RSD-HoG: A New Image Descriptor

Darshan Venkatrayappa(✉), Philippe Montesinos, Daniel Diep, and Baptiste Magnier

Ecole des Mines D'Ales, LGI2P, Parc Scientifique Georges Besses, 30035 Nimes, France
{darshan.venkatrayappa,philippe.montesinos,daniel.diep, baptiste.magnier}@mines-ales.fr

Abstract. In this paper we propose a novel local image descriptor called RSD-HoG. For each pixel in a given support region around a key-point, we extract the rotation signal descriptor(RSD) by spinning a filter made of oriented anisotropic half-gaussian derivative convolution kernel. The obtained signal has extremums at different orientations of the filter. These characteristics are combined with a HoG technique, to obtain a novel descriptor RSD-HoG. The obtained descriptor has rich, discriminative set of local information related to the curvature of the image surface. With these rich set of features, our descriptor finds applications in various branches of computer vision. For evaluation, we have used the standard Oxford data set which has rotation, brightness, illumination, compression and viewpoint changes. Extensive experiments on these images demonstrates that our approach performs better than many state of the art descriptors such as SIFT, GLOH, DAISY and PCA-SIFT.

Keywords: Image descriptor · Half gaussian kernel · Feature matching · Rotation signal descriptor · HoG

1 Introduction

In computer vision, problems related to object matching, tracking, panorama generation, image classification, structure and motion estimation are effectively addressed by the popular approach of image representation by a set of local image descriptors. The main purpose of local image descriptor is to capture the geometry of a support region around a key-point. In addition to this, the image descriptor should be invariant to certain image transformations such as rotation, brightness, blurring and scale changes. Scanning through the com-puter vision literature, one can come across the term image matching pipeline. The image matching pipeline has four stages. In the first stage, key-points or regions are selected using the popular detectors such as LoG[2], DoG [3] or Har-ris Affine [4]. This is followed by the extraction of features or feature description from the support region around the key-point. Next, various post processing

applied. The final block involves matching the descriptor using different distance measures such as euclidean distance[3], hamming distance[6], Earth Mover's Distance(EMD)[7].

We are mainly interested in the second stage of the above described pipeline. The popular approach to extract features from a support region is to use the Histogram of Gradients(HoG)[8]. SIFT[3] effectively makes use of HoG to generate a descriptor. In PCA-SIFT[1] the dimension of the SIFT descriptor is reduced by applying the PCA technique. GLOH[9] and DAISY[10] use radial and daisy binning strategies respectively to improve on the existing SIFT descriptor. The standard version of SURF[11] achieves speed-up over SIFT by using Haar wavelets. MROGH[12] and MRRID[12] improve on the rotational invariance aspect by pooling local features based on their intensity orders in multiple support regions. In RIFF[13], radial and tangential components are extracted from the support region to form the descriptor. CSLBP[14] and MRRID[12] use local binary patterns to encode the image structure in the support region. A detailed evaluation of these descriptors can be found in[25].

There is a line of research oriented towards object recognition and image matching on camera-enabled mobile devices (e.g. phones and tablets). Limited computational power and storage space on these devices has enabled the emergence of binary descriptors. The main idea behind binary descriptor is that each bit is independent and Hamming distance can be used as similarity measure. Some of the most popular binary descriptors are BRIEF[6], ORB[15], BRISK[16] and FREAK[17]. A detailed evaluation of these descriptors can be found in[18].

Filter responses has been used in abundance for image description. Schmid and Mohr[19] 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. Larsen et al.[20] 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.[21] have come up with a local image descriptor issued from a filtering stage made of oriented anisotropic half-gaussian smoothing convolution kernels. In their work the authors use euclidean invariance is achieved by FFT correlation between the two signatures/descriptors. Moderate deformation invariance is achieved using Dynamic Time Warping (DTW).

The contributions of this paper include a new approach for descriptor generation. We embed the response of an anisotropic half Gaussian kernel in a HoG framework. For each pixel in the support region a rotation signal descriptor(RSD) is extracted by spinning a filter made of oriented anisotropic half-Gaussian derivative convolution kernel. We select two orientations at which a global maxima and a global minima occurs. Thus, for each pixel in the support region we have two orientations and a magnitude. These orientations are binned by weighing them with the magnitude.

