

HSI-MSER: Hyperspectral Image Registration Algorithm Based on MSER and SIFT

Álvaro Ordóñez , Álvaro Acción , Francisco Argüello , and Dora B. Heras , *Member, IEEE*

Abstract—Image alignment is an essential task in many applications of hyperspectral remote sensing images. Before any processing, the images must be registered. Maximally stable extremal regions (MSER) is a feature detection algorithm that extracts regions by thresholding the image at different grey levels. These extremal regions are invariant to image transformations making them ideal for registration. The scale-invariant feature transform (SIFT) is a well-known keypoint detector and descriptor based on the construction of a Gaussian scale-space. This article presents a hyperspectral remote sensing image registration method based on MSER for feature detection and SIFT for feature description. It efficiently exploits the information contained in the different spectral bands to improve the image alignment. The experimental results over nine hyperspectral images show that the proposed method achieves a higher number of correct registration cases using less computational resources than other hyperspectral registration methods. Results are evaluated in terms of accuracy of the registration and also in terms of execution time.

Index Terms—Hyperspectral imaging, image registration, maximally stable extremal regions (MSER), remote sensing, scale-invariant feature transform (SIFT).

I. INTRODUCTION

CURRENTLY, hyperspectral image (HSI) from remote sensing are widely available thanks to the increased availability of sensors. This allows us to obtain images of the same region of the Earth taken from different viewpoints at different times. The series of remote-sensing images are used in applications where it is essential to compare, study, or find differences between images. Automatic change detection [1], environmental monitoring [2], or super-resolution image creation [3], among others are applications in which registration is a fundamental

prior task [4]. The images must be previously aligned in order to work with them afterwards.

Image registration algorithms can be classified into area-based and feature-based methods [4]. Methods in the first group work directly with image intensity, e.g., Fourier Transform [5] and mutual information [6], while those in the second group, feature-based methods, look for information at a higher level, e.g., at the level of regions, lines, or points [7]–[10]. This property makes feature-based methods more suitable for images with illumination changes, which it is the case of remote sensing HSIs of the earth surface. For these images, the atmospheric conditions usually vary from one capture to another.

Generally, feature-based methods consist of the following four stages: feature detection, feature description, feature matching, and image transformation [11]. These methods rely on extracting the same features in the images to be registered. Knowing a number of corresponding features, an image transformation that aligns one image with respect to the other can be calculated.

Maximally stable extremal regions (MSER) [7] is a feature-based method for region detection in images that can be used for extracting the features needed for a later registration process. This method extracts regions, called extremal regions (ERs), by thresholding the image at different grey levels and according to a stability criterion. If MSER is applied to a pair of images, the extracted and matched regions of both images can be used to compute an image transformation to register them. MSER is resilient to changes of scale, rotation, translation, and illumination conditions. Other well-known feature detector methods in the literature are features from accelerated segment test (FAST) [8] and speeded up robust features scale-invariant feature transform (SURF) [9]. Both build a scale-space where scale-invariant points are detected. But the most popular feature detector and descriptor algorithm is the scale-invariant feature transform (SIFT) [10]. SIFT extracts keypoints from a multiresolution pyramid of the images created performing Gaussian convolutions and interpolations. Its descriptor stands out for being highly distinctive and invariant to illumination and distortion changes, making it a widely used method in the literature [12], [13].

The literature indicates that MSER and SIFT are two of the best region detector [14] and descriptor algorithms [15], respectively. They are used not only for image registration [16] but also in many other applications such as object recognition [17], [18], image retrieval [19], [20], and robot localization [21], [22], among others. MSER is also often used in the literature with the local affine frames (LAF) descriptor [18], [23].

Manuscript received May 31, 2021; revised August 24, 2021 and October 8, 2021; accepted November 16, 2021. Date of publication November 18, 2021; date of current version December 8, 2021. This work was supported in part by the Ministerio de Ciencia e Innovación, Government of Spain, under Grant PID2019-104834GB-I00, in part by the Consellería de Cultura, Educación e Universidade under Grant ED431C 2018/19 and Grant accreditation 2019-2022 ED431G-2019/04, and in part by the Junta de Castilla y León under Project VA226P20 (PROPHET II Project). All are co-funded by the European Regional Development Fund (ERDF). The work of Álvaro Ordóñez was also supported by Ministerio de Universidades, Government of Spain, under a FPU Grant FPU16/03537. (*Corresponding author: Álvaro Acción.*)

Álvaro Ordóñez, Álvaro Acción, and Dora B. Heras are with the Centro Singular de Investigación en Tecnoloxías Intelixentes (CITIUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain (e-mail: alvaro.ordonez@usc.es; alvaro.accion.montes@usc.es; dora.blanco@usc.es).

Francisco Argüello is with the Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain (e-mail: francisco.arguello@usc.es).

Digital Object Identifier 10.1109/JSTARS.2021.3129099

SIFT and MSER-based methods have been previously proposed for image registration [24]–[27]. However, the exploitation of the whole spectral information of HSIs in order to improve the accuracy of the registration process has not been explored and analysed in-depth. To the best of authors' knowledge, the publications that used SIFT and MSER to register HSIs, as will be detailed in the following, separately reduce each of the images to a single band. New algorithms to deal with the spectral information in HSIs in an efficient way from the point of view of registration accuracy and computational cost need to be designed [28], [29]. The use of spectral information also allows registering images that cannot be registered considering only one band. A band selection method is required for selecting those bands that are most relevant for registration, i.e., it should avoid bands with redundant or low-quality information. Another feature of previous works is that they validate the algorithms considering a few scale factors and rotation angles. Unlike them, in this article, an exhaustive evaluation of the registration algorithms is carried out considering different hyperspectral datasets (urban, rural, crops, and nature scenes) and a wide variety of registration parameters (scale factors and rotation angles).

One example of a method that uses the multispectral image as a one-band image, and it is only evaluated on one pair of images is proposed in [24]. The authors propose a method to register a pair of multispectral and visible spectrum images using the keypoints detected and described by SIFT and the regions independently detected and described by MSER and LAF, respectively. Guo *et al.* [25] presented a multispectral remote sensing algorithm based on MSER for region detection. The regions are described twice using the SIFT descriptor and using a shape descriptor based on the Fourier transform. The method is evaluated on a SPOT2 multispectral and a SPOT2 pan image with different resolution. Moreover, the rotational invariance is not extensively tested, as the only angle difference considered is 30° . No exploitation of spectral information is performed. Zhang *et al.* [26] proposed a multisensor registration method that combines MSER and SIFT, and takes into account the number of matches used and their distribution in order to improve the final transformation. For the experimental analysis, they register two bands of Landsat multispectral images, and a panchromatic image with respect to a RADARSAT SAR image. The first pair has translation changes, and the second has a different scale factor and rotation angle. Liu *et al.* [27] presented edge-enhanced MSER, a method for multisensor image registration also based on MSER and SIFT. Before detecting and describing the regions, an edge enhancement is applied to the image pairs to obtain more stable matches. The method is evaluated on a pair of images with a scale factor of $0.7\times$ and a rotation angle of 35° . No use of the information from the different bands is detailed, as in previous works.

In this work, a registration method for HSIs based on MSER for region detection and SIFT for region description is presented. It exploits the spectral information available in the HSIs by performing feature detection and description in several pre-selected bands of the images to be registered and by incorporating spectral information into the descriptor. The algorithm was

designed to deal with extreme situations in terms of scale factor and rotation angle.

The main contributions of this article are the following.

- 1) An efficient registration method that adapts MSER and SIFT to efficiently exploits the spectral information available in HSIs is proposed. The method exploits spectral information in two ways. First, by extracting regions in several bands. Second, the regions are described with a descriptor composed of a spatial and a spectral part as it helps to discard false matches.
- 2) The proposal includes a band selection method specifically designed for the HSI registration problem. It selects bands according to their entropy and wavelength, and achieves better results than other band selection methods in the literature.
- 3) An exhaustive histogram-based search is used to estimate the registration parameters. All the possible combinations between the matched regions are taken into account. The method selects the best transformation considering all candidates.
- 4) An in-depth evaluation of the method is carried out. The evaluation is performed over nine pairs of HSIs taken by different sensors at different dates. Moreover, the set of images is extended by applying a wide range of scale factors and rotation angles.

