A Novel Image Descriptor Based on Anisotropic Filtering

Darshan Venkatrayappa, Philippe Montesinos, Daniel Dd Diep, Baptiste Magnier

A novel image descriptor based on anisotropic filtering

Darshan Venkatrayappa, Philippe Montesinos, Daniel Diep, and Baptiste Magnier

LGI2P - Ecole des Mines d'Ales, Nimes, France
{darshan.venkatrayappa,philippe.montesinos,daniel.diep,baptiste.magnier}@mines-ales.fr

Abstract. *In this paper, we present a new image patch descriptor for object detection and image matching. The descriptor is based on the standard HoG pipeline. The descriptor is generated in a novel way, by embedding the response of an oriented anisotropic derivative half Gaussian kernel in the Histogram of Orientation Gradient (HoG) framework. By doing so, we are able to bin more curvature information. As a result, our descriptor performs better than the state of art descriptors such as SIFT, GLOH and DAISY. In addition to this, we repeat the same procedure by replacing the anisotropic derivative half Gaussian kernel with a computationally less complex anisotropic derivative half exponential kernel and achieve similar results. The proposed image descriptors using both the kernels are very robust and shows promising results for variations in brightness, scale, rotation, view point, blur and compression. We have extensively evaluated the effectiveness of the devised method with various challenging image pairs acquired under varying circumstances.*

Keywords: Anisotropic half derivative Gaussian kernel, Anisotropic half derivative exponential kernel, image descriptor, HoG, image matching.

1 Introduction

Currently, most of the promising method for image content description use local features as their basis. Thus, image feature extraction has become an important and active research topic in the field of computer vision. Local image features form the foundation for many real time applications such as object detection and tracking, panorama stitching, video surveillance and image based retrieval. In addition to feature description these applications require invariance to rotation, scale, viewpoint and illumination changes. Many image description algorithms have been proposed to incorporate all these characteristics. To name a few: SIFT [1], GLOH [2], SURF [3], LIOP [4], ASIFT [5], DAISY [6] and many more. Most of these descriptors are based on Histogram of Oriented Gradients (HoG) [7]. SIFT, SURF and GLOH are robust to rotation and scale invariance characteristics. ASIFT improves on these descriptors by exhibiting good affine

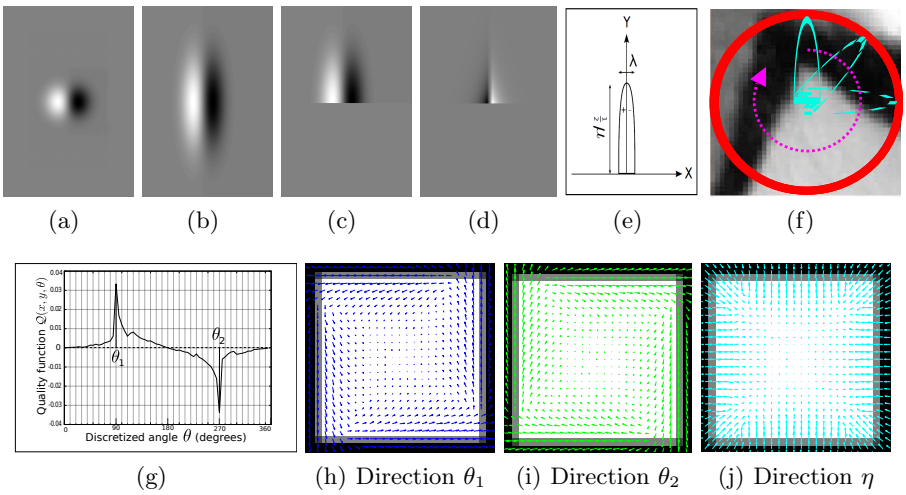
invariance properties. LIOP has succeeded in handling both the geometric distortions and monotonic changes in illumination. Some of these descriptors and their variants are compared and reviewed in detail in [2].

Another category of feature descriptors use filter response as their basis. Schmid and Mohr [8] use differential invariant responses to compute new local image descriptors. Differential invariant responses are obtained from a combination of Gaussian derivatives of different orders, which are invariant to 2-dimensional rigid transformations. Koenderink et al. [9] formulate a methodology based on Gaussian function and its derivatives to capture the local geometry of the image. Larsen et al. [10] follow a new approach for the construction of an image descriptor based on local k-jet, which uses filter bank responses for feature description. Palomares et al. [11] have come up with a local image descriptor issued from a filtering stage made of oriented anisotropic half-Gaussian smoothing convolution kernels. They achieve euclidean invariance in the matching stage. However, most of the descriptors based on filter response vaguely capture the geometry of the region around key-points, as a result they fail to compete with the HoG based descriptors.

Some researchers capture the geometry of the image patch using curvature information, and various edge related information. Monroy et al. [12] use descriptor based on curvature histograms for object detection. This descriptor gives good results for images with strong line segments. Authors in [13] use location, orientation and length of the edges to construct an image patch descriptor. Here, the descriptor encodes the presence or absence of edges using a binary value for a range of possible edge positions and orientations. Eigenstetter et al. [14] have proposed an object representation framework based on curvature self-similarity. This method goes beyond the popular approximation of objects using straight lines. However, like all descriptors using second order statistics, this approach also exhibits a very high dimensionality.

