Article

Geometric- and Optimization-Based Registration Methods for Long-Wave Infrared Hyperspectral Images

Alper Koz * and Ufuk Efe

Center for Image Analysis (OGAM), Middle East Technical University, 06800 Ankara, Turkey;
ufuk.efe@metu.edu.tr

* Correspondence: koz@metu.edu.tr

Abstract: Registration of long-wave infrared (LWIR) hyperspectral images with their thermal and emissivity components has until now received comparatively less attention with respect to the visible near and short wave infrared hyperspectral images. In this paper, the registration of LWIR hyperspectral images is investigated to enhance applications of LWIR images such as change detection, temperature and emissivity separation, and target detection. The proposed approach first searches for the best features of hyperspectral image pixels for extraction and matching in the LWIR range and then performs a global registration over two-dimensional maps of three-dimensional hyperspectral cubes. The performances of temperature and emissivity features in the thermal domain along with the average energy and principal components of spectral radiance are investigated. The global registration performed over whole 2D maps is further improved by blockwise local refinements. Among the two proposed approaches, the geometric refinement seeks the best keypoint combination in the neighborhood of each block to estimate the transformation for that block. The alternative optimization-based refinement iteratively finds the best transformation by maximizing the similarity of the reference and transformed blocks. The possible blocking artifacts due to blockwise mapping are finally eliminated by pixelwise refinement. The experiments are evaluated with respect to the (i) utilized similarity metrics in the LWIR range between transformed and reference blocks, (ii) proposed geometric- and optimization-based methods, and (iii) image pairs captured on the same and different days. The better performance of the proposed approach compared to manual, GPU-IMU-based, and state-of-the-art image registration methods is verified.

Keywords: long-wave infrared; hyperspectral image registration; temperature; emissivity; keypoints; optimization; GPU; IMU

Citation: Koz, A.; Efe, U. Geometric- and Optimization-Based Registration Methods for Long-Wave Infrared Hyperspectral Images. *Remote Sens.* **2021**, *13*, 2465. <https://doi.org/10.3390/rs13132465>

Academic Editor: Edoardo Pasolli

Received: 17 May 2021

Accepted: 19 June 2021

Published: 24 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image registration is a preprocessing stage in image analysis to geometrically align different sets of image data captured from different poses into the same coordinate system [1]. Such a process is used in various areas ranging from computer vision, medical imaging, and aerial and space research to military image analysis in order to increase the gathered information acquired from different cameras and sensors. In the case of hyperspectral images [2,3], the registration aims to align the three-dimensional (3D) hyperspectral data with two spatial dimensions and one spectral dimension, compared to the registration of two-dimensional (2D) images.

The previous studies on hyperspectral image registration [4–17] mainly achieve this goal in two different categories. The first category [4–12], namely feature-based methods, finds and matches the feature points on the given hyperspectral images and estimates the best geometrical transform between two images by using matched points. The main stream in this category applies a transformation to convert the 3D hyperspectral data cube into a 2D image, and then the registration is performed over the resulting 2D image correspondences. The second category of hyperspectral image registration methods [13–17] mainly originates

from medical image applications, which require the registration of different modalities [18]. This category defines the registration process as an optimization problem that iteratively maximizes a similarity metric defined over all pixels rather than feature points. Given two images, the iteration begins by applying a random geometric transformation, mostly taken as 2D planar projective transformation, to one of the images and computes the similarity between the transformed image and the second reference image. The parameters of the transformation are then updated with respect to a gradient search algorithm until a convergence criterion is reached.

A main characteristic of the previous studies [4–17] in hyperspectral image registration is that they deal mostly with the registration between visible near-infrared (VNIR) hyperspectral images, shortwave infrared (SWIR) hyperspectral images, and monochrome/RGB images. The registration of LWIR hyperspectral images is not widely studied, which can be explained by the scarcity and high cost of these data and related hyperspectral sensors in the past [19–21]. Nevertheless, with the proliferation of this technology in military and aerospace applications, registration has also become a necessary important stage for fundamental analysis and interpretations on LWIR hyperspectral images. For instance, the registration of two LWIR hyperspectral images taken at different times can improve the accuracy of temperature emissivity separation by using not only one radiance spectrum, but two radiance spectra of the same pixel taken at different times. With the additional spectrum information taken at a different time, the number of unknowns in the temperature emissivity separation problem would be alleviated. The extracted emissivity information for the pixels can then be used for the identification of different targets and materials in the scene. As another example, the alignment of the hyperspectral pixel spectra taken at different times enables change detection analysis in the scene. In particular, such an alignment is more crucial for the detection of buried objects in military relevance by means of examining the changes in the scene after the first capture. Last but not least, the noise suppressions can be more effective with the extra observations about the data after such alignments, which can all contribute to the performances of different detection, classification, and analysis applications.

Such an operation on LWIR hyperspectral images, however, entails both different and similar challenges compared to the registration of hyperspectral VNIR/SWIR images [4–6,10,11] and monochrome/RGB images [22–24]. First, as being the dominant component of the radiance data in the LWIR range, the thermal component of the hyperspectral image changes from time to time due to temperature differences, which causes significant variations in the hyperspectral images of the same scene taken at different instants. Second, the noise component on the thermal LWIR sensors is more severe compared to VNIR and SWIR images with a lower signal-to-noise ratio and more spikes on the captured image. Third, different regions on a captured scene can impose different transformations, in particular for the nonplanar surfaces. While the application of one affine or homography transformation to the whole image in aerial applications can be accepted up to an extent for the registration of RGB images in previous solutions, the alignment of the local regions is more critical in hyperspectral images in order to catch the spectral information at the same pixel location after the registration. Finally, the problems about the precisions of GPS and IMU devices [25–27] are also more critical for the registration of hyperspectral images as the acquisition time with pushbroom spectral cameras is much longer than with snapshot RGB cameras.

The adopted strategy in this paper is first to investigate the most appropriate features of hyperspectral pixels in the thermal LWIR range for feature extraction and matching and to convert the 3D hyperspectral images to 2D images by using those features. The registration of 3D hyperspectral images, therefore, turns into the registration of 2D images after such a conversion, and the utilization of the existing literature for 2D image registration [28–36], particularly recent deep learning-based methods, can be considered as a candidate solution. One of the major lacks of these methods is, however, that they perform the registration over a global transformation between the image pairs without taking the

local divergences into account. Second, they are very dependent on the generality of the training set as the main drawback of learning methods. Accordingly, their application to the 2D maps of LWIR hyperspectral images was unfortunately not very successful, as reported in this paper with the comparisons. Finally, the scarcity of the LWIR hyperspectral images and the challenges in the dynamic thermal modeling for synthetic LWIR image generation make the development of new learning-based solutions infeasible for the moment for such data.

Given the mentioned challenges, this article investigates the registration of aerial LWIR hyperspectral images taken at different times of the same scene at different positions. The proposed solutions are developed as an output of an experiment-driven research in a related NATO research panel with the provided LWIR hyperspectral data in order to reveal the registration performance of different LWIR spectral sensors. While the performed research first begins with the performances of the conventional registration pipeline for LWIR hyperspectral registration, it evolves to an ultimate method revealing the best features in the thermal range for registration and solving the problems regarding local distortions in the global registration. In this regard, the main contributions of the performed study can be summarized as follows:

- Considering the temperature variations in the scene from one acquisition to the other, the most appropriate and robust components among temperature, emissivity, and radiance features of hyperspectral pixels are investigated to convert the 3D hyperspectral images to 2D images.
- In order to solve the local misalignment problems in global and GPS-IMU-based registrations, two blockwise methods are developed. While the geometric-based solution searches for the best local homography from the keypoints in the neighborhood of each block, the alternative optimization-based solution takes the global transformation as the initial estimation for each block and iteratively improves the homography with respect to the similarity of the transformed and reference blocks.
- In order to account for the variations in the radiance data due to the temperature changes in the scene, the performances of different metrics in multimodal image registration, namely the geometric mean square error between the keypoints and their correspondences (geometric MSE), structural similarity index (SSIM) [37], and mutual information (MI) [38], are examined.
- The proposed method with local refinements is compared with manual and GPU-IMU-based registrations along with the state-of-the-art image registration methods by using the hyperspectral LWIR images captured on the same and different days.

The next section gives brief overviews of hyperspectral and image registration methods. Section 3 gives the utilized LWIR hyperspectral image data for the experiments. Section 4 then describes the proposed hyperspectral registration method with the details of the mentioned stages. Section 5 then presents the experimental results and comparisons. Then, the discussions are presented in Section 6, which is followed by the conclusions in Section 7.

2. Related Work

This section first provides a review of the hyperspectral image processing methods. Then, a brief summary of state-of-the-art image registration methods is given in connection with hyperspectral image registration.

2.1. Hyperspectral Image Registration Methods

Table 1 reveals an overview of the hyperspectral registration methods with respect to the spectral regions of the utilized images, which mostly select VNIR, SWIR, and visible RGB range. Among a few studies in the LWIR range, the registration is studied for broadband images of thermal infrared sensors in [19] and [20] for surveillance applications. As another work, Koz et al. [21] present some initial methods and results for the registration

of middle-wave infrared (MWIR) and LWIR hyperspectral images with some limited data, which also triggers the performed research in this paper in its early phase.

Table 1. An overview of hyperspectral image registration methods with the spectral range of image pairs.