The rest of this article is organized as follows. Section II describes the different stages of the proposed method; the results are discussed in Section III. Finally, Section IV concludes this article.

II. HYPERSPECTRAL MSER

In this section, we present hyperspectral MSER (HSI-MSER), a registration method to align two hyperspectral remote sensing images based on MSER as region detector followed by SIFT as feature descriptor. The method exploits the spectral information available in the images. First, the standard versions of MSER and SIFT are described. Then, the proposed method is presented.

A. Maximally Stable Extremal Region

MSER is a method for region detection in greyscale images [7]. It has been successfully applied to a large number of applications such as image recognition, tracking, and image registration [30], [31]. The algorithm extracts a number of regions called MSERs by thresholding the image at different grey levels and according to a stability criterion. An ER is a region in which all pixels within the ER have higher intensity values (for the bright ERs) or lower intensity values (for the dark ERs) than all the pixels on the outer boundary of the region. The outer boundary of a region is defined as the set of pixels that meet two conditions: being adjacent to one, two, or three pixels in the region and not belonging to the region. An ER is considered stable (an MSER) when it does not change substantially as the grey level threshold is varied [7].

MSER presents two properties that make it ideal for image registration [14]. First, linear or affine transformations do not

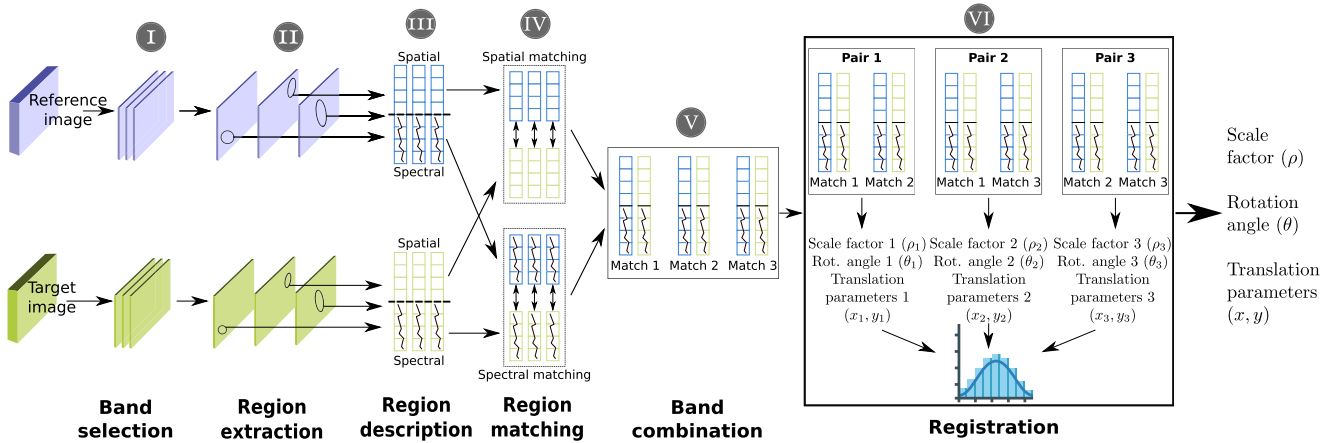


Fig. 1. Proposed HSI-MSER schematic for registration of two HSIs.

affect the extracted ERs because they only depend on pixel intensities that are preserved under these monotonic transformations. Second, a set of regions is preserved after applying geometric and photometric changes because an ER will continue to be an ER after these transformations.

A reference implementation of MSER can be found in [32] although it only considers RGB images in the range $[0,255]$. In this work, the original range of the input HSIs is considered and the algorithm is extended to an arbitrary number of bands.

B. SIFT Descriptor

The SIFT is one of the most popular feature detectors and descriptors. In this work, we use the descriptor part to calculate the description of each region previously detected by MSER.

The steps to compute the SIFT descriptor of each region are the following. First, the dominant angles are calculated for each region to achieve invariance to image rotation. An orientation histogram with 36 bins covering 360° is created from the gradient orientations within the surface of each region. Then, it is weighted by gradient magnitude and by a Gaussian-weighted circular window. The highest peak in the histogram is selected as the dominant orientation. Moreover, any peak above 80% of it is also taken into account. That means that we will have regions with the same location but different orientations.

The next step is the descriptor construction. First, an area of size 16×16 pixels centred around each region is selected. The area is divided into subareas of 4×4 pixels. For each subarea, an orientation histogram with 8 bins is created. Finally, a 128-parameter descriptor for each region is generated from this set of weighted histograms. To reduce the influence of boundary effects, brightness and illumination changes, the descriptor values are thresholded and normalized to unit length.

C. HSI-MSER

In this section, we present HSI-MSER, a registration algorithm for HSIs based on MSER and SIFT. HSI-MSER seeks to find a similarity transformation that successfully aligns two hyperspectral remote sensing images. One of the images is

called the reference image. The other, the image that we want to register with respect to the reference image, is called the target image. The method consists of the following six stages: band selection, region extraction, region description, region matching, band combination, and registration. A schematic of the proposed algorithm can be seen in Fig. 1.

1) *Band Selection*: It is common that contiguous bands do not differ in relevant information for the registration process. Band selection allows reducing the computational cost with respect to considering all the bands of the image but keeping only the relevant information. For this reason in the first stage, the most relevant spectral bands of the reference and target images are selected.

Out of the methods available in the literature for band selection of HSIs, principal component analysis (PCA) [33], Band-Clust [34], and ward's linkage strategy using mutual information (WaluMI) [35] were evaluated. All perform the selection by considering each HSI individually. The entropy-based selection method used in this work considers both images of the dataset jointly. The method consists in selecting the N bands with the highest entropy but separated by at least D consecutive bands. As the bands forming an HSI are ordered by wavelength, D is the minimum number of bands between each pair of selected ones. This ensures that the selected bands differ in both entropy and wavelength. We call this method entropy-based band selection (EBS).

First, the entropy of each band is calculated for both images, i.e., two entropy values are obtained for each band, and the minimum value of each pair is assigned to the band. Then, the bands are ordered according to decreasing values of entropy. The first band selected is the one with the maximum entropy. The next band selected will be the next with the highest entropy but separated by, at least, a distance of D bands, as indicated earlier. This step is repeated until N bands are selected. If it is not possible to find a band that fulfils this condition, D is reduced by one band and the process is restarted.

2) *Region Extraction*: The second stage consists in extracting regions from the HSIs. The region extraction is applied to each band selected in the previous stage.

Let I be an HSI consisting of $H \times W$ pixels indexed by the variable x and B spectral bands. Let $B_b(x)$ be the grey level value of a pixel x in the selected spectral band b . Let also $L = [\min(B_b(x)), \max(B_b(x))]$, $x \in I$ be the grey level range in the b band. The extracted regions in the b band for the greyscale level $l \in L$ are transformed into ellipses

$$\mu_l = \frac{1}{|R_l|} \sum_{x \in R_l} x, \quad \Sigma_l = \frac{1}{|R_l|} \sum_{x \in R_l} (x - \mu_l)^\top (x - \mu_l) \quad (1)$$

where μ_l and Σ_l are the mean and variance of the pixels composing the region, and R_l is an ER detected in this band for the greyscale level l [36].

The aim of this stage is to detect a large number of common structures in both HSIs that will then be used to calculate the transformation to align the images. HSI-MSER is specifically designed to deal with spectral information because some structures are only perceptible in some bands.

3) *Region Description*: In the third stage, the extracted regions are described. The algorithm is designed to register HSIs. This requires that the descriptor is made up of spatial and spectral information in order to achieve better alignments.

The SIFT descriptor is used for the spatial part. For each ellipse, the coordinates of its centre are considered as the coordinates of the region. The SIFT descriptor is computed on the surface of the region bounded by the ellipse as explained in Section II-B.

This spatial descriptor is enriched by a spectral part, in particular, the spectral signature of the keypoint, which is defined as the pixel vector of the centre of the ellipse since the regions extracted by MSER are homogeneous. Both parts are concatenated to form a descriptor that takes into account both the spatial and the spectral information. The descriptor is a vector made up of 128 components for the spatial part plus the number of bands of the original HSIs as the spectral part.