In our method rotation and affine normalized gray level image patch is extracted as in [19]. Then, we construct our descriptor on this image patch. Thus obtained descriptor is completely different than the above mentioned descriptors. It can be seen as a combination of all the above three approaches. In the remaining of this paper for simplicity we refer to anisotropic half Gaussian derivative kernel as AHGDK and anisotropic half exponential derivative kernel as AHEDK. In our descriptor :

1. The response of a rotating AHGDK or AHEDK around pixels is used to construct signatures.
2. The orientation of the edges and the anisotropic gradient directions are extracted from the signatures, thus capturing the geometry/curvature of the image patch.
3. These orientations are binned separately in different ways as in HoG to form different descriptors.
4. Additionally, we construct a faster variant of our descriptor by replacing the AHGDK with a computationally inexpensive AHEDK.



(a) (b) (c) (d) (e) (f)

(g) (h) Direction θ_1 (i) Direction θ_2 (j) Direction η

Fig. 1. (a) Isotropic gaussian kernel (b) Anisotropic gaussian kernel (c) AHGDK (d) AHEDK with $\alpha_\mu = 1$ and $\alpha_\lambda = .2$ (e) A thin Gaussian derivative half-filter. (f) AHGDK applied to a keypoint (x_p, y_p) in the image patch. (g) Extrema of a function $\mathcal{Q}(x_p, y_p, \theta)$. and $\Delta\theta = 2^o$. Note that the initial orientation of the filter is vertical, upwardly directed and steerable clockwise. ((h),(i),(j)) Synthetic square illustrating the image geometry captured by AHGDK. All the figures in this paper are best when viewed in color.

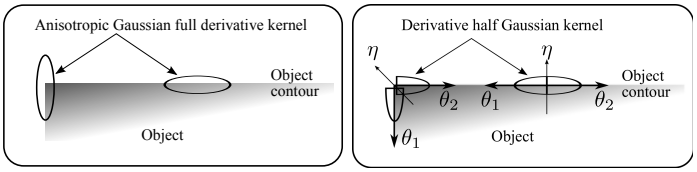
2 FILTERING STAGE

2.1 Anisotropic half Gaussian derivative kernel (AHGDK)

When compared to isotropic filters (Fig.1(a)), anisotropic filters (Fig.1(b)) have an advantage in detecting large linear structures. For anisotropic filters, at the corners the gradient magnitude decreases as the edge information under the scope of the filter decreases (Fig.2(a)). Consequently, robustness to noise, when dealing with tiny geometric structures weakens. This weakness can be nullified by using AHGDK [16]. In addition to this, thanks to the elongated and oriented half kernels (Fig.1(c), Fig.1(e)), we are able to estimate the two edge directions as in Fig.2(b). As illustrated in Fig.1(f), at the pixel coordinate (x, y), a derivative kernel is applied to obtain a derivative information $\mathcal{Q}(x, y, \theta)$ in a function of orientation $\theta \in [0; 360°[$:

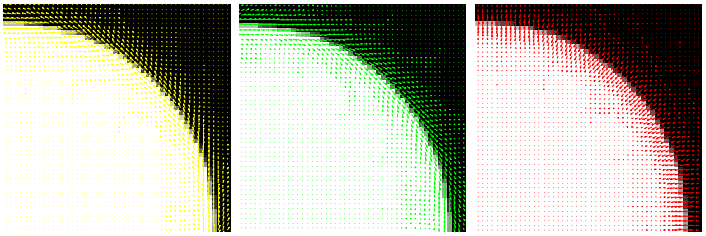
$$\mathcal{Q}(x, y, \theta) = I_\theta * C \cdot H(-y) \cdot x \cdot e^{-\left(\frac{x^2}{2\lambda^2} + \frac{y^2}{2\mu^2}\right)}, \tag{1}$$

where I_θ corresponds to a rotated image of orientation θ, C is a normalization coefficient [17] and (μ, λ) represent the standard deviation of AHGDK. Only the causal part of this filter along the Y axis is used, this is obtained by cutting the kernel in the middle, in an operation that corresponds to the Heaviside function



(a) Anisotropic full gaussian at a corner

(b) Edge directions and gradient direction extraction



(c) Direction θ_1 (d) Direction θ_2 (e) Direction η

Fig. 2. Synthetic quarter circle used to illustrate the image geometry captured by AHEDK. Enlarge the image to get a proper view of the orientations.

H. As in [16], we have chosen to rotate the image instead of filter, thereby reducing the algorithmic complexity by making use of a recursive Gaussian filter [17].

