LGHD: A FEATURE DESCRIPTOR FOR MATCHING ACROSS NON-LINEAR INTENSITY VARIATIONS

Cristhian A. Aguilera[*,†], *Angel D. Sappa*[*,‡], *Ricardo Toledo*[*,†]

[*]Computer Vision Center, Campus UAB, 08193, Bellaterra, Spain
[‡]Escuela Superior Politécnica del Litoral, Campus Gustavo Galindo, Guayaquil, Ecuador
[†]Computer Science Dpt., Universitat Autònoma de Barcelona, Campus UAB, Bellaterra, Spain
{caguilera, asappa, ricard}@cvc.uab.es

ABSTRACT

This paper presents a new feature descriptor suitable to the task of matching features points between images with non-linear intensity variations. This includes image pairs with significant illuminations changes, multi-modal image pairs and multi-spectral image pairs. The proposed method describes the neighbourhood of feature points combining frequency and spatial information using multi-scale and multi-oriented Log-Gabor filters. Experimental results show the validity of the proposed approach and also the improvements with respect to the state of the art.

Index Terms— Feature descriptor, multi-modal, multi-spectral, NIR, LWIR.

1. INTRODUCTION

Matching points between images is an important step in computer vision applications such as image registration, object recognition and 3D reconstruction, just to mention a few. In general, points to be matched are described by means of their surroundings providing a rich new representation. The goal of this process is to describe key points as distinctive as possible from other similar regions. Ideally, a description should be robust to different image transformations. In this paper, we focus on designing a new feature descriptor that is robust against non-linear intensity variation between image pairs that includes images from different modalities and different spectra.

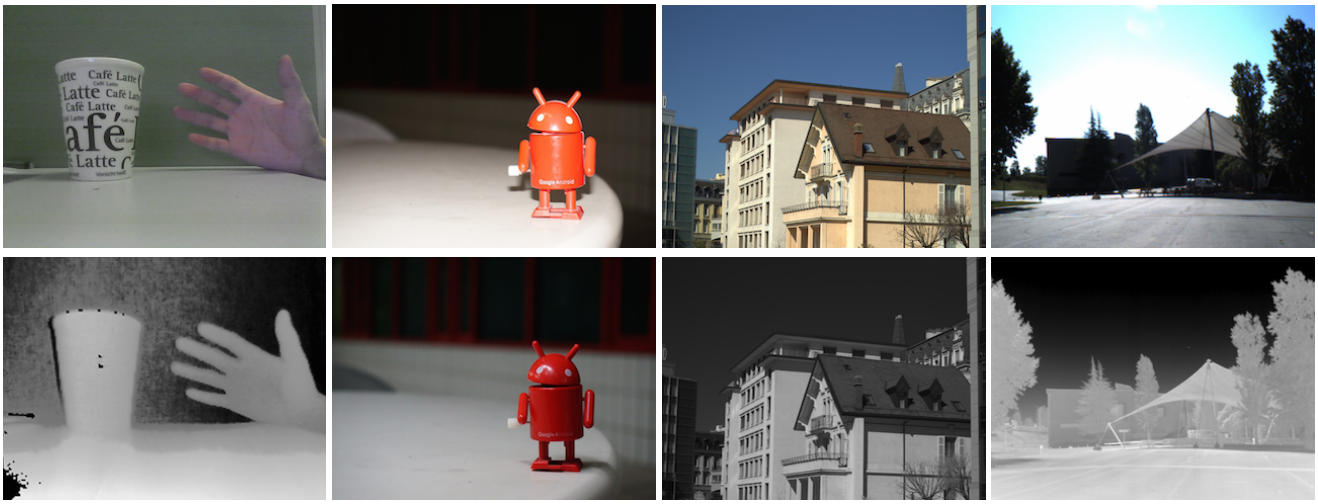
Recent advances in technology have opened new opportunities to develop novel solutions to tackle in a more efficient way classical computer vision problems. This includes working with images captured from different sensors, which result in devices that rely on multi-spectral/multi-modal technologies. One of the most widely used is the Kinect 3D motion sensing device that is based on the usage of a RGB and a near-infrared (NIR) cameras. The HeatWave system [1] is another example of such multi-spectral devices; it makes use of a RGB, a NIR and a long-wave-infrared (LWIR) cameras to create 3D thermal images of buildings. Figure 1 presents illustrations of image pairs from the same scenario acquired by different sensors or camera setup: (a) RGB and depth images; (b) RGB and NIR images; (c) flash and no-flash images; and (d) RGB and LWIR images—data sets from [2], [3] and [4].

