Á. Ordóñez, D. B. Heras and F. Argüello and, "SURF-Based Registration for Hyperspectral Images," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 63-99, 2019.
DOI: 10.1109/IGARSS.2019.8900462

SURF-BASED REGISTRATION FOR HYPERSPECTRAL IMAGES

Álvaro Ordóñez, Dora B. Heras, Francisco Argüello

Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS).
Universidade de Santiago de Compostela, Santiago de Compostela, Spain.
(e-mail: {alvaro.ordonez, dora.blanco, francisco.arguello}@usc.es)

ABSTRACT

The alignment of images, also known as registration, is a relevant task in the processing of hyperspectral images. Among the feature-based registration methods, Speeded Up Robust Features (SURF) has been proposed as a computationally efficient approach. In this paper HSI–SURF is proposed. This is a method to register hyperspectral remote sensing images based on SURF that takes advantage of the full spectral information of the images. In this sense, the proposed method selects specific bands of the images and adapts the keypoint descriptor and the matching stages to benefit from the spectral information, thus increasing the effectiveness of the registration.

Index Terms— Hyperspectral data, remote sensing, registration, SURF features, feature extraction.

1. INTRODUCTION

Image registration is a fundamental task in remote sensing hyperspectral imaging for different applications. In many cases, images that were acquired from different viewpoints need to be jointly analyzed. We consider the registration problem associated with images where the translation, rotation, and scaling parameters between a reference image and a target image of the same scene need to be calculated. Among the set of registration methods [1], feature–based methods extract significant features (regions, lines or points) from the images. These methods are especially robust to illumination changes, intensity changes introduced by noise, and changes introduced by the use of different sensors, which make them very appropriate for hyperspectral remote sensing images.

The most difficult task for the feature-based matching methods is to find discriminative feature descriptors that are invariant to the possible differences between the images. The Scale–Invariant Feature Transform (SIFT) [2] is the most popular feature–based algorithm. It consists of four stages: scale–space extrema detection, keypoint localization, orientation assignment, and keypoint description. The Speeded-Up Robust Registration Features (SURF) algorithm [3] is a more recent approach based on the same scheme as SIFT, but faster to compute. The reason is that the Gaussian scale-space is approximated in the case of SURF by building a pyramid of filters instead of a pyramid of images. This property removes the dependencies in the calculation of one level of the pyramid from the others making their parallel computation possible.

Most of the registration algorithms are designed to register greyscale or RGB images. The most direct approach for multiband images is to apply the algorithms to only one band of each image. However, the large amount of spectral information available in hyperspectral images can be used to improve the registration process [4, 5].

In this paper, we propose a method for registering images called HSI-SURF based on SURF as keypoint detector and descriptor, and that considers the spectral information of the images in the band selection, keypoint description, and keypoint matching stages.

2. REGISTRATION OF REMOTE SENSING HYPERSPECTRAL IMAGES USING SURF

In this section we present HSI–SURF. SURF [3] is a scale and in-plane rotation invariant feature detector and descriptor with better performance than SIFT. It is characterized by building the scale space using integral images, along with the use of Haar wavelets to build the descriptors. SURF consists of two stages: detector and descriptor. In the next sections, the SURF detector and descriptor are explained, the band selection methods are described and, finally, the resulting HSI–SURF method is presented.

2.1. SURF detector

SURF [3] uses the Hessian matrix to perform the detection. The maximum of the determinant of the Hessian is used for

This work was supported in part by the Consellería de Educación, Universidade e Formación Profesional [grant numbers GRC2014/008, ED431C 2018/19, and ED431G/08] and Ministerio de Economía y Empresa, Government of Spain [grant number TIN2016-76373-P] and by Junta de Castilla y Leon - ERDF (PROPHET Project) [grant number VA082P17]. All are co–funded by the European Regional Development Fund (ERDF). The work of Álvaro Ordóñez was also supported by Ministerio de Ciencia, Innovación y Universidades, Government of Spain, under a FPU Grant [grant number FPU16/03537].

the keypoint location and the scale selection. Given a point $\mathbf{x} = (x, y)$ in a image I, the Hessian matrix $\mathcal{H}(\mathbf{x}, \sigma)$ in $\mathbf{x}$ at scale σ is defined as follows

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}, \tag{1}$$

where $L_{xx}(\mathbf{x}, \sigma)$, $L_{xy}(\mathbf{x}, \sigma)$ and $L_{yy}(\mathbf{x}, \sigma)$ are the convolutions of the Gaussian second order derivative [3]. Due to the fact that these convolutions induce a high computational cost, SURF approximates them using integral images instead of the difference of Gaussians used by SIFT. An integral image $II(\mathbf{x})$ at a location $(\mathbf{x})$ for a image $I(i, j)$ is defined as

$$II(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j). \tag{2}$$

This way, the convolution can be efficiently approximated by Gaussian kernels of arbitrary size.

