quality optics used in drone-mounted cameras, although according to [15], the correction of aberrations, before combining images, has a positive effect on the registration of multimodal images from various devices.

Distortion correction of both lenses is possible thanks to the prior determination of the internal parameters of both cameras. If this is conducted using the same calibration target, it is possible to determine the position of both cameras in relation to each other and to find a common field of view for both cameras.

2.2. Calibration of Multimodal VIS–IR

Camera calibration is the process of determining a mathematical model that describes the course of rays in a camera's optical system for each pixel of its detector. The most common approximation of a real optical system is the pinhole model, in which all rays converge at one point. The calibration procedure is usually carried out, due to the simplicity of the method, using a flat pattern, according to the method proposed by Zhang [16]. A checkerboard pattern or a set of circular markers is commonly used to calibrate the camera's optical system.

In the case of the simultaneous calibration of two cameras with different spectra (IR + VIS), it is necessary to use a pattern with markers visible in both modalities. This prevents the use of standard patterns dedicated to calibrating VIS cameras because the markers on them are invisible in the IR band. Markers, in order to be visible in the IR band, must have a different emissivity factor or a different temperature than the rest of the target. Multimodal camera calibration methods can be divided into two groups in terms of the need for external target heating: passive and active.

Passive targets for multimodal calibration require a combination of two materials with different emissivity coefficients, which requires the combination of two different materials to build the target (e.g., plastic > 0.9 and aluminum < 0.03). In many cases, the combination of materials is conducted manually, for example, by sticking a pattern made of paper with holes onto an aluminum surface [17] or applying markers made of self-adhesive foil to an aluminum surface [18]. These approaches make it impossible to achieve a mounting accuracy greater than 0.5 mm [19]. To overcome this problem, Usamentiaga et al. [19] proposed printing the markers on a plate made of Dibond®, a composite consisting mainly of aluminum. The markers are visible in the IR spectrum because of the difference in emissivity between the target material and the ink. Another possible solution is to use a checkerboard made of two types of synthetic resin characterized by different emissivity levels [20]. The disadvantage of all passive solutions is that materials characterized by a low emissivity are, at the same time, characterized by a high reflectivity; therefore, they reflect the radiation of the environment. For this reason, they can only be used under strictly controlled conditions. For indoors, in most cases, the pattern needs to be heated to increase the contrast in the calibration images. However, under external conditions, natural solar radiation is often sufficient. It should be ensured that during calibration there are no sources of heat radiation or other objects in the environment, such as trees or buildings, whose heat images could reflect off the surface of the pattern [18].

Active multimodal calibration methods can be divided into two groups according to whether they require a continuous supply of thermal energy or only temporary heating of the target and rapid calibration. The second approach requires the immediate collection of calibration photos, as the contrast of IR images decreases instantly as the pattern cools. For this reason, the calibration performed may be inaccurate after a few seconds [21]. In 2015, Saponaro et al. [22] suggested placing a checkerboard, printed on sheet of paper, on a ceramic surface and heating the surface with a heating lamp. Their approach extended the time of calibration to several minutes, but it was tested only indoors under strictly controlled conditions. Systems with tungsten bulbs are also commonly used to calibrate multimodal systems because they emit radiation that is visible in both modalities [23,24]. The disadvantage of this approach is the fact that it is impossible to determine the exact position of the bulb with a high accuracy based on the image [18].

2.3. In Situ Calibration Correction

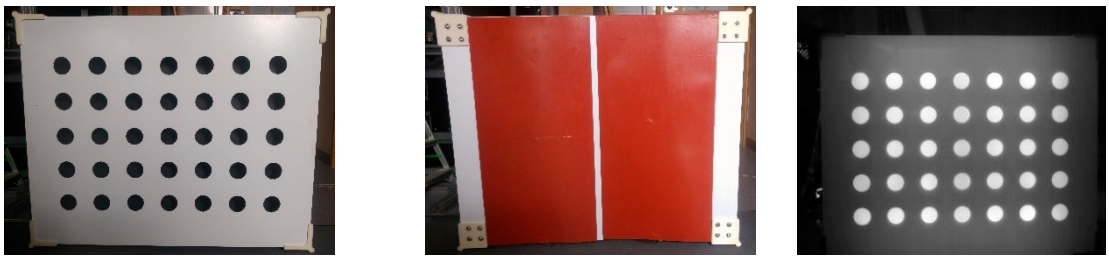
As rightly noted in [25], a multimodal system mounted on a movable platform can become uncalibrated over time (i.e., drift may occur between the previously determined camera positions), resulting in a lack of spatial synchronization between images. In addition, due to temperature differences, accelerations, vibrations, and so on, the camera system may move, which, additionally, necessitates the provision of an active compensation mechanism for such phenomena during flight. The cited publication proposes an algorithm based on the use of the Gaussian image pyramid in both modalities and image mapping using polynomials.

Miscalibration detection can be reduced to the problem of the image registration between a pair of VIS and IR images, but as noted in [12], analyzing image sequences in each modality separately may produce better results. In the same publication, the authors proposed two other methods for detecting the mutual position of both sensors: one based on combining multiple images and another based on depth maps. The first approach requires the computation of an orthophoto map of the whole area in both modalities, which requires the use of a computationally expensive bundle adjustment algorithm [26], so it cannot be used online. The second approach is based on computing depth maps between consecutive pairs of images and the subsequent determination and matching of characteristic points between them. However, this method may work only if the area over which the flight takes place is characterized by elements of different heights that are distinguishable in both modalities.

3. Materials and Methods

3.1. Design of the Calibration Target

The developed target for multimodal calibration consisted of two parallel composite panels made of Dibond®with dimensions of 1000 by 800 by 3 mm, which were placed in parallel at a distance of 80 mm. The plate located closer to the camera was white, and 35 circular holes with diameters of 60 mm were milled therein, showing a second black plate (Figure 2a). Thanks to this, the image from the VIS camera showed a very high contrast between the areas representing a bright front plane and those representing the rear plane, which allowed for accurate detection of the center of the circles in this band.



(a) (b) (c)

Figure 2. Calibration target: (**a**) frontal view—holes that are markers for both modalities are visible; (**b**) rear view—two red heating mats are visible; and (**c**) view of the pattern in the IR band, after warming up the pattern.

The separation of boards is important from the point of view of infrared imaging. The backplate was heated using two electric heating mats (Figure 2b), with a power of 800 W each, enabling the heating of the backplate without changing the temperature of the front plate. The established temperature difference is clearly visible in the pictures, recorded by the IR camera (Figure 2c). The proposed construction enabled the simultaneous detection of markers, both in the IR band and in the VIS band.

3.2. IR–VIS Image Registration Procedure

To input the data from two cameras into one coordinate system and determine the area observed by both cameras, stereo calibration was performed. Stereo calibration is the process of determining the relative position of a pair of detectors. Knowing the relative transformation between the coordinate systems of both cameras and their internal parameters, it is possible to determine the intersection points of rays from corresponding pixels and, thus, determine the depth in the scene. Stereo calibration of the camera system is most often performed by taking a series of photos that show the same pattern seen by both cameras. Based on information about the corresponding pixels (the same points on the pattern), it is possible to determine the mutual positions of the pair of detectors in the same coordinate system.

