Exploring the MSER-based hyperspectral remote sensing image registration

Ordóñez, Álvaro, Heras, Dora, Argüello, Francisco

Álvaro Ordóñez, Dora B. Heras, Francisco Argüello, "Exploring the MSER-based hyperspectral remote sensing image registration," Proc. SPIE 11533, Image and Signal Processing for Remote Sensing XXVI, 115330E (20 September 2020); doi: 10.1117/12.2574200

SPIE.

Event: SPIE Remote Sensing, 2020, Online Only

Exploring the MSER-based hyperspectral remote sensing image registration

Álvaro Ordóñez[a], Dora B. Heras[a], and Francisco Argüello[b]

[a]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain
[b] Departamento de Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

ABSTRACT

Image registration is an essential preprocessing task in many applications of hyperspectral images capturing the Earth surface. Maximally Stable Extremal Regions (MSER) is a feature–based method for image registration which extracts regions by thresholding the image at different grey levels. Its invariance to affine transformations makes it ideal for image registration. This method is usually employed in text detection and recognition as well as in the medical domain. Hyperspectral images contain spectral information that can be used for improving the image alignment. This article presents a first approach to a hyperspectral remote sensing image registration method based on MSER that efficiently exploits the information contained in the different spectral bands. The experimental results over four hyperspectral images show that the proposed method is promising as it achieves a higher number of correct registration cases than other feature–based methods.

Keywords: Image registration, remote sensing, hyperspectral, MSER

1. INTRODUCTION

Image registration is an essential task in many remote–sensing applications such as automatic change detection, environmental monitoring or super-resolution image creation, among others.[1] In these applications, different images of the same scene acquired at different times, from different viewpoints, and probably by different sensors, have to be aligned. Most registration algorithms are designed to register RGB or greyscale images and do not profit from the high spectral and spatial resolution available in hyperspectral images. New algorithms to deal with this information in an efficient way from the point of view of registration accuracy and computational cost need to be designed.

Image registration algorithms can be classified into area–based and feature–based methods.[1] Methods in the first group work directly with image intensity 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. This property makes feature–based methods more suitable for images with illumination changes which it is the case of remote sensing hyperspectral images of the Earth surface. For these images, the atmospheric conditions usually vary from one capture to another so feature–based methods are the adequate for processing them. Generally, feature–based methods consist of three stages: feature detection, description and matching. They rely on extracting the same features in the images to be registered. Knowing a number of corresponding features, an image transformation can be calculated that aligns one image with respect to the other.

Maximally Stable Extremal Regions (MSER)[2] 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, by thresholding the image at a certain grey level and according to a stability criterion. If MSER is applied to a couple of images, the extracted and matched regions of both images can be used to compute an image transformation to align them. MSER is resilient to changes of scale, rotation, shift and illumination conditions. Other well-known feature detector methods in the literature are KAZE[3] and SURF.[4] Both build

Further author information: E-mail: {alvaro.ordonez, dora.blanco, francisco.arguello}@usc.es

Image and Signal Processing for Remote Sensing XXVI, edited by Lorenzo Bruzzone, Francesca Bovolo
Emannuele Santi, Proc. of SPIE Vol. 11533, 115330E · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2574200

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).[5] SIFT extracts keypoints from a multi–resolution pyramid of the images created performing Gaussian convolutions and interpolations. Its descriptor stands out for being highly distinctive and invariant to changes in illumination and distortion, making it a widely used method in the literature.[6,7]

In this work, a first approach of a registration method for hyperspectral images based on MSER for region detection and SIFT for region description is presented. It exploits the spectral information available in the multi or hyperspectral images by performing feature detection and description in several bands of the images to be registered. The following sections present the different stages of the proposed method (Section 2) as well as the result discussion (Section 3) and conclusions (Section 4).

