Article

Multispectral Image Feature Points

Cristhian Aguilera [1],*, Fernando Barrera [2], Felipe Lumbreras [2,3], Angel D. Sappa [2] and Ricardo Toledo [2,3]

[1] Department of Electrical and Electronics Engineering, Collao 1202, University of Bío-Bío, 4051381 Concepción, Chile

[2] Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain; E-Mails: jfbarrera@cvc.uab.es (F.B.); felipe@cvc.uab.es (F.L.); angel.sappa@cvc.uab.es (A.D.S.); ricardo@cvc.uab.es (R.T.)

[3] Computer Science Department, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain

* Author to whom correspondence should be addressed; E-Mail: cristhia@ubiobio.cl; Tel.: +56-41-311-1287; Fax: +56-41-311-1013.

Abstract: This paper presents a novel feature point descriptor for the multispectral image case: Far-Infrared and Visible Spectrum images. It allows matching interest points on images of the same scene but acquired in different spectral bands. Initially, points of interest are detected on both images through a SIFT-like based scale space representation. Then, these points are characterized using an Edge Oriented Histogram (EOH) descriptor. Finally, points of interest from multispectral images are matched by finding nearest couples using the information from the descriptor. The provided experimental results and comparisons with similar methods show both the validity of the proposed approach as well as the improvements it offers with respect to the current state-of-the-art.

1. Introduction

The analysis of multispectral or multiband imaging has recently attracted the attention of the research community for applications in the areas of image and video processing (e.g., [1–4]). The

information provided by the images from different spectral bands, either directly or supplementary to the information provided by the visible spectrum, can help tackle different problems in an efficient way. 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 for the image and video processing community. Multispectral analysis has been widely studied in the remote sensing field, where satellite images are commonly 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., [5–7]); just to mention a few; where visible (VS: 0.4–0.7 µm) images are merged with Near-Infrared (NIR: 0.75–1.4 µm), Short-Wave Infrared (SWIR: 1.4–3 µm), Mid-Wave Infrared (MWIR: 3–8 µm) or Long-Wave Infrared (LWIR: 8–15 µm) ones.

Feature points are at the base of different computer vision problems and are generally studied at three levels—feature point: detection, descriptor and matching. One of the most successful and widely used approaches is the SIFT algorithm [8]. Over the last decade several variations of the original algorithm, as well as some novel approaches, have been proposed focusing on the perceived weaknesses of SIFT (e.g., [9,10]). All the approaches mentioned before are intended for applications that involve images from the same spectral band, generally Visible Spectrum (VS) images. Recently, applications that combine feature points from different spectral band images are being developed. These works are mainly based on the use of classical SIFT algorithm, or minor modifications to the classical approach. For instance [11] proposes a scale restriction criteria in order to reduce the number of incorrect matches of SIFT when it is adopted to tackle the multispectral case. Accurate matching results have been reported when the spectral bands of the pair of images are somehow near (VS-NIR); however, further improvements are needed for tackling those cases where the spectral bands are far away from each other (VS-LWIR). Actually, recent studies have shown that as the spectral bands 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., [12,13]).

The difficulty in finding correspondences between feature points from VS-LWIR images results from the nonlinear relationship between pixel intensities. Variations in LWIR intensities are related to 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, LWIR images appear smoother, with loss of detail and texture [14], 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, or otherwise they 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 describes 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 LWIR images other kinds 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., [6,15]).

Figure 1. (**a**) LWIR image; (**b**) Histogram of LWIR image; (**c**) VS image; (**d**) Histogram of VS image.

