Performance Evaluation of Feature Point Descriptors in the Infrared Domain

Pablo Ricaurte[1], Carmen Chilán[1], Cristhian A. Aguilera-Carrasco [2], Boris X. Vintimilla[1] and Angel D. Sappa[1,2]

[1]*CIDIS-FIEC, Escuela Superior Politécnica del Litoral (ESPOL), Campus Gustavo Galindo, Km 30.5 vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*
[2]*Computer Vision Center, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

Keywords: Infrared Imaging, Feature Point Descriptors.

Abstract: This paper presents a comparative evaluation of classical feature point descriptors when they are used in the long-wave infrared spectral band. Robustness to changes in rotation, scaling, blur, and additive noise are evaluated using a state of the art framework. Statistical results using an outdoor image data set are presented together with a discussion about the differences with respect to the results obtained when images from the visible spectrum are considered.

1 INTRODUCTION

In general, computer vision applications are based on the use of cameras that work in the visible spectrum. Recent advances in infrared imaging, as well as the reduction on the prices of these cameras, have opened new opportunities to develop novel solutions working in infrared spectral band or in the cross-spectral domain between infrared and visible images (e.g., (Barrera et al., 2013), (Aguilera et al., 2012), (Barrera et al., 2012) and (Felicísimo and Cuartero, 2006)).

The spectral band of infrared imaging goes from $0.75\mu m$ to 15 μm, which is split up into the following categories: Near-Infrared (NIR: 0.751.4 μm), Short-Wave Infrared (SWIR: 1.43 μm), Mid-Wave Infrared (MWIR: 38 μm) or Long-Wave Infrared (LWIR: 815 μm). Images from each one of these categories have a particular advantage for a given application; for instance NIR images are generally used in gaze detection and eye tracking applications (Coyle et al., 2004); SWIR spectral band has shown its usage in heavy fog environments (Hansen and Malchow, 2008); MWIR is generally used to detect temperatures somehow above body temperature in military applications; finally, LWIR images have been used in video surveillance and driver assistance (Krotosky and Trivedi, 2007). The current work is focussed on the LWIR domain, which corresponds to the farthest infrared spectral band from the visible spectrum.

Following the evolution of visible spectrum based computer vision, in the infrared imaging domain topics such as image registration, pattern recognition or stereo vision, are being addressed. As a first attempt, classical tools from the visible spectrum are just used or little adapted to the new domain. One of these tools is the feature point description, which has been a very active research topic during the last decade in the computer vision community. Due to the large amount of contributions on this topic there were several works on the literature evaluating and comparing their performance in the visible spectrum case (e.g., (Miksik and Mikolajczyk, 2012), (Mikolajczyk and Schmid, 2005), (Bauer et al., 2007) and (Schmid et al., 2000)).

Similarly to in the visible spectrum case, the current work proposes to study the performance of feature point descriptors when they are considered in the infrared domain. Since there is a large amount of algorithms in the literature, we decided to select the most representative and recent ones. Our study includes: SIFT (Lowe, 1999), SURF (Bay et al., 2006), ORB (Rublee et al., 2011), BRISK (Leutenegger et al., 2011), BRIEF (Calonder et al., 2012) and FREAK (Alahi et al., 2012). The study is motivated by the fact that although the appearance of LWIR images is similar to the ones from the visible spectrum their nature is different, hence we consider that conclusions from visible spectrum cannot be directly extended to the infrared domain.

