A Novel SIFT-Like-Based Approach for FIR-VS Images Registration

by Cristhian Aguilera*, Fernando Barrera‡, Angel D. Sappa‡ and Ricardo Toledo‡†

Dept. of Electrical and Electronics Engineering, Collao 1202 University of Bio-Bio, Concepcion–Chile, cristhia@ubiobio.cl
‡ Computer Vision Center, edifici O, campus UAB, 08193 Bellaterra, Barcelona – Spain
† Computer Science Department, edifici O, campus UAB, 08193 Bellaterra, Barcelona – Spain

Abstract

This paper presents a novel approach, based on a SIFT-like scale-space pyramidal scheme, for registering Far-Infrared and Visible Spectrum images. The proposed approach consists of 3 stages. Initially, a key point detector is used to find interest points. This first stage is a SIFT-like based scale space representation. Then, each key point is characterized using an edge-oriented-histogram (EOH) in its neighborhood. Finally, a matching stage finds the nearest pairs in the couple of multispectral images. Experimental results are provided showing the validity of the proposed approach for registering such a kind of multispectral pair of images.

1. Introduction

The analysis of multispectral or multiband imaging has recently attracted the attention of the research community for applications in the image and video processing (e.g., [1], [2], [3], [4]). The information provided by the images from different spectral bands, either directly or supplementary to the information provided by the visible spectrum, can help to tackle in an efficient way different problems. However, the processing of images from spectral bands outside the visible spectrum requires the development of new tools, or the adaptation of current ones, opening new challenges in the image and video processing community. Multispectral analysis has been widely studied in the remote sensing field, where satellite images are registered and fused. Recently, due to advances in the technology, multispectral imaging is being used in applications such as video surveillance or driver assistance (e.g., [5], [6], [7]); just to mention a few; where visible (VS) images are merged with Near-Infrared (NIR), Mid-Infrared (MIR) or Far-Infrared (FIR). This classification depends on the application field, in the current work the wavelength associated to every spectral band that corresponds to the technology of the sensors [8] are as indicated in Table 1.

Table 1. *General spectral bands based on atmospheric transmission and sensor technology*

Abbreviation	Designation	Wavelength (µm)
VS	Visible	0.4 – 0.7
NIR	Near InfraRed	0.75 - 1.4
SWIR	Short-Wave InfraRed (SWIR)	1.4 - 3
MWIR	Mid-Wave InfraRed (MWIR)	3 - 8
LWIR	Long-Wave InfraRed (LWIR) (Forward-looking infrared, FLIR, Thermal Infrared or Far Infrared, usually referred to as **FIR**)	8 - 15

Although all of them are tackled under the multispectral processing framework, different approaches need to be developed depending on the specific spectral bands. Recent studies have shown that as the spectral band go away from visible spectrum, classical feature descriptors generally used for finding matching and registration of images in the visible spectrum (e.g., SIFT, SURF, etc.) are useless (e.g., [9], [10])

This problem, which has been tackled in NIR-VS multispectral imaging (e.g. [11] [12]), becomes critical in the FIR-VS image registration, especially when the images are further away from the visible spectrum (LWIR-FIR). The difficulty in FIR-VS image registration corresponds to the nonlinear relationship between pixel intensities, which is one of the differences with respect to NIR-VS and MIR-VS images. Variations in FIR intensities are determined by variations in the temperature of the objects, while variations in VS intensities come from color object and light reflections. Therefore, this nonlinear relationship results in a lack of correlation between their respective gradients. Furthermore, FIR images appear smoother, with loss of detail and texture [12], so that the detection of corners, as candidates for local descriptor points, is also poorly favored. In conclusion, most of the image processing tools that use gradient of pixels based descriptors need to be adapted, otherwise become useless. Figure 1 shows a couple of images, from the same scenario, obtained with a camera working in the visible spectrum and a camera in the infrared spectrum; their corresponding histograms are provided in the right column showing the lack of contrast in the infrared image as well as showing how