Because of the noise appearing in the IR images, the calibration process was divided into two stages: geometric calibration of each camera and stereo calibration. In the first stage, 40 pairs of images of the calibration target in different orientations were acquired. Then, the centers of the holes were determined using the blob detection algorithm, implemented in the OpenCV library [27]. After that, each camera was calibrated using the Zhang algorithm [16], implemented in the same library. As a result, a matrix of the internal camera parameters (focal length, principal point) and distortion coefficients was determined. In the second stage, using the camera parameters computed in the previous step and extracting the centers of the holes, the mutual position of the cameras was determined. For this purpose, the stereo calibration algorithm, implemented in the OpenCV library, was used. The resulting coordinate system was hooked at the nodal point of the VIS camera and followed the axis direction of the local camera system. After stereo calibration, the transformation between the VIS and IR cameras was known.

The relationship between VIS and 3D data was determined by the information generated by the SfM algorithm. The transformation of the IR camera, relative to the VIS determined earlier during calibration, was used to map images to each other and to the point cloud. As a result, all modalities of the system (i.e., IR–VIS–3D) were interrelated and placed in one coordinate system.

3.2.1. Perspective Transformation

(1) Homography

Homography is a geometric operation that transforms any plane into another plane using projective transformation. It requires a minimum of 4 pairs of points to determine the homography between two planes. Knowing the corresponding points from stereo calibration, it is possible to determine a homographic transformation that minimizes the mean square error between the corresponding pairs of points. Then, this homography may be used to map the IR image to the VIS.

The first step of a mapping procedure is the calculation of the positions of the corners belonging to the IR image, mapped to the VIS. If these corners are outside the VIS area, the IR image is cropped accordingly. Then, the VIS image is cropped to the area occupied by the IR image. This approach allows the system to use the full area observed by both cameras.

Because homography transforms a plane into a plane, it can be used only to map photos from a drone flying at a high altitude, so that the height of the objects on the ground is small in relation to the flight altitude of the drone [28]. This assumption is not fulfilled during calibration, in which the camera is located just a few meters from the target. The second limitation of this approach is the fact that the feature points of the pattern in the images may not be distributed approximately evenly on the image plane, which may cause the determined homography to be inaccurate.

(2) Ray–Plane Intersections

A virtual plane P representing the ground level was set. This plane was parallel to the VIS camera sensor and away from it, according to the drone's flying altitude. Then, for each pixel $p(x, y)$ of a VIS camera, a ray $r(x, y)$ passing through the nodal point of the VIS camera and $p(x, y)$ was formed. Then, the intersection point between $r(x, y)$ and P was calculated. As a result, a set of 3D points was obtained. Next, the 3D points were projected onto the IR camera matrix, which was shifted relative to the VIS camera by the transformation determined by stereo calibration. Finally, the area covered by the IR image was selected from the VIS image, and both images were cropped accordingly (Figure 3).

(3) Depth Map

A depth map is an image in which the depth of the scene in each pixel is encoded. If the 3D geometry of the scene is known, it is possible to generate depth maps that correspond to the images used to create the model using the SfM technique.

The algorithm mentioned in the previous section may also use depth maps, with a single modification. Instead of examining the point of intersection between the ray and the plane, the point

lying at a distance from the nodal point of the camera that is equal to the depth at a given point was used.

The drawback of this approach is the fact that 3D reconstruction using SfM technology and related depth map generation cannot be determined from a single image and additionally requires a high computing power and time. Therefore, this approach may only be used to process data offline.

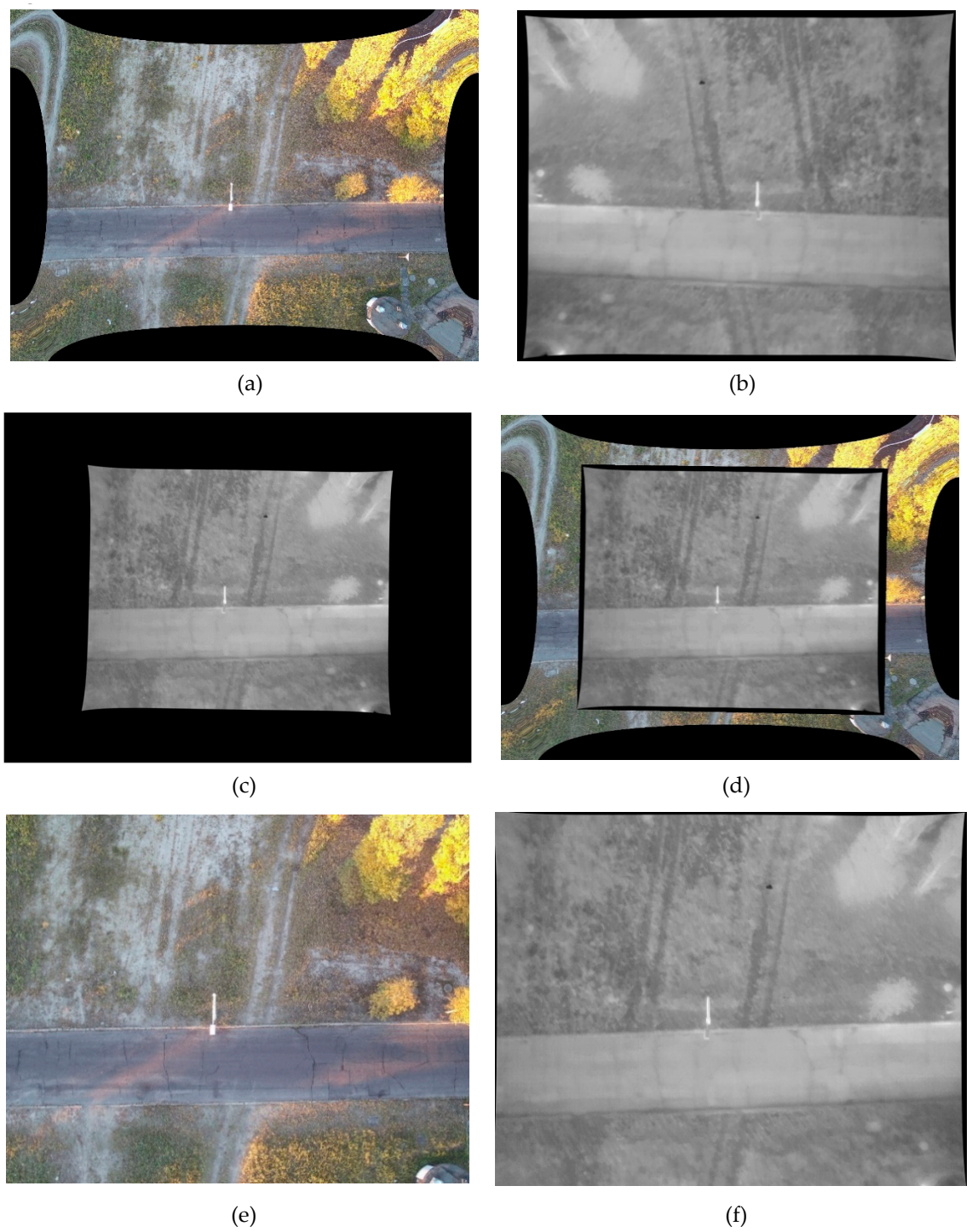


Figure 3. Mapping of a sample photo: (**a**) VIS image with distortion removed; (**b**) IR image with distortion removed; (**c**) IR image resized to VIS; (**d**) IR image superimposed on the VIS; (**e**) cropped VIS image; and (**f**) cropped IR image.