4) *Region Matching*: Then, in the fourth stage, the regions of each pair of bands (one band for each HSI) are matched independently, i.e., without taking into account the regions of the other pairs of selected bands. Although the bands are matched independently, the spectral information of the other bands has been taken into account. In particular, the Euclidean distance is used for the spatial part of the descriptor, and the cosine similarity for the spectral part.

The process for matching the regions of both images consists in calculating the distances between their descriptors. Given a region in the reference image, the best candidate match in the target image is the one with the closest distance. However, some regions are not detected in the target image, which means that we will get a false match. Therefore, a method is needed to discard false matches.

The method for region matching consists of two steps. First, the Euclidean distances between each region of the reference band and all of the regions of the target band are computed. The Euclidean distances are calculated on the spatial part of the descriptor, i.e., the SIFT descriptor. Given a region in the reference image, the region closest to it in the target image is considered a possible match if the ratio between the distances to

the two closest regions is smaller than a distance D_{spatial} . Second, to finally be considered a match, the cosine similarity between the spectral signatures of the centre of the regions must be higher than S_{spectral} . The spectral information allows discarding false matches in this second step.

D_{spatial} and S_{spectral} were experimentally set at 0.7 and 0.95, respectively. These are the tradeoff values that achieve good results in terms of number of successfully registered cases for the whole dataset with a moderate computational cost.

5) *Band Combination*: As some regions are only detected in some bands, all matched regions extracted from the selected bands are joined in the fifth stage, i.e., all the regions extracted from the different bands are considered in the same pool.

Thus, regions that are only present in some bands are used to compute the transformation, i.e., all the spectral information is considered together.

6) *Registration*: Finally, in the sixth stage, an exhaustive histogram-based search is performed to register the images. The method computes a possible transformation for each combination of two matched regions. A selection is then carried out based on all the rotation angles and scale factors obtained.

The procedure is as follows. First, a scale factor, a rotation angle, and translation parameters are computed from each combination of two matched regions, as can be seen in Fig. 1. Second, a histogram with the rotation values is calculated. As we want a robust method against rotations, the 360° have been divided into 72 bins, i.e., the bin size is 5° . It was selected based on experiments and following the recommendations by [37]. This allows obtaining bins with a considerable number of elements (transformations), higher accuracy in terms of rotation angle for a first estimation, and a well-defined peak. Bin sizes of 2.5° and 10° were also considered, obtaining worse results in terms of number of successfully registered cases.

Moreover, a 2.5° overlap between bins has been defined. The overlap of 2.5° allows having a flexible boundary between bins, so each angle could contribute to different bins, for example, an angle of 7.5° contributes to the bins centred in 5° and 10° .

Once the histogram is built, the elements (transformations) of the bin of highest frequency are sorted by the scale factor to obtain the median, which is a measure that is more robust to outliers than the mean. Finally, the scale factor ρ , rotation angle θ , and translation parameters (x, y) of the median element are selected to register the HSIs.

III. RESULTS

In this section, the results obtained by the HSI-MSER method using different hyperspectral remote sensing images are presented. First, the experimental conditions and test images are described in Section III-A. In Section III-B, the proposed band selection method is compared to other methods in the literature. Then, in Section III-C, an analysis exploiting different numbers of bands in the first stage of the algorithm (see Fig. 1) is carried out. In Section III-D, HSI-MSER is compared to other hyperspectral registration algorithms in the literature in terms of number of successfully registered cases,

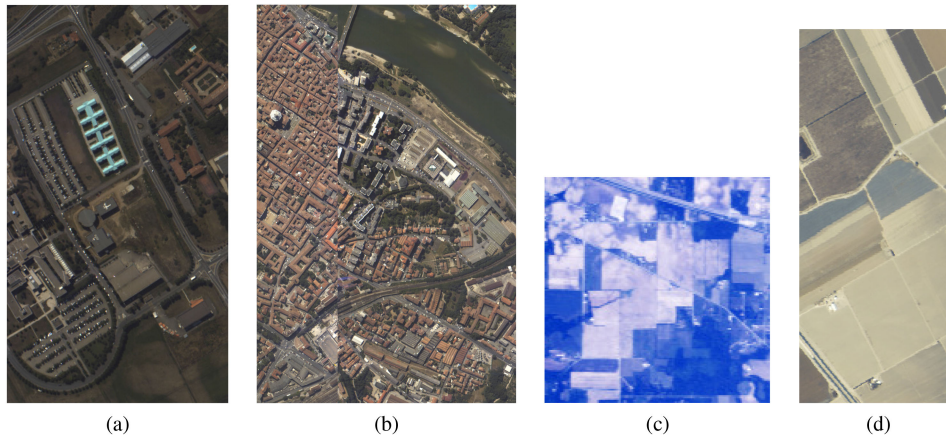


Fig. 2. HSIs commonly used for testing in remote sensing: (a) Pavia University, (b) Pavia Centre, (c) Indian Pines, (d) Salinas.

correct number of matches and root-mean-square error (RMSE) among other common measures of registration accuracy. Finally, in Section III-E, the execution times of HSI-MSER are presented and compared to those of other algorithms in the literature.

A. Experimental Conditions and Dataset

This section presents the experimental conditions and test images as well as some experimental results. The experiments were carried out on a PC with a quad-core Intel i7-4790 CPU at 3.60 GHz and 24 GB of RAM. The code was written in C and compiled using the gcc and the g++ 7.5.0 versions under Ubuntu 18.04.

The test procedure consists in registering one image, called reference image, with respect to a second image, called target image, which presents changes of scale, rotation, and translation. The evaluation of the proposed method was performed over nine hyperspectral remote-sensing scenes [38] that can be divided into two groups: images frequently used in the literature, for which only one image is available, and pairs of images taken by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor at different dates.

The first group contains scenes of rural places and cities. The target image, the image we want to align with respect to the original, is generated by scaling and rotating the original images (called reference images). In this way, we can investigate all the registration details in controlled conditions. The generation of the target images will be explained in more detail later. A colour composition of these images is presented in Fig. 2.

- 1) *Pavia University*: It is an urban area surrounding the University of Pavia, Italy, taken by the reflective optics system imaging spectrometer (ROSIS) sensor and it is made up of 103 spectral bands. The image size is 610×340 pixels with a spatial resolution of 1.3 m/pixel.
- 2) *Pavia Centre*: It is an ROSIS image of 102 bands of the city of Pavia, Italy. The image size is 1096×715 pixels with a spatial resolution of 1.3 m/pixel. A 381-pixel-wide black vertical band in the middle of the image was removed.

- 3) *Salinas*: It is a rural scene taken by the AVIRIS sensor in the Salinas Valley, California. This image has a size of 512×217 pixels with a spatial resolution of 3.7 m/pixel and 204 spectral bands because 20 bands covering the region of water absorption were removed [39]: 108–112, 154–167, and 224.
- 4) *Indian Pines*: It was collected by the AVIRIS sensor over a rural area in NW Tippecanoe County, Indiana. It is a 220-band image because four noisy bands were removed [40]: 1, 33, 97, and 161. The image size is 145×145 pixels with a spatial resolution of 20 m/pixel.

The second group consists of pairs of images of the same urban or rural scene taken at different dates by the AVIRIS sensor. Due to this, changes in infrastructure, vegetation, buildings, crops, and others, are present, as well as, different scale factors, rotation angles, and translations. The oldest image of each pair is used as reference image and the most recent one as target image. A colour composition of these images is presented in Fig. 3.