some details are missed. In addition to the lack of contrast and missed details it can be observed that in the infrared image there is a big difference in the transformation of intensities with respect to the one presented in the visible spectrum. This fact becomes critical to define the method that best describe this kind of images, since the transformations of intensity pixels in the infrared spectrum are non linear or non correlated with respect to the corresponding visible ones. Additionally, it should be noticed that in the FIR images other kind of information is available, which is not present in the visible one. The latter is an important characteristic where there is not enough illumination in the given scene, for instance when driving a car at night (e.g., [6], [13]).
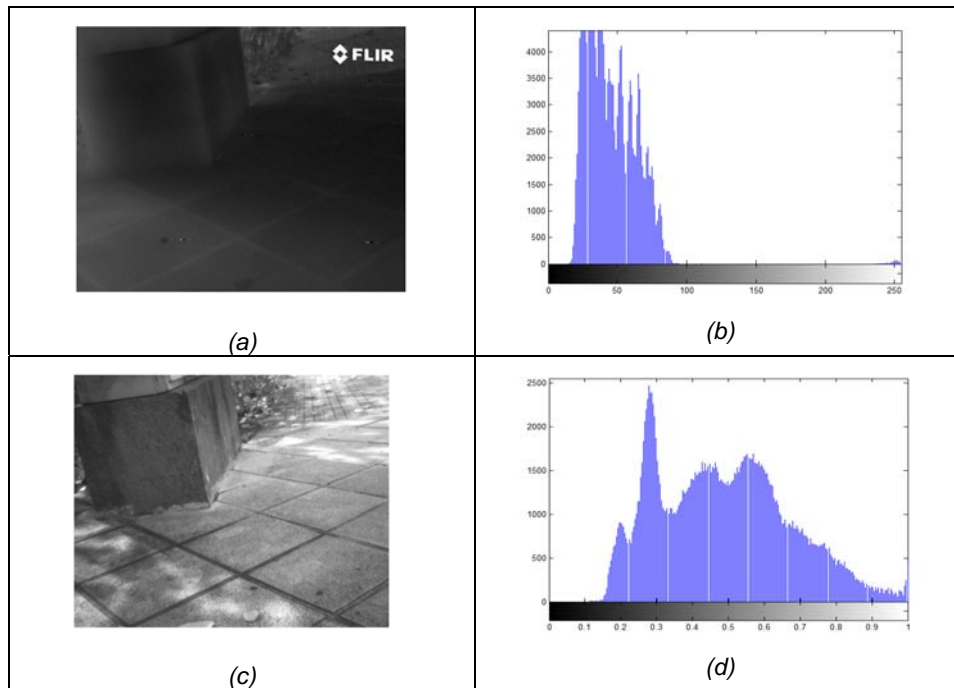

Fig. 1. *(a) FIR image. (b) Histogram of FIR image. (c) VS image. (d) Histogram of VS image.*

SIFT has proved to be a robust tool for detecting and describing local features in visible spectrum images; it is robust to changes in scale, rotation, affine transformations and lighting. In particular, the detector based on the sift space-scale has proven to be very stable and robust for the detection of keypoints. This robustness to handle VS images is exploited for image registration. A fundamental aspect of this descriptor is the large amount of detected feature points. Unfortunately, this number drops dramatically when low-contrast images are considered. A similar effect happen when FIR images are considered. The current work shows that the correct tuning of SIFT parameters is a critical task, not only in the parameters suggested by Lowe [14], but also other parameters that favor the generation of edges and corners. Our study shows very significant improvements in the number of feature points obtained from FIR images, and therefore the potential number of point that could be correctly matched (these points will be referred to as *potential keypoints*), thanks to a proper tuning of the space-scale pyramid generation in SIFT. On average, an improvement of about 91% of keypoint detected in FIR images and about 32% of potential keypoints generated due to a correct tuning of space-scale SIFT generation. Table 2 presents the average keypoints detected when a data set defined with 100 pairs of FIR-VS images is considered.

Table 2. *Average of keypoints and potential keypoints for 100 pairs of FIR-VS images*

	KeyPoint (Lowe [10])	KeyPoint (best-tune)	Percentage of Improvement
VS Images	2005	3651	82%
FIR Images	474	904	91%
Potential Keypoint	**33**	**43**	**32%**

In the literature, much of the effort devoted to improve the characteristics of local descriptors such as SIFT, is primarily focused at improving the quality during the matching stage [15] [16]. This can help to improve the matching only in those cases where the images are quite similar, for instance in the case of NIR-VS image registration, but not in the case of FIR-VS. Actually, even in the cases where feature points are correctly detected in both FIR and VS images, the results remain very poor during the registration due to the orientations of SIFT descriptors in FIR images are different and uncorrelated with those from VS images. In other words, the descriptors are different; hence it is not correct to try to use