Study	Spectral Range		Class
	Image 1	Image 2	
Rzhanov et al. [10]	RGB image	VNIR (377–1041 nm)	feature-based
Sima et al. [6]	RGB image	SWIR (1300–2500 nm)	feature-based
Dorado-Munoz [11]	VNIR (400–970 nm)	VNIR (400–970 nm)	feature-based
Mukherjee et al. [4]	VNIR + SWIR (357–2576 nm)	VNIR + SWIR (357–2576 nm)	feature-based
Gonçalves et al. [5]	VNIR + SWIR (400–2500 nm)	VNIR + SWIR (400–2500 nm)	feature-based
Hasan et al. [15]	RGB image (440, 540, 570 nm)	VNIR image (430–630 nm)	optimization-based
Zhou et al. [16]	RGB image	VNIR (430–860 nm)	optimization-based
Steven et al. [20]	RGB image	Broadband thermal (LWIR) sensor image	feature-based
Kim et al. [19]	Broadband thermal (MWIR) sensor image	Broadband thermal (LWIR) sensor image	feature-based
Koz et al. [21]	MWIR (3500–4900 nm)	LWIR (7800–11,200 nm)	feature-based

The methods in the table are also classified as feature-based methods and optimization-based methods. The feature-based methods mainly utilize a conversion from 3D hyperspectral cubes to 2D images and then use the resulting 2D images for further processing. Mukherjee et al. [4] perform such a transformation by using principal component analysis. The resulting principal components are ranked and selected with respect to smoothness criteria, and the selected components are summed up with a nonlinear function to obtain the 2D images. The scale-invariant feature transform (SIFT) features [39] are then utilized to find and match the correspondences on the resulting 2D images. The performance of this method is improved by Gonçalves et al. [5] by incorporating segmentation and outlier removal steps. Similar methods propose to regulate the parameter selection of the SIFT descriptors as in [6], to modify the orientation component of the SIFT [7,8], or to adopt a different descriptor such as speeded up robust features (SURF) [40] for better and more robust registration performances [9].

A different approach in feature-based methods [11,12] performs the registration by interpreting the 3D hyperspectral data cube as a vector image formed of vectors, instead of pixels. Rather than finding the feature points on 2D images, these methods utilize the changes in spectral dimensions to find feature points in three-dimensional hyperspectral cubes. To this end, Dorado-Munoz et al. [11,12] develop the modified version of the SIFT descriptor for 3D data. The modified method describes the scale-space of the difference of Gaussians over smoothed vector images rather than pixel images, finds the extreme points by defining the greater-than and smaller-than operations over vectors, and finally, performs the extrema localization by defining a Hessian matrix over vectors. A major challenge in this approach is, however, describing a proper ordering operation over vectors, whose components are in fact hyperspectral bands with different significance.

The optimization-based methods as the second category in the table utilize similarity metrics borrowed from medical imaging [13–17], which are based on statistical measures. For instance, Kern et al. [13] extend the description of mutual information from image pairs to a set of image pairs and utilize the extended metric for hyperspectral image registration. In [15], cross cumulative residual entropy is utilized to register the RGB images with the VNIR hyperspectral images by converting the images into grayscale 2D images. More sophisticated optimization-based methods are also proposed by Zhou et al. [16,17] for the registration of hyperspectral and color images having very large scale differences. Compared to the feature-based methods, the main disadvantage of iterative optimization methods is their computational complexity. Based on the utilized search algorithm, sticking in the local minima and performance variations with respect to the selected initial points can be regarded as another drawback of this category.

2.2. State-of-the-Art Image Registration Methods

The image registration pipeline mainly consists of feature detection, feature description, feature matching, and geometric transformation estimation stages. Recent developments in deep learning have also affected this pipeline with a significant amount of performance gain by focusing on one or more stages of the registration pipeline.

Table 2 gives a brief overview of the state-of-the-art image registration methods by mentioning the particular stages that each method handles in the chain of registration process along with the input and outputs of the algorithms. In accordance with Table 2, the first group [28–30] of these methods is able to find correspondences of image pair without explicitly detecting, describing, and matching the features. Deep feature matching (DFM) [28] benefits from a pretrained VGG-19 [41] extractor and, starting from terminal layers' features, hierarchically matches features until the first layers. NCNet [29] uses a neighbor consensus network to process the 4D correlation map. Patch2Pix [30] starts to match the features at the patch level and, using fine and mid-level regressors, finds the pixel level matches. The second group [31,32] in Table 2 consists of joint feature detectors and descriptors. SuperPoint [31] is an encoder–decoder network that outputs detected features and corresponding descriptors at the same time. D2-Net [32] also jointly extracts features by finding salient features in the trained network based on VGG-16 [41]. The third group focuses on producing good matches, either by rejecting outliers or learning to match well. LPM [33] removes mismatches benefitting from local neighborhood structures, while SuperGlue [34] learns to remake descriptors better for matching and learns to match features optimally. The last group of algorithms directly produces the homography between two images with regression. Deep image homography estimation [35] in this group inputs two grayscale images stacked and outputs the geometric relationship between them. Unsupervised deep homography [36] improves the former network to make it appropriate for unsupervised training.

With regard to the given overviews, the first observation is that there were no methods addressing the registration of LWIR hyperspectral images in the previous literature at a journal level. The second observation is that the local misalignment problems in global and GPS-IMU-based registrations are not studied in the previous literature of both image and hyperspectral image registration. The proposed algorithms in this paper will focus on these two aspects.

Table 2. A brief overview of state-of-the-art image registration methods. The particular stages that each method handles in the chain of the registration process are also presented along with the inputs and outputs. (+ indicates that the algorithm covers that stage.).

Methods	Year	Main Stages during Registration				Input	Output
		Feature Detection	Feature Description	Feature Matching	Geometric Estimation		
DFM [28]	2021	+	+	+		Image Pair	Putative Matches
NCNet [29]	2020	+	+	+		Image Pair	Putative Matches
Patch2Pix [30]	2020	+	+	+		Image Pair	Putative Matches
SuperPoint [31]	2018	+	+			Image	Features + Descriptors
D2-Net [32]	2019	+	+			Image	Features + Descriptors
LPM [33]	2018			+		Features + Descriptors	Putative Matches
SuperGlue [34]	2020			+		Features + Descriptors	Putative Matches
Deep Homog. Est. [35]	2016					Image Pair	Homography
Unsuper. Deep Homog. [36]	2018					Image Pair	Homography

3. Experimental Dataset

Three sets of LWIR hyperspectral images could be utilized for the experiments due to the scarcity of such publicly available hyperspectral LWIR data. Only the parts of the captured scenes without any sensitive information could be utilized for the experiments. Table 3 gives the details of the images with their abbreviations and Figure 1 illustrates a sample band from each image. The first and second sets are captured with SEBASS sensor [42] from a height of 500 m above ground level, and the third set is captured with TASI [43] from the heights of about 610 m (2000 ft) and 762 m (2500 ft). The spectral ranges of SEBASS and TASI sensors are 7.6–13.5 μm and 7.6–13.5 μm , involving 128 bands and 32 bands, respectively. All three sets consist of three images, two of which are captured on the same day at different times and the third of which is captured on a different day. The performances are reported for the image pairs on the same and different days.

Table 3. Abbreviations and other characteristics of the selected hyperspectral data set for the registration experiments.

Abbreviation	Spectral Range	No. of Bands	Capturing Day	Capturing Time	Capturing Height
LWIR1 _a	7.6–13.5 μm	128	20 August 2014	18:05	500 m
LWIR1 _b	7.6–13.5 μm	128	20 August 2014	16:35	500 m
LWIR1 _c	7.6–13.5 μm	128	12 August 2014	18:18	500 m
LWIR2 _a	7.6–13.5 μm	128	20 August 2014	18:05	500 m
LWIR2 _b	7.6–13.5 μm	128	20 August 2014	16:35	500 m
LWIR2 _c	7.6–13.5 μm	128	12 August 2014	18:18	500 m
LWIR3 _a	8.0–11.5 μm	32	12 August 2014	Not provided	2000 ft
LWIR3 _b	8.0–11.5 μm	32	12 August 2014	Not provided	2500 ft
LWIR3 _c	8.0–11.5 μm	32	19 August 2014	Not provided	2500 ft

Figure 1. Selected experimental dataset. Sample bands around $7.5\ \mu\text{m}$ for SEBASS and $8\ \mu\text{m}$ for TASI are used for illustration. (a) LWIR1_a , (b) LWIR1_b , (c) LWIR1_c , (d) LWIR2_a , (e) LWIR2_b , (f) LWIR2_c , (g) LWIR3_a , (h) LWIR3_b , (i) LWIR3_c . The information for the images is provided in Table 3.

4. Proposed Registration Methods

The proposed registration method is illustrated in Figure 2. The main stages of the method can be described as follows:

- **3D–2D Conversion:** The hyperspectral images with two spatial and one spectral dimension, namely HSI1 and HSI2, are transformed to 2D maps in this stage. Different conversions based on temperature and emissivity components in the LWIR range are proposed and compared.
- **Global Pose Estimation based on Keypoint Matching:** The keypoints of the 2D images are extracted and matched. The homography (2D planar projective) transformation [44] between the 2D images is estimated by random sample consensus (RANSAC) [45].
- **Blockwise Refinement:** The global planar projective transformation is refined over blocks to align the detail edges, objects, and contrast in local regions by using the proposed geometry-based or optimization-based methods.
- **Pixelwise Refinement:** Blocking artifacts that can occur due to the individual mapping of neighbor blocks are further refined by investigating the best homography among different candidates for each pixel at this final stage.
- **2D Mosaicing and Outputs:** Given the 2D images, one image is geometrically transformed to the coordinate system of the other (reference) image by using the estimated transforms for each pixel. The resulting mosaic image and the similarity metrics, namely mutual information (MI) and structural similarity index (SSIM), between the transformed and reference images are returned as outputs for visual inspection and performance evaluation.

Figure 2. General scheme of the proposed hyperspectral image registration system where the dashed blocks correspond to the parts concentrated on in the proposed methods.