- 1) *Jasper Ridge*: As their name implies, these images cover a region of the Jasper Ridge Biological Preserve, California. They include a forest, a lake, some buildings, vegetation, roads, and parts of bare soil, among others. The first image was taken in 2006 with a spatial resolution of 3.3 m/pixel and the second one in 2007 with a spatial resolution of 3.4 m/pixel. The second image has a drier appearance with respect the first one. The size of the images is 1286×588 pixels and 224 spectral bands.
- 2) *Santa Barbara Box*: These images were taken in a cultivation area near the city of Santa Barbara, California. The first image was taken in 2009 with clear weather and the second one in 2010 with a few clouds. Changes in crops are also present. There have been changes in the crops from one year to the next. Both images have a size of 1024×769 pixels with a spatial resolution of 15.2 m/pixel and 224 spectral bands.
- 3) *Santa Barbara Front*: These images were collected over the coastal city of Santa Barbara, California. They include an urban area with numerous buildings, some vegetation, as well as the coastline. They are 224-band images with a

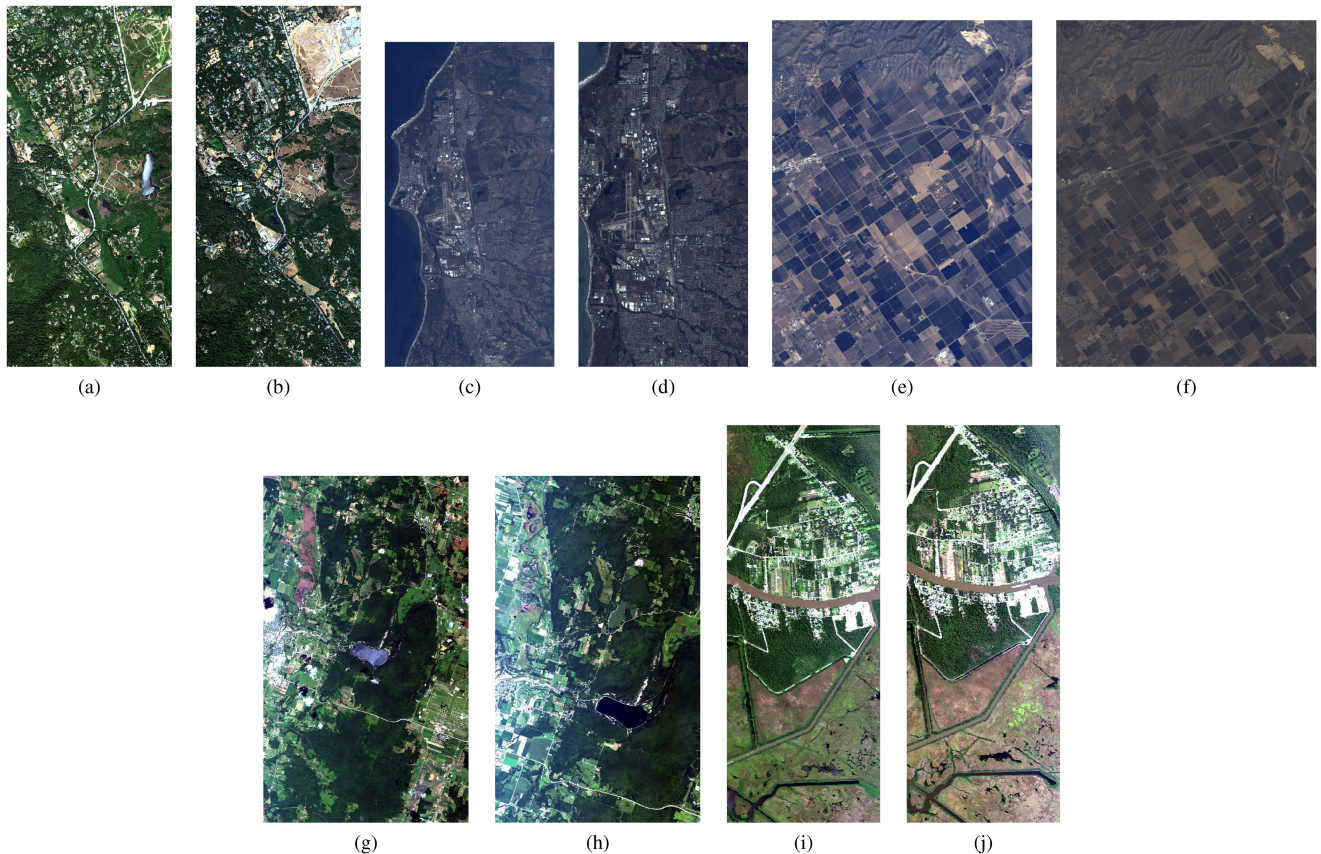


Fig. 3. Jasper Ridge, Santa Barbara, Baraboo Hills, and Crown Point scenes taken for the AVIRIS sensor. (a) Jasper image taken on May 12, 2006, (b) Jasper image taken on August 13, 2007, (c) Santa Barbara Front taken on March 30, 2009, and (d) fragment of the second Santa Barbara Front image taken on April 30, 2010, (e) Santa Barbara Box image taken on April 11, 2013, (f) Santa Barbara Box image taken on April 16, 2014, (g) Baraboo Hills image taken on July 13, 2008, (h) Baraboo Hills image taken on August 26, 2010, (i) Crown Point image taken on September 30, 2010, (j) Crown Point image taken on November 16, 2011.

size of 900×470 . The first image was taken in 2013 and the second one in 2014. Both images differ in the spatial resolution, which is 16.4 m/pixel for the first image and 11.3 m/pixel for the second one.

- 4) *Baraboo Hills*: As the name indicates, these images were taken in Baraboo, Wisconsin. In the image, we can see a part of the city of Baraboo, as well as the Devil's Lake State. They also include forests, vegetation, crops, clouds, and roads, among others. The first image was taken in 2008 with a spatial resolution of 16.9 m/pixel and the second one in 2010 with a spatial resolution of 11.7 m/pixel. Both are 224-band images with a size of 1200×710 .
- 5) *Crown Point*: These images were taken in Crown Point, which is a rural neighborhood located in Marrero, Louisiana. In addition to the urban area, they also include vegetation, a river, canals, and roads. Both images have a size of 1400×540 pixels and 224 spectral bands. The first image was taken in 2010 with a spatial resolution of 3.5 m/pixel and the second one in 2011 with a spatial resolution of 3.4 m/pixel.

In order to evaluate the registration capabilities under extreme scaling and rotating conditions, and investigate all the registration details also in controlled conditions, a comprehensive set of target images was created. This set was generated applying

a range of scale factors from $1/9 \times$ to $16.5 \times$ (40 scale factors) and rotation angles from 0° to 360° in increments of 5° (72 angles). From $1.0 \times$ to $16.5 \times$ the images are scaled up in steps of $0.5 \times$, as shown in Fig. 4, while from $1/2 \times$ to $1/9 \times$ the images are scaled down (scale factors of $1/i \times$, $i = 2, 3, \dots, 9$).

In the case of the images in the second group, these parameters are applied to the most recent image, as mentioned above. In all cases, the target images are trimmed on the central region to keep the same size as the original images. The test consists in registering each target image (the generated ones) with respect to the reference image.

The registration algorithm obtains angle, scale, and translation parameters as output. The registration is considered correct if the parameters obtained by the algorithm are the same as the original values.

B. Evaluation of Band Selection Methods

In this section, the evaluation of different band selection methods in the first stage of the proposed algorithm is presented. EBS is compared to PCA [33], BandClust [34], and WaluMI [35].

PCA is a well-known dimensionality reduction method. It generates a new set of linearly uncorrelated variables where the first few retain most of the data variation [33]. The idea is to

TABLE I
SUCCESSFULLY REGISTERED CASES FOR EACH SCENE USING DIFFERENT BAND SELECTION METHODS IN THE FIRST STAGE OF HSI-MSER: A RANDOM BAND OF EACH SCENE (IN THIS CASE BAND 88), PCA, BANDCLUST, WALUMI, AND EBS

Scene	Band 88	PCA	BandClust	Walumi	EBS 8
Pavia University	1/4× to 7.0× (16)	1/3× to 8.0× (17)	1/5× to 11.0× (25)	1/7× to 13.0× (31)	1/7× to 12.0× (29)
Pavia Centre	1/5× to 7.5× (18)	1/5× to 7.0× (17)	1/8× to 12.5× (31)	1/9× to 14.0× (35)	1/8× to 15.0× (36)
Indian Pines	1/2× to 5.0× (10)	3.0× (5)	1/3× to 8.0× (17)	1/2× to 8.0× (16)	1/3× to 8.0× (17)
Salinas	1/4× to 4.5× (11)	5.5× (10)	1/6× to 8.5× (21)	1/6× to 8.5× (21)	1/5× to 7.0× (17)
Jasper Ridge	(0)	1.5× (2)	1/4× to 7.0× (16)	1/3× to 5.0× (11)	1/3× to 7.0× (15)
Santa Barbara Front	1/2× to 4.5× (9)	1.5× (2)	1/3× to 7.5× (16)	1/2× to 7.5× (15)	1/3× to 7.0× (15)
Santa Barbara Line	1/2× to 2.0× (4)	1.5× (2)	1/5× to 6.5× (16)	1/4× to 6.0× (14)	1/5× to 7.0× (17)
Average number of scalings	(9.71)	(7.86)	(20.29)	(20.43)	(20.86)

The range indicates the scales successfully registered for the 72 angles. The numbers in parentheses summarize the number of scales that were correctly registered for all angles. If an angle is incorrectly registered, the whole scale factor is considered incorrect, i.e., this case is not included in the table. The registration is considered correct if the parameters obtained by the algorithm are the same as the original values.