Although working with images from different modalities helps to device novel solutions, new challenging and difficult problems need to be tackled since they cannot be faced up with the state of the art. In the case of feature descriptor, algorithms such as [5], or some of its variations, which were build to work with images rich in texture are not good enough in the multi-modal or multi-spectral domains since: *i)* color may change between images (see Figure 1(b)); *ii)* texture may be lost (see Figure 1 (a) and (b)); and *iii)* the direction of the intensity gradients may also change (see Figure 1(d)), just to mention some of the common problems.

Recently, some contributions have been proposed to work in the multi-spectral/multi-modal domains (e.g., [6, 4, 7]), some of them are briefly presented in Section 2. Unfortunately, their performance is far away from the one obtained when images from the same sensor and setup are considered. The goal of the current work is to investigate feature description and matching for image pairs where non-linear intensity variations appear. For that purpose we propose a new feature descriptor that uses multiple oriented Log-Gabor filters at different scales to describe image patches in a distinctive way. The proposed approach is evaluated and compared with state-of-art descriptors in four different scenarios.

The rest of the manuscript is organized as follows. Section 2 presents related works. Then, the proposed descriptor is introduced in Section 3. The evaluation methodology together with the system set up are presented in Section 4. Experimental results, including comparisons with four state-of-art descriptors using four different set of images, are presented in Section 5. Finally, conclusions are given in Section 6.

The code and set of images used in this article are available through the personal webpages of the authors.



(a) RGB/DEPTH (b) FLASH/NO-FLASH (c) RGB/NIR (d) RGB/LWIR

Fig. 1. Examples of pair of images from the four data sets evaluated in the current work. **This figure is best viewed in color**.

2. RELATED WORK

Finding point correspondences across non-linear intensity variations is a challenging task. The usage of classical descriptors between image pairs from different modalities or spectra tends to give very poor results. For instance, Cronje et al. [8] compare the usage of SIFT [5], SURF [9] and BRIEF [10] descriptors to register RGB/LWIR image pairs without achieving good results and in general with a low percentage of inlier correspondences. Nevertheless, recently some methods have been proposed in the literature that can be grouped into two categories, dense and local matching.

Ce Liu et al. [6] propose a dense correspondence algorithm based on the SIFT descriptor that works on different multi-spectral and multi-modal image pairs. The solution consists in matching dense SIFT representations of features using an objetive function similar to the one used to compute optical flow. More recently, Shen et al. [4] propose a different methodology for multi-modal and multi-spectral registration of natural images using a variational approach. The method consists in two phases, firstly a global matching estimates large position transform and then a local matching estimates residual errors.

On the contrary to previous approaches, the authors in [11] employ a local patch similarity function to find correspondences between images from the RGB and LWIR domain. The solution consists of the combination of a tuned version of the DoG detector [5] and a local EHD descriptor [12]. Although interesting results are obtained, the main problem with this approach lies in the reduced number of matches. A different approach is presented [13] where the authors choose the phase congruency (PC) model [14] as feature detector and a combination of frequency and spatial information as feature descriptor. This approach is similar to the one presented in [11] but including 24 bins of Log-Gabor components over the central pixel. This fact improves the results from [11] by increasing the number of matches. However, the total number of correct correspondences is still low in comparison to other cross-domain cases such as the RGB/NIR case [15]. Also a simpler version of the SIFT description has been proposed in [15] in order to introduce invariance to the gradient direction. Instead of computing gradient directions between $[0, 2\pi)$, the descriptor computes gradient directions between $[0, \pi)$.

3. PROPOSED DESCRIPTOR

The non-linear intensity variations between a pair of images can be the result of different configuration setups, which can affect the images in a different way. However, in spite of these intensity differences, the global appearance of the shape of the objects contained in the scene tends to remain constant. This fact makes us to think that a descriptor based on the distribution of high frequency components would be robust to different non-linear intensity variations, which is the idea behind the proposed approach.