The main differences of the proposed scheme with respect to the conventional image registration pipeline are highlighted in bold in Figure 2. The contributions of the performed research are concentrated on the investigation of the temperature, emissivity, and radiance components in thermal LWIR range for 3D–2D conversion for the purpose of feature extraction and matching, development of the geometric- and optimization-based blockwise local refinements after global transformation, and the elimination of blocking artifacts using pixelwise refinements. The other parts of the proposed scheme such as feature matching and RANSAC follow the well-settled processing in 2D image registration. The following sections describe these contributions in detail.

4.1. 3D–2D Conversions

As the radiance signal in the thermal range is mainly affected by the temperature and emissivity of the materials, different approaches are proposed to convert the 3D hyperspectral images to 2D images. These images are then investigated with respect to their performances for keypoint extraction and matching. First, a method to estimate the brightness temperatures of hyperspectral pixels is developed [46] and a 2D image is formed by using this estimate for each pixel. Second, the average spectral energy and the principal components of the radiance signal for each hyperspectral pixel are used to generate 2D images. Finally, the average spectral energy and the principal components of the emissivity signal for each pixel are utilized as the other main components of the signal captured in the thermal range. The descriptions of the algorithms are as follows.

4.1.1. Brightness Temperature Estimation for Hyperspectral Pixels

Let us assume that the hyperspectral image with radiance information and of size $m \times n \times p$ is denoted as $I(x, y, \lambda)$, where x and y indicate the spatial horizontal and vertical coordinates and λ is the spectral wavelength. The brightness temperature of the hyperspectral pixels is optimally estimated by minimizing the mean square error (MSE) between the radiance of the hyperspectral pixels and the Planck curves generated for different temperatures. The proposed estimation is as follows:

1. The radiance of the hyperspectral pixel is assigned to a vector for each pixel of the hyperspectral image:

$$v_{x,y}(\lambda) = I(x, y, \lambda). \quad (1)$$

2. Planck curves [47] for different temperatures are generated for the spectral range of LWIR camera from a minimum temperature, T_{min} , to a maximum temperature, T_{max} , with a step size of temperature, ΔT :

$$B_T(\lambda) = \frac{2hc^2\lambda^{-5}}{(e^{hc/k\lambda T} - 1)}, \quad (2)$$

where in cgs units, $h = 6.626068 \times 10^{-27}$ erg s (Planck's constant), $k = 6.626068 \times 10^{-16}$ erg deg⁻¹ (Boltzman's constant), $c = 2.997925 \times 10^{10}$ cm/s (speed of light in vacuum), and T is the object temperature in kelvin.

- 3 The mean square error (MSE) between the generated Planck curve for each temperature and the radiance of the hyperspectral pixel is computed. The temperature giving the minimum MSE is assigned as the brightness temperature of the hyperspectral pixel:

$$T_B(x, y) = \operatorname{argmin}_T \left(\frac{1}{p} \sum_{\lambda} (v_{x,y}(\lambda) - B_T(\lambda))^2 \right) \quad (3)$$

- 4 The algorithm takes the T_{min} , T_{max} , and ΔT values along with the hyperspectral image as inputs. The 2D image, $T_B(x, y)$, involving the estimated brightness temperature for each pixel, is returned as the output.

4.1.2. Average Spectral Energy and PCA Components of Radiance Spectra as 2D Maps

The average spectral energy of the radiance signal for each pixel of the hyperspectral image, $E_R(x, y)$, is defined as follows:

$$E_R(x, y) = \frac{1}{p} \sum_{\lambda} I(x, y, \lambda)^2 \quad (4)$$

In addition, principal component analysis (PCA) is applied to the LWIR images in accordance with the proposed registrations methods for VNIR and SWIR images [4,5]. The corresponding images for the first and second principal components are used as 2D maps.

4.1.3. Average Spectral Energy and PCA Components of Emissivity Signal as 2D Maps

The computation of the average spectral energy for the emissivity component of the hyperspectral image involves the following stages:

1. The given hyperspectral image, $I(x, y, \lambda)$, is first separated into temperature and emissivity components by using a temperature emissivity separation (TES) algorithm [48]:

$$[T_p(x, y), e(x, y, \lambda)] = TES(I(x, y, \lambda)), \quad (5)$$

where $T_p(x, y)$ is the estimate of the physical temperature for the hyperspectral pixels and $e(x, y, \lambda)$ is the estimate of the three-dimensional emissivity signal with a size of $m \times n \times p$.

2. The average emissivity energy for each pixel, $E_e(x, y)$, is computed as follows:

$$E_e(x, y) = \frac{1}{p} \sum_{\lambda} e(x, y, \lambda)^2. \quad (6)$$

The algorithm returns the 2D image, $E_e(x, y)$, as the output. In addition to the average spectral energy, the corresponding images for the first and second components of the emissivity signal, $e(x, y, \lambda)$ after the PCA transform are also investigated for feature extraction.

4.2. Global Pose Estimation Based on Keypoint Matching

The registration of the hyperspectral cubes turns into the registration of 2D images at this stage, which follows the conventional pipeline for pose estimation between images. The resulting 2D maps first undergo median filtering due to the dominant sensor noise in thermal range, which usually emerges as spikes on the generated maps. As a second preprocessing step, the 2D image pairs are brought into the same range of [0,255] by using the maximum and minimum of the images.

The widely known scale-invariant feature transform (SIFT) [39] is adopted for keypoint extraction. However, the initial tests on global pose estimation have revealed that the same conclusions are also valid for the other keypoint extraction method such as SURF [40], Harris [49], DOM-SIFT [7], and GOM-SIFT [8]. The keypoints from two images are matched with respect to the mean square error (MSE) between their descriptors. These initial matches

are inputted to the RANSAC algorithm to find the best planar projective transformation, namely the homography matrix of size 3×3 .

4.3. Blockwise Refinement

The resulting global homographies and georeferencing (GPS-IMU)-based registrations [25] might indicate local misalignments due to the inaccuracies in the selection of keypoints and instabilities in the positions of the GPS and IMU devices during capturing. Therefore, global transformation should be further refined to align the local regions.

4.3.1. Geometric-Based Local Refinement

The proposed geometric-based method in this section performs the refinement over each image block by defining a homography transformation with respect to the nearest keypoints to that block. Let us assume that $P(x', y')$ and $R(x, y)$ denote the utilized 2D maps for the two hyperspectral images, namely original and reference images, with the coordinate systems (x', y') and (x, y) . Let $\bar{P}(x, y)$ also denote the ultimate transformed image with the geometric-based local refinement and H_g be the obtained global homography from the coordinate system from (x', y') to (x, y) as the output of the global pose estimation. The proposed blockwise geometric refinement involves the following stages:

1. First, divide the reference image, $R(x, y)$ into the nonoverlapping blocks of size $N_B \times N_B$. Denote the resulting blocks as $r_{ij}(x, y)$, where i and j refer to the horizontal and vertical indices for the block number as illustrated in Figure 3.
2. For each block, $r_{ij}(x, y)$, find the spatial (Euclidian) distances of all keypoints to the center of that block. Find the closest N_K number of matched keypoints to the center of that block. As an example, the selected closest keypoints are illustrated with solid circles in Figure 3.
3. For each 4-point combination of the N_K points,
  - i. Derive a homography matrix by using their corresponding matches [44].
  - ii. Form the transformed block, $\bar{p}_{ij}(x, y)$, by finding the corresponding position, (x', y') , for each pixel position, (x, y) , inside the block, with the given homography matrix and by performing a bilinear interpolation on $P(x', y')$.
  - iii. Compute the distance (with respect to a metric) between $r_{ij}(x, y)$ and $\bar{p}_{ij}(x, y)$.
4. Assign the homography matrix, which gives the minimum distance among all the homography matrices including also the inverse of the global homography, H_g^{-1} , as the homography of the block $r_{ij}(x, y)$. The resulting homography for the block $r_{ij}(x, y)$ is denoted as H_{ij} .
5. The ultimate transformed image $\bar{P}(x, y)$ is formed by concatenating the transformed image blocks with the resulting homographies (H_{ij} 's).

Figure 3. An illustration for the division of the image into blocks and selection of the closest $N_K = 9$ points to the center of the block r_{ij} . The solid circles indicate the closest points to the center of the block (shown with blue x), while the crosses correspond to other keypoints.

The block size, N_B , and the number of selected keypoints for each block, N_K , are selected as the design parameters in the proposed algorithm. The performances of three metrics, namely SSIM, mutual information, and the average spatial (Euclidian) distance between the coordinates of the selected N_K keypoints after the transformation and the coordinates of the corresponding keypoints on the reference image, are compared between the transformed and reference blocks. Note that the negatives of the similarity metrics MI and SSIM are utilized as distance metrics during implementation to align with the distance definition. The overall similarity of the ultimate transformed image, $\bar{P}(x, y)$, and reference image, $R(x, y)$, is evaluated in terms of SSIM.

4.3.2. Optimization-Based Local Refinement

The optimization-based local refinement also works over the image blocks. Given a reference block, the algorithm tries to find the best homography, which minimizes the distance metric, D , between the reference block and the corresponding transformed block, which is formed by mapping the pixels of the other image with the given homography matrix. The optimization problem can be mathematically described as follows:

$$H_{ij} = \underset{H}{\operatorname{argmin}} \left(D \left(\bar{p}_{ij}(x, y), r_{ij}(x, y) \right) \right) \quad (7)$$

where H refers to the homography transformation and $r_{ij}(x, y)$ is the i th and j th block of the reference image $R(x, y)$ as mentioned in the previous section. $\bar{p}_{ij}(x, y)$ corresponds to the transformed block, which is formed by mapping the corresponding pixels of $P(x', y')$ to the coordinate system (x, y) with the given H . Therefore, the distance, D , is a function of H , which is a 3×3 matrix with 9 parameters. In the description of (7), it is not preferred to include the dependency of $\bar{p}_{ij}(x, y)$ to H in order not to make the equation too cumbersome. The distance metric, D , is selected as the negative of the SSIM between the transformed and reference block as the optimization problems are conventionally expressed as a minimization problem rather than a maximization problem.