(a)   (b)

(c)   (d)
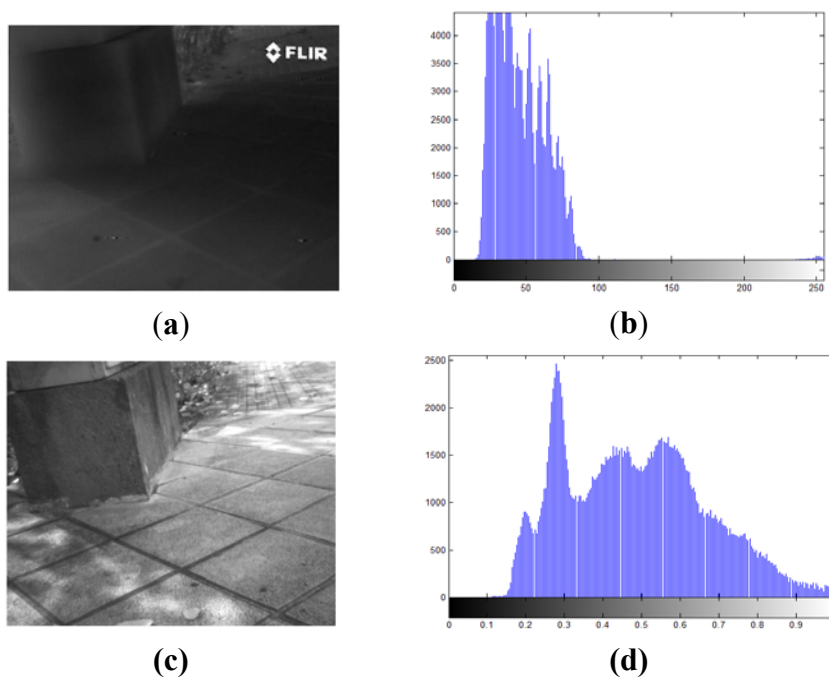
From the three levels mentioned above—feature point: detection, descriptor and matching—the feature point descriptor becomes *the key element* when images from VS and LWIR spectrum are considered. Even though the percentage of correct matching can be improved by introducing modifications at the matching stage, results remain very poor when SIFT, or modifications of it (e.g., [11,16]), are used as descriptors in the multispectral case (VS-LWIR), as will be presented in the Experimental Results section. This low correspondence rate is mainly due to the lack of descriptive capability of gradient in LWIR images, which in general appears smoother with loss of detail and texture. Actually, even in the cases where the detected feature points correspond to the same position in both images, the matching results remain quite poor due to the differences in their gradient orientation, which is used as a descriptor by SIFT. In other words, since the descriptors are by nature different it is not correct to try to use them for finding similarities and matches.

2. Proposed Approach

The proposed scheme 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 VS and LWIR images. Our proposal uses a descriptor based on the edge histogram. This edge orientation histogram describes the shapes and contours from LWIR images, keeping in the scale-space their invariance. Figure 2 presents a flow chart of the proposed method. It consists of three steps, namely detection, description and matching, as detailed next.

**Figure 2.** Flowchart of the proposed approach.
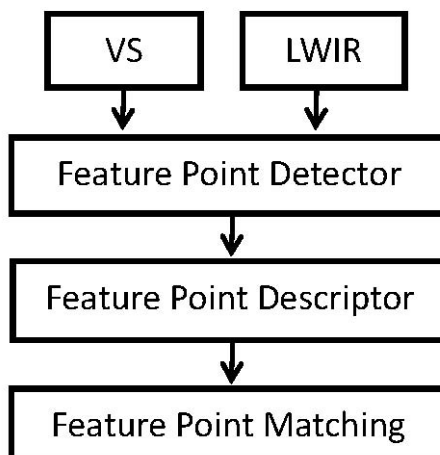
2.1. Feature Point Detection

Feature points, which will be also referred to as keypoints, are detected by using a scale-space pyramidal representation. They correspond to the maxima and minima of a difference of Gaussiass applied over a series of smoothed and resampled images [8]. The result from this first stage is 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. In the current implementation potential keypoints are obtained by setting the SIFT like detector with the following parameters: sigma = 1.2 and threshold = 40. The same parameters' setting is used in both multispectral images resulting in a set of $\mathbf{P}_{VS}$ and $\mathbf{P}_{LWIR}$ keypoints. Each keypoint is denoted by a vector (x_i, y_i, σ_i), where (x_i, y_i) correspond to the location and (σ_i) is the scale of that pyramidal representation where the keypoint appears.