Ricaurte P., Chilán C., Aguilera-Carrasco C., Vintimilla B. and Sappa A..
Performance Evaluation of Feature Point Descriptors in the Infrared Domain.
DOI: 10.5220/0004725305450550
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 545-550
ISBN: 978-989-758-003-1

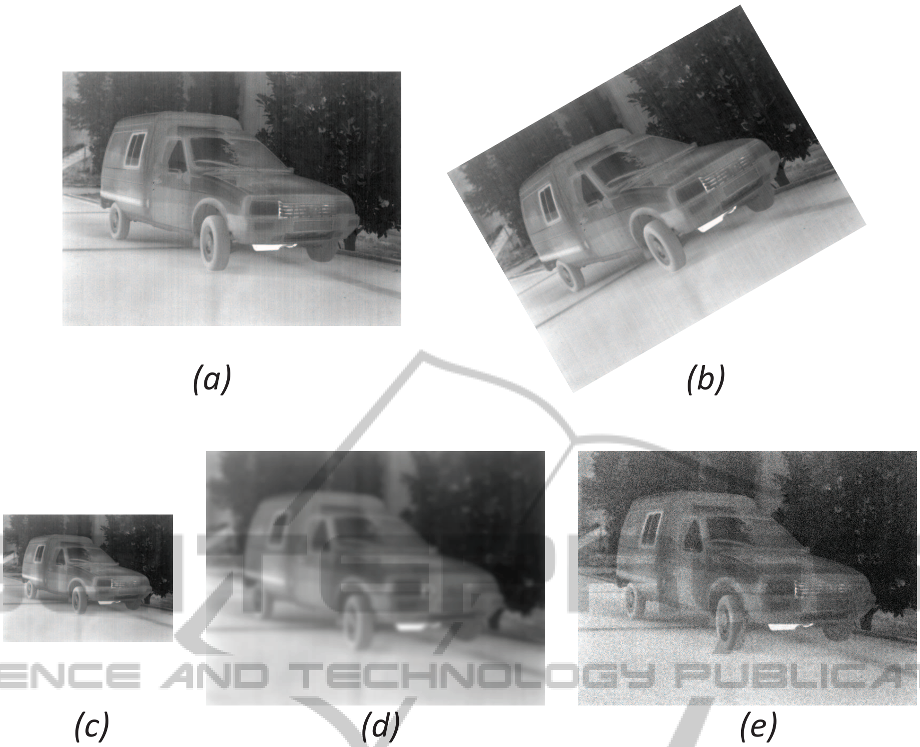

Figure 1: A LWIR image from the evaluation dataset together with some illustrations of the applied transformations: (a) original image; (b) rotation; (c) scale; (d) blur; (e) noise.

The manuscript is organized as follow. The evaluation methodology used for the comparison is presented in Section 2. Then, the data set and experimental results are detailed in Section 3. Finally, conclusions and discussions are given in Section 4.

2 EVALUATION FRAMEWORK

This section summarizes the framework used to evaluate the performance of the different approaches. It aims at finding the best descriptor for feature point correspondence when common image transformations are considered: rotation in the image plane, changes in the image size, blur and presence of noise in the images. In order to take into account all these possible changes, the given images are modified, then different descriptors are applied. The used framework has been proposed by Khvedchenia[1] for evaluating the performance of feature descriptors in the visible spectrum case. The algorithms are evaluated considering as a ground truth those points in the given image. A brute force strategy is used for finding the matchings, together with a L2 norm or Hamming distance, as detailed in Table 1. The percentage of correct matches between the ground truth image and the modified one is used as a criterion for the evaluation. The transformations applied to the given images are detailed below. Figure 1 shows an illustration of a given LWIR image together with some of the images resulting after applying the different transformations.

- **Rotation**: the study consists in evaluating the sensibility to rotations of the image. The rotations are in the image plane spanning the 360 degrees, a new image is obtained every 10 degrees.
- **Scale**: the size of the given image is changed and the repeatability of a given descriptor is evaluated. The original image is scaled in between 0.2 to 2 times its size with a step of 0.1 per test. Pixels of scaled images are obtained through a linear interpolation.
- **Blur**: the robustness with respect to blur is evaluated. It consists of a Gaussian filter iteratively applied over the given image. At each iteration the size of the kernel filter ($K \times K$) used to blur the image is update as follows: $K = 2t + 1$, where $t = \{1, 2, ..., 9\}$.
- **Noise**: this final study consists in adding noise to the original image. This process is implemented by adding to the original image a personalized