2. HYPERSPECTRAL REMOTE SENSING IMAGE REGISTRATION USING MSER

In this section, we present a first version of a registration method based on MSER as region detector followed by SIFT, as feature descriptor to register two hyperspectral remote sensing images. In this section, the standard versions of MSER and SIFT are described followed by a description of the resulting registration method, which exploits the spectral information available in the images.

2.1 MSER

The Maximally Stable Extremal Regions (MSER) is a method for region detection in greyscale images.[2] It has been successfully applied to a large number of applications such as image recognition, tracking and image registration.[8,9] The method extracts a number of regions called Maximally Stable Extremal Regions (MSERs) by thresholding the image at a certain grey level and according to a stability criterion. A Maximally Stable Extremal Region is a stable connected component through different grey levels of an image. An extremal region is considered stable when it does not change much as the grey level threshold is varied, i.e., all pixels within the region have higher intensity values (bright extremal regions or MSER+) or lower intensity values (dark extremal regions or MSER−) than the rest of the pixels outside the region. MSER presents two properties that makes it ideal for image registration.[10] First, linear or affine transformations do not affect the extracted extremal regions 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 extremal region will continue to be an extremal region after these transformations.

As a reference MSER implementation, we used the VLFeat code in which the stability criterion was slightly modified in relation to the original article.[11]

2.2 SIFT descriptor

The Scale-Invariant Feature Transform (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 SIFT descriptor is used in many applications due to its invariance to translations, rotations and scaling transformations.

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^o 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. Firstly, an area of size 16×16 centred around each region is selected. Each area is divided into 4×4 subareas. 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.

2.3 Proposal of an adapted MSER registration method

In this section, we present the proposed registration algorithm for hyperspectral images based on MSER and SIFT. It consists of six stages: band selection, region detection, region description, region matching, band combination and registration. A schematic of the proposed algorithm can be seen in Figure 1.

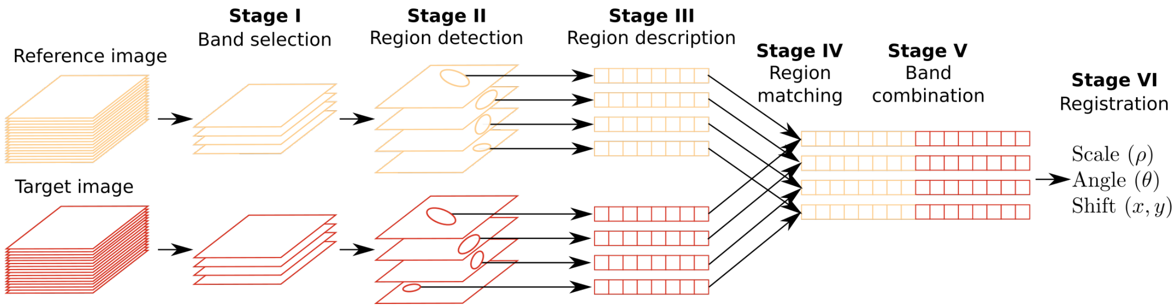


Figure 1: Schematic of the proposed method.

It is common that contiguous bands do not differ in relevant information for the registration process. Band extraction allows reducing the computational cost with respect to considering all the bands of the image but keeping the relevant information. For this reason in the first stage, the most relevant spectral bands of the reference and target images are extracted using BandClust.[12] This method is an unsupervised recursive binary band-splitting algorithm. It iteratively splits a band interval into two disjoint contiguous subbands 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 interval are averaged.

In the second stage, regions of each selected band are extracted using MSER. Let I be an hyperspectral image 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:

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

where μ_l and Σ_l are the mean and variance of the pixel composing the region, and R_l is an extremal region detected in this band for the greyscale level l.[13]

In the third stage, the regions are described. For each ellipse, the coordinates of its centre are considered as the coordinates of the region. In addition, the SIFT descriptor is computed on the surface of the region bounded by the ellipse as it is explained in Section 2.2. Then, in the fourth stage, the regions of both hyperspectral images are matched using the Euclidean distance. All matched regions extracted from the selected bands are joined in the fifth stage. This way, regions that are only present in some bands are used to compute the transformation, i.e., all the spectral information is considered together. Finally, in the sixth stage, an exhaustive search based on histograms is performed to register the images. All the possible pairs of matched regions are considered in the same way as in 14. A scale factor ρ, a rotation angle θ, and translation parameters (x, y) are finally selected to register the images.