3.3. Online Miscalibration Detection and Correction

It is assumed that as long as the positions of both cameras remain unchanged, the system is calibrated. However, a precalibrated system may become uncalibrated during flight due to numerous factors. If this occurs, the operator should be notified. Moreover, the device should try to perform calibration correction automatically to ensure that the images collected in different modalities are correctly superimposed. Therefore, a new method for determining the relative movement between multimodal cameras has been developed. If the developed algorithm detects the relative movement of the individual sensors, the calibration correction procedure is initiated. The algorithm has been divided into two separate stages:

- detection of miscalibration between cameras, and
- mutual calibration correction.

3.3.1. Miscalibration Detection Method

After the mapping process, multimodal images are registered. However, it is crucial to find a relative motion between them to detect miscalibration. A direct comparison of multimodal images may be challenging. Therefore, an image-filtering method for both modalities (IR and VIS) was used, which facilitates a comparison of both images, and on this basis it is possible to detect system miscalibration. Preprocessing was performed on each pair of images. Three methods of image preprocessing were tested: histogram equalization, gradient detection using a Sobel operator, and edge detection using the method proposed by Canny [29]. In the case of the latter, the thresholds were selected in accordance with [30]: the binarization threshold was determined using the Otsu method, the upper threshold was set to this value, and the lower one was set to half of it. Figure 4 presents a comparison of the available methods of image preprocessing.

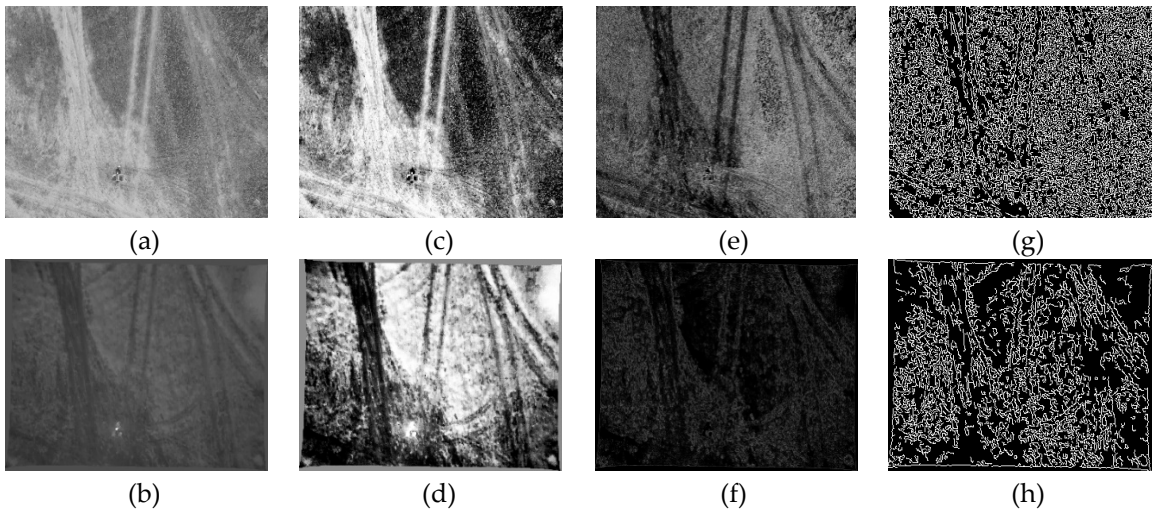


Figure 4. Available methods of image preprocessing for determining miscalibration: (**a**) VIS image converted to grayscale; (**b**) IR image; (**c**) VIS image after histogram equalization; (**d**) IR image after histogram equalization; (**e**) VIS image after applying a Sobel filter; (**f**) IR image after applying a Sobel filter; (**g**) VIS image after applying Canny edge detection; and (**h**) IR image after applying Canny edge detection.

In the registered pair of images, the shift vector (x, y) between them was computed. If its length was higher than the threshold, miscalibration was detected. It has been assumed that if the transformation between images can be described using homography (scale difference introduction, rotation in the plane and out of plane, etc.), the determined translation will certainly be significant. Image shifts can be detected by both the ECC (Enhanced Correlation Coefficient) algorithm [31], which relies on image

correlation, or based on a Fourier transform (phase correlation). The ECC algorithm is able to find a correct perspective transformation between images at a high computational cost and is currently not suitable for online applications. To translate images into one another, one can limit the degrees of freedom of the matrix calculated by the ECC algorithm to include translation only. In addition, the performance of this algorithm depends on the quality of the edges in the image, and those of the IR images are often blurred, which means that the algorithm does not guarantee that the images will be registered correctly. On the other hand, phase correlation enables the precise determination of the relative image shift. However, to determine the rotation and the scale change of the images, one must switch from Cartesian coordinates to log-polar coordinates. This approach is often used to determine the transformation between the image and its shifted copy [7]. Despite many attempts, it was not possible to obtain satisfactory results on multimodal data. Both approaches have their limitations and can be used for the detection of miscalibration, but not for active calibration correction.

3.3.2. Multimodal Calibration Correction

Since accurate image registration between multimodal images is challenging, we proposed an approach based on extracting feature points in subsequent IR and VIS images, that is, IR and VIS images were not compared with each other, but rather the current IR image was compared with the previous IR image, and similarly, the current VIS image was compared with a previous one separately. Then, in each pair, the SIFT feature points were extracted, and their descriptors were matched. After that, the optimal transformation in the matrix form between subsequent images was determined. The RANSAC (Random Sample Consensus) algorithm [32] was used to eliminate outliers (points that were matched incorrectly). Two different types of correction matrices are available: affine transformation and perspective transformation (homography). This approach is not suitable for the first frame in the sequence, as the previous frame is not available. Thus, we can detect miscalibration between the first image pair, but we cannot automatically correct it. The diagram of the proposed autocalibration procedure is shown in Figure 5.