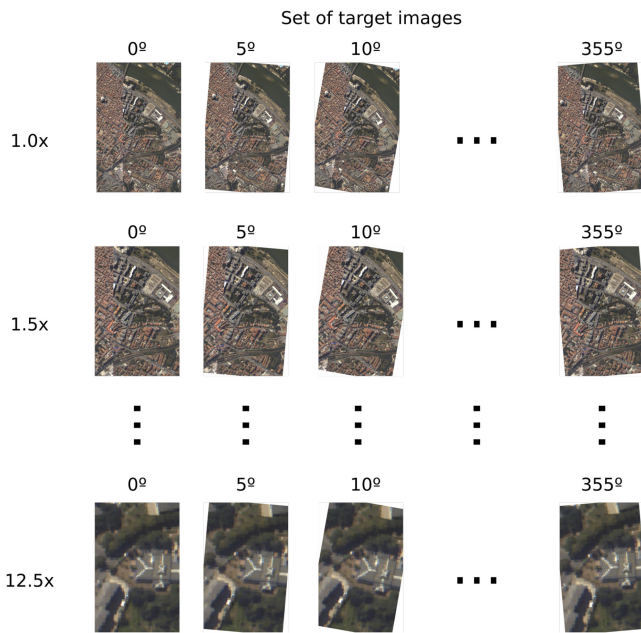


Fig. 4. Method for creating the comprehensive set of target images by applying different scales and rotations.

eliminate data redundancy while preserving relevant information.

BandClust is an unsupervised recursive binary band-splitting algorithm [34]. It iteratively splits a band interval into two disjoint contiguous sets based upon a criterion of minimization of the mutual information, i.e., the method automatically determines the optimal number of bands. Finally, the bands of each set are averaged.

Walumi performs a hierarchical clustering based on the Ward's linkage method [35]. It groups bands to minimize the intracluster variance and maximize the intercluster variance. The distance used is based on the mutual information between each pair of bands. In the end, Walumi chooses the most representative band of each cluster.

Table I summarizes the cases that were correctly registered for each scene using a single band, randomly selected for each image, or a set of bands extracted by these methods. As explained in Section III-A, the registration is considered correct

if the parameters obtained by the algorithm are the same as the original values. In the case of the band selection method used in the proposal, EBS, two parameters must be fixed: the number of bands to be selected N and the minimum distance between the selected bands D . N is set to 8, while D is set to 20, as we want to select bands with different wavelengths to keep all the relevant information. This configuration is called EBS 8. The same number of bands have been selected for the state-of-the-art methods, with the exception of BandClust in which the method itself determines the optimal number of bands. The last row of Table I shows the average number of scalings, i.e., the sum of the number of scales per scene that were correctly registered for all angles divided by the number of scenes.

As shown in Table I, better results are obtained when the spectral information is exploited by considering several bands. This allows detecting features that are only present in some bands. Fig. 5 illustrates this statement. It shows an example of matching for two pairs of bands selected by EBS for the Jasper Ridge images. It can be observed that some features are only present and detected in some spectral bands.

PCA is the exception to the rule. It provides worse results than using only one band. The reason is that PCA applies different transformations to the reference image and to the target images obtaining 8 different principal components (PCs) for each one. This results in a small number of common regions.

The results in Table I show that using the proposed band selection method, EBS, 20.86 cases are correctly registered on average (for all the scenes), more than twice the number of cases achieved using only one band (9.71 cases).

C. Results Exploiting Different Numbers of Bands

As explained in Section II-C, HSI-MSER exploits spectral information in two ways. First, by searching for different regions in selected bands, and second, by incorporating the spectral information into the descriptor. In this section, we analyze the effect of selecting a different number of bands in the first stage, i.e., different values for N in EBS will be evaluated. D is set to 20 as in the previous section.

The test procedure is as explained in the previous section. A total of 40 scale factors from 1/9× to 16.5× and 72 rotation angles from 0° to 360° are applied to the target image. Table II

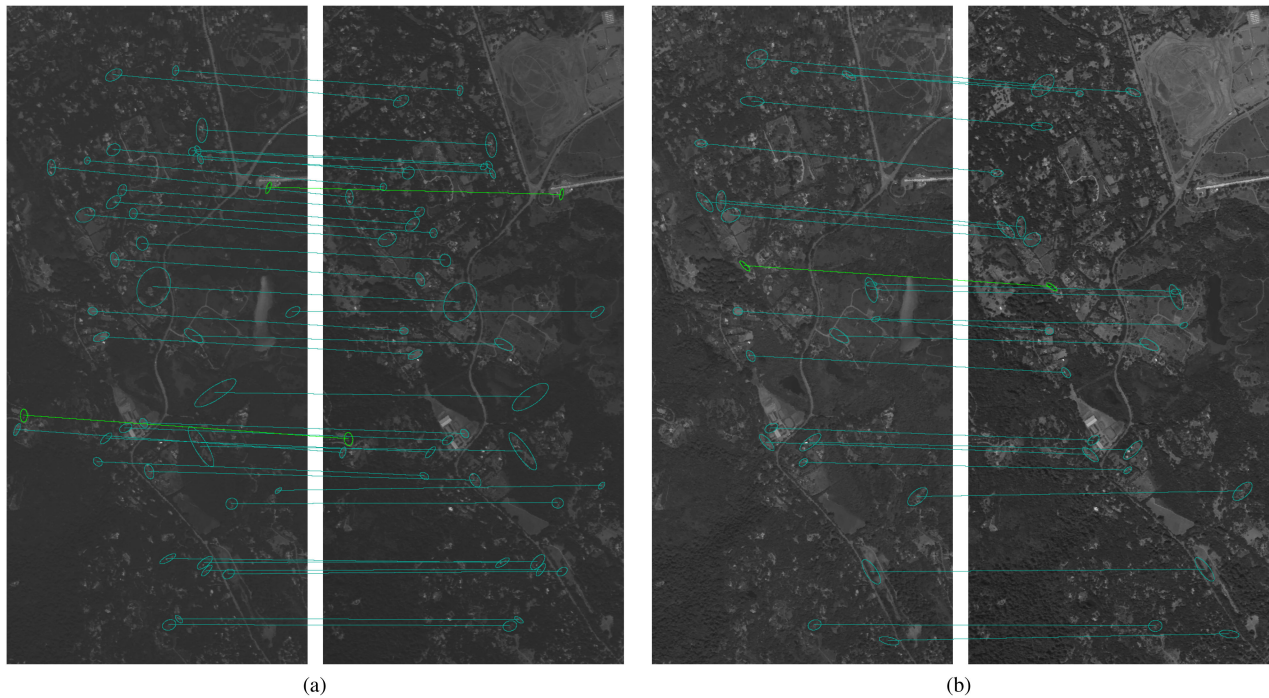


Fig. 5. Matched ellipses detected in two pairs of bands of the Jasper Ridge images. Matches (blue), and matches used in registration (green). (a) Band 8. (b) Band 20.