2 Methodology

2.1 Image-Processing

Isotropic Gaussian filters Fig.1(a), Fig.1(b) are widely used for smoothing images. But, it is well known that this type of filters blur the image features. Gaussian anisotropic filtering overcome this by preserving the features along a particular direction. Compared to isotropic filters, anisotropic filters are elongated along one of the directions. Fig.1(c) and Fig.1(d) shows smoothing and derivative anisotropic filters respectively. We use a filter made of anisotropic half Gaussian derivative kernel as in [22] [23]. An anisotropic half-Gaussian derivative filter 1(f) is designed with a Gaussian half-smoothing filter in Y direction and a derivative Gaussian filter in X direction. The filtering in the Y direction acts as a causal integration, as the filtering in X direction is a full derivative. Then, rotating the filter around pixels generates a description around the pixels. Applying such filters on an edge point will produce a minima and a maxima in two directions. On a straight line edge these two direction are exactly opposite (180°), but generally the difference of angles reflects the curvature at edge points.

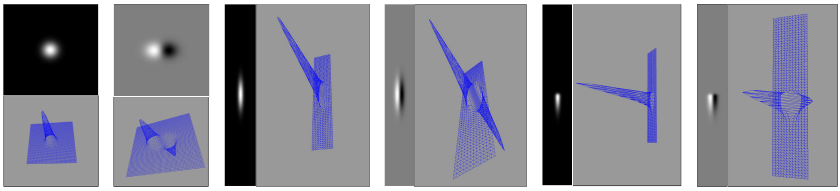


Fig. 1. (a) Isotropic smoothing filter. (b) Isotropic derivative filter. (c) Anisotropic smoothing filter. (d) Anisotropic derivative filter. (e) Half anisotropic smoothing filter. (f) Half anisotropic derivative filter.

As shown in Fig 2(a). at pixel (x, y), a derivative kernel is applied to obtain a derivative information $\varphi(x, y, \theta)$ in a function of orientation $\theta \in [0; 360°[$:

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

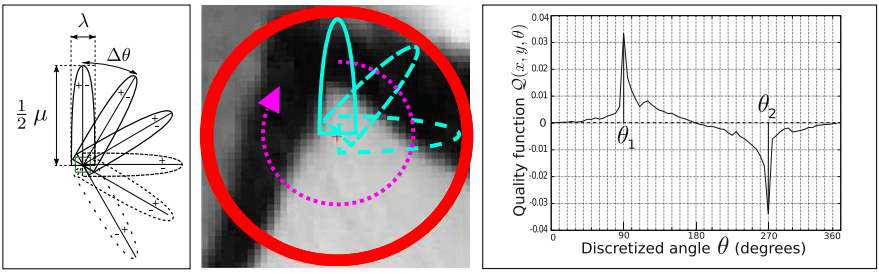
Where I_θ corresponds to a rotated image of orientation θ, C is a normalization coefficient, (x, y) are pixel coordinates, and (μ, λ) the standard deviation of the anisotropic Gaussian kernel. Only the causal part of this filter along the Y axis is used. This is obtained by cutting the kernel in the middle, in an operation that corresponds to the Heaviside function H. As in [23] we have chosen to rotate the image instead of the filter, there by reducing the algorithmic complexity by making use of a recursive Gaussian filter [24].

2.2 Rotational Signal Descriptor(RSD)

RSD is obtained by rotating the above described filter around a key-point. Fig. 2(c) shows a sample RSD obtained by applying the Gaussian derivative half filter at the the pixel location (x_p, y_p) in steps of 2°. We compute the gradient $\|\nabla I\|$ and the two angles for the key-point/pixel at location (x_p, y_p) by considering the global extrema of the function $\mathcal{Q}(x_p, y_p, \theta)$. The two angles θ_1 and θ_2 define a curve crossing the pixel(an incoming and outgoing direction). Two of these global extrema are combined to maximize $\|\nabla I\|$, i.e :

$$\begin{cases} \|\nabla I\| = \max\limits_{\theta \in [0,360°[} \mathcal{Q}(x_p, y_p, \theta) - \min\limits_{\theta \in [0,360°[} \mathcal{Q}(x_p, y_p, \theta) \\ \theta_1 = \arg\max\limits_{\theta \in [0,360°[} (\mathcal{Q}(x_p, y_p, \theta)) \\ \theta_2 = \arg\min\limits_{\theta \in [0,360°[} (\mathcal{Q}(x_p, y_p, \theta)) \end{cases} \tag{2}$$

We can spin an anisotropic half Gaussian derivative kernel at a key-point and the resulting response(RSD) can be considered as a descriptor. But, RSD alone gives a weak description, as it fails to capture the geometry around the key-point. This is the main motivation for combining the HoG technique with RSD.