The current work is based on the local EHD descriptor presented by Aguilera et al. [11]. The EHD descriptor describes the spatial edge distribution around a point computing an orientation histogram of 80 bins. For each interest point a region of SxS is defined and further divided into 16 smaller subregions (4x4). Within each subregion, an orientation histogram of 5 bins is computed using the strongest pixel value for one of 5 different oriented Sobel filters (horizontal, vertical, 35 degrees, 135 degrees and non-oriented).

The Log-Gabor Histogram Descriptor (LGHD), which is the main contribution of the current work, describes local

patches in a similar way to EHD but instead of using multi-oriented Sobel descriptor it uses multi-oriented and multi-scale Log-Gabor filters. Log-Gabor filters are the keystone of several computer vision algorithms (e.g., [13, 14]); they can be constructed with any arbitrary bandwidth and, by definition, they do not have a DC component.

The proposed descriptor is obtained as follows:

- Convolve the image with a Log-Gabor filter bank as in [14]. The bank is composed by 24 different filters (6 orientations between $[0, \pi)$ and 4 scales).
- Select a region of SxS centered around a point of interest. The resulting region is divided into 16 smaller subregions (4×4).
- Build a histogram of oriented Log-Gabor filters in each subregion at each scale. Every pixel of each subregion contributes to a bin on the histogram according to the orientation of the filters. We use the magnitude of the filter response to determine the dominant filter.
- Join the histogram of each subregion at the four scales (96×4) and combine the 4 resulting histograms to obtain a 384 bin feature vector.

As noted in [11] the selection of the right region size (S) is a key factor in the performance of the descriptor.

4. EVALUATION METHODOLOGY

The proposed approach has been evaluated using four different set of image pairs: 58 RGB/NIR pairs taken in an urban environment from [2]; 44 RGB/LWIR outdoor pairs specially acquired for the current work; 120 FLASH/NO-FLASH pairs of images from [3] and 4 RGB/DEPTH pairs from [4].

In addition to the evaluation mentioned above, the proposed approach has been compared with four state-of-art descriptors: *1)* the EHD descriptor that was originally proposed for the RGB/LWIR case [11]; *2)* the gradient invariant version of SIFT (GISIFT) [15]; *3)* the PCEHD descriptor [13]; and *4)* the SIFT descriptor [5] that is used as a reference of classical descriptors.

In order to evaluate the performance of feature descriptors, avoiding bias due to feature detector performance, we follow a similar approach to [10]. We detect features just in one image using the FAST detector [16], and then we project them into the corresponding pair using the homography information. This process is done for 3 sets of the image pair; for the remaining one (RGB/DEPTH) we use 100 points manually selected (provided by [4]), since the images cannot be represented by a unique homography.

The performance of the different descriptors is evaluated using the resulting matching precision:

$$Precision = \frac{C}{T}, \quad (1)$$

where C is the number of correct matches and T is the total number of correspondences.

In our experiments we convolve the different images with Log-Gabor banks using the Matlab implementation of [14]. We set *nscale=4*, *norient=6*, *minWaveLength=3*, *mult=1.6* and *sigmaOnf=0.75*. Additionally, Table 1 shows the different patch sizes used to evaluate the EHD, PCEHD and LGHD descriptors (these sizes were empirically obtained in order to have a fair evaluation). The matches are found by using Euclidean distance (SSD).

Descriptor	RGB/DEPTH	Other cases
EHD	32×32	80×80
PCEHD	32×32	80×80
LGHD (Ours)	32×32	80×80

Table 1. Patch sizes used to evaluate the EHD, PCEHD and LGHD descriptors.

5. EXPERIMENTAL RESULTS

Results are shown in Table 2, where each cell of the table indicates the average matching precision for the corresponding descriptor computed over the whole data set. The proposed approach, LGHD, obtained the best performance in every category when compared with all the other descriptors evaluated in the current work. Regarding computational times, the proposed LGHD descriptor has a similar performance to other approaches with respect to the feature description estimation, but its matching cost is the most expensive one due to the size of the description vector (384 elements).