TABLE II
SUCCESSFULLY REGISTERED CASES FOR EACH SCENE USING HSI-MSER EXPLOITING DIFFERENT NUMBERS OF BANDS: 2, 4, 6, 8, 10, 12, 14, AND 16

Scene	EBS 2	EBS 4	EBS 6	EBS 8
Pavia University	1/6× to 9.0× (22)	1/6× to 12.0× (28)	1/6× to 12.5× (29)	1/7× to 12.0× (29)
Pavia Centre	1/7× to 12.5× (30)	1/8× to 15.0× (36)	1/9× to 14.0× (35)	1/8× to 15.0× (36)
Indian Pines	1/2× to 8.0× (16)	1/3× to 9.0× (19)	1/3× to 8.0× (17)	1/3× to 8.0× (17)
Salinas	1/4× to 7.0× (16)	1/5× to 7.0× (17)	1/6× to 7.5× (19)	1/5× to 7.0× (17)
Jasper Ridge	(0)	3.5× (6)	1/2× to 6.5× (13)	1/3× to 7.0× (15)
Santa Barbara Front	1/2× to 4.5× (8)	1/2× to 5.0× (9)	1/2× to 7.0× (14)	1/3× to 7.0× (15)
Santa Barbara Line	1/2× to 5.5× (11)	1/4× to 6.5× (15)	1/4× to 7.0× (16)	1/5× to 7.0× (17)
Baraboo Hills	(0)	3.5× (6)	1/2× to 3.5× (7)	1/2× to 4.0× (8)
Crown Point	2.0× (3)	1/3× to 5.5× (12)	1/4× to 6.0× (14)	1/7× to 6.0× (17)
Average number of scalings	(11.78)	(16.44)	(18.22)	(19.00)

Scene	EBS 10	EBS 12	EBS 14	EBS 16
Pavia University	1/7× to 12.5× (30)	1/6× to 13.0× (30)	1/6× to 13.5× (31)	1/8× to 13.5× (33)
Pavia Centre	1/9× to 16.0× (39)	1/9× to 16.5× (40)	1/9× to 16.5× (40)	1/9× to 16.5× (40)
Indian Pines	1/3× to 8.5× (18)	1/3× to 9.0× (19)	1/3× to 8.5× (18)	1/3× to 8.5× (18)
Salinas	1/5× to 8.0× (19)	1/6× to 7.5× (19)	1/6× to 8.0× (20)	1/6× to 8.5× (21)
Jasper Ridge	1/4× to 7.0× (16)	1/4× to 7.0× (16)	1/5× to 5.5× (14)	1/4× to 5.5× (13)
Santa Barbara Front	1/3× to 7.0× (15)	1/3× to 7.5× (16)	1/3× to 7.5× (16)	1/4× to 7.5× (17)
Santa Barbara Line	1/5× to 7.0× (17)	1/5× to 7.5× (18)	1/5× to 6.5× (16)	1/5× to 8.5× (20)
Baraboo Hills	4.0× (7)	1/2× to 4.0× (8)	1/2× to 5.0× (10)	1/3× to 5.0× (11)
Crown Point	1/7× to 6.0× (17)	1/5× to 6.5× (16)	1/5× to 6.5× (16)	1/8× to 6.0× (18)
Average number of scalings	(19.78)	(20.22)	(20.11)	(21.22)

The range indicates the scales successfully registered for the 72 angles. The numbers in parentheses summarize the number of scales that were correctly registered for all angles. If an angle is incorrectly registered, the whole scale factor is considered incorrect, i.e., this case is not included in the table. The registration is considered correct if the parameters obtained by the algorithm are the same as the original values.

summarizes the cases that were correctly registered for each scene by exploiting different numbers of bands selected by EBS, from 2 to 16 bands in steps of 2. It shows that the more bands we used, the better the results. The HSI-MSER using 16 selected bands by EBS provides the best results on average, correctly registering 21.22 cases as compared to 11.78 cases registered

using 2 bands. For EBS 8, the number of cases correctly registered is 19.00, which are very close in quality to the results obtained by EBS 16 but with lower computational cost. For that reason, we chose EBS 8 as the default configuration for the proposed method. The computational cost will be evaluated in Section III-E.

TABLE III
COMPARISON OF DIFFERENT HYPERSPECTRAL REMOTE-SENSING IMAGE REGISTRATION METHODS REGARDING THE SUCCESSFULLY REGISTERED CASES FOR EACH SCENE: HYFM, HSI-SURF, AND THE PROPOSED METHOD HSI-MSER

Scene	HYFM 8	HSI-SURF 8	HSI-MSER 8
Pavia University	1/4× to 5.5× (13)	1/7× to 10.5× (26)	1/7× to 12.0× (29)
Pavia Centre	1/5× to 7.5× (18)	1/9× to 11.5× (30)	1/8× to 15.0× (36)
Indian Pines	1/2× to 4.0× (8)	1/3× to 5.5× (12)	1/3× to 8.0× (17)
Salinas	1/2× to 4.5× (9)	1/6× to 7.0× (18)	1/5× to 7.0× (17)
Jasper Ridge	1/5× to 3.0× (9)	1/2× to 4.0× (8)	1/3× to 7.0× (15)
Santa Barbara Front	1/4× to 3.5× (9)	1/6× to 4.0× (12)	1/3× to 7.0× (15)
Santa Barbara Line	1/4× to 6.0× (14)	1/4× to 3.5× (9)	1/5× to 7.0× (17)
Baraboo Hills	1/2× to 2.5× (5)	1/2× to 3.0× (6)	1/2× to 4.0× (8)
Crown Point	1.0× (1)	1/5× to 4.5× (12)	1/7× to 6.0× (17)
Average number of scalings	(9.56)	(14.78)	(19.00)

The range indicates the scales successfully registered for the 72 angles. The numbers in parentheses summarize the number of scales that were correctly registered for all angles. If an angle is incorrectly registered, the whole scale factor is considered incorrect, i.e., this case is not included in the table. The registration is considered correct if the parameters obtained by the algorithm are the same as the original values.

TABLE IV
REFERENCE REGISTRATION PARAMETERS FOR THE SECOND GROUP OF TEST HSIS

Scene	Scale factor	Rotation angle	Translation (x,y)
Jasper	0.97	-6.05°	(-9, 25)
Santa Barbara Front	1.45	3.52°	(2, 6)
Santa Barbara Box	1.00	0.00°	(32, 41)
Baraboo Hills	1.44	0.00°	(46, 132)
Crown Point	1.03	-0.70°	(-24, 48)

TABLE V
COMPARISONS OF HSI-SURF AND THE PROPOSED METHOD HSI-MSER (8 BANDS) REGARDING THE NUMBER OF MATCHES OBTAINED FOR EACH SCENE

	HSI-SURF		HSI-MSER	
Jasper	Number of matches	346	251	
	Number of correct matches	302	186	
	Correct match ratio	0.873	0.741	
Santa Barbara Front	Number of matches	996	991	
	Number of correct matches	875	739	
	Correct match ratio	0.879	0.746	
Santa Barbara Box	Number of matches	1962	971	
	Number of correct matches	1609	750	
	Correct match ratio	0.820	0.772	
Baraboo Hills	Number of matches	238	127	
	Number of correct matches	188	96	
	Correct match ratio	0.790	0.756	
Crown Point	Number of matches	1908	682	
	Number of correct matches	409	154	
	Correct match ratio	0.214	0.226	

HYFM is not included as it does not look for matches.

D. Comparison to Other Methods in the Literature

In this section, the proposed method is compared to other hyperspectral registration algorithms in the literature: the hyperspectral Fourier-Mellin (HYFM) [41], and the hyperspectral SURF (HSI-SURF) algorithm [42]. Both algorithms exploit 8 spectral bands to register two HSIs. The comparison is made in terms of range of successfully registered cases for each scene, number of matches, number of correct matches, RMSE, registration error, and computational time.