3. RESULTS

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 evaluation of the proposed method was performed over a dataset of four hyperspectral remote sensing images frequently used in the literature corresponding to rural spots and cities.[15] A colour composition of these images is presented in Figure 2:

- 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 1.3m/pixel. In this image the optimal number of subbands according to BandClust is 2: 1–64 and 65–103.
- Pavia Centre. It is a ROSIS image of 102 bands of the city of Pavia, Italy. The image size is 1096 × 715 pixels with a spatial resolution of 1.3m/pixel. A 381-pixel-wide black vertical band in the middle of the image was removed. In this image the optimal number of subbands according to BandClust is 3: 1–23, 24–60, and 61–102.
- Salinas. It is a rural scene taken by the Airbone Visible/Infrared Imaging Spectometrer (AVIRIS) sensor in the Salinas Valley, California. This image has a size of 512 × 217 pixels with a spatial resolution of 3.7m/pixel and 204 spectral bands. In this image the optimal number of subbands according to BandClust is 14: 1–11, 12–23, 24–33, 34–44, 45–61, 62–68, 69–78, 79–83, 84–104, 105–109, 110–121, 122–150, 151–155, and 156–204.
- Indian Pines. It was collected by the AVIRIS sensor over a rural area in NW Tippecanoe County, Indiana. It is a 220-band image with a size of 145 × 145 and a spatial resolution of 20m/pixel. In this image the optimal number of subbands according to BandClust is 9: 1–38, 39–60, 61–65, 66–79, 80–102, 103–116, 117–149, 150–160, 161–220.

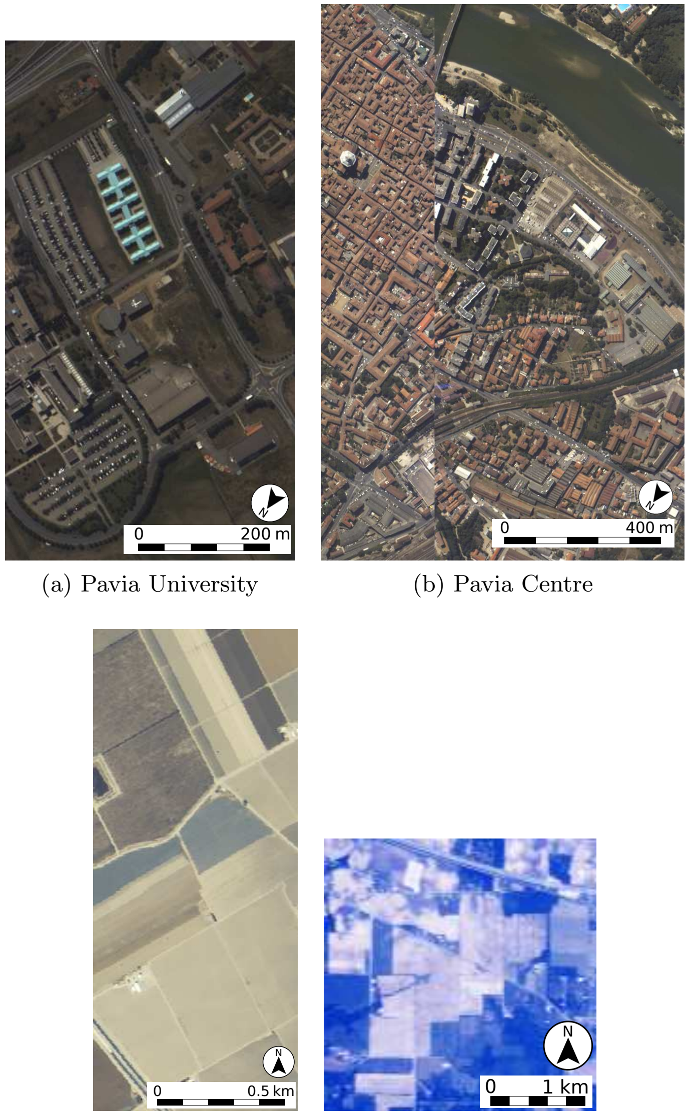
The procedure test consists in registering each scene (reference image) with respect to a scaled and rotated version of this image (target image). The set of target images is generated applying a range of scale factors from 1/8× to 12.5× (31 scale factors) in increments of 0.5× and rotation angles from 0 to 360 degrees in increments of 5 degrees (72 angles). This way, we can investigate all the registration details in controlled conditions and evaluate the registration capabilities under extreme scaling and rotating conditions.