them for registration. Figure 2 shows this situation; in Fig. 2(a) a small area of a given FIR image is presented, the corresponding gradients are depicted in Fig. 2(b). The same area, but from the VS image, is shown in Fig. 2(c); it can be observed how the corresponding gradients are completely different with respect to the previous ones, which avoids finding correlations between the images. Actually, all the techniques that use gradient based descriptors will fail when trying to find correspondences between these kinds of images. Recently, some work have been proposed to tackle this lack of correlation; for instance [17] proposes *Gradient Orientation Modification* (GOM-SIFT) and *Scale Restriction* (SR-SIFT) to improve both the descriptor and the matching rate between the multispectral images. Similarly, another modification of SIFT has been presented in [18], *Orientation Restricted* (OR-SIFT), trying to improve the matching performance. Finally, in [9] an *Edgelets and Histogram of Oriented Gradient* features (EHOG) scheme has been proposed for pedestrian detection to improve the descriptor. Unfortunately, none of these approaches is good enough to produce a good result when the spectral bands of the images are far away from each other.
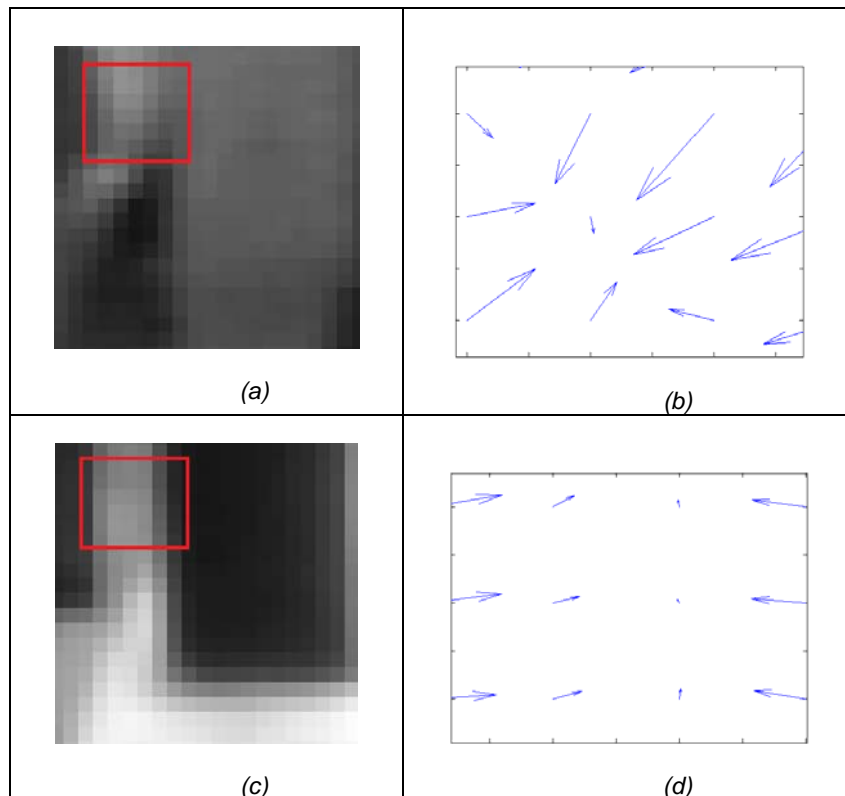

Fig. 2. *(a) FIR image. (b) Gradients of FIR image. (c) VS image. (d) Gradients of VS image.*

In order to tackle this problem a new scheme for FIR-VS image registration based on a SIFT-like detector and descriptor is proposed in the current work as a second contribution, in addition to the study mentioned above.

2. Proposed Approach

The proposed scheme, which will be referred to as *Edge Oriented Histogram descriptor* (EOH-SIFT), consists of a scale-space pyramid, like the one used by SIFT. Similarly, invariant features are used but by modifying the feature vector in such a way to incorporate spatial information from the contours of each keypoint without using gradient information. This allows us to generate a correlated parameter space in both the FIR and VS images. Our proposal uses a descriptor based on the edge histogram. This edge orientation histogram describes the shapes and contours from FIR images, keeping in the scale-space their invariance.

Figure 3 presents a flow chart of the proposed method. It consists of three steps as detailed next. Firstly, keyPoints are detected by using a scale-space representation. It results in a set of stable keypoints, similar to those resulting from classical SIFT. Note that this result is invariant to scale, position and orientation, which are needed for registration applications. It should be also highlighted that finding the right tuning for the scale-space representation is a crucial task for detecting the potential keypoints, in particular due to the low contrast of the FIR images. The current work shows that the best generation of potential keypoints is obtained by setting the SIFT detector with the following parameters: sigma=1.2 and threshold=40.
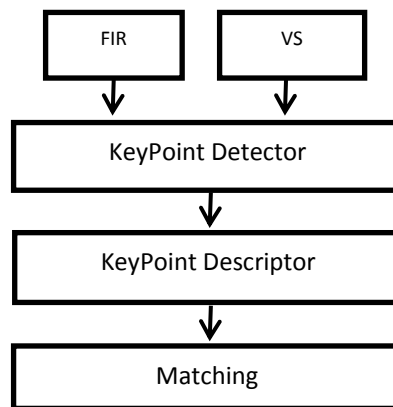
Fig. 3. *Flowchart of EOH-SIFT algorithm.*