HYFM is an area-based method, which performs a PCA to reduce the dimensionality and extracts 8 PCs for each HSI [41]. One log-polar grid for each pair of PCs is computed using the

TABLE VI
REGISTRATION RESULTS FOR HSI-SURF AND HSI-MSER (8 BANDS) REGARDING THE NUMBER OF MATCHES REALLY USED TO COMPUTE THE REGISTRATION PARAMETERS

	HSI-SURF		HSI-MSER	
Jasper	Number of matches used for registration	2	2	
	Correct match ratio	1.00	1.00	
	RMSE	0.80	0.78	
	Registration error (pixels)	0.78	0.60	
Santa Barbara Front	Number of matches used for registration	2	3	
	Correct match ratio	1.00	0.667	
	RMSE	1.18	1.97	
	Registration error (pixels)	1.16	1.95	
Santa Barbara Box	Number of matches used for registration	4	2	
	Correct match ratio	1.00	1.00	
	RMSE	1.20	0.99	
	Registration error (pixels)	1.19	0.99	
Baraboo Hills	Number of matches used for registration	2	2	
	Correct match ratio	0.50	1.00	
	RMSE	2.23	1.19	
	Registration error (pixels)	1.68	1.18	
Crown Point	Number of matches used for registration	4	4	
	Correct match ratio	0.00	0.50	
	RMSE	4.92	3.42	
	Registration error (pixels)	4.81	2.73	

HYFM is not included as it does not look for matches.

adaptable multilayer fractional Fourier transform. The different log-polar grids are combined to integrate the information from the different PCs. The highest peaks in the combined log-polar grid are examined to determine the scaling, rotation, and translation parameters.

HSI-SURF is a feature-based method [42]. It extracts keypoints in 8 selected bands for each image. The method is based on SURF [9] algorithm as keypoint detector and descriptor, and considers the spectral information of the images in the band selection, keypoint description, and keypoint matching stages. It uses a band selection method based on entropy as in the HSI-MSER.

A comparison between HYFM, HSI-SURF, and HSI-MSER regarding the number of successfully registered cases is presented in Table III for the same number of extracted bands.

TABLE VII
EXECUTION TIME IN CPU FOR EACH SCENE CONSIDERING THE LAST COMMON SCALE SUCCESSFULLY REGISTERED FOR ALL METHODS IN
TABLE III (8 BANDS) AND HSI-MSER EXPLOITING 8 AND 16 BANDS

Scene	# data values in millions (10^6)	HYFM 8 (s)	HSI-SURF 8 (s)	HSI-MSER 8 (s)	HSI-MSER 16 (s)
Indian Pines	4.63	29.08	7.37	15.99	32.65
Pavia University	21.36	1.55	0.51	5.26	8.57
Salinas	22.67	6.47	3.41	6.58	12.29
Pavia Centre	79.93	127.96	36.55	28.97	56.91
Santa Barbara Front	94.75	29.85	58.13	33.48	59.22
Crown Point	169.34	146.29	208.63	35.92	61.81
Jasper Ridge	169.38	132.88	143.61	41.94	79.71
Santa Barbara Box	176.39	30.13	57.79	29.98	53.2
Baraboo Hills	190.85	135.54	252.93	38.69	99.26
Average	103.26	71.08	85.44	26.31	51.51

The first column indicates the number of data values (height \times width \times bands). HSI-MSER 8 and HSI-MSER 16 correspond to the EBS 8 and EBS 16 configurations in Table II, respectively.

Feature-based methods (HSI-SURF and HSI-MSER) achieve better results than the area-based method (HYFM) because of their resilience to illumination and intensity changes introduced by noise. HSI-MSER is the method that correctly registers more cases on average, specifically, 19.00 cases. The most notable improvement occurs in the case of the second group of images for which almost twice the scale factors are correctly registered, for example, for Jasper Ridge up to $7.0\times$ compared to $3.0\times$ or $4.0\times$ for HYFM and HSI-SURF, respectively.

The registration accuracy must also be evaluated. Extracting control points manually for this evaluation is a time-consuming task that depends on the user decision. As an alternative, the regions extracted by MSER are considered as control points, as proposed in [16]. The reference registration parameters are applied to the matched regions detected in the target image to calculate how much they differ from the regions in the reference image. The reference registration parameters can be seen in Table IV. The original scale factors and angular rotations in the AVIRIS database for each image were used as a starting point [43]. Then, an expert refined these scale factors and rotation angles by hand and obtained the translation parameters to be considered as the correct values.

These experiments are carried out on the second group of test images presented in Section III-A. No additional scale factor, angle of rotation, and translation parameters are applied, only the original transformations already presented, because they were taken on different flights and dates.

Table V compares the number of matches for each scene using the HSI-SURF and HSI-MSER methods. The first row displays the number of keypoint matches extracted in the case of HSI-SURF and the number of region matches in the case of HSI-MSER. It is interesting to compare the number of keypoint or region matches found by a method, although it does not influence the quality of the registration, but it does affect the computation time. In the second row, the number of correct matches is shown. A match is considered incorrect if the error measured as the Euclidean distance between the coordinates of the features obtained for the reference image and the corresponding ones for the target image after applying the correct transformation is higher than 2 pixels. In the case of HSI-SURF, the features are the coordinates of the keypoints, while in the case

of HSI-MSER, the features are the coordinates of the centres of the extracted regions.

Table VI shows the number of matches really used to compute the final registration parameters. These matches are a subset of those shown in Table V. Both HSI-SURF and HSI-MSER use the exhaustive search method explained in Section II-C to compute the registration parameters. It takes into account all the possible combinations between the matched features. Then, a histogram representing all obtained angles of rotation is created. The bin with the maximum number of elements is selected. The final registration parameters are obtained from the median of the scale factors of the parameters of this bin.

Two of the most frequently used error measures in the literature are also presented in Table VI: RMSE [44] and registration error. The registration error is computed as the average Euclidean distance between the features of the reference image used to compute the final registration parameters and the features of the target image used to compute the final registration parameters after the application of the reference transformation shown in Table IV. It is measured in pixels.

Let \vec{r}_i represent the coordinates of the centre of the region i in the reference image, \vec{t}_i the coordinates of the centre of its matched region in the target image after the application of the reference transformation; and M , the number of matches

$$\text{registration error} = \frac{\sum_{i=1}^M \|\vec{r}_i - \vec{t}_i\|_2}{M}.$$

The features are the coordinates of the keypoints in the case of HSI-SURF.

As shown in Table VI, HSI-MSER achieves smaller errors than HSI-SURF in all scenes with the exception of Santa Barbara Front in which one of the selected regions has an RMSE of 2 pixels. The high error values obtained for both methods in the Crown Point scene are related to the large distortion present in these images. More degrees of freedom are needed to better align this scene, i.e., an affine or nonrigid transformation.

E. Computation Times

The time performance is crucial in real-time applications or when large datasets are available. In this section, we present an

analysis of the computation times of HYFM, HSI-MSER, and HSI-SURF.

Table VII shows the execution times in CPU for each scene considering the last common scale successfully registered for all methods in Table III. It also includes the HSI-MSER version exploiting 16 selected bands by EBS called HSI-MSER 16 in Table VII.

The lowest average execution time, 26.31 s, is achieved by HSI-MSER exploiting 8 bands. This method is less computationally expensive. Although HYFM and HSI-SURF obtain better times than HSI-MSER for smaller images (Pavia University, Indian Pines, and Salinas), for larger images HSI-MSER is more efficient. The reason is that, as it is shown in Table V, HSI-SURF needs a larger number of features than HSI-MSER to register the HSIs. This larger number of features increments the computational costs of the matching stage.

Thanks to the additional bands exploited by HSI-MSER 16, it obtains better results than any other in terms of registration precision as seen in Table II even though the execution time is twice as long compared to HSI-MSER 8, but lower than those of HYFM and HSI-SURF.

IV. CONCLUSION

In this article, HSI-MSER, a feature-based method for registering pairs of hyperspectral remote sensing images, is presented. In particular, the method uses MSER to detect regions and the SIFT descriptor to describe them. To improve the image alignment, the method exploits the spectral information available in the images by detecting features in several preselected bands as well as by including spectral information in the descriptor.

The proposed algorithm is evaluated for a wide variety of scale and rotation parameters, as well as compared in terms of registration precision to other methods in the literature, HSI-SURF and HYFM that also exploit spectral information. Nine HSIs taken by the AVIRIS and the ROSIS sensors are used to evaluate the method. They include urban or rural scenes and changes in different spatial structures and illumination.

Our proposal achieves competitive results when compared to HSI-SURF and HYFM in terms of registration precision and execution time, especially on larger images. Thanks to the exploitation of the spectral information, the method achieves correct alignment of up to $15.0\times$ and a registration time of 26.31 s on average when 8 bands are exploited. In the case of 16 bands, successful alignments of up to $16.5\times$ and an execution time of 51.51 s on average are achieved.