The optimization algorithm takes the global homography H_g as an initial guess for each block. The quasi-Newton method is adopted for the solution of the minimization problem in (7) as one of the standard methods preferred in optimization toolboxes. Given that the distance D in (7) is a function of H , the main stages of the proposed algorithm are as follows:

1. First, divide the reference image, $R(x, y)$ into the nonoverlapping blocks of size $N_B \times N_B$. Denote the resulting blocks as $r_{ij}(x, y)$, where i and j refer to the horizontal and vertical indices for the block number, respectively.
2. For each block, $r_{ij}(x, y)$,
  - i. Assign H as H_g at the initialization.
  - ii. Form the transformed block, $\bar{p}_{ij}(x, y)$, by finding the corresponding position, (x', y') , for each pixel position, (x, y) , inside the block, with the given homography matrix H and by performing a bilinear interpolation on $P(x', y')$.
  - iii. Compute the distance between $r_{ij}(x, y)$ and $\bar{p}_{ij}(x, y)$, namely $D(\bar{p}_{ij}(x, y), r_{ij}(x, y))$.
  - iv. Compute the gradient ∇D and Hessian $\nabla^2 D$ of the cost function D with respect to H .
  - v. Update the H with respect to the quasi-Newton algorithm:

$$H_{k+1} = H_k - (\nabla^2 D(H_k))^{-1} \nabla D(H_k) \quad (8)$$

where k is the number of iteration.

- vi. If the number of iteration is smaller than the maximum number of iteration and the change in the cost function, $D(H_{k+1}) - D(H_k)$, is greater than a threshold value, then update the homography H as H_{k+1} and go to step ii. Otherwise, finish the iteration.

vii. The ultimate homography for the block $r_{ij}(x, y)$ is denoted as H_{ij} .

The distance metric, D , is selected as the negative of the SSIM between the transformed and reference block as the optimization problems are conventionally expressed as a minimization problem rather than a maximization problem. The transformed image $\bar{P}(x, y)$ is formed by concatenating the transformed image blocks with the resulting homographies (H_{ij} 's).

4.4. Pixelwise Refinement

Although blockwise refinement improves the performance of the global mapping, it can also produce blocking artifacts in some regions, as illustrated in Figure 12a–c, which is similar to the blocking problem in motion-compensated video coding [50]. As the algorithm maps all the pixels inside the blocks with the same homography, the neighbor blocks having different but wrong transformations can produce blocking distortions. The refinement is therefore further improved at pixel level. The proposed pixelwise refinement updates the homography for each pixel by searching for the best homography from a candidate set of homographies. This set is formed of the homographies of the blocks giving the best-matched results over all the blocks in the image in terms of the resulting SSIM score between the transformed and reference blocks after the blockwise refinement.

Let us assume that the best M blockwise homographies with respect to the distance of the transformed and reference blocks are denoted as $H_1, H_2, \dots, H_M$ and $\bar{P}_{pw}(x, y)$ is the resulting image after pixelwise refinement. The proposed pixelwise refinement is described as follows:

1. Apply the homographies $H_1, H_2, \dots, H_M$ to the original image, $P(x', y')$, resulting a set of transformed images, $L_1(x, y), L_2(x, y), \dots, L_M(x, y)$.
2. Compute the SSIM map between each of the transformed image and the reference image, $R(x, y)$, resulting in the SSIM maps, $S_1(x, y), S_2(x, y), \dots, S_M(x, y)$. Note that $S_i(x, y)$ indicates the similarity of the pixels $R(x, y)$ and $L_i(x, y)$ with respect to SSIM.
3. Apply an averaging filter to the resulting SSIM maps in order also to include the effect of neighbor pixels in the similarity computation.
4. For each pixel (x, y) ,
  - i. Find the highest score among $S_1(x, y), S_2(x, y), \dots, S_M(x, y)$.
  - ii. Assuming that $S_k(x, y)$ is the highest score, assign the pixel value of the transformed image, $L_k(x, y)$, as the value of the $\bar{P}_{pw}(x, y)$.

The algorithm ultimately computes the MI and SSIM between the $\bar{P}_{pw}(x, y)$ and $R(x, y)$ for the evaluation.

4.5. Outputs: 2D Mosaic and Objective Similarity Metrics

A 2D mosaic is formed by combining the transformed and reference images and visually inspecting whether the lines and texture exhibit a continuous behavior or not after the applied transformations. In order to perform an objective assessment, the mutual information and SSIM between the transformed and reference images are also computed for the comparison of proposed 2D conversion methods and algorithms.

Given that u and v correspond to the intensity values of $I(x, y)$ and $R(x, y)$ respectively, the mutual information (MI) [37] between the images is computed as follows:

$$MI = \sum_{u=1}^L \sum_{v=1}^L P(u, v) \log\left(\frac{P(u, v)}{P_I(u)P_R(v)}\right) \quad (9)$$

where $P_I(u, v)$ and $P_R(u, v)$ are the marginal distributions of the images, $P(u, v)$ is the joint distribution of the images, and L is the maximum intensity of both images. Mutual information basically measures the dependence of two images by calculating the distance between the joint distribution of intensity values of two images $P(u, v)$ and the product of the marginal distributions $P_I(u)P_R(v)$.

The structural similarity index measure (SSIM) [38] between the two images, of $I(x, y)$ and $R(x, y)$, is computed as follows:

$$SSIM = \frac{(2\mu_I\mu_R + c_1)(2\sigma_{IR} + c_2)}{(\mu_I^2 + \mu_R^2 + c_1)(\sigma_I^2 + \sigma_R^2 + c_2)} \quad (10)$$

where μ_I and μ_R are averages, σ_I^2 and σ_R^2 variances, and σ_{IR} is the covariance of the images $I(x, y)$ and $R(x, y)$. c_1 and c_2 are defined as $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$, where L stands for the dynamic range of the intensity values with $k_1 = 0.01$ and $k_2 = 0.03$. SSIM is basically an image quality assessment measure inspired by the human visual system's capability of extracting structural information. It is a weighted combination of three comparison measurements: luminance, contrast, and structure [38].

MI metric is maximized when the images are precisely aligned. It is minimized when their statistics are independent, resulting in zero information. Therefore, MI could take values greater than or equal to zero and less than or equal to the minimum of the entropy values of two images [37]. On the other hand, SSIM could take values between and equal to zero and one. Larger values of MI and SSIM between the transformed and reference images mean a better registration performance. Hence, the cost function for the minimization problem is chosen as negative of SSIM throughout the article as in the conventional description of optimization problems.

5. Results

The experiments were conducted in MATLAB 16 environment, using a 2.5 GHz computer with 16 GB RAM. The Vlfeat library [51] and hyperspectral library by Greg [52] were utilized for the baseline operations regarding feature matching and hyperspectral image processing. The quasi-Newton algorithm in MATLAB optimization toolbox was adopted for the optimization-based registration method by selecting the tolerance in the change of the cost function as 10^{-4} and the maximum number of iterations as 500. The results are presented in five subsections investigating the following main aspects of the proposed registration method:

- First, the performances of the proposed 3D–2D conversions are discussed and the best conversions for the global pose estimation are determined in Section 5.1.
- This is followed by the description of controlled experiments conducted to select the design parameters for the proposed blockwise local refinement in Section 5.2.
- The improvements of the pixelwise refinement in comparison with the blockwise refinement are then discussed in Section 5.3.
- The next subsection, Section 5.4, compares the performances of the proposed geometric-based and optimization-based local refinements.
- Finally, the improvements of the proposed registration method are revealed with respect to the conventional approaches, such as manual and direct georeferencing-based registration, and state-of-the-art image registration methods, including deep learning-based approaches, in Section 5.5.

5.1. Selection of 3D to 2D Conversion Method for Global Pose Estimation

Figure 4 illustrates the 2D images obtained with the proposed 3D–2D conversions for the images $LWIR1_a$ and $LWIR1_b$, which were captured on the same day. The 2D images were scaled to the range of 0–255 by using the maximum and minimum values. The resulting images for the brightness temperature (a), average energy of the radiance spectra (b), and first principal component of the radiance spectra (c) have similar contrast. On the other hand, the images for the average emissivity maps (d) possess relatively low contrast. Therefore, histogram equalization was performed on these images to enhance the contrast, as illustrated in Figure 4e. The 2D image for the first principal component of the emissivity spectra also has comparatively less contrast with respect to the radiance-based transformations and brightness temperature. This can be linked with the fact that the

temperature and related 2D maps for radiance spectra correspond to the coarse part of the spectral changes, whereas the emissivity component can be interpreted as a detailed component from the point of signal decomposition theory. Therefore, the spatial contrast is comparatively lower for the emissivity-based 2D maps.

Figure 4. Generated 2D maps for the sample pair $LWIR1_a$ (left) and $LWIR1_b$ (right). (a) Brightness temperature estimate; (b) average energy of radiance spectrum for each pixel; (c) 1st PCA component of radiance spectra; (d) average energy of emissivity component for each pixel; (e) average energy of emissivity components for each pixel after histogram equalization; (f) 1st PCA component of emissivity spectra.