2.2 Anisotropic half exponential derivative kernel (AHEDK)

We modify the Shen Castan filter [18] to imitate the AHGDK. We call this modified filter as anisotropic half exponential derivative kernel (AHEDK). AHEDK shows similar characteristics and produces similar results as that of the above explained AHGDK. We use the recursive implementation of half exponential kernel which is of order 1. So, it is approximately 5 times faster than the recursive implementation of anisotropic half derivative kernel. By construction AHEDK exhibits derivative characteristics along the X direction and smoothing characteristics along the Y direction.

$$\mathcal{E}(x, y, \theta) = I_\theta * C_1 \cdot H(y) \cdot e^{(-\alpha_\mu \cdot y)} \cdot sign(x) \cdot e^{(-\alpha_\lambda \cdot |x|)} \tag{2}$$

The derivative information $\mathcal{E}(x, y, \theta)$ in Eq.2 is obtained by spinning the AHEDK around a pixel (x, y), C_1 is a normalization coefficient and $(\alpha_\mu, \alpha_\lambda)$ the height and width of the anisotropic half exponential kernel (see Fig.1(d)). Similar to Eq.1 only the causal part of this filter along the Y axis is used by cutting the kernel in the middle, in an operation that corresponds to the Heaviside function H.

2.3 Anisotropic gradient magnitude and direction estimation

We construct the signature $\mathcal{Q}(x_p, y_p, \theta)$ by considering the response of the above described rotating filter at a pixel location (x_p, y_p) as in Fig.1(f). Fig.1(g) shows a sample signature obtained by applying the AHGDK at the the pixel location (x_p, y_p) in steps of 2°. Anisotropic gradient magnitude $\|\nabla I\|_a$ and its associate direction η for the key-point at location (x_p, y_p) are obtained by considering the global extrema G_{max} and G_{min} of the function $\mathcal{Q}(x_p, y_p, \theta)$ along with θ_1 and θ_2. The two angles θ_1 and θ_2 define a curve crossing the pixel (an incoming and outgoing direction) thus representing the two edge directions. Two of these global extrema are combined to maximize the gradient $\|\nabla I\|_a$, i.e:

$$\begin{cases} G_{max} = \max\limits_{\theta\in[0,360°[} \mathcal{Q}(x_p, y_p, \theta) \quad \text{and} \quad \theta_1 = \arg\max\limits_{\theta\in[0,360°[} (\mathcal{Q}(x_p, y_p, \theta)) \\ G_{min} = \min\limits_{\theta\in[0,360°[} \mathcal{Q}(x_p, y_p, \theta) \quad \text{and} \quad \theta_2 = \arg\min\limits_{\theta\in[0,360°[} (\mathcal{Q}(x_p, y_p, \theta)) \\ \|\nabla I\|_a = G_{max} - G_{min} \\ \eta = \dfrac{\theta_1 + \theta 2}{2} \end{cases} \tag{3}$$

By construction, our filter naturally detects curvature information with θ_1 and θ_2 using only 1st order derivatives. The arrows on the circle in Fig.2 and on the square in Fig.1 illustrates this property. We follow the same procedure to obtain the signature $\mathcal{E}(x_p, y_p, \theta)$ and extract the edge and gradient directions using the AHEDK.

3 DESCRIPTOR CONSTRUCTION

Descriptor construction process is shown in Fig.3. As in [19], we follow the standard procedure to obtain the rotation and affine normalized gray level image patch. This standard procedure is followed in the construction of all the descriptors (SIFT, DAISY, GLOH) used in our experiments. As in Fig.3, for each pixel in the image patch, we spin the AHGDK and obtain a signature (for simplicity and proper viewing, in Fig.3 signatures are not shown). From this signature we extract : i) Angle at the maxima, θ_1 and the response G_{max} at θ_1. ii) Angle at the minima, θ_2 and the response $\|G_{min}\|$ at θ_2. iii) Anisotropic gradient angle η and its magnitude $\|\nabla I\|_a$. The next step is to construct 3 intermediate descriptors by forming the HoG of all the 3 angles separately. The angle θ_1 is weighed by G_{max}, θ_2 by $\|G_{min}\|$, η by $\|\nabla I\|_a$ and binned as in Eq.4.

$$\begin{cases} HoG_{\theta_1} = \{\theta_{1_{bin1}}, \theta_{1_{bin2}}, \theta_{1_{bin3}}, \theta_{1_{bin4}} ..... \theta_{1_{bin128}}\} \\ HoG_{\theta_2} = \{\theta_{2_{bin1}}, \theta_{2_{bin2}}, \theta_{2_{bin3}}, \theta_{2_{bin4}} ..... \theta_{2_{bin128}}\} \\ HoG_{\eta} = \{\eta_{bin1}, \eta_{bin2}, \eta_{bin3}, \eta_{bin4} ..... \eta_{bin128}\} \end{cases} \tag{4}$$

Finally, we combine the three intermediate descriptors in 4 different ways, i) DESCT1-theta1-eta : A 256 dimension(length) descriptor obtained by concatenating HoG_{θ_1} and HoG_{η}. ii) DESCT2-theta2-eta : A 256 dimension descriptor obtained by concatenating HoG_{θ_2} and HoG_{η}. iii) DESCT3-theta1-theta2 : A

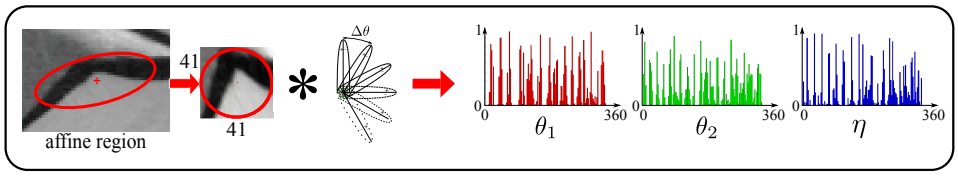


Fig. 3. Histogram construction process using 3 directions θ_1, θ_2 and η obtained using anisotropic half Gaussian kernel.