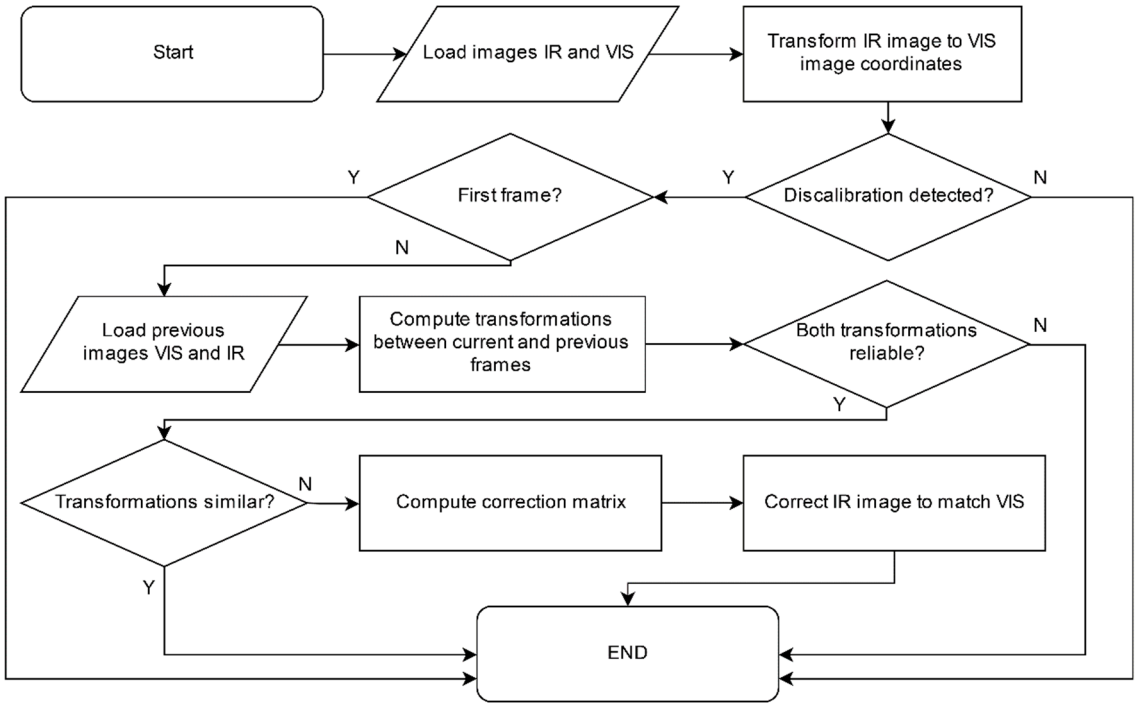


Figure 5. A block diagram of the autocalibration procedure for each pair of IR and VIS images.

(1) Assessment of the Reliability of the Determined Transformation

The transformations determined in the previous step may be erroneous (Figure 6). These transformations must be detected, and the frames must be rejected from further processing. Therefore, the criteria for assessing the reliability of the determined transformation are introduced. Regardless of the type of transformation, the determinant of the first four elements of the matrix was checked. If it was negative, the matrix introduced the mirror effect to the image, so the determined transformation could not be correct. Otherwise, the scale factor was also tested. Affine transformation is decomposed into a translation vector $\left[t_x,\ t_y\right]$, rotation angle $\propto$, and scale factor $s_x = s_y$, according to Equations (1)–(5):

$$\begin{bmatrix} a & b & t_x \\ c & d & t_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} s_x cos \propto & -s_x sin \propto & t_x \\ s_y sin \propto & s_y cos \propto & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$s_x = sign(a)\sqrt{a^2 + b^2} \quad (2)$$

$$s_y = sign(d)\sqrt{c^2 + d^2} \quad (3)$$

$$tan \propto = -\frac{b}{a} = \frac{c}{d} \quad (4)$$

$$\propto = atan2(-b, a) = atan2(c, d) \quad (5)$$

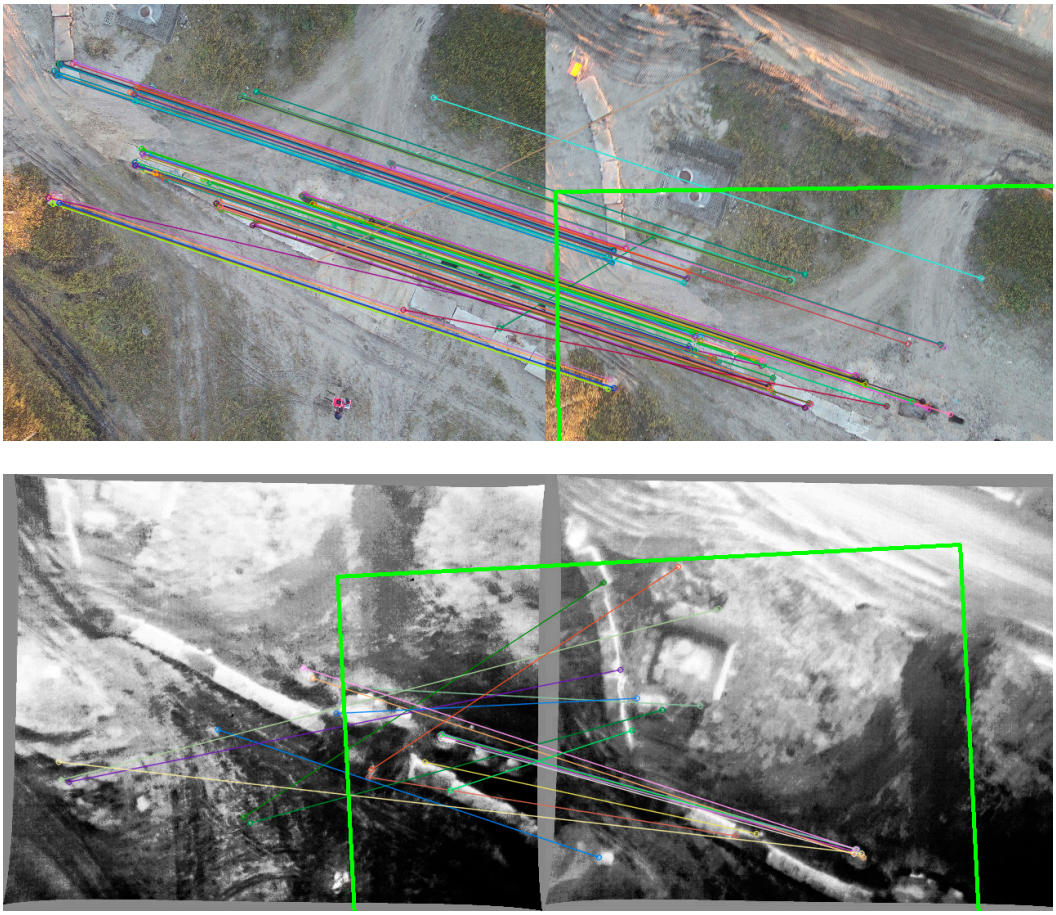


Figure 6. VIS images matched reliably (scale factor = 0.983) and IR images matched unreliably (scale factor = 1.148). A green rectangle marks the determined transformation of the left image, relative to the right one, and the colored lines show matched feature points (not all correctly).

We assumed that the scale factor should be close to 1 because the drone's flying altitude was approximately constant. If the scale factor differed from 1 by less than the set threshold (0.05 was chosen empirically), then the given transformation was marked as reliable. In the case of homography, four points were taken as the vertices of the square and were transformed by the determined homography (6).

$$p'_i = H \cdot p_i \tag{6}$$

where p_i—ith-vertex of the square, before transformation; p'_i—ith-vertex of the square, after transformation; and H—homography matrix.

In the case of perspective transformation, the determined scale factor was calculated as the ratio of the perimeters of both figures, before and after transformation. The reliability threshold of 0.10 was chosen empirically for this case. We decided to set this threshold to a higher value than in the affine transformation because perspective effects may occur that need to be taken into account. Figure 6 presents the case where one of the transformations is unreliable.