The second step, which represents the main contribution of current work, is the development of a robust descriptor using information from the neighborhood of the detected keypoints. The idea behind the proposed descriptor is to keep in mind the non linear relationships between image intensities so that it should be mainly based on global region information instead of local one. The proposed descriptor is based on the use of histograms of the contours in the neighborhood of the given keypoints. Figure 4 illustrates the proposed scheme, where it can be observed how the proposed descriptor is composed. Initially, centered in every keypoint, a region of NxN pixels is created. This region is split up into 4x4=16 subregions. Each one of these subregions is represented by a histogram of contours of 5 bins, corresponding to contours' orientations of: 0 degree, 45 degrees, 90 degrees, 135 degrees and a bin with no-orientation (no) that corresponds to those subregions that do not a contain contour. The result from this second stage is a normalized vector of characteristics, which will be also referred to as descriptor vector, of 16x5=80 elements for every keypoint.
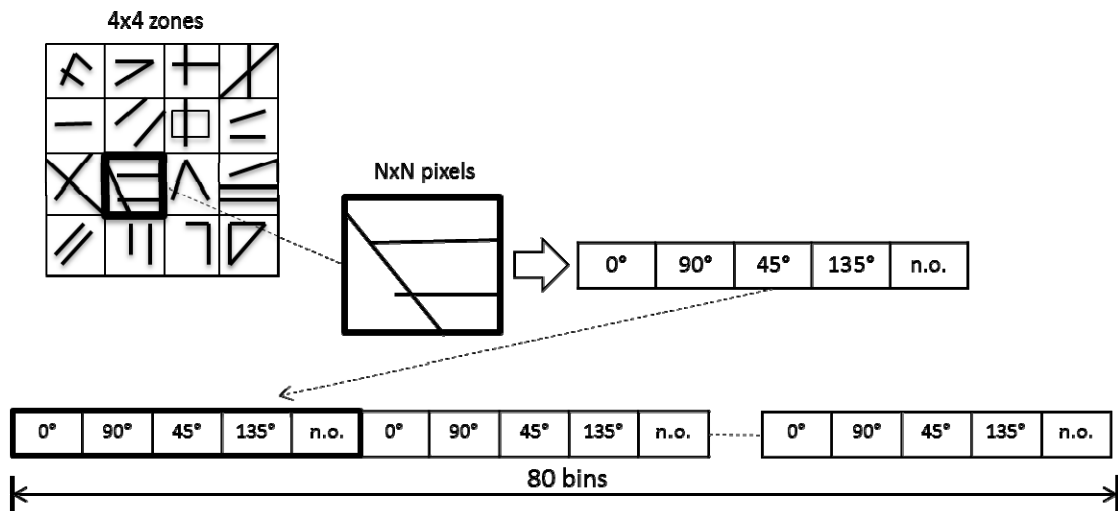

Fig. 4. *Proposed EOH-SIFT image descriptor to register VS-FIR images.*

The current work shows that the selection of the right window's size (N) for every keypoint is an important factor of the proposed scheme. A window with a small or big size will increase the number of wrong matching of keypoints. Table 3 presents the performance of the proposed approach when windows with different sizes are considered to compute the vector of characteristics mentioned above; these values represent the average obtained with the whole data set, which contains 100 pair of multispectral images. It can be seen that the best performance is obtained when windows with sizes in between 80x80 and 100x100 are used.

Table 3. Average correct matching from the whole data set; images of 408x506 pixels

Window size	20x20	40x40	60x60	80x80	100x100	120x120	140x140
SIFT	6%	6%	6%	6%	6%	6%	6%
EOH-SIFT	2%	17%	27%	36%	37%	35%	32%