256 dimension descriptor obtained by concatenating HoG_{θ_1} and $HoG_{\theta 2}$. iv) DESCT4-theta1-theta2-eta : A 384 dimension descriptor obtained by concatenating HoG_{θ_1}, HoG_{θ_2} and HoG_{η}. We follow the same procedure to construct the descriptors using AHEDK and call them EXP-DESCT1-theta1-eta, EXP-DESCT2-theta2-eta, EXP-DESCT3-theta1-theta2 and EXP-DESCT4-theta1-theta2-eta.

4 Experiments and Results

4.1 Dataset and Evaluation

The entire code was implemented on Matlab platform. Harris affine key points [19] were used for image patch extraction. Key points obtained from other detectors can also be used for extracting the image patches. We evaluate and compare the performance of our descriptor against the state of the art descriptors on the standard dataset using the standard protocol provided by Oxford group. The binaries and dataset are obtained from website linked to [2] (http://www.robots.ox.ac.uk/ vgg/research/affine/). The dataset used in our experiments has different geometric and photometric transformations such as change of scale and image rotation (boat), viewpoint change (graf), image blur (bike), JPEG compression (compression) and illumination change (Leuven). For each type of image transformation there is a set of six images with established ground truth homographies.

We use the evaluation criterion as proposed by [2]. The evaluation criteria is based on the number of correspondences, number of correct matches and the number of false matches between two images. Here, we test the descriptors using Similarity based method and Nearest neighbour method. Due to lack of space we restrain from going in to the details of these methods. A detailed description of these methods can be found in [2]. The results are presented using the recall vs 1-precision curves. As in eq.5, recall is defined as the total number of correctly matched regions over the number of corresponding regions between two images of the same scene. From eq.6, 1-precision is represented by the number of false matches relative to the total number of matches. In all our experiments, Euclidean distance is used as the distance measure.

$$\text{recall} = \frac{\text{Total No of correct matches}}{\text{No of correspondences}} \tag{5}$$

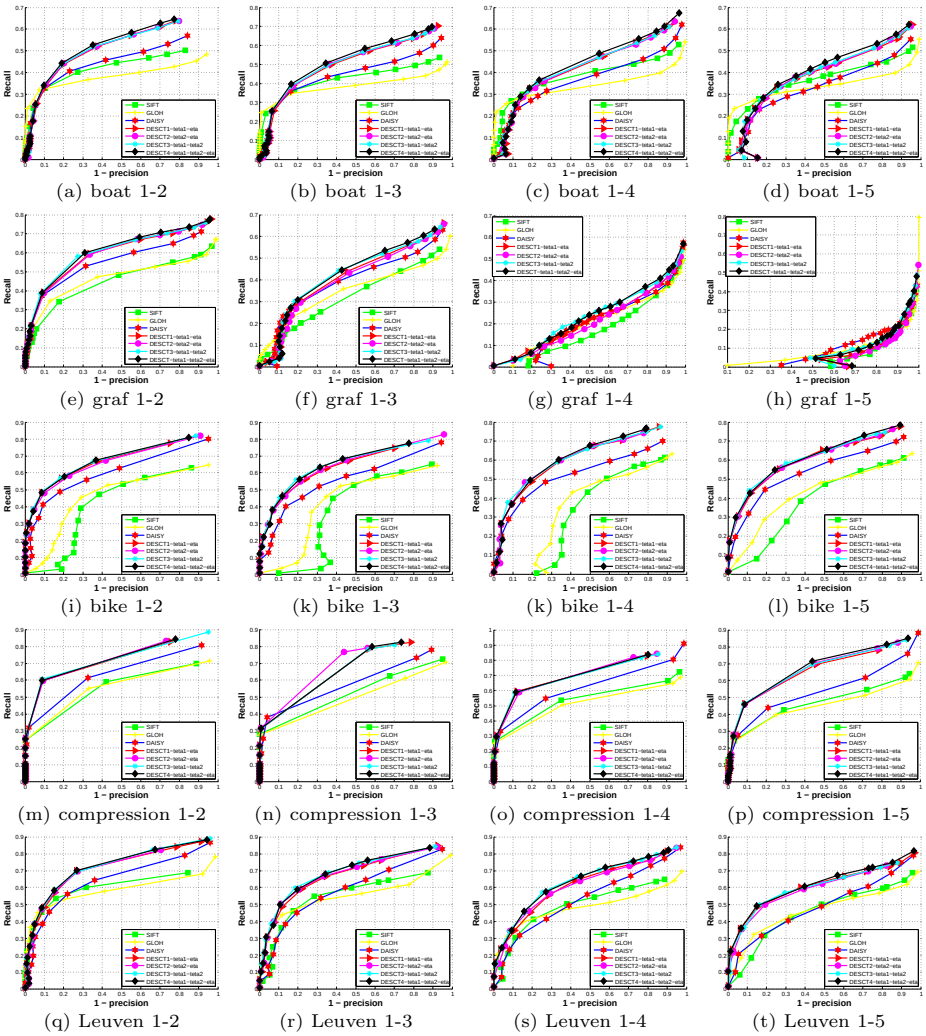
$$\text{1-precision} = \frac{\text{No of false matches}}{\text{No of correct matches} + \text{No of false matches}} \tag{6}$$