(2) Assessment of Transformation Similarity

If transformation matrices between consecutive VIS (T_{vis}) and IR (T_{ir}) images were reliable, it was necessary to check if they were similar. If so, no recalibration is required. If not, based on T_{vis} and T_{ir}, the correction matrix C is calculated according to Equation (7):

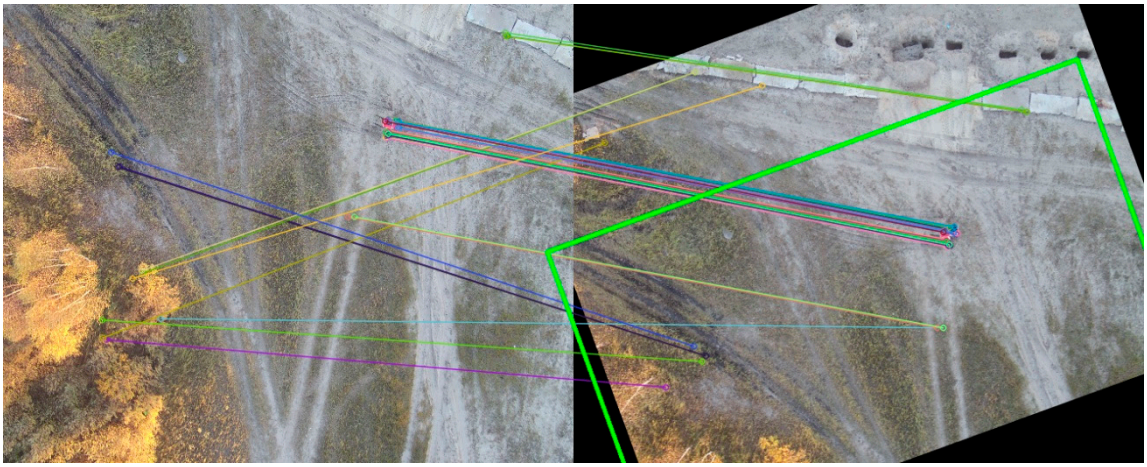
$$C = T_{vis}^{-1} \cdot T_{ir} \tag{7}$$

The T_{vis} and T_{ir} transformations were not similar if they differed in translation, rotation, or scale. Four points p, forming a square, were taken and multiplied by the determined matrix C, forming points p'. The length of the vectors between p_i and p'_i was examined. After a series of experiments, we assumed that if any of them was greater than 3% of the side of the square, the transformations were not similar. This condition ensured that if the transformation was similar, two corresponding points from both images were closer than 3% of the resolution of the images. This condition eliminated the cases when the perspective effect introduced by homography results in a large shift between the original and transformed vertices.

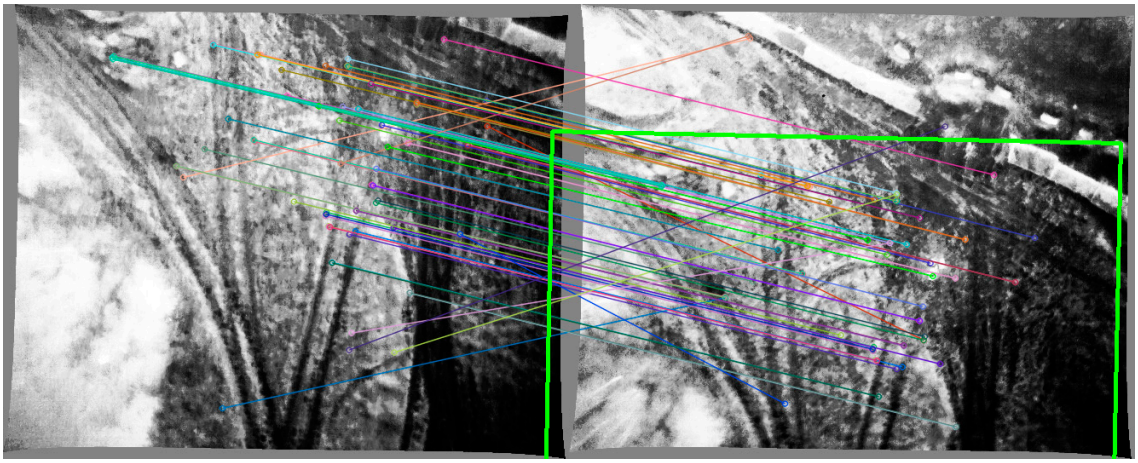
Then, the scale factor was calculated as the quotient of the areas of the square and the quadrangle, formed after transforming the vertices of the square. Similarly, if it was not in the 0.95–1.05 range, the transformations were not similar. The last step was to approximate the determined matrix with an affine transformation (ignoring the perspective coefficients in homography) and determine the rotation angle according to Equation (5). We also assumed that if it was greater than 2°, the transformations were not similar. This resulted in a maximal shift of the image corners up to 5% of the image resolution. However, this still enabled the user to correctly perceive the spatial relationships between multimodal images.

The proposed solution was designed to work on UAVs, so it should not consume energy. It should be noted that we intentionally set all of the thresholds to quite high values to invoke our correction algorithm only when it was really necessary to save energy. However, one may set lower threshold values to correct smaller misalignments between multimodal images.

If the transformations were not similar, calibration correction is required, where the IR image is multiplied by the computed correction matrix C (Figure 7). As we focused only on providing a correct registration between IR and VIS images, the calibration between cameras remained unchanged because it was used only to provide an initial registration between multimodal images. Only the IR image was transformed, relative to the VIS, so the relationship between the VIS and 3D data was not modified.

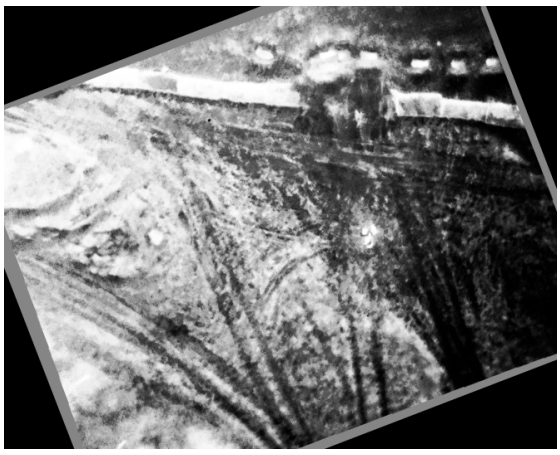


(a)



(b)

(c)



(d)

Figure 7. Example of calibration correction: (**a**) reliably matched VIS images, despite the given transformation; (**b**) reliably matched IR images; (**c**) a VIS image with a given transformation; and (**d**) the corresponding IR image, after calibration correction.

4. Experimental Results

4.1. Calibration and IR–VIS Image Mapping

Three sets of IR and VIS cameras were calibrated using the developed calibration target:

1. FLIR Duo Pro R camera, consisting of a pair of IR and VIS sensors in one case,
2. Sony Alpha a6000 camera and IR FLIR Vue Pro R camera,
3. Logitech C920 webcam and FLIR A35 camera.

In each set, the cameras were positioned so that their optical axes were approximately parallel and the distance between them was as small as possible (up to 10 cm between the optical axes of the cameras). The results are presented in Table 1. The first two sets were mounted on the drone, and the last set was checked only under laboratory conditions. However, there was a problem associated with the proper temporal synchronization between the cameras used in the second set due to hardware issues. This resulted in a random delay between the acquisition of each camera. Therefore, we only tested our method using the first setup.

Table 1. Calibration and stereo calibration results for different sets of VIS and IR cameras.