REFERENCES

[1] D. Marinelli, F. Bovolo, and L. Bruzzone, "A novel change detection method for multitemporal hyperspectral images based on binary hyperspectral change vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4913–4928, Jul. 2019.

[2] P. G. Bascosy, A. S. Gareca, D. B. Heras, F. Argüello, and A. Ordóñez, "Texture-based analysis of hydrographical basins with multispectral imagery," *Proc. SPIE*, vol. 11149, 2019, Art. no. 111490Q.

[3] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.

[4] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[5] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 12, pp. 1156–1168, Dec. 1994.

[6] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Nonrigid image registration using conditional mutual information," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 19–29, Jan. 2009.

[7] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.

[8] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 430–443.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[11] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A survey," *IEEE Access*, vol. 7, pp. 6424–6434, Dec. 2018.

[12] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, Apr. 2009.

[13] H. Goncalves, L. Corte-Real, and J. A. Goncalves, "Automatic image registration through image segmentation and SIFT," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2589–2600, Jul. 2011.

[14] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.

[15] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[16] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2004, pp. 506–513.

[17] F. Alharwarin, C. Wang, D. Ristić-Durrant, and A. Gräser, "Improved sift-features matching for object recognition," in *Proc. Vis. Comput. Sci.-BCS Int. Academic Conf.*, 2008, pp. 179–190.

[18] Š. Obdržálek and J. Matas, "Object recognition using local affine frames on maximally stable extremal regions," in *Toward Category-Level Object Recognition*. Berlin, Germany: Springer, 2006, pp. 83–104.

[19] N. Ali *et al.*, "A novel image retrieval based on visual words integration of SIFT and SURF," *PLoS One*, vol. 11, no. 6, 2016, Art. no. e0157428.

[20] H. A. Elnemr, "Combining SURF and MSER along with color features for image retrieval system based on bag of visual words," *J. Comput. Sci.*, vol. 12, no. 4, pp. 213–222, 2016.

[21] H. Tamimi, H. Andreasson, A. Treptow, T. Duckett, and A. Zell, "Localization of mobile robots with omnidirectional vision using particle filter and iterative SIFT," *Robot. Auton. Syst.*, vol. 54, no. 9, pp. 758–765, 2006.

[22] A. Ramisa, A. Tapus, R. L. De Mántaras, and R. Toledo, "Mobile robot localization using panoramic vision and combinations of feature region detectors," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2008, pp. 538–543.

[23] J. Matas, T. Obdržálek, and O. Chum, "Local affine frames for wide-baseline stereo," in *Proc. Int. Conf. Pattern Recognit.*, vol. 4, 2002, pp. 363–366.

[24] H. Sun and C. Wang, "Combining local affine frames and sift for remote sensing image registration," *Proc. SPIE*, vol. 7494, 2009, Art. no. 74941B.

[25] J. Guo, H. Sun, and C. Zhu, "Multispectral remote sensing image registration based on maximally stable extremal regions," *Proc. SPIE*, vol. 7494, pp. 265–270, 2009. [Online]. Available: <https://doi.org/10.1117/12.832949>

[26] Y. Zhang, Y. Guo, and Y. Gu, "Robust feature matching and selection methods for multisensor image registration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 3, 2009, pp. III-255–III-258.

[27] L. Liu, H. Tuo, T. Xu, and Z. Jing, "Multi-sensor image registration using edge-enhanced maximally stable extremal region," in *Proc. 5th IEEE Int. Congr. Image Signal Process.*, 2012, pp. 901–905.

[28] M. Teke and A. Temizel, "Multi-spectral satellite image registration using scale-restricted SURF," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2310–2313.

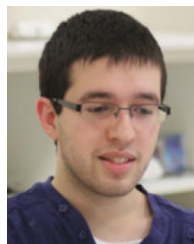
[29] L. P. Dorado-Muñoz, M. Velez-Reyes, A. Mukherjee, and B. Roysam, "A vector SIFT detector for interest point detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4521–4533, Nov. 2012.

- [30] X. Yin, X.-C. Yin, H.-W. Hao, and K. Iqbal, "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost," in *Proc. 21st Int. Conf. Pattern Recognit.*, 2012, pp. 725–728.
- [31] L. Gómez and D. Karatzas, "MSER-based real-time text detection and tracking," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 3110–3115.
- [32] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008. [Online]. Available: <http://www.vlfeat.org/>
- [33] J. A. Richards, *Remote Sensing Digital Image Analysis*, vol. 3. Berlin, Germany: Springer, 1999.
- [34] C. Cariou, K. Chehdi, and S. Le Moan, "BandClust: An unsupervised band reduction method for hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 565–569, May 2010.
- [35] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
- [36] S. Obdrzalek and J. Matas, "Object recognition using local affine frames on distinguished regions," in *Proc. Brit. Mach. Vis. Conf.*, vol. 1, 2002, pp. 113–122.
- [37] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.
- [38] Á. Ordóñez, D. B. Heras, and F. Argüello, "Repository of hyperspectral images for image registration," Accessed: Jul. 30, 2021. [Online]. Available: <https://gitlab.citius.usc.es/hiperespectral/RegistrationRepository>
- [39] Grupo de Inteligencia Computacional de la Universidad del País Vasco (UPV/EHU), "Hyperspectral remote sensing scenes," Accessed: Jul. 30, 2021. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes
- [40] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 band AVIRIS hyperspectral image data set: Jun. 12, 1992 Indian Pine Test Site 3," Sep 2015, Accessed: Jul. 30, 2021. [Online]. Available: <https://purr.purdue.edu/publications/1947/1>
- [41] Á. Ordóñez, F. Argüello, and D. B. Heras, "Fourier-Mellin registration of two hyperspectral images," *Int. J. Remote Sens.*, vol. 38, no. 11, pp. 3253–3273, 2017.
- [42] Á. Ordóñez, D. B. Heras, and F. Argüello, "SURF-Based registration for hyperspectral images," in *Proc. IEEE Int. Geosci Remote Sens. Symp.*, 2019, pp. 63–66.
- [43] "AVIRIS database," Accessed: Jul. 30, 2021. [Online]. Available: <http://aviris.jpl.nasa.gov/data/index.html>
- [44] H. Goncalves, J. A. Goncalves, and L. Corte-Real, "Measures for an objective evaluation of the geometric correction process quality," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 292–296, Apr. 2009.



Álvaro Ordóñez received the B.S. degree in computer science and the M.S. degree in big data technologies in 2015 and 2016, respectively, from the University of Santiago de Compostela, Santiago, Spain, where he is currently working toward the Ph.D. degree in computer science.

He is currently an Assistant Researcher of Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Santiago de Compostela, Spain. His main research interests include image analysis and processing, parallel algorithms, and big data technologies.



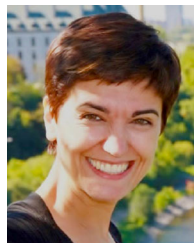
Álvaro Acción received the B.S. degree in computer science and the M.S. degree in big data technologies in 2014 and 2017, respectively, from the University of Santiago de Compostela, Santiago de Compostela, Spain, where he is currently working toward the Ph.D. degree in computer science and classification of hyperspectral remote sensing images.

He is currently an Assistant Researcher of Centro Singular de Investigación Intelixentes (CiTIUS), Santiago de Compostela, Spain. His main research interests include image analysis and processing.



Francisco Argüello received the B.S. and Ph.D. degrees in physics from the University of Santiago de Compostela, Santiago, Spain, in 1988 and 1992, respectively.

He is currently an Associate Professor with the Department of Electronic and Computer Engineering, University of Santiago, Santiago, Spain. His current research interests include signal and image processing, computer graphics, parallel and distributed computing, and quantum computing.



Dora B. Heras (Member, IEEE) received the M.S. degree in physics in 1994 and the Ph.D. degree in physics in 2000 from the University of Santiago de Compostela, Santiago, Spain.

She is currently an Associate Professor with the Department of Electronics and Computer Engineering, University of Santiago de Compostela. Her research interests include parallel and distributed computing, software optimization techniques for emerging architectures, computer graphics, and image processing.