(a) (b) (c)

Fig. 2. (a) A thin rotating Gaussian derivative half-filter. (b) Half Gaussian kernel applied to a keypoint x_p, y_p. (c) Extrema of a function $\mathcal{Q}(x_p, y_p, \theta)$. $\mu = 10$, $\lambda = 1$ and $\Delta\theta = 2^o$. Note that the initial orientation of the filter is vertical, upwardly directed and steerable clockwise.

2.3 Pre-Processing

Prior to feature extraction, the image is smoothed with a Gaussian filter. Then, key-points(regions) need to be localized. We have used Harris affine feature detector. It can also be noted that any of the existing feature detector can be used for this step. As in Fig. 3, we first extract image patches surrounding the key-points. Then, depending on the key-point attributes such as scale and orientations, we rotate the patch by its orientation. This is followed by scale normalizing the oriented patch to a fixed size of 41x41. More details about the process for patch extraction and normalization can be found in [4].

2.4 RSD-HoG Construction

The framework of the RSD-HOG extraction is illustrated in Fig. 3. On each pixel of the normalized patch, we apply the rotating semi Gaussian filter to obtain the RSD. From this RSD, we extract two angles and a magnitude for each pixel as explained above. Then we bin the two angles separately as in Eq.3 and Eq.4. The image patch is divided in to 16 blocks. Since the image patch is of size 41x41, most of the blocks are of size 10x10 (blocks on the extreme right and extreme bottom are of size 11x11). Each of these block contributes 8 bins to the final descriptor. We fuse the two intermediate descriptors to form the final descriptor as in Eq.5. The intermediate descriptors in Eq.3 and Eq.4 alone can be used as descriptors. But, fusing these two descriptors as in Eq.5 results in a more robust description. The performance of two intermediate descriptor and the final descriptor for the boat dataset(Rotation changes) can be seen in the first row of Fig.4.

$$RSD - HoG - Theta1 = \{\theta_{1_{bin1}}, \theta_{1_{bin2}}, \theta_{1_{bin3}}, \theta_{1_{bin4}} ..... \theta_{1_{bin128}}\} \quad (3)$$

$$RSD - HoG - Theta2 = \{\theta_{2_{bin1}}, \theta_{2_{bin2}}, \theta_{2_{bin3}}, \theta_{2_{bin4}} ..... \theta_{2_{bin128}}\} \quad (4)$$

$$RSD - HoG = \{\theta_{1_{bin1}}, \theta_{1_{bin2}}, ... \theta_{1_{bin128}}, \theta_{2_{bin1}}, \theta_{2_{bin2}}, ... \theta_{2_{bin128}}\} \quad (5)$$

3 Experiments, Discussions and Results

3.1 Dataset and Evaluation

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

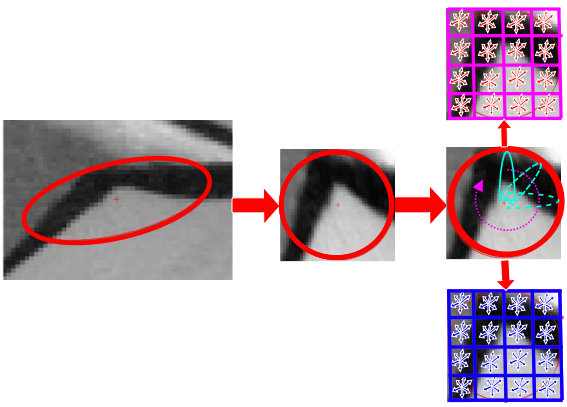


Fig. 3. Affine region is normalized to a square of size 41x41. On each pixel of the patch, RSD is generated and the two angles θ_1 and θ_2 are extracted. These angles are binned separately.

by Oxford group. The binaries and dataset are obtained from website linked to [25] (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, image rotation, viewpoint change, image blur, JPEG compression and illumination change. For each type of image transformation there is a set of six images with established ground truth homographies.