2.2. Feature Point Description

Detected feature points are described through the use of an Edge-Oriented-Histogram, which is the main contribution of current work and will be referred to henceforth as EOH. These EOHs incorporate spatial information from the contours in the neighbourhood of each feature point. They describe the shapes and contours from both VS and LWIR images. The idea behind the proposed descriptor is motivated by the nonlinear relationships between image intensities so that the descriptor should be mainly based on region information instead of pixel information. Hence, the proposed descriptor is based on the use of histograms of contours' orientations in the neighbourhood of the given keypoints. Initially, both images (VS and LWIR) are represented by means of their edges, which are extracted using the Canny edge detector algorithm [17]. In all the cases Canny's thresholds have been automatically set relative to the highest value of the gradient magnitude of the image and $\sigma = 4$. Once both images are represented by means of their edges, feature points detected in Section 2.1 are described as illustrated in Figure 3. The different steps of the feature point descriptor are detailed below.

Figure 3. Proposed EOH based keypoint image descriptor for VS-LWIR images.
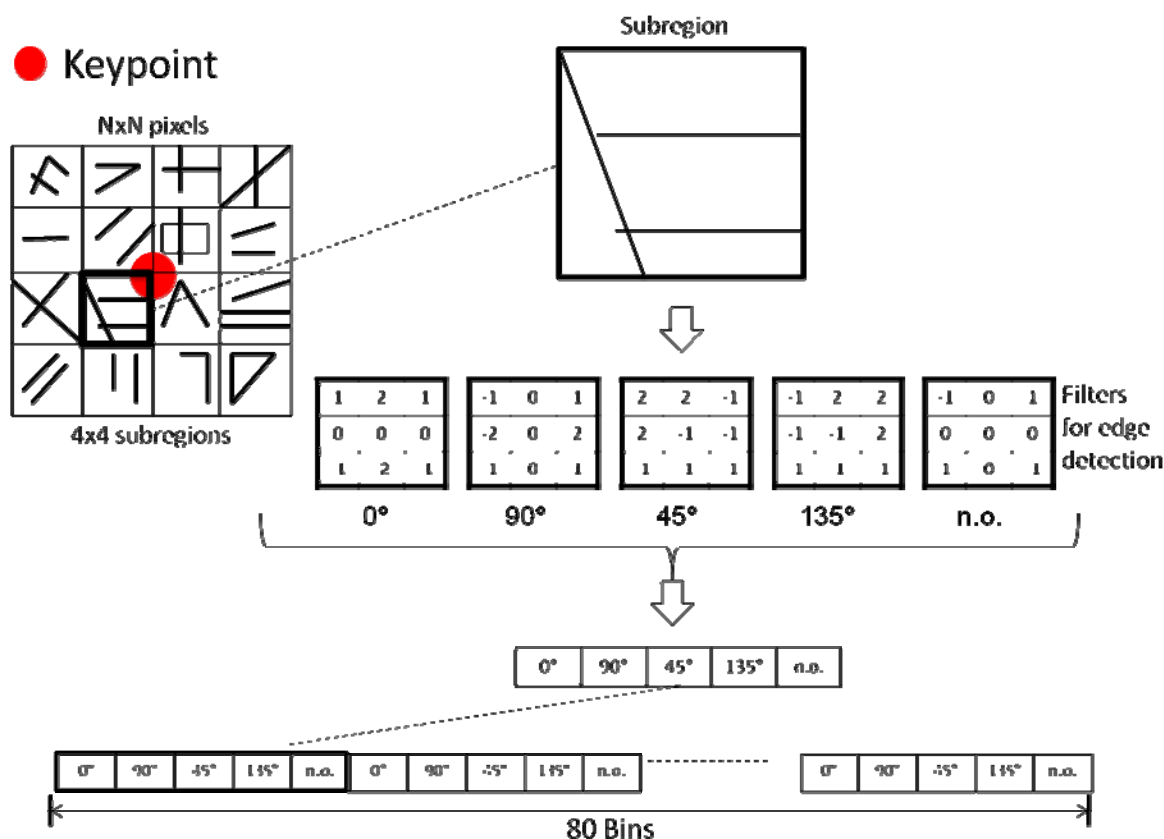
Firstly, a region of $N \times N$ pixels, centered at the given keypoint, is obtained. Then, this region is split up into $4 \times 4 = 16$ subregions. Finally, each one of these subregions is represented by a histogram of contours computed following the Edge Histogram Descriptor (EHD) [18] of the MPEG-7 standard [19]. This histogram represents the spatial distribution of four directional edges and one non-directional edge (five bins in total); these bins correspond to contour orientations of 0, 45, 90, and 135 degrees; additionally a bin with no orientation (n.o.) is considered, which corresponds to those areas that do not contain a contour. Every pixel of each subregion contributes to a bin of the histogram according to the five filters, of 3×3 pixels, shown in Figure 3. The filter with the largest value is used as a criterion to vote in the corresponding bin. After processing all the elements of the 16 subregions a vector with 80 elements is obtained ($16 \times 5 = 80$). This vector is normalized (by dividing each one of its components by the Euclidean norm) and used as the vector of characteristics, which will be also referred to as the descriptor vector for every keypoint.