In the same way as SIFT, SURF builds a scale space organized into octaves and scales in order to achieve scale invariance. The difference lies in that SURF builds a pyramid of filters instead of a pyramid of images to approximate the Laplacian-of-Gaussian operator. SURF changes the scale of box filters rather than reducing the image size as in SIFT. It allows obtaining the different scales by filtering the image with gradual masks and without any dependence between the computations at consecutive scales, unlike SIFT. The reason is that SURF directly computes the filtered image at each scale from the original image.

2.2. SURF descriptor

The SURF descriptors are based on the Haar wavelet and have only 64 components as compared to SIFT. First, in order to achieve rotational invariance, an unique orientation is assigned to each keypoint. To this end, the Haar wavelet responses in the x and y directions, of size $4s$ being s the keypoint scale, are calculated for the neighboring pixels within a $6s$ radius. Each wavelet response is weighted using a Gaussian with $\sigma = 2s$ centered at the keypoint. Afterwards, the dominant orientation is calculated. All wavelet responses in the x and y directions within a sliding orientation window are summed. This window is a sector of angle $\frac{\pi}{3}$. The vector of wavelet responses with the largest norm defines the dominant orientation.

The next stage is the descriptor construction. Firstly, a region of size $20s \times 20s$ centered around each keypoint is constructed. Each region is split into 4×4 subregions. For each subregion, Haar wavelet responses of size $2s$ are computed at 5×5 equally spaced sample points. Next, these Haar wavelet responses are weighted using a Gaussian with $\sigma = 3.3s$ centered at the keypoint in order to increase the robustness to geometric deformations and localization errors. Finally, the wavelet responses in both directions are summed up over each subregion and are concatenated forming a descriptor vector of length 64. To achieve invariance to contrast, the descriptor vector is converted into a unit vector.

2.3. Band selection methods

Trying to save computational costs in terms of memory requirements and computational time, dimensionality reduction methods are applied to the hyperspectral images as a first stage of the registration process. This reduction could be achieved using feature extraction or feature selection techniques. In the first case, we obtain a reduced image representing the transformed initial one, whereas in the last one, we extract a subset of relevant bands from the image without modifying them.

In this paper, we apply Principal Component Analysis (PCA) as a statistical feature extraction technique that produces transformed bands of the image, that can be ordered in terms of increasing amount of statistical relevance. Another feature reduction method considered is BandClust [6]. This method performs unsupervised clustering of the image bands according to a mutual information measure. The group of bands is iteratively clustered minimizing the measure. Each cluster is finally represented by the average of its bands.

The applied band selection methods are WaluMI [7] and the Entropy-based Band Selection (EBS) [5] method. WaluMI is also based on clustering as BandClust, but selecting one band of the image belonging to the cluster as representative of each cluster. The clustering uses the minimum dissimilarity difference as a measure for cluster evaluation. This measure is based on calculating the mutual information between each pair of bands.

Finally, EBS [5] selects bands based on entropy and inter-band distance and considers both images unlike the methods described above. First, the entropy of each band of each image is computed. Next, the minimum entropy of each band considering the two values (one per image) is selected and the bands of both images equally ordered in increasing order of entropy value. Finally, the highest entropy bands are chosen. Each candidate pair of bands to be selected must have, at least, an inter–band distance greater than or equal to an empirically fixed value with respect to the previous selected pair.

2.4. HSI–SURF

Figure 1 outlines the registration method proposed that consists of the following steps. First, a band selection method is applied in order to select one or more bands from each image representing the most relevant information and reducing the computational time. The different methods described in Section 2.3 are proposed and compared.

Features for each selected band of each image are extracted and described in the second and third stages. An interpolation is applied to the original images in order to highlight

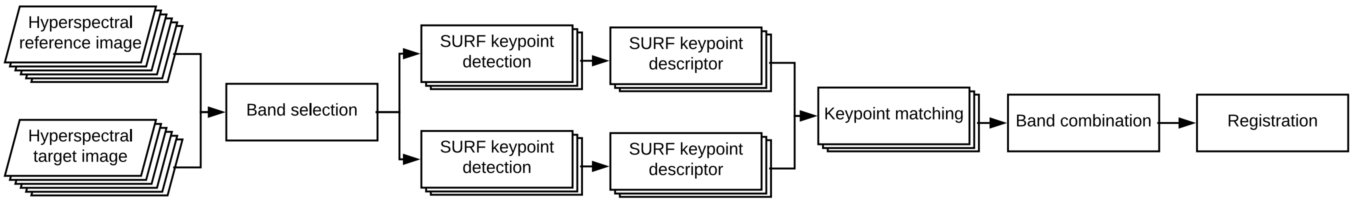


Fig. 1. Proposed HSI–SURF scheme for the registration of two remote sensing hyperspectral images.

details and extract a large number of keypoints in a similar way as in [2, 5]. HSI–SURF uses a descriptor made up of a spatial and a spectral part: the SURF descriptor and the spectral signature. This signature is made up of the spectral components resulting from the band selection stage.