[1] http://computer-vision-talks.com/2011/08/feature-descriptor-comparison-report/

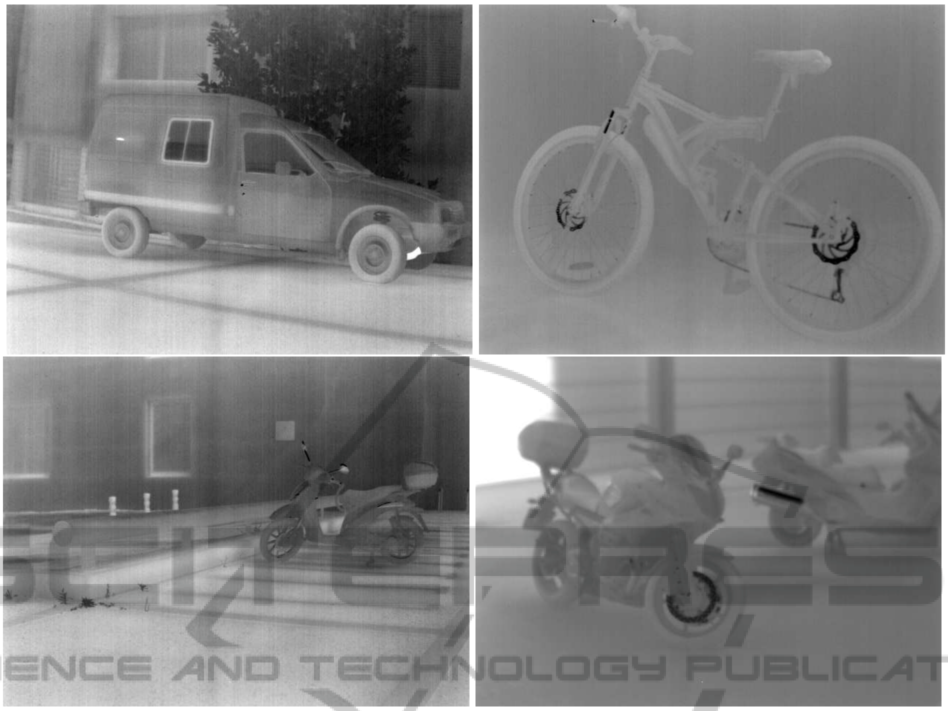

Figure 2: Illustration of infrared images considered in the evaluation framework.

image. The value of the pixels of the personalized image are randomly obtained following a uniform distribution with $\mu = 0$ and $\sigma = t$, where $t = \{0, 10, 20, ..., 100\}$.

In the original framework proposed by Khvedchenia, lighting changes were also considered, since that study was intended for images in the visible spectrum. In the current work, since it aims at studying the LWIR spectrum, changes in the image intensity values won't follow the same behavior all through the image (like lighting changes in the visible spectrum). Intensity values in LWIR images are related with the material of the objects in the scene. In summary, a study similar to the lighting changes is not considered in the current work.

3 EXPERIMENTAL RESULTS

A set of 20 LWIR images has been considered with the evaluation framework presented above. For each one of the LWIR images the corresponding image from the visible spectrum is also provided. These visible spectrum images are used to compare the results obtained in the LWIR domain. Figure 2 presents some of the LWIR images contained in our dataset; it is publicly available for further research through our Web site[2]. For each algorithm and transformation the number of correct matches out of the total number of feature points described in the original image (the number of correspondences used as ground truth) is considered for the evaluation, similar to (Mikolajczyk and Schmid, 2005):

Table 1: Algorithms evaluated in the study.