The third and last step corresponds to the kepoint matching process between the two images. This process finds the matches between keypoints using the Euclidean distance between the corresponding descriptor vectors. Two keypoints from different images are matched if the Euclidean distance between their vectors is smaller than a given threshold *th1* and corresponds to the minimum distance of the whole set of keypoints. Similarly to in the SIFT algorithm, to increase the matching robustness, two keypoints are matched only if the ratio between the first and second best matches is smaller than a second threshold *th2*. Figure 5 shows the different steps of the proposed algorithm for a pair of keypoints from VS and FIR images.
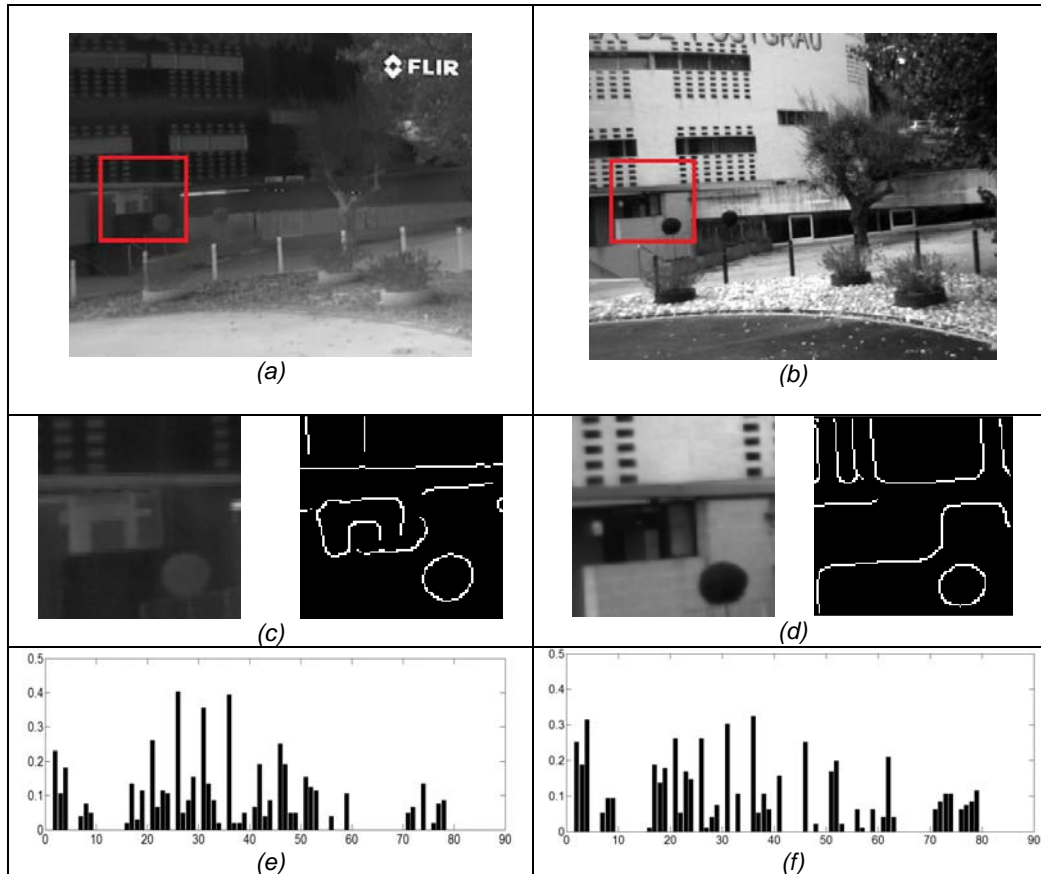

(a)   *(b)*

(c)   *(d)*

(e)   *(f)*

Fig. 5. *Steps of the proposed VS-FIR registration method: (a) and (b) keypoints and neighborhood; (c) and (d) images of contours; (e) and (f) Histograms used as descriptor vectors.*

In addition to the threshold mentioned above, the matching robustness is increased by discarding those keypoints that have some of their subregions without information (i.e., subregions only containing a few contours); finally, scale restrictions such as the one proposed in [17] are also used to improve the performance of the proposed approach.

3. Experimental Results

The proposed registration approach has been evaluated with a data base containing 100 pair of VS-FIR images. These images have been obtained using the cameras detailed in Table 4.

Table 4. Camera specifications

Specifications	VS	FIR
Image sensor type	CCD	Thermal
Resolution	640×480	164×129
Wavelength	0.4 to 0.7µm	7.5 to 13.5µm
Focal length	6 mm	18 mm