Figure 5a–g illustrates the matched points for different 3D–2D conversions for the same-day pair $LWIR1_a$ and $LWIR1_b$. The inliers after the RANSAC algorithms are illustrated. All the 2D maps have enough matches to estimate the geometric transformation between the pair $LWIR1_a$ – $LWIR1_b$. Table 4 gives the ratios of the inliers after the matching to the total number of matched points for all the pairs including the same and different day capturing. While brightness temperature maps and average energy give high scores for the same-day pairs, namely $LWIR1_a$ – $LWIR1_b$, $LWIR2_a$ – $LWIR2_b$, and $LWIR3_a$ – $LWIR3_b$, the principal components for the radiance components and the emissivity-based maps do not achieve a sufficient number of matching points for the pose estimation, particularly for the pair $LWIR2_a$ – $LWIR2_b$. Another important observation regarding the matching performances is a severe decrease in the inlier ratios for different days for all the 2D maps. While the ratios for the brightness temperatures and average radiance energy fall significantly down to 10%, the matching results for the principal components of the radiance and emissivity-based 2D maps do not indicate stable performances for different pairs on different days.

Figure 5. Resulting matched points after different 3D–2D conversions for the pair $LWIR1_a$ – $LWIR1_b$. (a) Brightness temperature; (b) average energy of radiance spectrum; (c) 1st PCA component of radiance spectra; (d) 2nd PCA component of radiance spectra; (e) average energy of emissivity component for each pixel; (f) 1st PCA component of emissivity spectra; (g) 2nd PCA component of emissivity spectra.

Table 4. Ratio and percentage of inliers after RANSAC for different 2D maps for the images captured on the same (S) and different (D) days. “X” indicates the cases where there is not a sufficient number of matches to estimate the geometric transform between the images. (BT: brightness temperature estimate, E_R : average energy of radiance spectrum, PCA1_R: 1st PCA component of radiance spectra, PCA2_R: 2nd PCA component of radiance spectra, E_e : average energy of emissivity component, PCA1_e: 1st PCA component of emissivity spectra, PCA2_e: 2nd PCA component of emissivity spectra.).

Pairs	S/D	BT	E_R	PCA1 _R	PCA2 _R	E_e	PCA1 _e	PCA2 _e
$LWIR1_a$ – $LWIR1_b$	S	86/176 (49%)	89/163 (57%)	89/173 (51%)	80/167 (48%)	60/130 (46%)	109/175 (62%)	71/156 (50%)
$LWIR1_a$ – $LWIR1_c$	D	11/81 (14%)	15/104 (14%)	X	X	15/87 (17%)	13/101 (13%)	X
$LWIR2_a$ – $LWIR2_b$	S	25/117 (21%)	27/119 (23%)	X	X	X	X	X
$LWIR2_a$ – $LWIR2_c$	D	9/89 (10%)	11/89 (12%)	13/96 (14%)	X	X	25/103 (24%)	X
$LWIR3_a$ – $LWIR3_b$	S	88/134 (66%)	94/138 (68%)	101/142 (71%)	X	20/50 (40%)	X	X
$LWIR3_a$ – $LWIR3_c$	D	39/91 (43%)	41/83 (49%)	45/104 (43%)	14/53 (26%)	X	X	X

Figure 6a–g gives the mosaic images obtained by overlapping the transformed image with the reference image. In particular, the continuity of the horizontal and vertical lines while passing from one image to the other indicates the success of the registration. The geometry of the transformed images is similar for all the proposed 2D conversion methods for the $LWIR1_a$ – $LWIR1_b$ pair. Figure 6h–k gives the mosaic images for the 2D maps, which have enough matches for registration, for the pair captured on different days, $LWIR1_a$ – $LWIR1_c$, as indicated in Table 4. The mosaic images are coarsely aligned for different-day

image pairs compared to the same-day results. Note that the mosaic images for the other 2D maps denoted with “X” in Table 4 cannot be obtained due to an insufficient number of matches to estimate the transform.

Figure 6. (a–g) Resulting mosaic images for the pair $LWIR1_a$ - $LWIR1_b$ for (a) brightness temperature, (b) average energy of radiance spectrum for each pixel, (c) 1st PCA component of radiance spectra, (d) 2nd PCA component of radiance spectra, (e) average energy of emissivity component for each pixel, (f) 1st PCA component of emissivity spectra, and (g) 2nd PCA component of emissivity spectra. (h–k) Resulting mosaic images for the pair $LWIR1_a$ - $LWIR1_c$ for (h) brightness temperature, (i) average energy of radiance spectrum for each pixel, (j) average energy of the emissivity spectrum, and (k) 1st PCA component of emissivity spectra. The mosaic images are given for the 2D maps that have enough matches, as indicated in Table 4.

Table 5 gives the mutual information and SSIM between the transformed images and reference images, respectively. Both the metrics indicate similar behaviors for different pairs. While the brightness temperature and average radiance energy give the best results for the same pair, their performances significantly decrease for different-day image pairs. The results for the emissivity-based maps are not very stable both for the same-day and different-day captures. Contrary to an initial assumption that the emissivity maps can survive temperature changes and achieve a better matching, the experiments reveal that they do not possess enough contrast for feature extraction and matching. However, the temperature maps can still be used to extract and match features even though the temperatures vary from time to time. For the sake of completeness, Figure 7 give also the mosaic images for the brightness temperature maps for the other four pairs, namely LWIR2a–LWIR2b, LWIR2a–LWIR2c, LWIR3a–LWIR3b, and LWIR3a–LWIR3c. In the rest of the experiments, the brightness temperature maps and SSIM were utilized for the global transformation and performance evaluation to further investigate the proposed local refinements.

Table 5. Mutual information and SSIM results after the registration for different 2D maps for the images captured on the same (S) and different (D) days. “X” indicates the cases where there is not a sufficient number of matches to estimate the geometric transform between the images. (BT: brightness temperature estimate, E_R : average energy of radiance spectrum, PCA1_R: 1st PCA component of radiance spectra, PCA2_R: 2nd PCA component of radiance spectra, E_e : average energy of emissivity component, PCA1_e: 1st PCA component of emissivity spectra, PCA2_e: 2nd PCA component of emissivity spectra.).

Pairs	S/D	Mutual Information						
		BT	E_R	PCA1 _R	PCA2 _R	E_e	PCA1 _e	PCA2 _e
LWIR1 _a –LWIR1 _b	S	0.52	0.56	0.59	0.95	0.66	0.85	0.68
LWIR1 _a –LWIR1 _c	D	0.26	0.29	X	X	0.40	0.40	X
LWIR2 _a –LWIR2 _b	S	0.50	0.59	X	X	X	X	X
LWIR2 _a –LWIR2 _c	D	0.50	0.62	0.64	X	X	0.94	X
LWIR3 _a –LWIR3 _b	S	1.07	1.10	1.10	X	0.25	X	X
LWIR3 _a –LWIR3 _c	D	0.82	0.84	0.86	0.30	X	X	X
Pairs	S/D	SSIM						
		BT	E_R	PCA1 _R	PCA2 _R	E_e	PCA1 _e	PCA2 _e
LWIR1 _a –LWIR1 _b	S	0.18	0.22	0.22	0.25	0.12	0.17	0.17
LWIR1 _a –LWIR1 _c	D	0.08	0.06	X	X	0.04	0.07	X
LWIR2 _a –LWIR2 _b	S	0.12	0.18	X	X	X	X	X
LWIR2 _a –LWIR2 _c	D	0.07	0.08	0.08	X	X	0.14	X
LWIR3 _a –LWIR3 _b	S	0.31	0.32	0.32	X	0.05	X	X
LWIR3 _a –LWIR3 _c	D	0.29	0.25	0.28	0.04	X	X	X

Figure 7. Resulting mosaic images for the brightness temperature maps for the four pairs, namely (a) $LWIR2_a-LWIR2_b$, (b) $LWIR2_a-LWIR2_c$, (c) $LWIR3_a-LWIR3_b$, and (d) $LWIR3_a-LWIR3_c$.

5.2. Selection of Design Parameters and Distance Metrics for Blockwise Local Refinement

The global transformation defined in the previous section coarsely matches the images without considering a fine registration in local regions. As an example, the given transformations in Figures 6 and 7 indicate that the registration is in fact mostly achieved throughout the vertical and horizontal roads on the given scenes, while the performance for the other regions can still be improved. This section gives the results for the proposed blockwise local refinement for such a purpose along with the selection of the design parameters.

The design parameters in the proposed geometric-based blockwise local refinement in Section 4.3 are the number of neighbor points, N_K , to determine the homography for a given block and the size of the blocks, N_B . Figure 8a illustrates the change of the global SSIM between the reference image and the transformed image after the proposed blockwise refinement as a function of N_K , when the block size is kept fixed as 27. It is quite expected that as the number of neighbor points increases, the number of combinations of four points to generate a better homography transformation for a given block increases as well. However, such a performance increase is limited with the execution time of the method, which is given in Figure 8b as a function of N_K . In order to hold the duration at the level of a few minutes, an N_K of 9 was selected in the experiments. The resulting SSIM scores in the given figure are all higher than the one obtained with the global homography, as the algorithm forces the selection of the better local SSIM for each candidate homography obtained with any four points. However, the selection of N_K as 6 or 7 can give local black regions, particularly in smooth areas due to similarities of those regions in terms of SSIM.

Figure 8. (a) Global SSIM between transformed and reference images vs. number of neighbor points, N_K . (b) Duration (s) vs. number of neighbor points, N_K .