Descriptor Alg.	**Matcher norm type**
SIFT	L2 Norm
SURF	L2 Norm
ORB	Hamming Distance
BRISK	Hamming Distance
BRIEF (SURF detector)	Hamming Distance
FREAK (SURF detector)	Hamming Distance

$$recall = \frac{\#correct \quad matches}{\#correspondences}. \qquad (1)$$

The algorithms evaluated in the current work are presented in Table 1. In the cases of BRIEF and FREAK the SURF algorithm is used as a detector. In ORB, BRISK, BRIEF and FREAK the Hamming distance is used, instead of L2 norm, for speeding up the matching. For each transformation (Section 2) a set of images is obtained; for instance, in the rotation case 36 images are evaluated.

[2]http://www.cvc.uab.es/adas/projects/simeve/

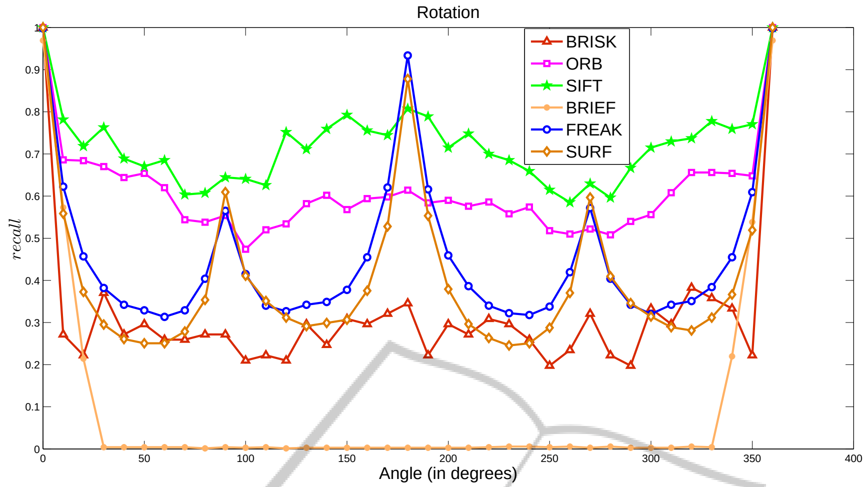

Figure 3: Rotation case study: average results from the evaluation data set.

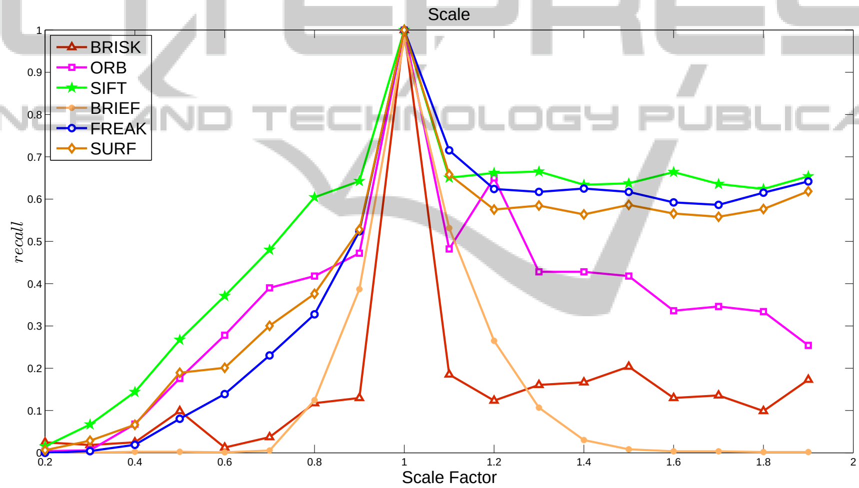

Figure 4: Scale case study: average results from the evaluation data set.

Figure 3 depicts results obtained when the given image is rotated 360 degrees. It can be observed that the most robust algorithm is SIFT, which is the same conclusion obtained when images from the visible spectrum are considered. On the other hand, the BRIEF algorithm (using SURF as a detector) is the most sensible to rotations; actually, its performance drop to zero just after applying ± 25 degrees to the given image. This behavior was also observed in the visible spectrum. Regarding the algorithms in between, a slightly better performance was appreciated in the LWIR case. In summary, the ranking from the best to the worst is as follow: SIFT, ORB, FREAK, SURF, BRISK, BRIEF (the same ranking was observed in both spectrums).