Figure 6 presents the results obtained by matching a pair of FIR-VS images with SIFT and the proposed EOH-SIFT method (both with and without the additional constraints mentioned above). Note that since this pair corresponds to rectified images the matching should correspond to keypoints lying in the same rows; in other words the segments that connect keypoints should be horizontal lines. It can be appreciated that the classical SIFT (Fig. 6*(a)*) only matches correctly 32 keypoints, from a total of 69 keypoints, which represents a 46% of success. On the contrary, in the case of EOH-SIFT (Fig. 6*(b)*) 39 keypoints, from a total of 47, are correctly matched. This represents a 83% of success, which can be increased just by impossing additional constraints during the matching process. Since most of wrong matches correspond to keypoints where the descriptor vector contains some subregions without information, just by removing keypoints with more than 2 subregions without information and by applying scale restrictions [17] the percentage of wrong matches is considerably reduced. Actually, 28 keypoints are correctly matched by applying these constraints (Fig. 6*(c)*), from a total of 31 keypoints, which represents a 90% of success.
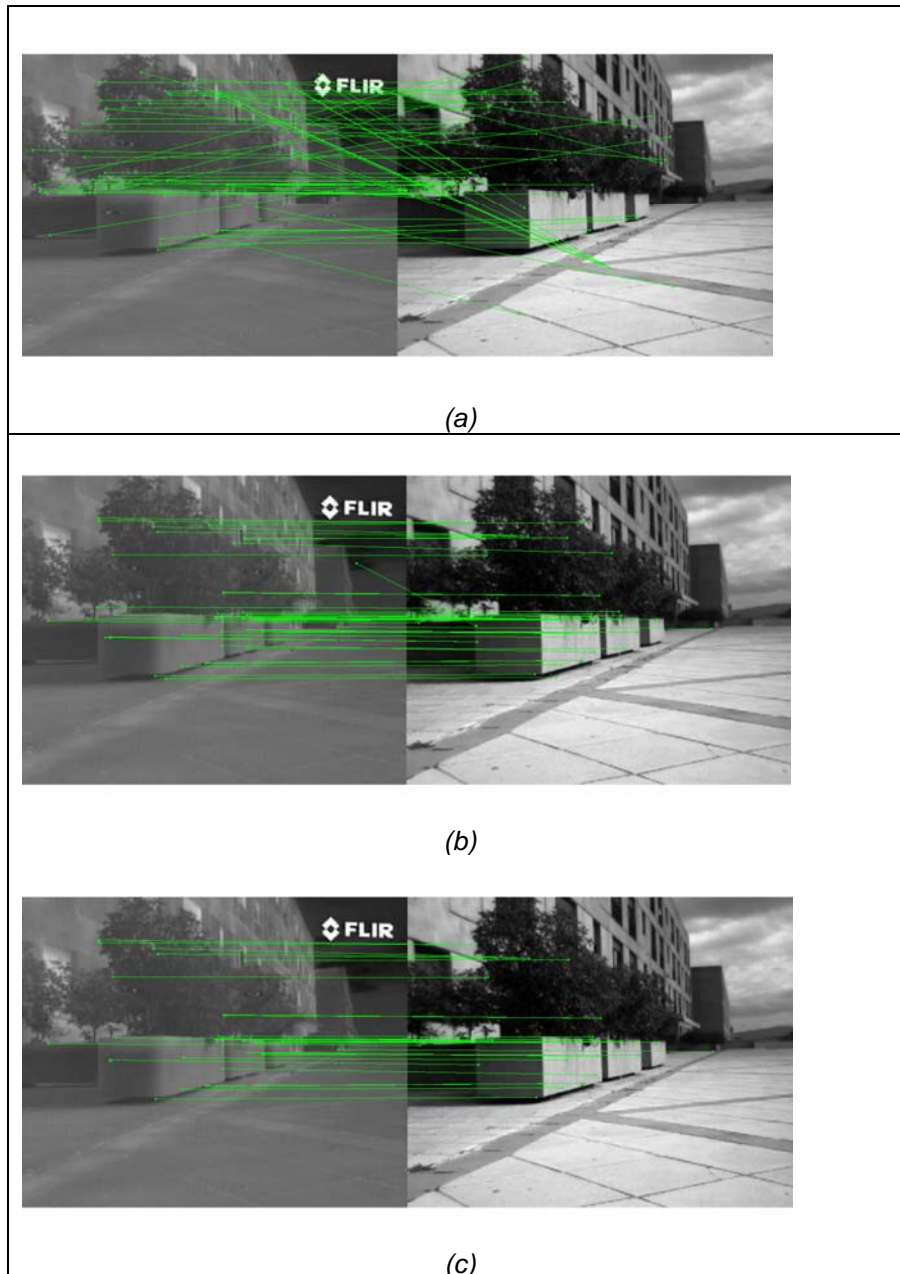

(a)

(b)

(c)

Fig. 6. *Keypoints matched (green segments) with: (a) SIFT; (b) EOH-SIFT without constraints; (c) EOH-SIFT with constraints.*

The proposed method has been applied to data base of 100 multispectral image pairs captured using a VS and a FLIR cameras. In all the cased the images were correctly rectified and aligned so that matches should be found in horizontal lines. The data base contains outdoor images of different urban scenarios. In order to evaluate the performance of the proposed approach a *Precision and Recall* scheme has been used, as suggested by Tan [16]. Table 5 shows average results obtained with different descriptors.