Table 1 summarizes the cases that were correctly registered for each scene using different information as input: a single band randomly selected for each image or several bands selected by BandClust. As shown in the table, 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. Figure 3 illustrates this statement. It shows an example of subband matching for three groups of bands selected by BandClust for the Pavia Centre image when a scale of 2.0× are considered. It can be observed that some features are only present and detected in some spectral bands. The results in Table 1 show that using BandClust for selecting several bands, 23.50 cases (scales) are correctly registered for all the angles, almost twice the number of cases achieved using only one band (13.75 cases).

Table 1: Successfully registered cases for each scene using a random band of each image (in this case band 88) and BandClust as band selection method. 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.

Image	**Band 88**	**BandClust**
Pavia University	1/4× to 7.0× (16)	1/5× to 11.0x (25)
Pavia Centre	1/5× to 7.5× (18)	1/8× to 12.5x (31)
Salinas	1/4× to 4.5× (11)	1/5x to 8.5× (21)
Indian Pines	1/2× to 5.0× (10)	1/3× to 8.0× (17)
Number of scales (average)	(13.75)	(23.50)

We compared our proposal to HSI–SURF,[16] a method to register hyperspectral remote sensing images based on SURF. It also exploits the spectral information using different bands for registering and not only one, as it is common in the literature. Table 2 presents the cases that were correctly registered for each scene using HSI–SURF and the proposal, both using BandClust as band selection method. The best results are achieved



(a) Pavia University (b) Pavia Centre

(c) Salinas (d) Indian Pines

Figure 2: Hyperspectral images used for testing.[15]

using the proposed method based on MSER. Specifically, 23.50 cases are correctly registered on average, while in the case of HSI–SURF only 20.25 correct cases are achieved for the same images.

Table 2: Successfully registered cases for each image using HSI–SURF and our proposal with BandClust as band selection method.

Image	**HSI–SURF**	**Adapted MSER**
Pavia University	1/6× to 8.5× (21)	1/5× to 11.0× (25)
Pavia Centre	1/7× to 11.5× (28)	1/8× to 12.5× (31)
Salinas	1/6×to 7.0× (18)	1/6× to 8.5× (21)
Indian Pines	1/4× to 6.0× (14)	1/3× to 8.0× (17)
Number of scalings (average)	(20.25)	(23.50)

4. CONCLUSIONS

In this article, a feature–based method for registering couples of hyperspectral remote sensing images and based on the MSER method for region extraction is proposed. In particular, the method uses MSER to detect regions and the SIFT descriptor to describe them. It exploits the spectral information available in the images by detecting features in several pre-selected bands.

The proposed algorithm is evaluated for a wide variety of scale and rotation parameters and compared in terms of registration precision to another method in the literature, HSI–SURF, that also exploits spectral information. Four hyperspectral images taken by the AVIRIS (Airbone Visible/Infrared Imaging Spectometrer) and the Reflective Optics System Imaging Spectrometer (ROSIS) sensors are used to evaluate the method. Our proposal produces competitive results when compared to HSI–SURF. Thanks to the exploitation of the spectral information, the method achieves correct alignment up to 12.5× for the Pavia Centre image.