Recall versus precision curves as proposed by [25] are used to evaluate our descriptor. This is based on the number of correspondences, correct matches and false matches obtained for an image pair. Number of correspondences is defined as the number of pairs of features whose euclidean distance is below a certain threshold. We vary the threshold in steps to obtain the recall vs precision curves. A correct match is recorded when the pair of features with the smallest euclidean distance coincides with the pair of points/regions matched by the ground-truth. As in Eq.6, 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.7, 1-precision is represented by the number of false matches relative to the total number of matches.

$$\text{recall} = \frac{\text{Total number of correct matches}}{\text{Number of correspondences}} \tag{6}$$

$$\text{1-precision} = \frac{\text{Number of false matches}}{\text{Number of correct matches} + \text{Number of false matches}} \tag{7}$$

3.2 Parameter Selection

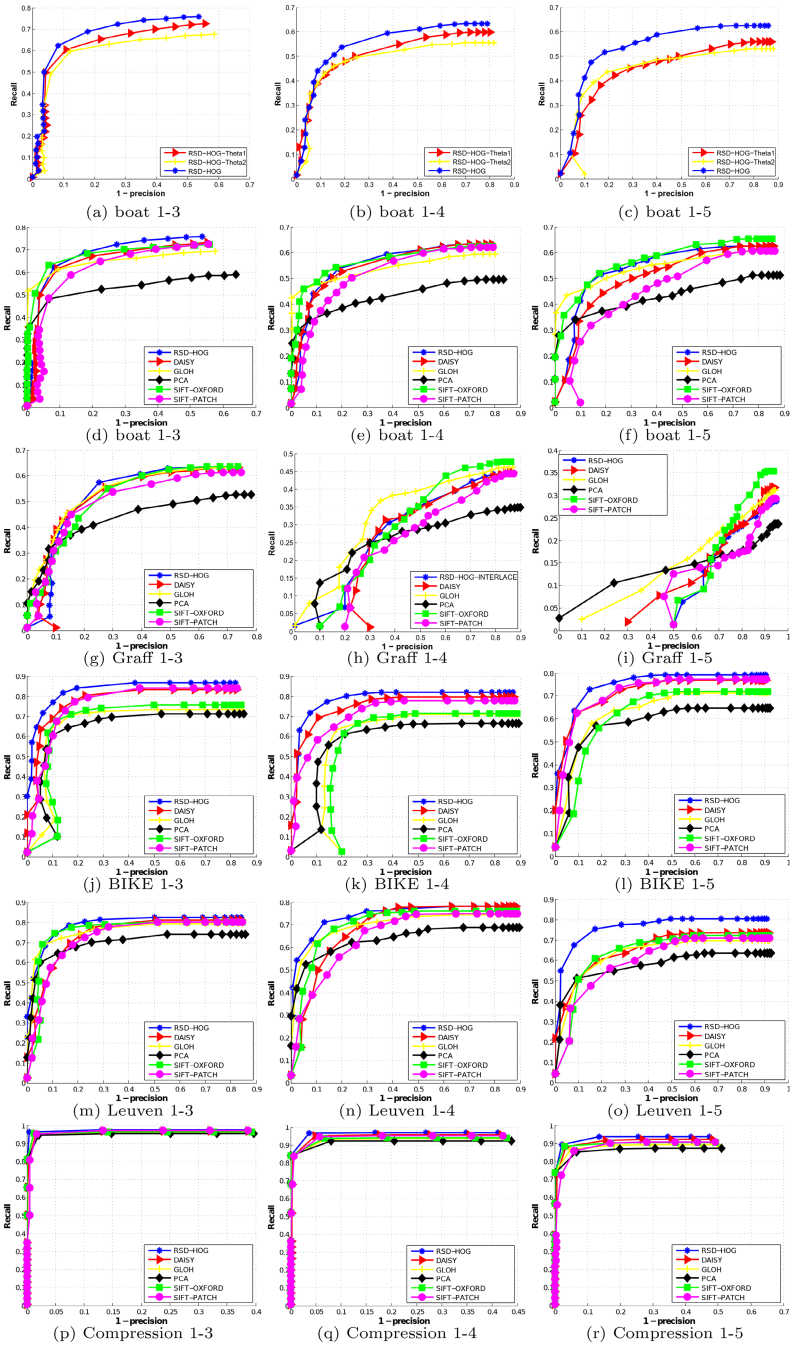
Our descriptor has 4 different parameters that are tabulated in the table 1. The rotation step of the filter is fixed to 5°. Increasing the rotation step results in loss of information. We have fixed the number of bins to 8 per block, resulting in a $8 * 16 = 128$ bins for 16 blocks. Increasing the number of bins results in the same performance but, increases the dimensionality of the descriptor. All the parameters are chosen empirically.