Camera	FLIR Duo Pro R (VIS)	FLIR Duo Pro R (IR)	Sony Alpha a6000	FLIR Vue Pro R	Logitech C920	FLIR A35
Resolution	4000 × 3000	640 × 512	6000 × 4000	640 × 512	1920 × 1080	320 × 256
Focal length [mm]	7.64	13.20	26.6	13.12	3.47	12.99
Principal point (x, y) [mm]	(3.80, 2.81)	(5.26, 4.40)	(11.97, 7.52)	(5.53, 4.18)	(2.37, 1.34)	(4.13, 3.89)
Reprojection error [px]	0.25	0.22	0.18	0.13	0.13	0.24
Stereo reprojection error [px]	0.48		0.79		0.78	
Translation (x, y, z) [mm]	(39.4, 1.1, −7.5)		(−46.8, −13.2, −35.9)		(−95.2, −7.1, −63.7)	
Rotation (x, y, z) [°]	(0.11, 1.77, −0.29)		(1.44, −0.09, −0.35)		(3.79, 1.50, 2.37)	

4.2. Automatic Calibration Correction

Tests of the calibration correction in-flight were performed using a dataset containing 801 pairs of mapped IR and VIS images, collected by the FLIR Duo Pro R camera. Mapped images were then checked by the operator to ensure that no frame pairs were miscalibrated. After that, in every tenth frame, a random rotation of the IR camera was added to each axis in the range of –5° to 5°. It was assumed that larger errors would be rare and would result, for example, from the collision of the drone with the ground during landing, so full sensor recalibration was required. Tests were carried out on the correctness of miscalibration detection, depending on the method, and the results are presented in Table 2. Two standard indicators, precision (8) and recall (9), were used to assess the correctness of the miscalibration detection:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (8)$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (9)$$

where *TruePositives*—the number of correctly detected miscalibrated IR–VIS image pairs; *FalsePositives*—the number of erroneously detected miscalibrated IR–VIS image pairs; and *FalseNegatives*—number of undetected miscalibrated IR–VIS image pairs.

Table 2. Comparison of different miscalibration detection methods (precision/recall).

Preprocessing Method	ECC	ECC with Histogram Equalization	Phase Correlation	Phase Correlation with Histogram Equalization
None	0.38/0.66	0.32/0.70	0.53/0.18	0.85/0.36
Gradient	0.49/0.33	0.80/0.20	**0.95**/**0.94**	**0.95**/**0.93**
Canny	0.50/0.18	0.33/0.01	0.66/**0.98**	0.77/**0.98**

The quality of the calibration correction using homography and the affine transformation was also examined on the same dataset. Table 3 presents the results of two available calibration correction methods.

Table 3. Impact of the selected transformation type on calibration correction.

Calibration Correction Method	Detected Miscalibrations	Reliable Transformations	Calibration Correction Required	Erroneously Detected the Need for Correction
Affine transform	79	79	63	0
Perspective transform	79	78	75	2

5. Discussion

The fusion of data from different sources may be beneficial in remote sensing using drones, especially in surveillance applications, where the system mounted on the drone must be able to detect threats that cannot be detected using one modality. However, it is challenging to present data from different sensors in the same coordinate system. Therefore, in this paper, we presented a method to automatically correct an IR image to make it consistent with VIS and 3D, when the IR camera moves, relative to the VIS.

The proposed calibration target enabled the simultaneous geometric calibration of both the IR and VIS cameras in the same coordinate system. Both the intrinsic and extrinsic parameters of each camera were consistent with those declared by the manufacturers. The translation values and rotation angles of each stereo-pair indicated that the relative position of the sensors was determined correctly. Small rotation angles suggested that the cameras were placed approximately parallel. In each case, the biggest translation value was in the x axis, which indicated that the cameras were positioned side-by-side. The sign of the x axis indicated that, in the case of the first set, the VIS sensor was on the right of the IR, opposite to the two other sets.

The FLIR Duo Pro R camera has an IR and VIS photo fusion algorithm implemented by the manufacturer. Unfortunately, it does not correct distortion, which, in this configuration, was clearly visible in the images, especially in those collected using a VIS camera (see Figure 3a). In addition, the correction operation requires manual adjustment of the parameters. Using the proposed calibration target, we were able to correct distortion in both modalities and automatically map images from both sensors.

Because the solution was used on drones that operated at a fairly high altitude, mapping with rays and using a depth map produced very similar results (Figure 8). Depth map generation using the SfM algorithm is a computationally demanding operation and, therefore, cannot be used online. In offline mapping, it also does not improve the results if the objects in the picture are much lower than the drone's flying altitude. Mapping using homography, computed directly during calibration, may be inaccurate because, during calibration, the distance between the sensors and the target is usually a few meters. Therefore, homography can cause errors when flying at a high altitude (Figure 8a). However, in the case of manual homography determination, the method works in a similar way to the previous two.

The phase correlation algorithm used to detect miscalibration has an advantage over the ECC algorithm, which in many cases did not find a correct translation because of the lack of sharp edges in the images. The detection of miscalibration by a Fourier transform was able to detect over 90% of miscalibrated image pairs. When using Canny's algorithm for image preprocessing, the recall factor reached 98%. However, this has been paid for by the over-sensing of the method and a significant number of false alarms. In most cases, applying histogram equalization resulted in improved results if the Fourier transform was used later to detect miscalibration. As the results of the research show, the preferred method of miscalibration detection used image gradients in both modalities and then compared them using phase correlation.

The proposed calibration correction method used SIFT features to detect the relative motion between consecutive frames in both modalities separately. IR images are often blurry, so calculating motion from features extracted from them is prone to error. To decide whether the extracted motion of each camera was reliable, the scale factor of the transformation was analyzed. If it was lower than a certain threshold, the extracted motion was marked as reliable. Setting this value to a value that is too low (e.g., 0.01) may result in unreliable transformations between frames, and as a consequence, the

algorithm would be unable to correct transformations. On the other hand, setting this threshold to a value that is too high (e.g., 0.2) may result in accepting transformations that are computed wrongly.