In the scale study, although similar results were obtained in both spectrums, the algorithms' performance was better in the infrared case, in particular in the case of BRISK, which being the worst algorithm in both cases it has a better performance in the LWIR case. The ranking of algorithms' performance is as follow (from the best to the worst): SIFT, FREAK, SURF, ORB, BRIEF and BRISK. Figure 4 shows these results.

Figure 5 presents the study of robustness of the different algorithms when the given images are degraded using a Gaussian filter of increasing size. Similarly to in the previous case all the algorithms have a better performance in the LWIR case than in the visible spectrum. In this case the BRIEF algorithm is the most robust one, the other algorithms are sorted as follow: FREAK, SURF, ORB, SIFT and BRISK; being BRISK the algorithm less robust to noise.

Finally, Fig 6 shows the curves obtained when additive noise is considered. In this case, differently than in the previous studies, the algorithm has a bet-

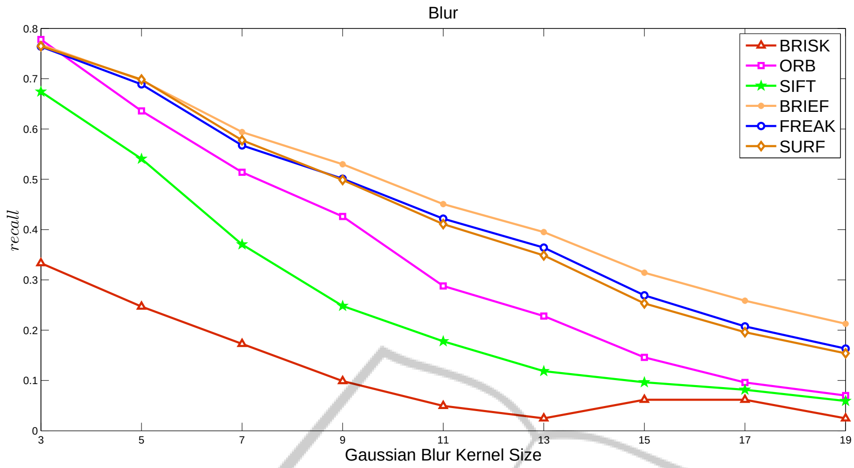

Figure 5: Blur case study: average results from the evaluation data set.

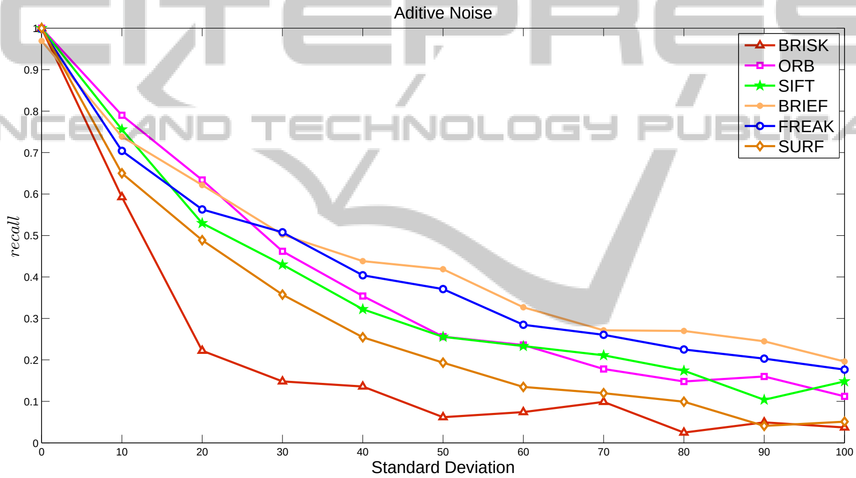

Figure 6: Noise case study: average results from the evaluation data set.

ter performance in the visible spectrum. Additionally, the performance of SIFT in the LWIR spectrum is not as bad as in the visible spectrum. The ranking of algorithms' performance is as follow (from the best to the worst): BRIEF, FREAK, ORB, SIFT, SURF and BRISK.