The matching precision for the FLASH/NO-FLASH and the RGB/NIR cases was considerably higher than in the other two scenarios. This fact is mainly due to the spectral band closeness of the image pairs: *i)* the NIR spectrum is the closest infrared band to the visible spectrum; *ii)* while in the FLASH/NO-FLASH dataset, the pairs of images correspond both to the same spectral band (the visible one). On the other hand, lower precision rates were obtained for the RGB/DEPH and RGB/LWIR cases. The LWIR band is the most distant infrared band from the visible spectrum. Image pairs mostly share shape information, while most of the texture information is missed. The depth case is even worse since all texture information is missed; in this case just a limited number of visual similarities between visible and depth images is kept.

6. CONCLUSIONS

We have presented a new feature descriptor that can be used to the task of matching features between images with non-linear intensity variations such as multi-spectral and multi-modal images. Results show that the proposed algorithm out-

Descriptor	RGB/DEPTH	RGB/LWIR	FLASH/NO-FLASH	RGB/NIR
SIFT	0.19	0.08	0.76	0.77
GISIFT	0.26	0.19	0.74	0.74
EHD	0.13	0.13	0.66	0.77
PCEHD	0.04	0.08	0.69	0.74
LGHD (Ours)	**0.30**	**0.24**	**0.81**	**0.85**

Table 2. Average precision values for the methods evaluated on the different set of images.

performs state-of-art algorithms in the four data sets considered in the evaluation. The RGB/LWIR and the RGB/Depth were the most challenging cases. Results show that in both cases matching descriptors generate an elevated number of mismatches. These mismatches can be reduced using a robust formulation such as RANSAC.

7. ACKNOWLEDGMENT

This work has been partially supported by the Spanish Government under Project TIN2014-56919-C3-2-R and PROMETEO Project of the "Secretaría Nacional de Educació Superior, Ciencia, Tecnología e Innovación de la República del Ecuador". Cristhian A. Aguilera was supported by a grant from Universitat Autònoma de Barcelona.

8. REFERENCES

[1] P. Moghadam and S. Vidas, "Heatwave: the next generation of thermography devices," in *SPIE Sensing Technology+ Applications*, Baltimore, USA, May 2014, pp. 91050F–91050F.

[2] M. Brown and S. Susstrunk, "Multi-spectral sift for scene category recognition," in *CVPR*, Colorado Springs, USA, Jun 2011, pp. 177–184.

[3] H. Shengfeng and L. Rynson, "Saliency detection with flash and no-flash image pairs," in *ECCV*, Zurich, Switzerland, Sep 2014, pp. 110–124.

[4] X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and Multi-spectral Registration for Natural Images," in *ECCV*, Zurich, Switzerland, Sep 2014, pp. 309–324.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[6] Ce Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *PAMI*, vol. 33, no. 5, pp. 978–994, May 2011.

[7] F. Barrera, F. Lumbreras, and A. Sappa, "Multispectral piecewise planar stereo using manhattan-world assumption," *PRL*, vol. 34, no. 1, pp. 52–61, Jan. 2013.

[8] J. Cronje and J. de Villiers, "A comparison of image features for registering lwir and visual images," in *PRASA*, Pretoria, South Africa, Nov 2012.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *CVIU*, vol. 110, no. 3, pp. 346 – 359, 2008.

[10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *ECCV*, Heraklion, Greece, Sep 2010, pp. 778–792.

[11] C. Aguilera, F. Barrera, F. Lumbreras, A. Sappa, and R. Toledo, "Multispectral image feature points," *Sensors*, vol. 12, no. 9, pp. 12661–72, Jan. 2012.

[12] B.S. Manjunath, J.-R. Ohm, V.V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 703–715, Jun 2001.

[13] T. Mouats, N. Aouf, A.D. Sappa, C. Aguilera, and R. Toledo, "Multispectral stereo odometry," *ITS*, vol. PP, no. 99, pp. 1–15, Sep 2014.

[14] P. Kovesi, "Phase congruency detects corners and edges," in *DICTA*, Sydney, Dec. 2003, pp. 309–318.

[15] D. Firmenichy, M. Brown, and S. Süsstrunk, "Multispectral interest points for RGB-NIR image registration," in *ICIP*, Brussels, Belgium, Sept. 2011, pp. 181–184.

[16] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *PAMI*, vol. 32, no. 1, pp. 105–119, Jan 2010.