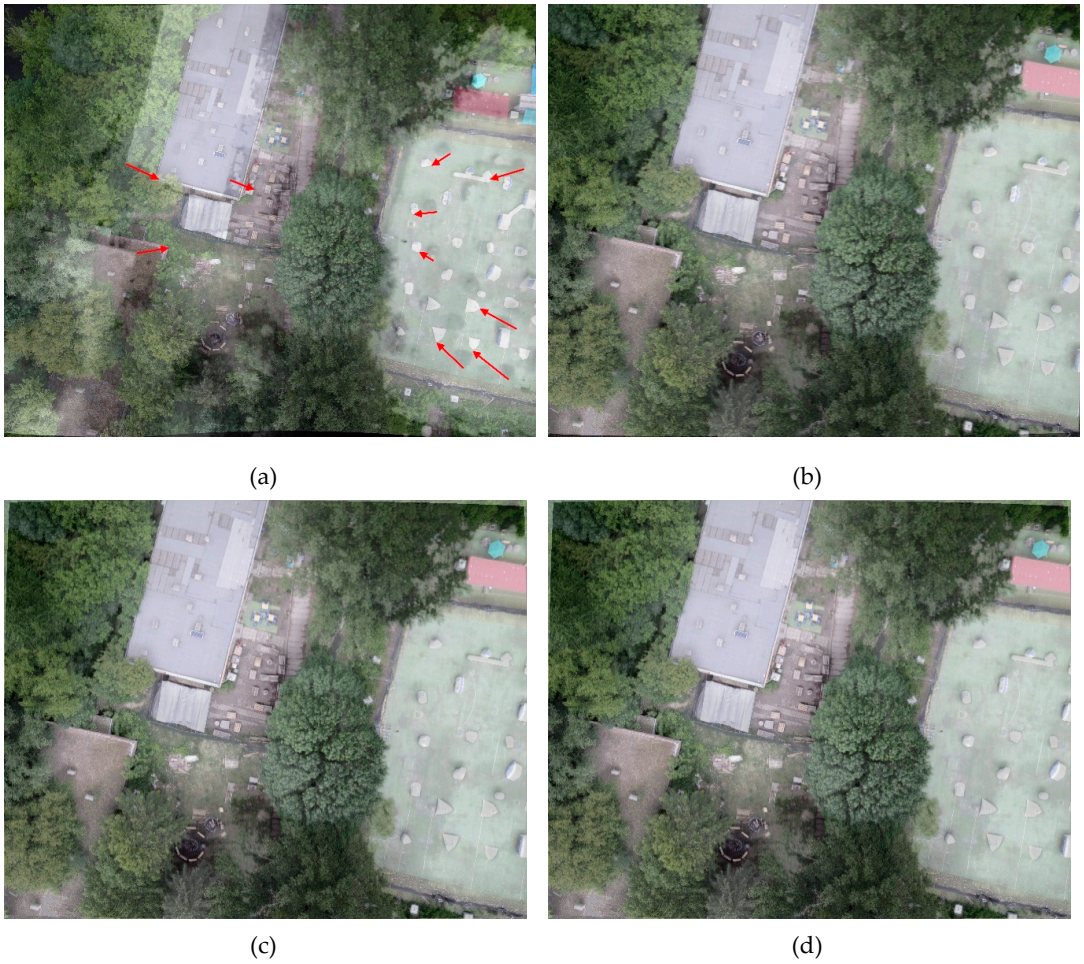


Figure 8. Comparison of the available photo mapping methods. (**a**) Mapping using homography, computed directly from the corresponding points of the calibration pattern, and the pictures are registered wrongly. Corresponding points between IR and VIS images are connected with red arrows; (**b**) mapping using homography computed manually; (**c**) mapping using ray–plane intersections; and (**d**) mapping using a depth map.

Next, both transformations were checked for similarity. If they were similar, recalibration was not required. Similarity was assessed based on the translation, rotation, and scale thresholds. We set the scale threshold to a similar value to that presented in the previous paragraph. Moreover, we also recommend setting values for the angle and translation thresholds within a range of 0.5°–2° and 1%–3%, respectively. If they are too strict (for example 0.1° or 0.5%), the computed transformations will be marked as dissimilar, and recalibration will need to be performed. It is vital to underline that transformations calculated from IR images may be slightly inaccurate, so if the threshold is set at a value that is too strict, the recalibration procedure will be constantly executed, which consumes energy.

Moreover, as the proposed approach relied on one previous frame per camera only, continuous computed motion may result in error accumulation and, as a consequence, in the divergence of images. This approach may be solved using the windowed bundle-adjustment process on a set of past frames at the price of a much higher computational complexity and power requirements.

Setting relatively large rotation angles between calibrated cameras (>3°) resulted in a large shift of the IR image, relative to the VIS, and perspective effects were revealed that could not be modeled in the case of affine transformation (Figure 9). In this case, it was better to use homography to recalibrate the system.

(a) (b) (c)

Figure 9. Example of calibration correction: (**a**) miscalibrated IR and VIS images (given additional rotations (−4 °, −4 °, 3 °)). Corresponding points between IR and VIS images are connected with red arrows; (**b**) correction by affine transformation (visible ghosting, especially in the upper left corner); and (**c**) correction using homography—images registered correctly.

In the case of small rotation angles and the occurrence of quasi-periodic structures in the images, the phase correlation algorithm may fall into the local minimum and not detect any miscalibration in the system (Figure 10). In addition, as shown in Table 3, the proposed algorithm was very conservative when using affine transformation and did not detect the need for calibration correction. This problem did not occur when homography was used to correct the calibration.

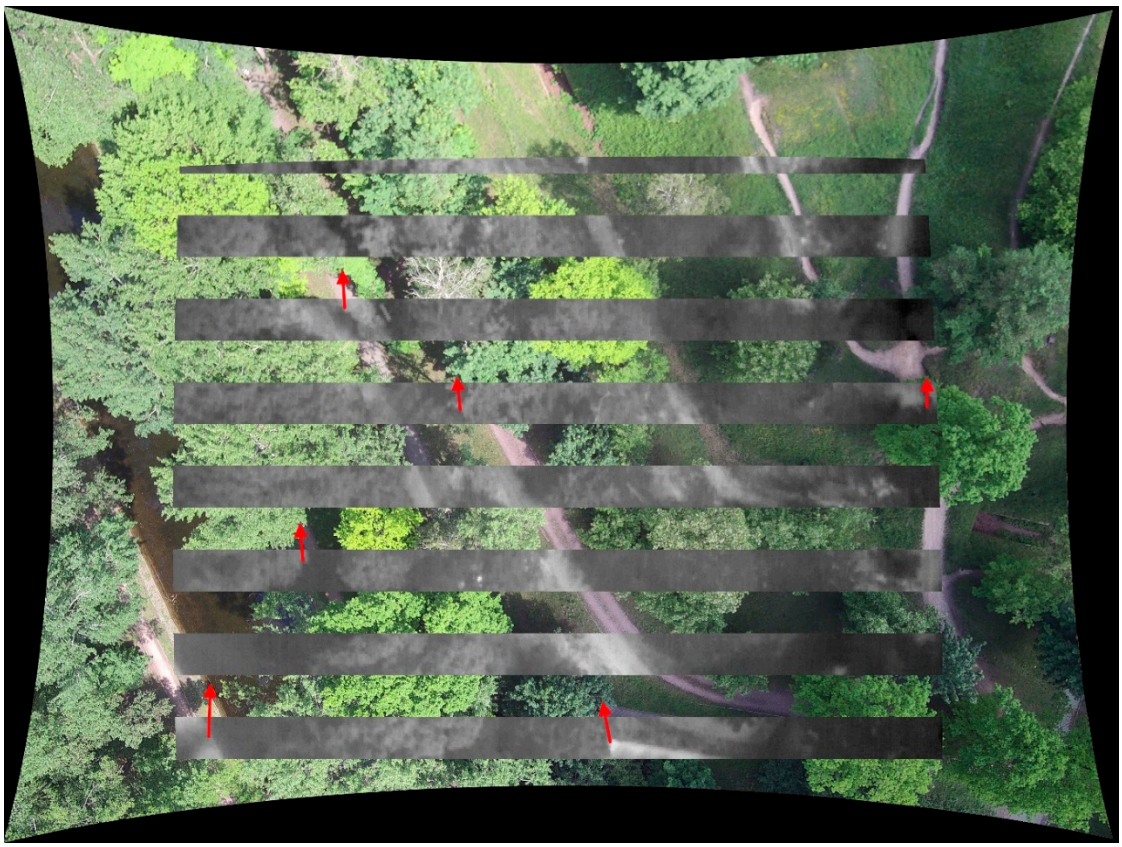


Figure 10. One of the failure cases of the proposed algorithm. Fragments of the IR image overlap over the VIS image. No miscalibration was detected (additional rotations in the x, y, and z axes were 2°, 0°, and 1°, respectively). Red arrows show exemplary matching points between the pair of images.

6. Conclusions

This article presents calibration and detection of miscalibration in a multimodal measuring system consisting of a VIS and IR camera mounted on a drone. A new calibration target was presented to determine the internal and external parameters of a pair of IR and VIS cameras. Then, three methods of mapping IR images on VIS were compared using a determined stereo calibration between multimodal camera pairs as well as an algorithm to detect and correct the spatial miscalibration of a pair of detectors in real time. To achieve this, two methods for determining the miscalibration of the system and two methods for mutual calibration correction were compared. The article demonstrates that the approach based on spectral analysis of both images produced better results than the approach using correlation in determining miscalibration. In addition, it has been shown that it is advisable to use homography to correct system miscalibration, even in the case of relatively small shifts of the sensors.