As future work, we plan to analyse more alternatives for considering the spectral information provided by the different bands of the images. It will be also necessary to analyse the performance of the proposed method with hyperspectral images in real registering conditions, i.e., obtained at different times and, as a result, at different illumination conditions and with different spatial structures present in the image. The comparative analysis will also be extended considering other methods and quality measures. The analysis of the computational cost is also part of the future work as it is a relevant factor to select the adequate registration method for applications with time constraints.

ACKNOWLEDGMENTS

This work was supported in part by Ministerio de Ciencia e Innovación, Government of Spain [grant number PID2019-104834GB-I00], and Consellería de Educación, Universidade e Formación Profesional [grant numbers ED431C 2018/19, and accreditation 2019-2022 ED431G-2019/04]. 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 [grant number FPU16/03537].

REFERENCES

[1] Zitova, B. and Flusser, J., "Image registration methods: a survey," *Image and vision computing* **21**(11), 977–1000 (2003).

[2] Matas, J., Chum, O., Urban, M., and Pajdla, T., "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing* **22**(10), 761–767 (2004).

[3] Alcantarilla, P. F., Bartoli, A., and Davison, A. J., "KAZE features," in [*European Conference on Computer Vision*], 214–227, Springer (2012).

[4] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L., "Speeded-up robust features (SURF)," *Computer vision and image understanding* **110**(3), 346–359 (2008).

[5] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *International journal of computer vision* **60**(2), 91–110 (2004).

[6] Sirmacek, B. and Unsalan, C., "Urban-area and building detection using sift keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing* **47**(4), 1156–1167 (2009).

[7] Goncalves, H., Corte-Real, L., and Goncalves, J. A., "Automatic image registration through image segmentation and SIFT," *IEEE Transactions on Geoscience and Remote Sensing* **49**(7), 2589–2600 (2011).

[8] Yin, X., Yin, X.-C., Hao, H.-W., and Iqbal, K., "Effective text localization in natural scene images with mser, geometry-based grouping and adaboost," in [*Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*], 725–728, IEEE (2012).

[9] Gómez, L. and Karatzas, D., "MSER-based real-time text detection and tracking," in [*2014 22nd International Conference on Pattern Recognition*], 3110–3115, IEEE (2014).

[10] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Van Gool, L., "A comparison of affine region detectors," *International journal of computer vision* **65**(1-2), 43–72 (2005).

[11] Vedaldi, A. and Fulkerson, B., "VLFeat: An open and portable library of computer vision algorithms," in [*Proceedings of the 18th ACM international conference on Multimedia*], 1469–1472 (2010).

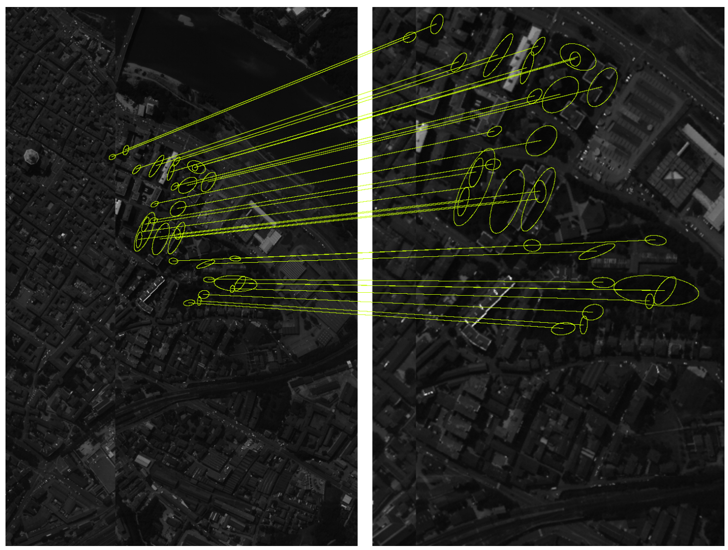
[12] Cariou, C., Chehdi, K., and Le Moan, S., "BandClust: An unsupervised band reduction method for hyperspectral remote sensing," *IEEE Geoscience and Remote Sensing Letters* **8**(3), 565–569 (2010).