The fourth stage is the keypoint matching between pairs of bands. The matching is based on the distance between the SURF descriptors and the cosine similarity between the spectral signatures. The spectral information allows refining the matching process discarding outliers. In the sixth stage, called band combination, all pairs of matched keypoints for the different pairs of bands are considered together.

The process finishes with an exhaustive search for the registration. All the possible pairs of matched keypoints are considered in a same way as in [5].

3. 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++ 5.4.0 versions under Ubuntu 16.04. For WaLuMI the original software was used, whereas for BandClust the groups of bands were provided by its authors[1].

The evaluation of the algorithm was performed over a set of seven hyperspectral scenes. Table 1 provides detailed information for each scene. The first four scenes are single images commonly used for testing in the field of remote sensing. The remaining scenes are pairs of images taken at different dates. Each pair of images presents changes in spatial structures and in illumination, as well as different scale factors, orientation and translation. A false–colour composite of the Jasper Ridge pair of images is displayed in Figure 2.

The procedure used to test the method is as follows. The first image of each pair is used as a reference while the target image is rotated and scaled. We have carried out exhaustive search with scale factors ranging from 1/10$\times$ to 11.5$\times$ in increments of 0.5$\times$ (31 scale factors) and rotation angles from 0 to 360 degrees in increments of 5 degrees (72 angles). Consequently, we have carried out an exhaustive evaluation of 2232 cases for each scene. In all cases, the target images are trimmed on the central region to keep the same size as the original images.

[1]The authors would like to thank Claude Cariou at Université de Rennes 1, France, for providing the band limits of BandClust used in this work.

Table 1. Sensor, size, number of spectral bands, and spatial resolution (m/pixel) for the test hyperspectral images.

Image	Sensor	Size	Bands	Spatial Resolution
Pavia University	ROSIS–03	610×340	103	1.3
Pavia Centre	ROSIS–03	1096×715	102	1.3
Indian Pines	AVIRIS	145×145	220	20
Salinas Valley	AVIRIS	512×217	204	3.7
Jasper Ridge 2006	AVIRIS	1286×588	224	3.3
Jasper Ridge 2007	AVIRIS	1286×588	224	3.4
Santa Barbara Box 2013	AVIRIS	1024×769	224	15.2
Santa Barbara Box 2014	AVIRIS	1024×769	224	15.2
Santa Barbara Front 2009	AVIRIS	900×470	224	16.4
Santa Barbara Front 2010	AVIRIS	900×470	224	11.3

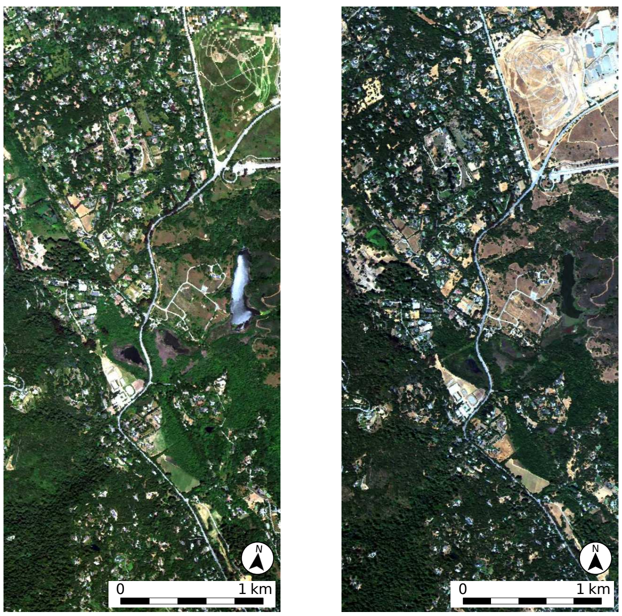
For all the methods, 8 bands have been selected in order to make a fair comparison among proposals, with the exception of BandClust for which the band limits are different for each cluster [5]. Table 2 summarizes the cases that were correctly registered for each scene using HSI–SURF considering different feature reduction methods and band selection methods. As a baseline for comparison, a random band of each image was selected, in this case band 28 (first column in the table). For EBS the band separation is empirically fixed to 20.