Our descriptors depends on 4 different parameters: $\Delta\theta$, $No{-}of{-}bins$, height and width (μ, λ for AHGDK and α_μ, α_λ for AHEDK). The rotation step $\Delta\theta$ is fixed to 10°. Increasing the rotation step results in loss of information. The image patch is divided into 16 blocks. All blocks are of size 10x10(Since we are using a patch of size 41x41 the blocks in the extreme right and bottom have 11x11 size). The number of bins ($No{-}of{-}bins$) is fixed to 8 per block, resulting in a $8 * 16 = 128$ bins for 16 blocks. Increasing the number of bins results in almost the same performance but, increases the dimensionality of the descriptor. AHGDK height μ and width λ is fixed to 6 and 1 respectively. AHEDK height α_μ and width α_λ is fixed to 1 and 0.2 respectively. Width and height parameters are chosen to have a ratio sharpness length suitable for robust edge detection [16], which generally gives good results in most cases. This ratio is compatible with the angle filtering step.

4.2 Descriptor Performance

The performance of descriptors obtained using both AHGDK and AHEDK are compared against SIFT, GLOH and DAISY. For SIFT and GLOH, the descriptors are extracted from the binaries provided by Oxford group (`http://www.robots.ox.ac.uk/~vgg/research/affine/`) [2]. DAISY descriptor for patches is extracted from the code provided by [6]. We have compared our descriptor on all the images in the dataset. Due to lack of space, we omit showing the quantitative results for the image pair (1-6). Qualitative results using nearest neighbour approach for the image pair (1-4) using descriptors obtained from both AHGDK and AHEDK is shown in Fig.7.

1. Rotation changes (boat).
   Using similarity matching strategy, all of our descriptors obtained from both AHGDK and AHEDK have a clear advantage over SIFT, GLOH and DAISY. This can be seen from the 1st row of Fig.4 and Fig.5. When the descriptors obtained using AHGDK and AHEDK are compared against each other they exhibit similar results as shown in the 1st row of Fig.6. The same can be said, when experimented with nearest neighbour matching strategy. Due to lack of space, for all the dataset we omit showing the quantitative results using nearest neighbour matching method.
2. Viewpoint changes (graff).
   For image pair (1-2), (1-3) and (1-4), using similarity approach our descriptors and its variants performs similar to or better than that of SIFT, DAISY and GLOH. Image pair (1-5) is very challenging and all the descriptors including SIFT, DAISY and GLOH fail. The results can be seen from the



(a) boat 1-2 (b) boat 1-3 (c) boat 1-4 (d) boat 1-5

(e) graf 1-2 (f) graf 1-3 (g) graf 1-4 (h) graf 1-5

(i) bike 1-2 (k) bike 1-3 (k) bike 1-4 (l) bike 1-5

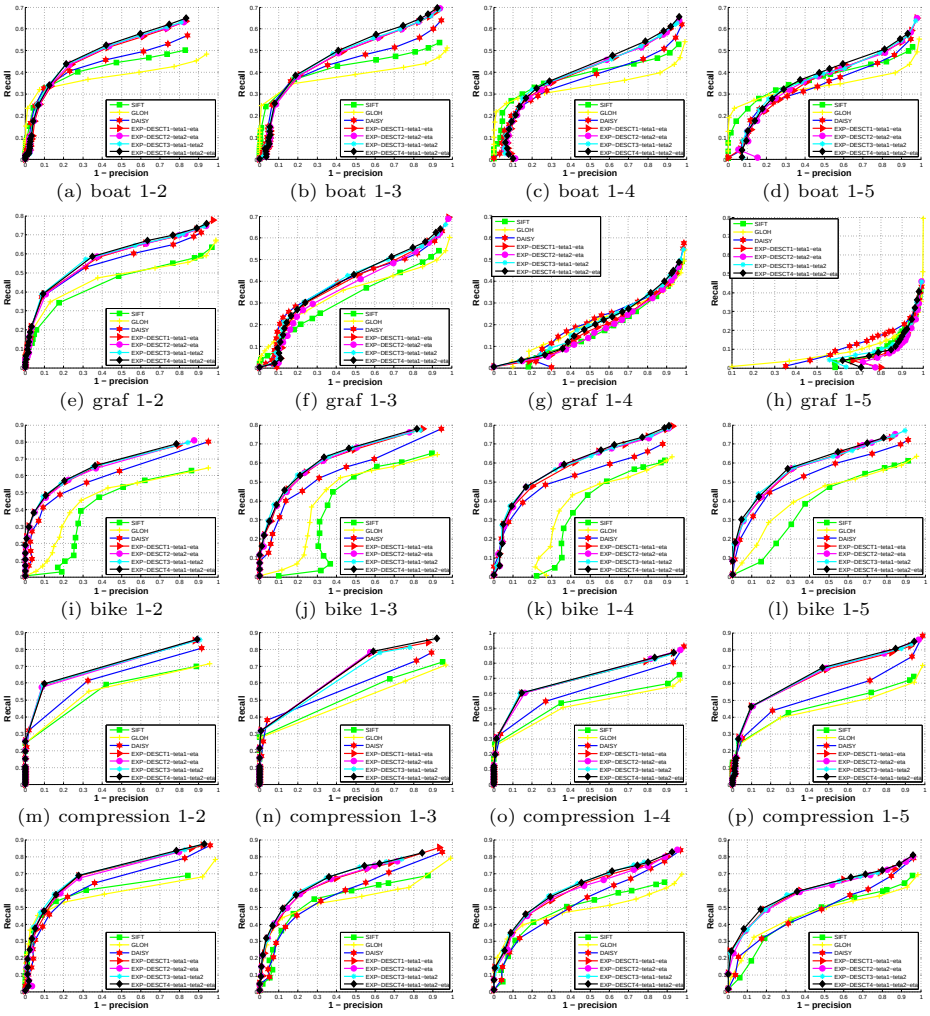
(m) compression 1-2 (n) compression 1-3 (o) compression 1-4 (p) compression 1-5