Table 1. Parameters

filter Height (μ)	filter Width (λ)	Rotation step ($\Delta\theta$)	No of BINS
6	1	5°	8

3.3 Descriptor Performance

The performance of RSD-HoG is compared against SIFT-OXFORD, SIFT-PATCH, GLOH, DAISY and PCA-SIFT. For SIFT-OXFORD, PCA-SIFT and GLOH the descriptors are extracted from the binaries provided by Oxford group (http://www.robots.ox.ac.uk/~vgg/research/affine/)[25]. DAISY descriptor is



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

(d) boat 1-3 (e) boat 1-4 (f) boat 1-5

(g) Graff 1-3 (h) Graff 1-4 (i) Graff 1-5

(j) BIKE 1-3 (k) BIKE 1-4 (l) BIKE 1-5

(m) Leuven 1-3 (n) Leuven 1-4 (o) Leuven 1-5

(p) Compression 1-3 (q) Compression 1-4 (r) Compression 1-5

Fig. 4. Recall vs Precision curves for Oxford dataset using nearest neighbour matching method

extracted from the code provided by [10]. SIFT-PATCH is the SIFT descriptor applied on a patch using the VLFEAT library [26]. Due to lack of space we restrict ourself to image pairs(1-3),(1-4) and (1-5).

1. Rotation changes.
   First, we compare the performance of $RSD-HoG$ (Eq.5) with the intermediate descriptors $RSD-HOG-Theta1$ (Eq.3)and $RSD-HOG-Theta2$ (Eq.4). From the graphs in the first row of Fig.4 it is clear that $RSD-HoG$ performs better than the two intermediate descriptors. This is same for all the images with different transformations. Due to lack of space we restrict ourselves to graphs for rotational changes. When $RSD-HOG$ is compared with other descriptors it can be seen that our descriptor performs better than other descriptors. This can be seen in the 2nd row of Fig.4.
2. Viewpoint and Blur changes.
   Graphs in the third row of Fig.4 represent Recall vs Precision curves for the Viewpoint changes. It can be seen that for the first 2 cases (1-2),(1-3)and (1-4) $RSD-HoG$ performs similar to other descriptors. For the final case the performance of $RSD-HoG$ deteriorates. This is a challenging sequence and all the other descriptors perform badly. Graphs in the fourth row of Fig.4 represent Recall vs Precision curves for the Blur changes. From the graphs it is clear that $RSD-HoG$ outperforms other descriptors.
3. Brightness and Compression changes.
   Based on graphs in the fifth row of Fig.4, we conclude that the performance of $RSD-HoG$ is superior to the performance of other descriptors when it comes to handling Brightness variations. Similarly, graphs in the last row of Fig.4 illustrate the dominance of our descriptor for compression variations.

4 Conclusion

This paper proposes a new image descriptor called RSD-HoG. It also proposes a new approach to construct the descriptor by interlacing the bins of the two intermediate descriptors. On the standard dataset provided by the Oxford group, RSD-HoG outperforms other state of the art descriptors. Currently, high complexity and the dimension of the descriptor are a major drawback. In the future we would like to reduce both the complexity and dimension of our descriptor. Here, we have used a fixed set of parameters for the anisotropic half Gaussian kernel. In the future, we would like to experiment with other variations of the anisotropic half Gaussian kernel. We would also like to focus on the real time implementation of our descriptor using parallel programming techniques. Another direction of our future work would be to test our approach on tasks related to object detection, writer classification and shape retrieval.