The selection of the right window size (N) for every keypoint is an important factor of the proposed scheme. A window with a too small or big size will increase the number of wrong matching of keypoints. Table 1 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 80×80 and 100×100 are used.

Table 1. Average correct matching from the whole data set; images of 408 × 506 pixels.

Window Size	20 × 20	40 × 40	60 × 60	80 × 80	100 × 100	120 × 120	140 × 140
SIFT	6%	6%	6%	6%	6%	6%	6%
EOH-SIFT	2%	17%	27%	36%	37%	35%	32%

2.3. Feature Point Matching

This stage finds nearest keypoints from different spectral images, in the descriptor space by filtering feature vectors with low descriptive elements. This process is based on the Euclidean distance between the corresponding descriptor vectors. Like the SIFT algorithm, in order to increase the matching robustness, two keypoints are matched only if the ratio between the first and second best matches is smaller than a given threshold. Additionally, 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, the scale restriction proposed in [11], and used in [16], is also considered in the current work to improve the performance of the proposed approach. This scale restriction process consists in discarding incorrect matches using the scale difference (*SD*) of the given pair of keypoints $P_{VS}(x_{VS}, y_{VS}, \sigma_{VS})$ and $P_{LWIR}(x_{LWIR}, y_{LWIR}, \sigma_{LWIR})$:

$$SD(P_{VS}, P_{LWIR}) = \sigma_{VS} - \sigma_{LWIR} \tag{1}$$

as proposed in [11], the match is rejected if it does not satisfy the following scale restriction criteria:

$$a < SD < b \tag{2}$$

where the values *a* and *b* are obtained by first computing a histogram of *SDs* of all matches; then, the peak in that *SDs* histogram, which is noted as $\overline{SD}$, is extracted and used to define the sough values:

$$a = \overline{SD} - 0.9 \qquad b = \overline{SD} + 0.9 \tag{3}$$

Figure 4 shows the different steps of the proposed algorithm for a pair of keypoints from VS and LWIR images.

Figure 4. Steps of the proposed VS-LWIR keypoint description method: (**a**) and (**b**) keypoints on VS and LWIR images together with their corresponding neighborhoods; (**c**) and (**d**) images of contours of the given keypoints; (**e**) and (**f**) histograms used as descriptor vectors.