Table 5. Average (*1-precision*) and *recall* for 100 VS-FIR images

FIR-VS	*Recall*	*(1-Precision)*
SIFT	6%	93%
SURF	45%	96%
GOM-SIFT	14%	86%
OR-SIFT	5%	97%
EOH-SIFT	74%	59%

From Table 5 we can appreciate that the proposed EOH-SIFT has the best performance when compared with all the other methods (i.e., SIFT[14], SURF[19], GOM-SIFT [17] that also includes scale restrictions, and OR-SIFT[18]). Figures 7 and 8 also show these results but by depicting the maximums, minimums and averave values for each one of these algorithms.
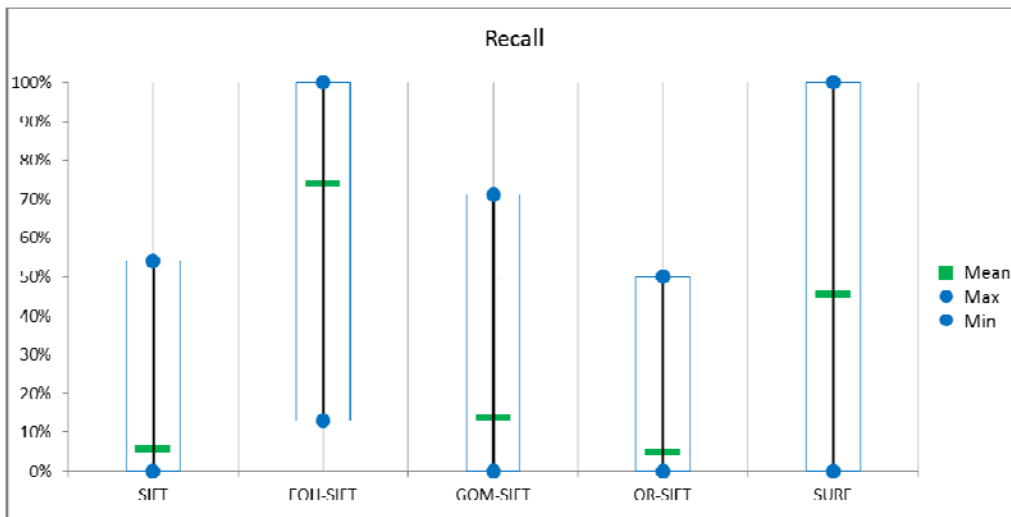

Fig. 7. *Recall for the methods evaluated with the 100 image pairs from the data set.*
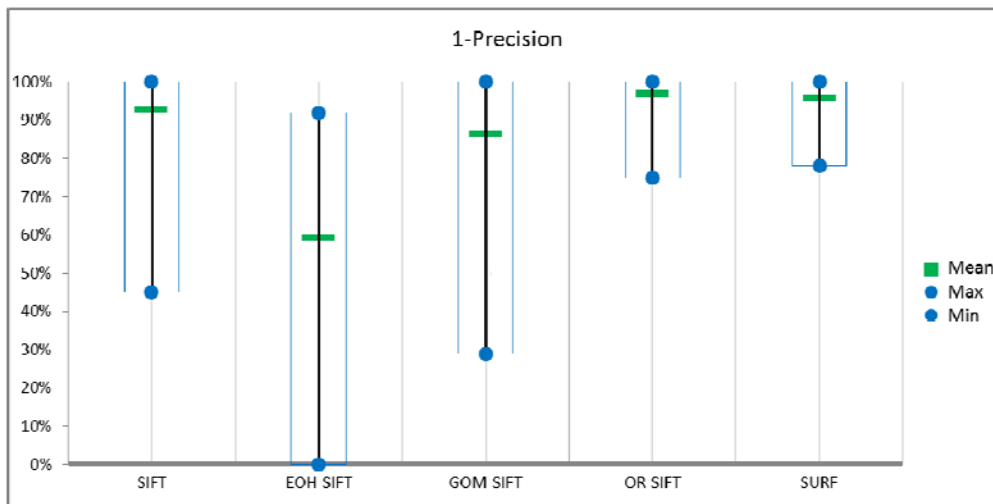

Fig. 8: *1-Presicion for the methods evaluated with the 100 image pairs from the data set.*