(q) Leuven 1-2 (r) Leuven 1-3 (s) Leuven 1-4 (t) Leuven 1-5

Fig. 4. Recall vs 1-Precision curves for SIFT, GLOH, DAISY and 4 descriptors obtained using AHGDK. Similarity matching is used for evaluation.

graphs in the 2nd row of Fig.4 and Fig.5. When the descriptors obtained using AHGDK and AHEDK are compared against each other they exhibit similar results as shown in the 2nd row of Fig.6.

3. Variations in blur, compression and brightness.
   For the blur (bike), compression, and brightness (Leuven) changes, using similarity approach all our descriptors outperform SIFT, DAISY and GLOH. The performance can be seen in the third, fourth and fifth rows of Fig.4 and Fig.5 respectively. When the descriptors obtained using AHGDK and



(a) boat 1-2 (b) boat 1-3 (c) boat 1-4 (d) boat 1-5

(e) graf 1-2 (f) graf 1-3 (g) graf 1-4 (h) graf 1-5

(i) bike 1-2 (j) bike 1-3 (k) bike 1-4 (l) bike 1-5

(m) compression 1-2 (n) compression 1-3 (o) compression 1-4 (p) compression 1-5

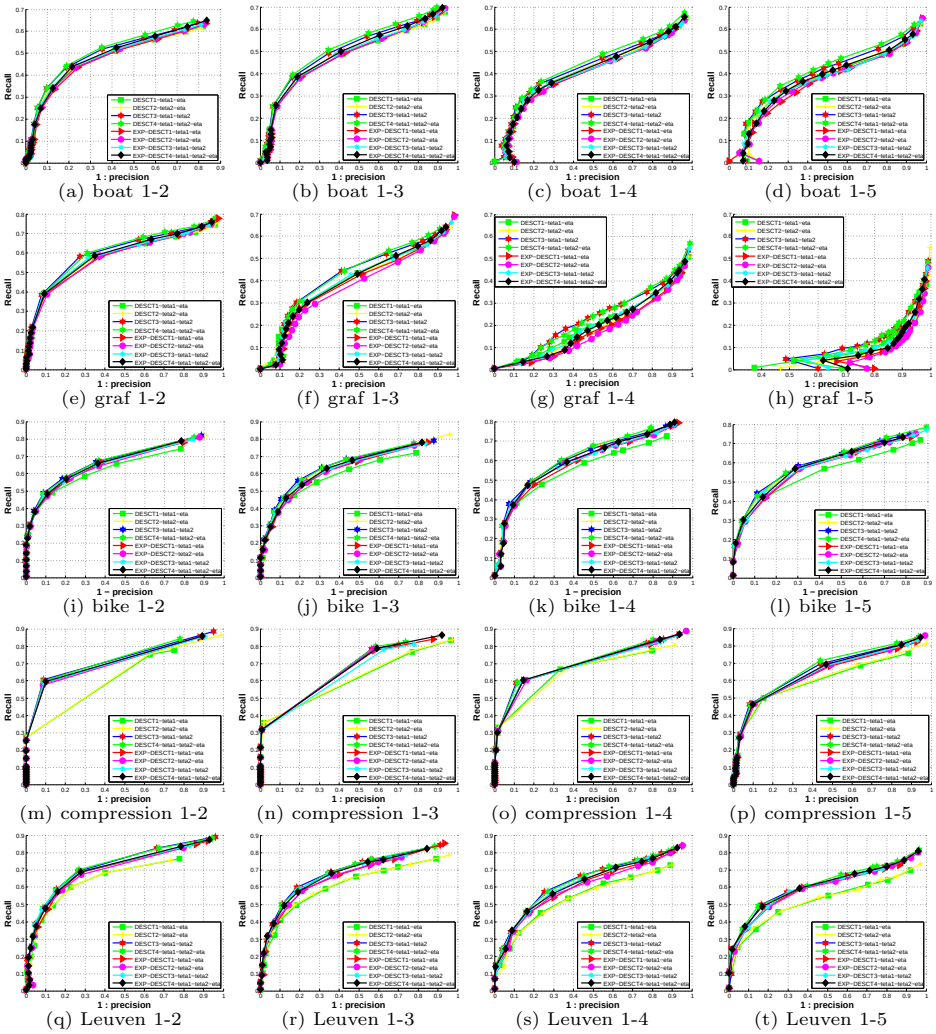
(q) Leuven 1-2 (r) Leuven 1-3 (s) Leuven 1-4 (t) Leuven 1-5

Fig. 5. Recall vs 1-Precision curves for SIFT, GLOH, DAISY and 4 descriptors obtained using AHEDK. Similarity matching is used for evaluation.