(**a**) (**b**)

Figure 4. *Cont.*

(c) (d)

(e) (f)
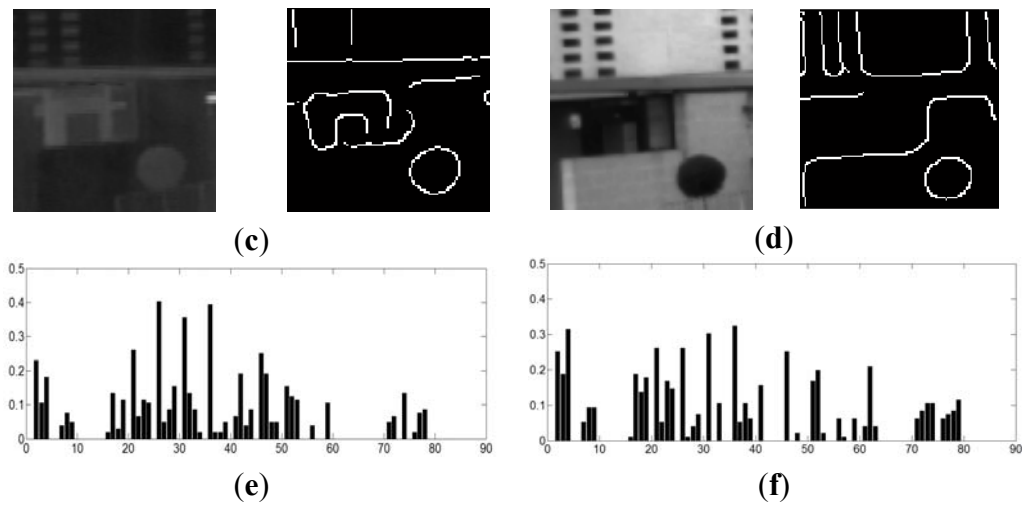
3. Experimental Results

The proposed approach has been evaluated with a data set containing 100 pairs of VS-LWIR images. These images have been obtained using the cameras detailed in Table 2.

Table 2. Camera specifications.

Specifications	VS	LWIR
Image sensor type	CCD	Thermal
Resolution	640 × 480	320 × 240
Wavelength	0.4 to 0.7 μm	8 to 14 μm
Focal length	6 mm	19 mm

In order to make the comparisons easier, in all the cases the images were rectified and aligned so that matches should be found in horizontal lines. This rectification and alignment process is just applied to facilitate the evaluation of the performance of the proposed approach as well as for further comparisons with other algorithms. The data set contains outdoor images of different urban scenarios. Figure 5 shows three pairs of multispectral images contained in the data set.

Figure 5. Illustrations of some of the VS-LWIR image pairs from the evaluated data set, which contains 100 pairs.

Figure 5. *Cont.*

This data set is available through our website (http://www.cvc.uab.es/adas/projects/simeve/) for evaluation and comparison with other multispectral feature point detectors and descriptors. The performance of the proposed approach is evaluated with a Precision and Recall scheme. Furthermore, results are compared with other implementations. Table 3 shows average results, computed over the whole data set, obtained with different descriptors; we can appreciate that the proposed approach has the best performance when compared with all the other methods.

Table 3. Average (1-Precision) and Recall for 100 VS-LWIR images.

VS\LWIR	*Recall*	*(1-Precision)*
SIFT [8]	6%	93%
SURF [10]	45%	96%
GOM-SIFT [11]	14%	86%
OR-SIFT [16]	5%	97%
Proposed Approach	**74%**	**59%**

Figure 6 just illustrates the results obtained with only two pairs of VS-LWIR images using both SIFT and the proposed EOH based descriptors (quantitative evaluation over the whole data set is presented in Table 3). Note that since these pairs correspond to rectified images the matching should correspond to keypoints lying in the same row; in other words the segments that connect keypoints should be horizontal lines.