4 CONCLUSIONS

This work presents an empirical evaluation of the performance of the state of the art descriptors when they are used in the LWIR domain. The main objective was to study whether conclusions obtained in the visible spectrum are also valid for the LWIR spectrum. Although results are similar to those obtained in the visible spectrum it can be appreciated that the performance of the algorithm BRIEF (using SURF as a detector) is better in LWIR spectrum when compared with its performance in the visible spectrum. The relative ranking in between algorithms keep the same in both spectral bands.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Government under Research Project TIN2011-25606 and PROMETEO Project of the "Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación de la República del Ecuador". Cristhian A. Aguilera-Carrasco was supported by a grant from "Universitat Autònoma de Barcelona". The authors would like to thanks to Mr. Khvedchenia Ievgen for providing them with the evaluation framework.

REFERENCES

Aguilera, C., Barrera, F., Lumbreras, F., Sappa, A., and Toledo, R. (2012). Multispectral image feature points. *Sensors*, 12(9):12661–12672.

Alahi, A., Ortiz, R., and Vandergheynst, P. (2012). FREAK: Fast retina keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21*, pages 510–517.

Barrera, F., Lumbreras, F., and Sappa, A. (2012). Multimodal stereo vision system: 3d data extraction and algorithm evaluation. *IEEE Journal of Selected Topics in Signal Processing*, 6(5):437–446.

Barrera, F., Lumbreras, F., and Sappa, A. (2013). Multispectral piecewise planar stereo using manhattan-world assumption. *Pattern Recognition Letters*, 34(1):52–61.

Bauer, J., Snderhauf, N., and Protzel, P. (2007). Comparing several implementations of two recently published feature detectors. In *Proceedings of the International Conference on Intelligent and Autonomous Systems, Toulouse, France*.

Bay, H., Tuytelaars, T., and Gool, L. J. V. (2006). SURF: Speeded Up Robust Features. In *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, May 7-13*, pages 404–417.

Calonder, M., Lepetit, V., Özuysal, M., Trzcinski, T., Strecha, C., and Fua, P. (2012). BRIEF: Computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(7):1281–1298.

Coyle, S., Ward, T., Markham, C., and McDarby, G. (2004). On the suitability of near-infrared (NIR) systems for next-generation braincomputer interfaces. *Physiological Measurement*, 25(4).

Felicísimo, A. and Cuartero, A. (2006). Methodological proposal for multispectral stereo matching. *IEEE Trans. on Geoscience and Remote Sensing*, 44(9):2534–2538.

Hansen, M. P. and Malchow, D. S. (2008). Overview of swir detectors, cameras, and applications. In *Proceedings of the SPIE 6939, Thermosense, Orlando, FL, USA, March 16*.

Krotosky, S. and Trivedi, M. (2007). On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 8:619–629.

Leutenegger, S., Chli, M., and Siegwart, R. (2011). BRISK: Binary Robust Invariant Scalable Keypoints. pages 2548–2555.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the IEEE International Conference on Computer Vision, Kerkyra, Greece, September 20-27*, pages 1150–1157.

Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630.

Miksik, O. and Mikolajczyk, K. (2012). Evaluation of local detectors and descriptors for fast feature matching. In *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012, Tsukuba, Japan, November 11-15*, pages 2681–2684.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. R. (2011). ORB: An efficient alternative to SIFT or SURF. In *IEEE International Conference on Computer Vision, Barcelona, Spain, November 6-13*, pages 2564–2571.

Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.