References

1. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: Computer Vision and Pattern Recognition(CVPR), pp. 506–513. IEEE Computer Society (2004)

2. Lindeberg, T.: Feature Detection with Automatic Scale Selection. International Journal of Computer Vision(IJCV) **30**, 79–116 (1998)
3. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision(IJCV) **60**, 91–110 (2004)
4. Mikolajczyk, K., Schmid, C.: Scale & Affine Invariant Interest Point Detectors. International Journal of Computer Vision(IJCV) **60**, 63–86 (2004)
5. Chandrasekhar, V., Takacs, G., Chen, D.M., Tsai, S.S., Reznik, Y.A., Grzeszczuk, R., Girod, R.: Compressed Histogram of Gradients: A Low-Bitrate Descriptor. International Journal of Computer Vision(IJCV) **96**, 384–399 (2012)
6. Calonder, M., Lepetit, V., Özuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: Computing a Local Binary Descriptor Very Fast. IEEE Trans. Pattern Anal. Mach. Intell.(PAMI) **34**, 1281–1298 (2012)
7. Lazebnik, S., Schmid, C., Ponce, J.: A Sparse Texture Representation Using Local Affine Regions. IEEE Trans. Pattern Anal. Mach. Intell.(PAMI) **27**, 1256–1278 (2005)
8. Dalal, N., Triggs, T.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition(CVPR), pp. 886–893 (2005)
9. Mikolajczyk, M., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell.(PAMI) **27**, 1615–1630 (2005)
10. Tola, E., Lepetit, V., Fua, P.: DAISY: An Efficient Dense Descriptor Applied to Wide Baseline Stereo. IEEE Trans. Pattern Anal. Mach. Intell.(PAMI) **32**, 815–830 (2010)
11. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.J.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding **110**, 346–359 (2008)
12. Fan, B., Wu, F., Hu, Z.: Rotationally Invariant Descriptors Using Intensity Order Pooling. IEEE Trans. Pattern Anal. Mach. Intell.(PAMI) **34**, 2031–2045 (2012)
13. Takacs, G., Chandrasekhar, V., Tsai, S.S., Chen, D.M., Grzeszczuk, R., Girod, B.: Rotation-invariant fast features for large-scale recognition and real-time tracking. Sig. Proc.: Image Comm. **34**, 334–344 (2013)
14. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recognition **42**, 425–436 (2009)
15. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.R.: ORB: An efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision, ICCV, pp. 2564–2571 (2009)
16. Leutenegger, S., Chli, M., Siegwart, R.: BRISK: binary robust invariant scalable keypoints. In: IEEE International Conference on Computer Vision, ICCV, pp. 2548–2555 (2011)
17. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: fast retina keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition, ICCV, pp. 510–517 (2012)
18. Figat, J., Kornuta, T., Kasprzak, W.: Performance evaluation of binary descriptors of local features. In: Chmielewski, L.J., Kozera, R., Shin, B.-S., Wojciechowski, K. (eds.) ICCVG 2014. LNCS, vol. 8671, pp. 187–194. Springer, Heidelberg (2014)
19. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 530–535 (1997)
20. Larsen, A.B.L., Darkner, S., Dahl, A.L., Pedersen, K.S.: Jet-based local image descriptors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 638–650. Springer, Heidelberg (2012)
21. Palomares, J.L., Montesinos, P., Diep, D.: A new affine invariant method for image matching. In: 3DIP Image Processing and Applications, vol. 8290 (2012)

22. 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)
23. Montesinos, P., Magnier, B.: A new perceptual edge detector in color images. In: Blanc-Talon, J., Bone, D., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2010, Part I. LNCS, vol. 6474, pp. 209–220. Springer, Heidelberg (2010)
24. Deriche, R.: Recursively implementing the gaussian and its derivatives. In: ICIP, pp. 263–267 (1992)
25. Mikolajczyk, K., Schmid, C.: A Performance Evaluation of Local Descriptors. IEEE Trans. Pattern Anal. Mach. Intell., 1615–1630 (2005)
26. Vedaldi, A., Fulkerson, B.: VLFeat: An Open and Portable Library of Computer Vision Algorithms (2008)