Using the SfM technique for the 3D reconstruction of the observed area, it is also possible to associate 3D data with IR images because 3D data are always consistent with VIS images. In future work, the authors plan to extend the proposed calibration algorithm and calibration correction to other 3D data sources, such as LIDAR (Light Detection And Ranging), to reduce the time needed for 3D reconstruction and allow the system to work completely in real time.

Author Contributions: Formal analysis, S.P.; Funding acquisition, M.M.; Investigation, P.S. and K.M.; Software, P.S.; Supervision, S.P.; Validation, K.M.; Writing—original draft, P.S.; Writing—review and editing, S.P. and M.M.

Funding: The authors gratefully acknowledge financial support from the European Union Funds under the Smart Growth Operational Programme (agreement: POIR 04.01.04-00-0105/16-00), National Centre of Research and Development. This work was also partially supported by scientific subsidy of Warsaw University of Technology.

Acknowledgments: The support of the GISS Sp. z o.o. employees is also greatly appreciated.

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

References

1. Rutkiewicz, J.; Malesa, M.; Karaszewski, M.; Foryś, P.; Siekański, P.; Sitnik, R. The method of acquiring and processing 3D data from drones. In *Speckle 2018: VII International Conference on Speckle Metrology*; Józwik, M., Jaroszewicz, L.R., Kujawińska, M., Eds.; SPIE: Bellingham, DC, USA, 2018; p. 97.
2. Özyeşil, O.; Voroninski, V.; Basri, R.; Singer, A. A survey of structure from motion. *Acta Numer.* **2017**, *26*, 305–364. [CrossRef]
3. Maes, W.H.; Huete, A.R.; Steppe, K. Optimizing the processing of UAV-based thermal imagery. *Remote Sens.* **2017**, *9*, 476. [CrossRef]
4. Padró, J.C.; Muñoz, F.J.; Planas, J.; Pons, X. Comparison of four UAV georeferencing methods for environmental monitoring purposes focusing on the combined use with airborne and satellite remote sensing platforms. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *75*, 130–140. [CrossRef]
5. Turner, D.; Lucieer, A.; Malenovský, Z.; King, D.H.; Robinson, S.A. Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over antarctic moss beds. *Remote Sens.* **2014**, *6*, 4003–4024. [CrossRef]
6. Putz, B.; Bartyś, M.; Antoniewicz, A.; Klimaszewski, J.; Kondej, M.; Wielgus, M. Real-time image fusion monitoring system: Problems and solutions. In *Advances in Intelligent Systems and Computing*; Springer: Berlin, Germany, 2013; Volume 184, pp. 143–152.
7. Pohit, M.; Sharma, J. Image registration under translation and rotation in two-dimensional planes using Fourier slice theorem. *Appl. Opt.* **2015**, *54*, 4514. [CrossRef] [PubMed]
8. Huang, X.; Chen, Z. A wavelet-based multisensor image registration algorithm. In Proceedings of the 6th International Conference on Signal Processing (ICSP), Beijing, China, 26–30 August 2002; IEEE: Piscataway, NJ, USA, 2002; Volume 1, pp. 773–776.
9. Li, H.; Ding, W.; Cao, X.; Liu, C. Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing. *Remote Sens.* **2017**, *9*, 441. [CrossRef]

10. Huang, Q.; Yang, J.; Wang, C.; Chen, J.; Meng, Y. Improved registration method for infrared and visible remote sensing image using NSCT and SIFT. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 2360–2363.
11. Liu, L.; Tuo, H.Y.; Xu, T.; Jing, Z.L. Multi-spectral image registration and evaluation based on edge-enhanced MSER. *Imaging Sci. J.* **2013**, *62*, 228–235. [CrossRef]
12. Yahyanejad, S.; Rinner, B. A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 189–202. [CrossRef]
13. Li, H.; Zhang, A.; Hu, S. A registration scheme for multispectral systems using phase correlation and scale invariant feature matching. *J. Sensors* **2016**, *2016*, 3789570. [CrossRef]
14. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
15. Kelcey, J.; Lucieer, A.; Kelcey, J.; Lucieer, A. Sensor Correction of a 6-Band Multispectral Imaging Sensor for UAV Remote Sensing. *Remote Sens.* **2012**, *4*, 1462–1493. [CrossRef]
16. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
17. Lagüela, S.; González-Jorge, H.; Armesto, J.; Herráez, J. High performance grid for the metric calibration of thermographic cameras. *Meas. Sci. Technol.* **2012**, *23*, 015402. [CrossRef]
18. Luhmann, T.; Piechel, J.; Roelfs, T. Geometric calibration of thermographic cameras. In *Remote Sensing and Digital Image Processing*; Springer: Dordrecht, The Netherlands, 2013; Volume 17, pp. 27–42.
19. Usamentiaga, R.; Garcia, D.F.; Ibarra-Castanedo, C.; Maldague, X. Highly accurate geometric calibration for infrared cameras using inexpensive calibration targets. *Measurement* **2017**, *112*, 105–116. [CrossRef]
20. Shibata, T.; Tanaka, M.; Okutomi, M. Accurate Joint Geometric Camera Calibration of Visible and Far-Infrared Cameras. *Electron. Imaging* **2017**, *2017*, 7–13. [CrossRef]
21. Vidas, S.; Lakemond, R.; Denman, S.; Fookes, C.; Sridharan, S.; Wark, T. A mask-based approach for the geometric calibration of thermal-infrared cameras. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 1625–1635. [CrossRef]
22. Saponaro, P.; Sorensen, S.; Rhein, S.; Kambhamettu, C. Improving calibration of thermal stereo cameras using heated calibration board. In Proceedings of the International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; IEEE: Piscataway, NJ, USA, 2015; Volume 2015, pp. 4718–4722.
23. Lagüela, S.; González-Jorge, H.; Armesto, J.; Arias, P. Calibration and verification of thermographic cameras for geometric measurements. *Infrared Phys. Technol.* **2011**, *54*, 92–99. [CrossRef]
24. Yang, R.; Yang, W.; Chen, Y.; Wu, X. Geometric calibration of IR camera using trinocular vision. *J. Light. Technol.* **2011**, *29*, 3797–3803. [CrossRef]
25. Heather, J.P.; Smith, M.I. Multimodal image registration with applications to image fusion. In Proceedings of the 2005 7th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1, pp. 372–379.
26. Brown, M.; Lowe, D.G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *74*, 59–73. [CrossRef]
27. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *25*, 120–125.
28. Yahyanejad, S.; Quaritsch, M.; Rinner, B. Incremental, orthorectified and loop-independent mosaicking of aerial images taken by micro UAVs. In Proceedings of the ROSE 2011-IEEE International Symposium on Robotic and Sensors Environments, Montreal, QC, Canada, 17–18 September 2011; pp. 137–142.
29. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *PAMI-8*, 679–698. [CrossRef]
30. Fang, M.; Yue, G.; Yu, Q. The study on an application of otsu method in canny operator. In Proceedings of the 2009 International Symposium on Information Processing (ISIP 2009), Huangshan, China, 21–23 August 2009; pp. 109–112.
31. Evangelidis, G.; Psarakis, E. Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1858–1865. [CrossRef] [PubMed]
32. Fischler, M.A.; Bolles, R.C. Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]