Figure 6. Keypoints matched (green segments) in two image pairs (the whole data set contains 100 pairs of **VS**-LWIR images) using: (**top**) SIFT descriptor; (**bottom**) proposed EOH based descriptor.
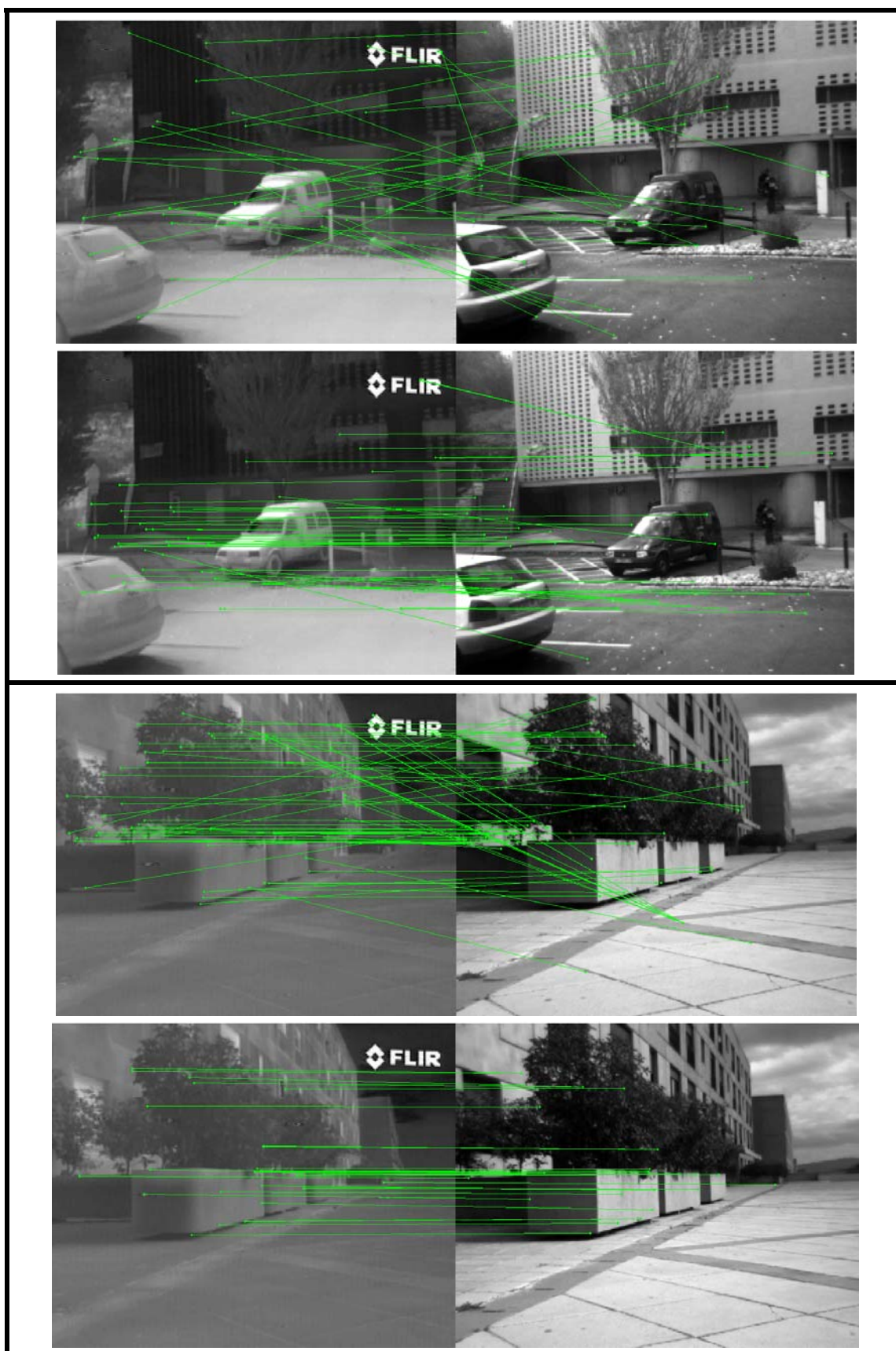
It can be appreciated that in the first case (top illustration) SIFT only matches correctly a few keypoints (two points), from a total of 34 keypoints, which represents about 5% success. On the contrary, when the proposed EOH based approach is considered, 36 keypoints, from a total of 51, are correctly matched (about 68% success). In the second case (bottom illustration), the classical SIFT algorithm matches 32 keypoints, from a total of 69 keypoints (46% success); while the proposed approach reaches the 90% success (28 keypoints are correctly matched from a total of 31).