4.- Conclusions and Further Work

This paper presents a novel feature descriptor method for the registration of multispectral image pairs (VS-FIR). It is based on the use of SIFT detector, which is a robust and stable approach for keypoint detection. The results obtained in the current work show that classical approaches have a low performance due to the transformations suffered by the intensities of images pixels in the infrared spectrum and consequently their corresponding gradients. The latter makes the use of classical descriptors not appropriated to this kind of problems. In the current work a new descriptor, based on the use of edge histograms (EOH), has been proposed. It is evaluated with respect to the state of the art; results show that the proposed approach (EOH-SIFT) is a valid option for registering multispectral VS-FIR images. Furthermore, the incorporation of additional constraints for the matching process considerably improves the performance of the proposed method. It is evaluated with a large data set of pair of VS-FIR images showing the validity of the proposed approach. As a future work research in three topics will be carried out. First, improvements in the keypoint detector, in order to increase the number of selected points in this kind of multispectral images, will be considered. Additionally, the proposal of new descriptors, looking for an invariant affine transformation, will be tackled. Finally, studies on further improvements in the matching process will be considered.

5.- Acknowledgement

REFERENCES

[1] Mahmudul H., Xiuping J., Robles A., "Multi-spectral remote sensing image registration via spatial relationship analysis on SIFT keypoints". IGARSS, pp. 1011-1014, 2010.

[2] Brown M., Su S., "Multi-spectral SIFT for Scene Category Recognition". Computer Vision and Pattern Recognition, pp. 177-184, 2011.

[3] Abdelrahman M., Ali A., Aly A., Farag A., "Precise Change Detection In Multi-Spectral Remote Sensing Imagery Using SIFT-based Registration". Multimedia Technology, pp. 6238-6242, 2011.

[4] Lalonde M., Byrns D., Gagnon L., Teasdale N., Laurendeau D., "Real-time eye blink detection with GPU-based SIFT tracking". Computer and Robot Vision, pp. 481-487, 2007.

[5] Leykin A., Hammoud R., "Pedestrian tracking by fusion of thermal-visible surveillance videos". Machine Vision and Applications, vol. 21, pp. 587–595, 2010.

[6] Krotosky S., Trivedi M., "On color-, infrared-, and multimodal-stereo approaches to pedestrian detection". IEEE Transactions on Intelligent Transportation Systems, vol.8, no.4, pp.619-629, 2007.

[7] Jung S., Eledath J., Johansson S., Mathevon V., "Egomotion estimation in monocular infra-red image sequence for night vision applications". IEEE Workshop on Applications of Computer Vision, WACV '07., pp.8, 2007.

[8] Barrera C.,"Towards a multimodal stereo rig for ADAS: State of art and algorithms". Master tesis Centre de Visi´o per Computador. Universitat Autònoma de Barcelona, 2009.

[9] Istenic R., Heric D., Ribaric S, Zazula D., "Thermal and Visual Image Registration in Hough Parameter Space". Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services, pp.106-109, 2007.

[10] Lee J., Kim Y., Lee D., Kang D., Ra J. , "Robust CCD and IR Image Registration Using Gradient-Based Statistical Information". IEEE Signal Processing Letters, vol.17, no.4, pp.347-350, 2010.

[11] Sadeghipoor Z., Lu Y., Süsstrunk S., "Correlation-Based Joint Acquisition and Demosaicing of Visible and Near-Infrared Images". IEEE Int. Conf. on Image Processing, pp. 3226-3229, 2011.

[12] Zhang L., Wu B., Nevatia R., "Pedestrian Detection in Infrared Images based on Local Shape Features". IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[13] Trivedi M., Cheng S., Childers E., Krotosky S., "Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation". IEEE Transactions on Vehicular Technology, vol. 53, no. 6, pp. 1698–1712, 2004.

[14] Lowe D., "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision, pp. 91-110, 2004.

[15] Yingdan W., Ming Y., "A Multi-sensor Remote Sensing Image Matching Method Based on SIFT Operator and CRA Similarity Measure". Intelligence Science and Information Engineering, pp. 115-118, 2011.

[16] Chunlin T., Wang H., Pei D., "SWF-SIFT Approach for Infrared Face Recognition". Tsinghua Science and Technology, vol. 15, no. 3, pp. 357-362, 2010.

[17] Yi Z., Zhiguo C., Yang X., "Multi-Spectral Remote Image Registration Based on SIFT". Electronics Letters, vol. 44, no. 2, pp. 107-108, 2008.

[18] Vural M., Yardimci Y., Temizel A., "Registration of Multispectral Satellite Images with Orientation-Restricted SIFT". Geoscience and Remote Sensing Symposium, vol. 3, pp. 243-246, 2009.

[19] Bay H., Ess A., Tuytelaars T., Van Gool T., "SURF: Speeded Up Robust Features". Computer Vision and Image Understanding (CVIU), Vol. 110, no. 3, pp. 346--359, 2008