Figure 9a gives the change of the global SSIM between the reference image and the transformed image after the blockwise refinement as a function of the block size N_B for the fixed $N_K = 9$. As the block size increases, the probability that there is a better homography for the selected blocks decreases. Note that the ultimate case for the selection of a block is just the whole image, where the global homography is applied. Therefore, the curve is decreasing through the indicated line for global homography as the block size is increasing. In order to hold the execution time shown in Figure 9b to a few minutes, the block size of 27 was selected, which gives a global SSIM score of about 0.35. Note that as the block size increases, the number of blocks decreases in the whole image. Accordingly, it can be observed that the decay in the computational time in Figure 9b is inversely proportional to the area of the blocks, N_B^2 .

Figure 9. (a) Global SSIM between transformed and reference images vs. block size, N_B . (b) Duration (s) vs. block size, N_B .

As the last experiment, the performances of different distance metrics for the selection of blockwise homography matrices in the proposed local refinement were compared. The metrics selected for measuring the distance between the reference block and the transformed block were SSIM, mutual information, and geometric mean square error (MSE) between the matched points as in the conventional RANSAC algorithm. Figure 10 gives the resulting global SSIM as a function of block size for the fixed $N_K = 9$ for the mentioned distance metrics. The performance of the geometric MSE is the worst among the three methods, which indicates that the localization of the SIFT points is not very accurate for better registration of the thermal images. As a second conclusion, the performance of the mutual information indicates a concave characteristic, which decreases for the lower block sizes due to the lower accuracies in the estimation of the probability density functions (pdfs) with fewer pixels. The mutual information then shows a coherent behavior with the SSIM for the higher block sizes. In the rest of the experiments, SSIM was selected as the

distance metric for the blockwise refinement, the number of neighbor points was fixed to 9, and the block size of 27 was selected.

Figure 10. Global SSIM between transformed and reference images vs. block size for different metrics in blockwise local refinement, namely SSIM, mutual information, and geometric MSE.

5.3. Comparison of Blockwise and Pixelwise Local Refinements

Figure 11 illustrates the visual results for the pair of *LWIR1a*–*LWIR1b*, which include the transformed images with the global transformation, blockwise refinement, and pixelwise refinement, and the corresponding SSIM maps between the reference image (*LWIR1b*) and the transformed images. Note that the brighter regions indicate higher SSIM scores in the given maps. As can be observed in Figure 11c,f, the global homography registers the parts throughout the road in the scene with high scores, but there are significant shifts in the regions shown with the red rectangles on the SSIM map in Figure 11f. The blockwise refinement on the other hand significantly improves the matching and SSIM scores on those regions as illustrated in Figure 11d and the corresponding SSIM map in Figure 11g.

Figure 11. (a,b) Brightness temperature (BT) maps for *LWIR1a* and *LWIR1b*. (c–e) Transformed images with global homography, after blockwise refinement, and after pixelwise refinement, respectively. (f–h) Corresponding SSIM maps between the transformed image and reference image (BT map of *LWIR1b*) for global homography, blockwise refinement, and pixelwise refinement, respectively.

However, a major drawback of the blockwise refinement is the blocking artifacts as the algorithm maps all the pixels inside the blocks with the same homography. Consequently, the neighbor blocks having different but wrong transformations can produce blocking distortions, which is similar to the blocking problem in motion-compensated video coding [50]. The red rectangles in Figure 11d illustrate some of those distortions, whose zoomed versions are shown in Figure 12a–c. The proposed pixelwise refinement given in Section 4.4 significantly compensates them, as shown in Figures 11e and 12d–f. Accordingly, the global SSIM results given in Figure 13 as a function of block size indicate better performances for the pixelwise refinement compared to blockwise refinement.

Figure 12. Some examples for the patches before (a–c) and after (d–f) pixelwise refinement. The artifacts for blockwise mapping are eliminated after pixelwise refinement.

Figure 13. Global SSIM between transformed and reference images vs. block size for blockwise and pixelwise local refinements.

5.4. Comparison of Geometric-Based and Optimization-Based Local Refinements

An alternative for the geometric feature-based blockwise refinement is optimization-based blockwise refinement, as described in Section 4.3. The optimization algorithm takes the global homography as an initial point and then iteratively updates this matrix for each block to maximize the SSIM between the reference block and transformed block. Figure 14a illustrates the change of the local SSIM for a sample block with respect to the iteration numbers. The saturation is quite observable as the iteration continues.

Figure 14. (a) An example for the change of the local SSIM for a sample block with respect to the iteration number in the optimization-based refinement. (b) Global SSIM vs. block size for optimization- and geometric-based refinements for the same-day pair, $LWIR1_a-LWIR1_b$. (c) Duration vs. block size for the optimization-based refinement and (d) global SSIM vs. block size for a different-day pair, $LWIR1_a-LWIR1_c$.

Figure 14b gives the resulting global SSIM with respect to the block size both for the optimization-based and geometric-based approaches for the pair of $LWIR1_a-LWIR1_b$. While the optimization-based approach is better than the result for the global homography, it gives comparable results with the geometric approach. However, the duration illustrated in Figure 14c, for instance for the block size of 27, is about 5 times the one for the geometric approach. The geometric approach is very appropriate with its tolerable execution time for practical usage. However, a possible limitation for the geometric-based method is its dependency on the number of extracted keypoints in the scene for the implementations in practice.

Figure 14d also gives the global SSIM results with respect to the block size for the pair of $LWIR1_a-LWIR1_c$, which is a pair captured on different days. There are no significant improvements in the registration results for both approaches compared to the global homography. The ultimate results for the same- and different-day pairs are discussed in the next section.

5.5. Improvements of the Proposed Refinements with Respect to the Baseline and State-of-the-Art Methods

In the presentation of the proposed method in previous sections, the disadvantages of the mainstream global registration methods are first revealed. Afterward, we eliminate the misalignments in global registration by means of the blockwise refinement, which is followed by the pixelwise refinement at the last stage. The ultimate proposed method is formed of all these stages, namely global registration, blockwise registration, and pixelwise registration, as indicated in Figure 2. Note that the geometric-based or optimization-based

method could be selected for the blockwise refinement in the whole registration chain. The experimental results are provided for both alternatives.

Manual registration, which is performed in typical remote sensing software, and georeferencing-based registration, which uses geo files obtained with the GPS-IMU system, were selected as the baseline methods for the experiments. The manual registration is realized by manually selecting a number of corresponding points in two images, as shown in Figure 15. These points are then used to estimate the transformation between two images. In the case of direct georeferencing, a corresponding pixel in the first image for a given pixel in the second image is found by performing a bilinear interpolation over the nearest pixels, whose corresponding distances are computed by using the longitudes and latitudes in the geo files.

Figure 15. An example for the manually selected matched points from the 2D temperature maps of $LWIR1_a$ – $LWIR1_b$. Note that the images are rotated horizontally. (+ indicates the selected points for matching.).

The SuperGlue [34], D2-Net [33], and locality preserving matching (LPM) [32] were selected as state-of-the-art methods due to their high performance and popularity in recent years. SuperGlue performs the feature matching by designing a neural network, which jointly finds correspondences and rejects outliers on local features. While the algorithm can also use regular keypoints extracted by classical extractors, the best-reported version, which uses the features extracted by SuperPoint algorithm [31], is preferred in the implementation. D2-Net is another popular deep learning-based network designed for joint detection and description of features. LPM, on the other hand, improves the classical registration pipeline by integrating a mismatch removal phase using putative correspondences.

A common characteristic of the given methods and the other state-of-the-art methods in Table 2 is that they mainly focus on the improvement of feature extraction and matching rather than the improvement of the alignment after the registration of the images. Therefore, the extracted features were further processed by RANSAC to estimate the ultimate homography between the image pairs in the case of SuperGlue and D2-Net. LPM was adapted to the given framework by first eliminating the outliers and then again using RANSAC to estimate the homography. The parameters such as the maximum number of keypoints, keypoint extraction threshold, and matching thresholds were selected as default values for the implementation of SuperGlue. The standard pretrained networks provided by the authors were employed for the implementation of both SuperGlue and D2-Net. As a critical parameter, the number of nearest neighbors during outlier rejection was also set to the default value for the implementation of LPM.

It should be noted that all the experiments were performed on the normalized 8-bit images obtained after 3D–2D conversions. The experiments have revealed that the application of learning-based SuperGlue and D2-Net algorithms, with the provided pretrained networks, to the converted 8-bit images in the LWIR range achieves registration as in the case of standard RGB images. One complementary option in the experiments could be to further train the pretrained networks with the converted 2D images. However, the scarcity of publicly available hyperspectral LWIR data was one challenge for such a direction. The complexity of the synthetic data generation for LWIR hyperspectral images was another challenge as it requires not only the modeling of the emissivity, temperature, and radi-

ance relations, but also the dynamic temperature modeling for heat transmission between different objects and regions.

Given the selected baseline and state-of-the-art methods, our claims for the experimental validation of the proposed improvements in this section are threefold. First, the manual and georeferencing-based methods can produce local misalignments. Second, a direct adaptation of the mentioned state-of-the-art methods to image registration by using conventional RANSAC and global homography-based transformation can produce local misalignments. Third, the proposed algorithm with the local refinements eliminates the local misalignments, which are not explicitly handled in the baseline and state-of-the-art methods.

Table 6 gives the overall SSIM results for the proposed methods, baseline methods, and state-of-the-art methods for the image pairs captured on the same and different days. The proposed geometric-based local refinement is the best approach among all the methods. The performance of the state-of-the-art methods, namely SuperGlue, D2-Net, and LPM, are in comparable ranges with respect to the manual and global registration. As a specific case, LPM could not match the image pair 2a–2c due to the insufficient number of matches to estimate a regular homography after outlier rejection. As all these compared methods are mainly focused on feature extraction and matching rather than the registration performance between the reference and transformed images, similar results with global registration are quite expected in the experiments. Accordingly, the proposed geometric- and optimization-based refinements after global transformation improve the registration performances in local regions and provide better scores.