Finally, although out of the scope of current work, Figure 7 shows how the proposed approach can also be used as a feature point descriptor when images from the same spectral band are considered. In these illustrations in can be seen that the proposed approach reaches 86.2% of success when a pair of VS-VS images is considered and 84.6% of success in the LWIR-LWIR case. Note that these two cases are just presented as illustrations; more rigorous evaluations and comparisons are needed, over a large set of image pairs, to obtain more solid conclusions.

Figure 7. Illustrations of the results obtained with the proposed approach when images from the same spectral band are considered: (**top**) VS-VS; (**bottom**) LWIR-LWIR.
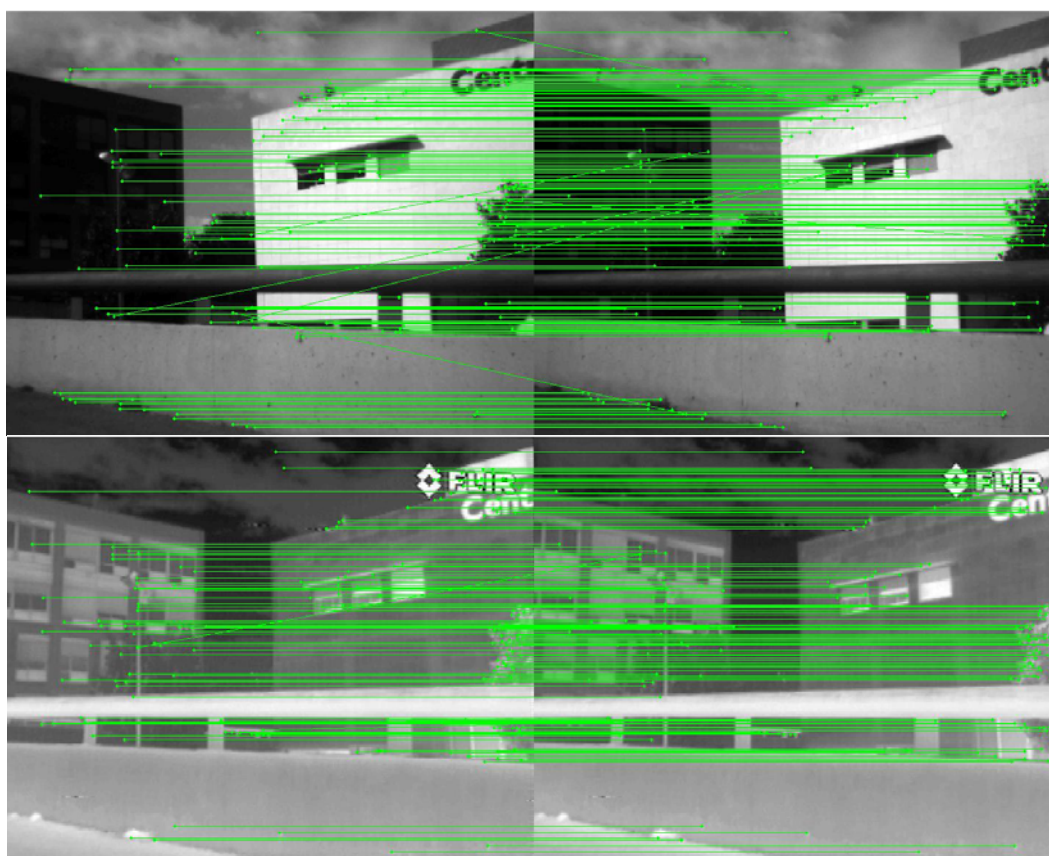
4. Conclusions

A novel multispectral feature descriptor method is presented, which is useful to register VS-LWIR images. It is based on the use of a SIFT-like detector to extract feature points. Then, an Edge Oriented Histogram (EOH) based approach is proposed as a robust descriptor for characterizing multispectral keypoints. Finally, matches are obtained by finding nearest couples in the feature description space. The proposed approach has been evaluated with a large set of multispectral pairs of images and

compared with the current state-of-the-art. Finally, it is shown that the proposed approach could also be also used when pairs of images from the same spectral band are considered. Future work will focus on studying the use of more elaborated edge descriptor techniques in order to improve the matching percentages.