As shown in Table 2, feature selection methods (WaLuMI and EBS) obtain better results than PCA. The reason is that they have the advantage over feature extraction methods, such as PCA, of preserving the relevant original information from the data. In contrast, PCA transforms the data to a different domain where the spatial information of the image is distorted, so the relevant structures in the keypoint detection are degraded. The best results are achieved using WaLuMI, 16.71 cases (scales) are correctly registered on average, and EBS, that registers 16.57 cases, while in the case of BandClust only 15.86 correct cases are achieved.

Regarding the computational cost, EBS can be considered the best band selection option in the experiments. For the largest pair of images (Santa Barbara Line) the execution time for the registration process using EBS for the case corresponding to the highest scale and angle that is correctly registered is 139.98 seconds. 2.26 seconds of this time are required by band selection. In the case of WaLuMI the time

Table 2. Successfully registered cases for each scene. The number in parentheses summarizes 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.

Scene	Band 28	8 PCs	BandClust	WaLuMI	EBS
Pavia University	$1/5\times$ to $7.0\times$ (17)	$1/2\times$ to $5.0\times$ (10)	$1/6\times$ to $8.5\times$ (21)	$1/7\times$ to $10.5\times$ (26)	$1/7\times$ to $10.5\times$ (26)
Pavia Centre	$1/6\times$ to $11.0\times$ (26)	$1/6\times$ to $7.5\times$ (19)	$1/7\times$ to $11.5\times$ (28)	$1/10\times$ to $11.5\times$ (31)	$1/9\times$ to $11.5\times$ (30)
Indian Pines	$1/2\times$ to $3.5\times$ (7)	$1.0\times$ to $2.0\times$ (3)	$1/4\times$ to $6.0\times$ (14)	$1/4\times$ to $6.5\times$ (15)	$1/3\times$ to $5.5\times$ (12)
Salinas	$1/4\times$ to $5.0\times$ (12)	$1.0\times$ to $3.5\times$ (6)	$1/6\times$ to $7.0\times$ (18)	$1/6\times$ to $7.0\times$ (18)	$1/6\times$ to $7.5\times$ (19)
Jasper Ridge	$1.0\times$ (1)	(0)	$1/2\times$ to $4.0\times$ (8)	$1/2\times$ to $3.5\times$ (7)	$1/2\times$ to $4.0\times$ (8)
Santa Barbara Front	$1/2\times$ to $2.5\times$ (5)	$1.0\times$ (1)	$1/5\times$ to $4.5\times$ (12)	$1/4\times$ to $4.0\times$ (10)	$1/6\times$ to $4.0\times$ (12)
Santa Barbara Box	$1/2\times$ to $2.0\times$ (4)	$1.0\times$ to $1.5\times$ (2)	$1/5\times$ to $3.5\times$ (10)	$1/4\times$ to $4.0\times$ (10)	$1/4\times$ to $3.5\times$ (9)
Number of scalings (average)	(10.29)	(5.86)	(15.86)	(16.71)	(16.57)



(a) (b)

Fig. 2. *Jasper Ridge* scene taken by the AVIRIS sensor. (a) Fragment of the first *Jasper* image of size 588×1286 pixels taken on 5/12/2006, and (b) fragment of the same size of the second *Jasper* image taken on 8/13/2007.

is 269.59 seconds, most of which, 158.91 seconds, are spent by the band selection stage.

4. CONCLUSIONS

In this paper, a method for registering remote sensing hyperspectral images called HSI-SURF is proposed and evaluated over seven hyperspectral scenes. The method is based on SURF as keypoint detector and descriptor, and it is especially adapted to hyperspectral images by incorporating the spectral information in different stages of the method. First, the algorithm works not only on a band of each image but with information provided by different bands. The spectral information is also included together with the SURF descriptor for each keypoint. Finally, all the keypoints for the different pairs of bands are considered in the exhaustive search for the registration parameters.

The results show that the number of correctly recovered scales for all angles is high, achieving correct registrations at scales of $11\times$ for some images. As future work, the effectiveness of HSI–SURF should be analyzed in detail evaluating more quality measures such as number of correct matches, correct match ratio, RMSE or registration error. The computational cost of this registration method comparing it to other methods such as SIFT should also be analyzed.

5. REFERENCES

[1] Barbara Zitova and Jan Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[2] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[4] Leidy P Dorado-Muñoz, Miguel Velez-Reyes, Amit Mukherjee, and Badrinath Roysam, "A vector SIFT detector for interest point detection in hyperspectral imagery," *IEEE transactions on Geoscience and Remote sensing*, vol. 50, no. 11, pp. 4521–4533, 2012.

[5] Álvaro Ordóñez, Francisco Argüello, and Dora B. Heras, "Alignment of hyperspectral images using KAZE features," *Remote Sensing*, vol. 10, no. 5, 2018.

[6] Claude Cariou, Kacem Chehdi, and Steven Le Moan, "BandClust: An unsupervised band reduction method for hyperspectral remote sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 565–569, 2011.

[7] Joe H Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.