AHEDK are compared against each other they exhibit similar results as shown in the 3rd, 4th and 5th row of Fig.6.

5 Conclusion

This paper proposes novel image patch descriptors based on anisotropic half Gaussian derivative kernel (AHGDK) and anisotropic half exponential derivative kernel (AHEDK). The originality of our method is that it captures the



(a) boat 1-2 (b) boat 1-3 (c) boat 1-4 (d) boat 1-5

(e) graf 1-2 (f) graf 1-3 (g) graf 1-4 (h) graf 1-5

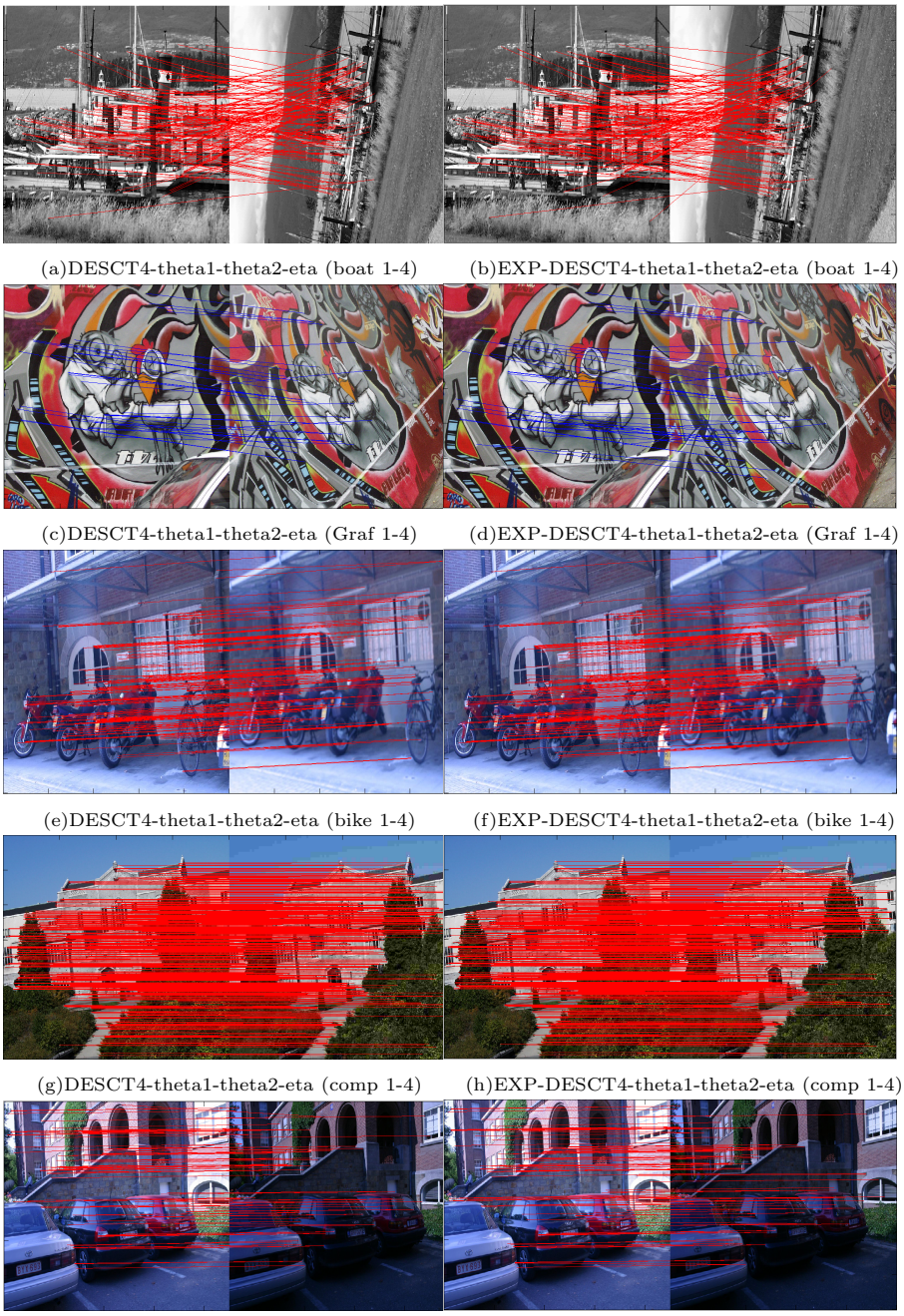
(i) bike 1-2 (j) bike 1-3 (k) bike 1-4 (l) bike 1-5