[13] Obdrzalek, S. and Matas, J., "Object recognition using local affine frames on distinguished regions.," in [*BMVC*], **1**, 3, Citeseer (2002).

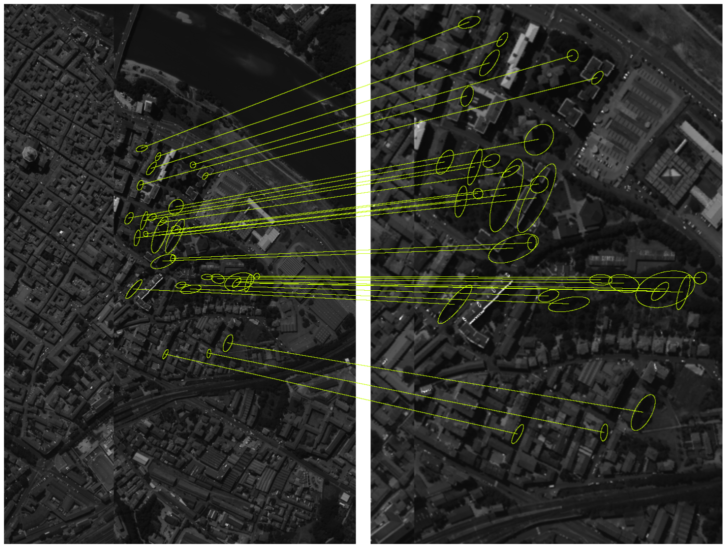
[14] Ordóñez, Á., Argüello, F., and Heras, D. B., "Alignment of hyperspectral images using KAZE features," *Remote Sensing* **10**(5), 756 (2018).

[15] "Repository of hyperspectral images." https://gitlab.citius.usc.es/hiperespectral/RegistrationRepository (2017). [Online; accessed 26-06-2020].

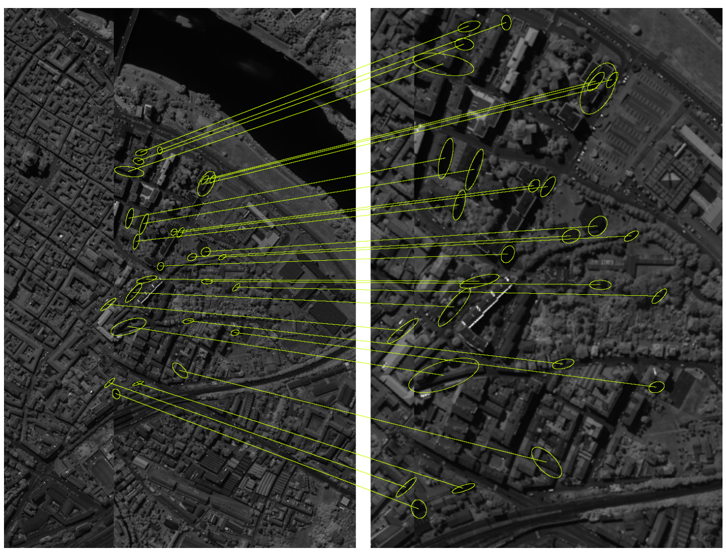
[16] Ordóñez, Á., Heras, D. B., and Argüello, F., "SURF-based registration for hyperspectral images," in [*IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*], 63–66, IEEE (2019).



(a) BandClust subband 1 (1–23)



(b) BandClust subband 2 (24–60)



(c) BandClust subband 3 (61–102)

Figure 3: Matched ellipses detected in three pairs of BandClust subbands of the Pavia Centre image (scale 2.0×).