Acknowledgments

References

1. Mahmudul, H.; Xiuping, J.; Robles, A. Multi-Spectral Remote Sensing Image Registration via Spatial Relationship Analysis on SIFT Keypoints. In *Proceedings of IEEE International Geoscience & Remote Sensing Symposium*, Honolulu, HI, USA, 25–30 July 2010; pp. 1011–1014.
2. Brown, M.; Su, S. Multi-Spectral SIFT for Scene Category Recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA, 20–25 June 2011; pp. 177–184.
3. Abdelrahman, M.; Ali, A.; Aly, A.; Farag, A. Precise Change Detection In Multi-Spectral Remote Sensing Imagery Using SIFT-based Registration. In *Proceedings of IEEE International Conference on Multimedia Technology*, Hangzhou, China, 26–28 July 2011; pp. 6238–6242.
4. Lalonde, M.; Byrns, D.; Gagnon, L.; Teasdale, N.; Laurendeau, D. Real-Time Eye Blink Detection with GPU-Based SIFT Tracking. In *Proceedings of Fourth Canadian Conference on Computer and Robot Vision*, Montreal, QC, Canada, 28–30 May 2007; pp. 481–487.
5. Leykin, A.; Hammoud, R. Pedestrian tracking by fusion of thermal-visible surveillance videos. *Mach. Vis. Appl.* **2010**, *21*, 587–595.
6. Krotosky, S.; Trivedi, M. On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 619–629.
7. Jung, S.; Eledath, J.; Johansson, S.; Mathevon, V. Egomotion Estimation in Monocular Infra-Red Image Sequence for Night Vision Applications. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, Austin, TX, USA, 20–21 February 2007; p. 8.
8. Lowe, D.G. Local Feature View Clustering for 3D Object Recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, HI, USA, 8–14 December 2001; pp. 682–688.
9. Lazebnik, S.; Schmid, C.; Ponce, J. Semi-Local Affine Parts for Object Recognition. In *Proceedings of British Machine Vision Conference*, London, UK, 7–9 September 2004; pp. 779–788.
10. Bay, H.; Tuytelaars, T.; Gool, L.V. SURF: Speeded Up Robust Features. In *Proceedings of European Conference on Computer Vision*, Graz, Austria, 7–13 May 2006; pp. 404–417.
11. Yi, Z.; Zhiguo, C.; Yang, X. Multi-spectral remote image registration based on SIFT. *Electron. Lett.* **2008**, *44*, 107–108.

12. Istenic, R.; Heric, D.; Ribaric, S.; Zazula, D. Thermal and Visual Image Registration in Hough Parameter Space. In *Proceedings of the 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, Maribor, Slovenia, 27–30 June 2007; pp. 106–109.

13. Lee, J.; Kim, Y.; Lee, D.; Kang, D.; Ra, J. Robust CCD and IR image registration using gradient-based statistical information. *IEEE Signal Process. Lett.* **2010**, *17*, 347–350.

14. Zhang, L.; Wu, B.; Nevatia, R. Pedestrian Detection in Infrared Images Based on Local Shape Features. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8.

15. Trivedi, M.; Cheng, S.; Childers, E.; Krotosky, S. Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation. *IEEE Trans. Veh. Technol.* **2004**, *53*, 1698–1712.

16. Vural, M.; Yardimci, Y.; Temizel, A. Registration of Multispectral Satellite Images with Orientation-Restricted SIFT. In *Proceedings of IEEE International Symposium on Geoscience and Remote Sensing*, Cape Town, South Africa, 12–17 July 2009; pp. 243–246.

17. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *8*, 679–698.

18. Manjunath, B.; Ohm, J.; Vasudevan, V.; Yamada, A. Color and texture descriptors. *IEEE Trans. Circuits Syst. Video Technol.* **2001**, *11*, 703–715.

19. Sikora, T. The MPEG-7 visual standard for content description—an overview. *IEEE Trans. Circuits Syst. Video Technol.* **2001**, *11*, 696–702.