Table 6. SSIM results between transformed and reference images for the proposed global, geometric-based, and optimization-based approaches with the baseline and state-of-the-art image registration methods. Note that there are no geo-coordinates provided for the images 3a, 3b, and 3c.

	SSIM for Same-Day Pairs			SSIM for Different-Day Pairs		
	1a–1b	2a–2b	3a–3b	1a–1c	2a–2c	3a–3c
Baseline Methods						
Global	0.1833	0.1244	0.3091	0.0772	0.1054	0.2624
Manual	0.1687	0.1384	0.1933	0.0368	0.1122	0.1302
Georeferencing-based	0.3111	0.3029	-	0.0119	0.0031	-
State of the Art methods						
SuperGlue	0.2255	0.1466	0.2787	0.0620	0.0617	0.2233
D2-Net	0.1788	0.1064	0.2455	0.0325	0.0447	0.1371
LPM	0.1873	0.1465	0.3130	0.0388	x	0.2322
Proposed Refinements						
Geometric-based	0.4223	0.3704	0.3811	0.1097	0.1404	0.3000
Optimization-based	0.3603	0.2569	0.3382	0.0996	0.1676	0.3118

Table 6 also gives the results for the image pairs captured on different days. In fact, all the methods give similar SSIM results, however, with much lower values compared to the ones on the same day. The results indicate that while the proposed global homography-based registration can still be used as an alternative to the manual and georeferencing-based approaches for different-day pairs, the proposed local refinements with geometric and optimization methods are not very useful due to the dissimilarity of the utilized 2D maps in different days.

The proposed geometric- and optimization-based refinements use the SSIM metric between the transformed and reference blocks to find the best blockwise homographies and then present the overall registration quality in terms of the SSIM metric between trans-

formed and reference images, as shown in Table 6. In order to perform a cross-evaluation with a different metric, Table 7 also gives the results in terms of mutual information between the transformed and reference images. Although the gains in terms of MI are not as large as the case of SSIM, over which the optimization is performed, they are still at significant levels allowing similar conclusions regarding the performances of the proposed, baseline, and state-of-the-art methods.

Table 7. Mutual information (MI) results between transformed and reference images for the proposed global, geometric-based, and optimization-based approaches with the baseline and state-of-the-art image registration methods. Note that there are no geo-coordinates provided for the images 3a, 3b, and 3c.

	MI for Same-Day Pairs			MI for Different-Day Pairs		
	1a–1b	2a–2b	3a–3b	1a–1c	2a–2c	3a–3c
Baseline Methods						
Global	0.5300	0.4814	1.0665	0.2453	0.6479	0.8039
Manual	0.5317	0.5132	0.9375	0.2531	0.5642	0.6376
Georeferencing-based	0.5532	0.4777	-	0.1648	0.4383	-
State of the Art methods						
SuperGlue	0.5376	0.4914	1.0460	0.2134	0.5664	0.7702
D2-Net	0.4852	0.4634	0.9663	0.2668	0.5794	0.7070
LPM	0.5292	0.4851	1.0818	0.2176	x	0.7680
Proposed Refinements						
Geometric-based	0.7745	0.6162	1.1075	0.2501	0.6761	0.8384
Optimization-based	0.7260	0.5569	1.1030	0.2592	0.6762	0.8428

6. Discussion

Different 3D–2D conversion methods have been proposed for the registration of LWIR hyperspectral images by considering the radiance, temperature, and emissivity components in the thermal infrared spectrum. The experiments first revealed that the 2D maps of brightness temperature and average radiance energy of hyperspectral pixels are more convenient for extracting and matching points and then estimating the global transformation to align hyperspectral LWIR images. In contrary to the preliminary assumption that the emissivity-based 2D maps can be more invariant in response to temperature changes in the fields, the resulting 2D images obtained from emissivity components were found to have low spatial contrast and, therefore, to be unsuitable for properly extracting and matching the feature points. As another observation, the number of extracted and matched points, and consequently the registration performance, significantly decreased for the pairs of images captured on different days with respect to the pairs of images captured on the same day. The sensor technology for thermal LWIR hyperspectral images is still evolving and is not yet mature. The quality of the hyperspectral images is very dependent on the sensor heatings and other circumstances during the capturing as well as calibration and normalization processes performed at different times on such sensors.

After globally aligning the 2D maps of LWIR images, the proposed method is further improved with blockwise and pixelwise registrations. The experiments for blockwise refinement first indicated that the best similarity metric for the registration of the local blocks is SSIM compared to the MSE and mutual information. While the performance of the MSE was always the lowest for different block sizes, the performance of the mutual information also decreased for lower size blocks due to the lower number of pixels for the estimation of density functions. The pixelwise improvement as a follow-up stage significantly decreased the blocking artifacts in blockwise refinement by properly assigning the correct transformation for each pixel inside the blocks. As another comparison in the

tests, the performance of the proposed geometric-based registration method was found to be better than or comparable to the proposed optimization-based method, while providing a more suitable execution time for practical usage. However, it should be noted that the performance improvement of the geometric-based refinement is directly related to the number of extracted keypoints for registration, as one of its limitations with respect to the optimization-based refinement.

Finally, the improvements in the proposed local refinements were presented in terms of SSIM with respect to the baseline and state-of-the-art methods. The cross-evaluation of the method with a different metric than the utilized metric for optimization also supported the conclusions regarding the improvements. In order to further discuss the contribution of the proposed method with respect to the previous literature, a detailed visual inspection was also performed on different patches of the test images. Table 8 illustrates these examples for different methods by overlapping the transformed patches with the reference patches. While the columns represent different patches, the rows indicate different methods. Gray regions in the overlapped patches indicate where the two images have similar values. Magenta and green colors correspond to the regions where the intensities are different. The regions in green or magenta, indicating the nonoverlapping parts for the patches, are quite visible for manual, global, and GPU-IMU-based registration. The same observation is also valid for the compared methods, SuperGlue [34], D2-Net [32], and LPM [33], which perform the registration with a global transformation. Accordingly, while SuperGlue and D2-Net give mismatches in the lower part of the test images, they are better in aligning the middle parts, as the global homography is estimated with respect to those regions. The reverse situation, where the lower parts are aligned regularly but the upper parts are mismatched, can also occur, as in the given example for LPM.

Table 8. Examples for the overlapped reference and transformed patches for the proposed, baseline, and state-of-the-art methods. The first row indicates the original patches in the reference image.

	Patch1	Patch 2	Patch 3	Patch 4
Reference Patch				
Global				
Manual				

Table 8. Cont.

	Patch1	Patch 2	Patch 3	Patch 4
Georeferencing-based				
SuperGlue				
D2-Net				
LPM				
Geometric-Based Refinement				
Optimization-Based Refinement				

The experiments reveal that the performances of the state-of-the-art methods such as SuperGlue, D2-Net, and LPM are in similar ranges compared to the global registration as all these methods focus on feature extraction and matching. The local refinements after global transformation were one of the neglected aspects of these methods. The proposed methods on the other hand successfully achieve registration on local regions as indicated with overlapped images by performing automatic refinements on local regions after global transformation. Without loss of generality, the proposed refinements can also take the resulting global homographies from the other methods as input and improve the performances of those methods as well. However, the performances of those methods were in comparable ranges, as given in Tables 6 and 7. Therefore, the presentation of the improvements with respect to the performed global transformation was found to be sufficient in the experiments to draw main conclusions.

7. Conclusions

The experiments on the registration of LWIR hyperspectral images in the performed research first indicate that the temperature components are more convenient for feature extraction and matching in thermal range compared to the emissivity component. The application of the mainstream global registration methods to the resulting 2D maps, however, produces local misalignments on the registered images. The proposed blockwise and pixelwise refinements eliminate those misalignments and give better results than manual, global, and GPU-IMU-based registration. In addition, the improvements of the proposed method compared to the state-of-the-art methods such as SuperGlue, D2-Net, and LPM are also reported both visually and in terms of objective similarity metrics, namely SSIM and mutual information. In the view of the current trend of image analysis, the proposed framework for local refinements can be extended with deep learning-based methods. Future work will focus on the potential challenges for the extension of the proposed optimization framework to deep networks.

Author Contributions: A.K. carried out method development, implementation, analysis, interpretation, and manuscript writing. U.E. contributed to implementation of state-of-the-art methods, analysis, interpretation, and manuscript writing. All authors have read and agreed to the published version of the manuscript.

Funding: No funding to be reported.

Acknowledgments: The authors are grateful to the NATO SET 240 Panel Members for the provided LWIR hyperspectral data and discussions regarding local refinements.