(m) compression 1-2 (n) compression 1-3 (o) compression 1-4 (p) compression 1-5

(q) Leuven 1-2 (r) Leuven 1-3 (s) Leuven 1-4 (t) Leuven 1-5

Fig. 6. Recall vs 1-Precision curves for all descriptors obtained using both AHGDK and AHEDK. Similarity matching is used for evaluation.

geometry of the image patch by embedding the response of the AHGDK or AHEDK in a HoG framework. Our method incorporates edge direction as well as anisotropic gradient direction for generating the descriptor and its variations. On the standard dataset provided by the Oxford group our descriptor and its variants outperform SIFT, GLOH and DAISY. In the future, we would like to learn the parameters by introducing a learning stage. We would also like to use this approach for object classification and image retrieval purpose. The speed of the descriptor generation can be boosted by parallel programming.



(a)DESCT4-theta1-theta2-eta (boat 1-4) (b)EXP-DESCT4-theta1-theta2-eta (boat 1-4)

(c)DESCT4-theta1-theta2-eta (Graf 1-4) (d)EXP-DESCT4-theta1-theta2-eta (Graf 1-4)

(e)DESCT4-theta1-theta2-eta (bike 1-4) (f)EXP-DESCT4-theta1-theta2-eta (bike 1-4)

(g)DESCT4-theta1-theta2-eta (comp 1-4) (h)EXP-DESCT4-theta1-theta2-eta (comp 1-4)

(i)DESCT4-theta1-theta2-eta (leuven 1-4) (j)EXP-DESCT4-theta1-theta2-eta (leuven 1-4)

Fig. 7. Qualitative matching results for the Oxford dataset using the nearest neighbour matching method. Matches in the Left column are obtained using AHGDK and the matches in the right column are obtained using AHEDK.

6 Acknowledgements

This work is funded by L'institut mediterraneen des metiers de la longevite (I2ML), Nimes, France.

References

1. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. In:International Journal of Computer Vision (IJCV), vol. 60, pp. 91-110. (2004)
2. Mikolajczyk, K., Schmid, C.: A Performance Evaluation of Local Descriptors. In: IEEE Trans. Pattern Anal. Mach. Intell, pp. 1615-1630. (2005).
3. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.J.: Speeded-Up Robust Features. In: Computer Vision and Image Understanding, vol. 110, pp. 346-359. (2008)
4. Fan, B., Wu, F., Hu, Z.: Rotationally Invariant Descriptors Using Intensity Order Pooling. In: Trans. Pattern Anal. Mach. Intell, vol. 34, pp. 2031-2045. (2012)
5. Morel, J. M., and Yu, G.: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. In: Journal on Imaging Sciences, vol. 2, pp. 438–469. (2009)
6. Tola, E., Lepetit, V., Fua, P.: DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo. In: TPAMI, vol. 32, pp. 815-830. (2010)
7. Dalal, N., Triggs, T.: Histograms of Oriented Gradients for Human Detection. In: Computer Vision and Pattern Recognition (CVPR), pp. 886-893. (2005)
8. Schmid, C., Mohr, R.: Local Gray-value Invariants for Image Retrieval. In: IEEE Trans. Pattern Anal. Mach. Intell, vol. 9, pp. 530-535. (1997)
9. Koenderink, J. J., Van Doorn, A.J.: Representation of local geometry in the visual system. In: Biological cybernetics, vol. 55, pp. 367-375. (1987)
10. Larsen, A. B. L., Darkner, S., Dahl, A. L., Pedersen, K. S.: Jet-Based Local Image Descriptors. In: European Conference on Computer Vision, pp. 638-650. (2012).
11. Palomares, J.L., Montesinos, P., Diep, D.: A new affine invariant method for image matching. In: 3DIP Image Processing and Applications, 8290. (2012).
12. Monroy, A., Eigenstetter, A., Ommer, B.: Beyond straight lines - Object detection using curvature . In: ICIP, pp. 3561-3564. (2011).
13. Zitnick, C.L.: Binary Coherent Edge Descriptors. In: 11th European Conference on Computer Vision (ECCV), pp. 170-182. (2010).
14. Eigenstetter, A., Ommer, B.: Visual Recognition using Embedded Feature Selection for Curvature Self-Similarity, Curran Associates, Inc. (2012).
15. Magnier, B., Montesinos, P.: Evolution of image regularization with PDEs toward a new anisotropic smoothing based on half kernels. In: SPIE, Image Processing: Algorithms and Systems XI. (2013)
16. Montesinos, P., Magnier, B.: A New Perceptual Edge Detector in Color Images. In: ACIVS, pp. 209-220. (2010).
17. Deriche, R.: Recursively implementing the gaussian and its derivatives. In: ICIP, pp. 263-267. (1992).
18. Shen, J., Castan, S.: An optimal linear operator for step edge detection. In: Graphical Model and Image Processing (CVGIP), vol. 54, pp. 112-133. (1992)
19. Mikolajczyk, K., Schmid, C.: Scale & Affine Invariant Interest Point Detectors. In:International Journal of Computer Vision (IJCV), vol. 60, pp. 63-86. (2004)