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

References

1. Brown, L.G. A survey of image registration techniques. *ACM Comput. Surv.* **1992**, *24*, 325–376. [\[CrossRef\]](#)
2. Liao, R. Deformable Image Registration for Hyperspectral Images. Master's Thesis, Dept. of Electrical and Systems Eng., McKelvey School of Engineering, Washington University, St. Louis, MO, USA, May 2019.
3. Veste, S. Registration in Hyperspectral and Multispectral Imaging. Master's Thesis, Department of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway, January 2017.
4. Mukherjee, A.; Velez-Reyes, M.; Roysam, B. Interest Points for Hyperspectral Image Data. *IEEE Trans. Geosci. Remote. Sens.* **2009**, *47*, 748–760. [\[CrossRef\]](#)
5. Goncalves, H.; Corte-Real, L.; Goncalves, J.A. Automatic Image Registration Through Image Segmentation and SIFT. *IEEE Trans. Geosci. Remote. Sens.* **2011**, *49*, 2589–2600. [\[CrossRef\]](#)
6. Sima, A.A.; Buckley, S.J. Optimizing SIFT for Matching of Short Wave Infrared and Visible Wavelength Images. *Remote. Sens.* **2013**, *5*, 2037–2056. [\[CrossRef\]](#)
7. Yi, Z.; Zhiguo, C.; Yang, X. Multi-spectral remote image registration based on SIFT. *Electron. Lett.* **2008**, *44*, 107–108. [\[CrossRef\]](#)
8. Vural, M.F.; Yardimci, Y.; Temizel, A. Registration of Multispectral Satellite Images with Oriented-Restricted SIFT. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July 2009.
9. Ordóñez, A.; Heras, D.B.; Argüello, F. Surf-Based Registration for Hyperspectral Images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 63–66.
10. Rzhannov, Y.; Pe'Eri, S. Pushbroom-Frame Imagery Co-Registration. *Mar. Geod.* **2012**, *35*, 141–157. [\[CrossRef\]](#)
11. Dorado-Munoz, L.P.; Velez-Reyes, M.; Mukherjee, A.; Roysam, B. A Vector SIFT Detector for Interest Point Detection in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4521–4533. [\[CrossRef\]](#)
12. Dorado-Munoz, L.P.; Velez-Reyes, M.; Roysam, B.; Mukherjee, A. Interest point detection for hyperspectral imagery. In Proceedings of the SPIE Defense, SPIE-Intl Soc Optical Eng: Bellingham, WA, USA, 2009; Volume 7334, p. 73340O.
13. Kern, J.P.; Pattichis, M.; Stearns, S.D. Registration of image cubes using multivariate mutual information. In Proceedings of the The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003.
14. Luo, X.; Guo, L.; Yang, Z. Registration of Remn ote Sensing Hyperspectral Imagery Using Mutual Information and Stochastic Optimization. *Remote Sens. Technol. Appl.* **2006**, *21*, 61–66.
15. Hasan, M.; Pickering, M.; Robles-Kelly, A.; Zhou, J.; Jia, X. Registration of hyperspectral and trichromatic images via cross cumulative residual entropy maximization. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 3 December 2010; pp. 2329–2332.
16. Zhou, Y.; Rangarajan, A.; Gader, P.D. Nonrigid Registration of Hyperspectral and Color Images with Vastly Different Spatial and Spectral Resolutions for Spectral Unmixing and Pansharpening. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1571–1579.

17. Zhou, Y.; Rangarajan, A.; Gader, P.D. An Integrated Approach to Registration and Fusion of Hyperspectral and Multi-spectral Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3020–3033. [\[CrossRef\]](#)
18. Pluim, J.P.W.; Maintz, J.B.A.; Viergever, M.A. Mutual Information based registration of medical images. *IEEE Trans. Med Imaging* **2003**, *22*, 986–1004. [\[CrossRef\]](#)
19. Kim, B.H.; Kim, M.Y.; Chae, Y.S. Background Registration-Based Adaptive Noise Filtering of LWIR/MWIR Imaging Sensors for UAV Applications. *Sensors* **2017**, *18*, 60. [\[CrossRef\]](#)
20. Steven, V.; Hoecke, V.; Nele, T.; Bart, M.; Bart, S.; Peter, L.; Charles-Frederik, H.; Rik, V.D.W. Hot Topics in Video Fire Surveillance. In *Video Surveillance*; InTech: London, UK, 2011.
21. Koz, A.; Çalışkan, A.; Aydın, A.; Alatan, A. Registration of MWIR-LWIR Band Hyperspectral Images. In Proceedings of the 8th Work-shop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Los Angeles, CA, USA, 21–24 August 2016.
22. Yang, K.; Pan, A.; Yang, Y.; Zhang, S.; Ong, S.H.; Tang, H. Remote Sensing Image Registration Using Multiple Image Features. *Remote Sens.* **2017**, *9*, 581. [\[CrossRef\]](#)
23. Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.; Tian, J. Robust Feature Matching for Remote Sensing Image Registration via Locally Linear Transforming. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6469–6481. [\[CrossRef\]](#)
24. Abbas, M.; Saleem, S.; Subhan, F.; Bais, A. Feature points-based image registration between satellite imagery and aerial images of agricultural land. *Turk. J. Electr. Eng. Comput. Sci.* **2020**, *28*, 1458–1473. [\[CrossRef\]](#)
25. Mostafa, M.M.R.; Hutton, J. Direct Positioning and Orientation Systems. In How do they work? What is the attainable accuracy? In Proceedings of the American Society of Photogrammetry and Remote Sensing Annual Meeting, St. Louis, MO, USA, 24–27 April 2001.
26. Cramer, M.; Stallmann, D.; Haala, N. Direct Georeferencing Using GPS/Inertial Exterior Orientations for Photogrammetric Applications. In Proceedings of the XIXth International Society for Photogrammetry and Remote Sensing Congress, Amsterdam, The Netherlands, 16–23 July 2000.
27. Yuan, X.; Zhang, X. Theoretical Accuracy of Direct Georeferencing with Position and Orientation System in Aerial Photogrammetry. In Proceedings of the XXith, International Society for Photogrammetry and Remote Sensing Congress, Beijing, China, 3–11 July 2008.
28. Efe, U.; Gokalp, K.; Ince, A.; Alatan, A. DFM: A Performance Baseline for Deep Feature Matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Virtual, 25 June 2021; pp. 4284–4293.
29. Rocco, I.; Cimpoi, M.; Arandjelovic, R.; Torii, A.; Pajdla, T.; Sivic, J. Ncnet: Neighbourhood consensus networks for estimating image correspondences. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Zhou, Q.; Sattler, T.; Leal-Taixe, L. Patch2pix: Epipolar-guided pixel-level correspondences. Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021.
31. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Self-supervised interest point detection and description. In Proceedings of the Computer Vision and Pattern Recognition Workshop, Salt Lake City, Utah, 18–22 June 2018.
32. Dusmanu, M.; Rocco, I.; Pajdla, T.; Pollefeys, M.; Sivic, J.; Torii, A.; Sattler, T. D2-Net: A Trainable CNN for Joint Description and Detection of Local Features. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 16–20 June 2019; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA; pp. 8084–8093.
33. Ma, J.; Zhao, J.; Guo, H.; Jiang, J.; Zhou, H.; Gao, Y. Locality Preserving Matching. *Int. J. Comput. Vis.* **2019**, *127*, 512–531. [\[CrossRef\]](#)
34. Sarlin, P.-E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching with Graph Neural Networks. In Proceedings of the 2020 Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4937–4946.
35. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Deep image homography estimation. In Proceedings of the RSS Workshop: Limits and Potentials of Deep Learning in Robotics, Ann Arbor, MI, USA, 18 June 2016.
36. Nguyen, T.; Chen, S.W.; Shivakumar, S.S.; Taylor, C.J.; Kumar, V. Unsupervised Deep Homography: A Fast and Robust Homography Estimation Model. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2346–2353. [\[CrossRef\]](#)
37. Viola, P.; Lii, W.M.W. Alignment by Maximization of Mutual Information. *Int. J. Comput. Vis.* **1997**, *24*, 137–154. [\[CrossRef\]](#)
38. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error measurement to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
40. Bay, H.; Ess, A.; Tuytelaars, T.; Gool, L.V. SURF: Speeded Up Robust Features. *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [\[CrossRef\]](#)
41. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
42. Hackwell, J.A.; Warren, D.W.; Bongiovi, R.P.; Hansel, S.J.; Hayhurst, T.L.; Mabry, D.J.; Sivjee, M.G.; Skinner, J.W. LWIR/MWIR imaging hyperspectral sensor for airborne and ground-based remote sensing. In Proceedings of the Imaging Spectrometry II; SPIE-Intl Soc Optical Eng: Bellingham, WA, USA, 1996; Volume 2819, pp. 102–108.

43. Pipia, L.; Aragüés, F.P.; Tardà, A.; Martinez, L.; Palà, V.; Arbiol, R. Thermal Airborne Spectrographic Imager for Temperature and Emissivity Retrieval. In Proceedings of the 3rd International Symposium on the Recent Advances in Quantitative Remote Sensing, Valencia, Spain, 27 September–1 October 2010.
44. Hartley, R.; Zisserman, A.; Faugeras, O. *Multiple View Geometry in Computer Vision*; Cambridge University Press (CUP): Cambridge, UK, 2004.
45. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Read. Comput. Vis.* **1987**, *24*, 726–740. [[CrossRef](#)]
46. Koz, A.; Soydan, H.; Duzgun, H.S.; Alatan, A.A. A local extrema based method on 2D brightness temperature maps for detection of archaeological artifacts. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2016; pp. 92–95.
47. Smith, R.B. Computing the Planck Function. 2003. Available online: <http://yceo.yale.edu/sites/default/files/files/ComputingThePlanckFunction.pdf> (accessed on 23 June 2021).
48. Dorado-Munoz, L.P.; Velez-Reyes, M.; Mukherjee, A.; Roysam, B. A Temperature and Emissivity Separation Algorithm for Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) images. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1113–1126.
49. Harris, C.; Stephens, M. A Combined Corner and Edge Detector. In Proceedings of the Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.
50. Terriberry, T.B. Adaptive motion compensation without blocking artifacts. In Proceedings of the Visual Information Processing and Communication VI, San Francisco, CA, USA, 10–12 February 2015; Volume 9410. [[CrossRef](#)]
51. Vedaldi, A.; Fulkerson, B.M. VLFeat: An open and portable library of computer vision algorithms. In Proceedings of the 18th ACM international conference on Multimedia, Firenze, Italy, 25–29 October 2010.
52. Matlab Hyperspectral Toolbox by Isaac Greg. Available online: <https://github.com/isaacgreg/matlabHyperspectralToolbox/tree/master/hyperspectralToolbox> (accessed on 23 June 2021).