

# KID: Kirsch Directional Features Based Image Descriptor

B.H. Shekar<sup>1</sup>, K. Raghurama Holla<sup>1</sup>, and M. Sharmila Kumari<sup>2</sup>

<sup>1</sup> Department of Computer Science, Mangalore University, Karnataka, India  
{bhshekar, raghu247}@gmail.com

<sup>2</sup> Department of Computer Science and Engineering, P.A. College of Engineering,  
Mangalore, Karnataka, India  
sharmilabp@gmail.com

**Abstract.** In these days we have seen the development of local image descriptors for several computer vision applications in order to perform reliable matching and recognition. In this direction, we have made an attempt to propose a new local descriptor which uses the Kirsch's four directional edge features to describe the neighbourhood of the interest point. The descriptor computation mainly consists of two stages: feature detection (identification of interest points) and feature description. In the first stage, the interest points are detected using *Features from Accelerated Segment Test* (FAST) algorithm where interest points are identified by comparing the pixels on a circle of fixed radius around the interest point. In the second stage, the directional features for horizontal, vertical, right-diagonal and left-diagonal directions are extracted from the local region around the interest point using Kirsch masks. In order to achieve rotation invariance, the descriptor window coordinates are rotated with respect to the dominant orientation of the interest point. Experiments have been conducted on several image datasets to reveal the suitability of the proposed approach for various image processing applications. A comparative analysis with the other well known descriptors such as SIFT, SURF and ORB is also provided to exhibit the performance of the proposed local image descriptor.

**Keywords:** Interest point, feature detection, local descriptor, object recognition.

## 1 Introduction

The task of matching the local regions between two images affected by different kinds of geometric or photometric transformations such as translation, rotation, change in viewpoint, poor illumination etc, is a challenging problem in computer vision. In recent years, several researchers have focussed their attention on developing a local descriptor which is robust against the above challenges. We can find that the local descriptor based matching has many applications such as object recognition, face recognition, image retrieval, object tracking, video retrieval, stereo correspondence problem etc. The local features provide a powerful

way of image representation and allows to recognize the objects or scenes within an image without performing image segmentation (see [20]).

Generally, the descriptor computation begins with the feature detection phase, where invariant features are extracted. These features must be distinctive and stable. In the next phase, the feature description is performed. It is important to choose the appropriate descriptor to represent the local region in the neighbourhood of the interest point. This descriptor must be robust against noise, geometric and photometric deformations. Then the descriptors are matched between different images. Hence we are motivated to take up the work on designing a new local image descriptor which is simple to implement, robust and efficient also.

The rest of the paper is organized as follows. Section 2 presents the review of existing algorithms for feature detection and feature description. The proposed model is introduced in section 3. Experimental results and comparison with other models are presented in section 4 and conclusion is provided in section 5.

## 2 A Review of Related Works

An overview of the literature for interest point detectors and feature descriptors is given below.

### 2.1 Feature Detectors: A Review

We can see that several algorithms for interest point detection have been reported in the literature [20]. The popularly used Harris corner detector proposed by Harris and Stephens [9] is based on the eigenvalues of the second moment matrix. Lowe(2004) [13] suggested a method for interest point detection by selecting the local extrema of an image, filtered with differences of Gaussians. Bay et.al(2006) [6] proposed a faster method of obtaining the Hessian detector using integral images and box filters for approximating the derivatives.

Rosten and Drummond [16] [17] developed a method for extracting interest points from an image using *Features from Accelerated Segment Test*(FAST) algorithm, which has become popular in recent years. The keypoints are identified by comparing the pixels in a circle of 16 pixels around the current pixel against a threshold. As the algorithm had some weaknesses, a machine learning approach is used to overcome the drawbacks of the initially developed algorithm. Finally, a non maximal suppression is performed to remove the multiple keypoints which are adjacent to one another. This detector is found to be more efficient when compared to DoG.

### 2.2 Local Descriptors: A Review

Several algorithms for computing the descriptors are available in the literature. A detailed comparative analysis can be seen in [14] [12] [8]. Lowe(2004) [13] introduced SIFT descriptor which consists of four stages, namely, scale-space peak

selection, keypoint localization, orientation assignment and descriptor building, resulting in a 128-dimensional descriptor. Ke and Sukthakar(2004) [10] developed PCA SIFT to improve the matching performance of SIFT. An extension of SIFT called *Gradient location-orientation histogram* (GLOH) [14] is proposed to increase the robustness and distinctiveness of SIFT, by computing the descriptor in log-polar location grid. Bay et al.(2008) [6] proposed SURF descriptor, which is computed from the sum of the Haar wavelet responses for  $x$  and  $y$  directions in the neighborhood of each keypoint.

To improve the computational efficiency of the high dimensional descriptors, binarization technique is suggested [7], where the descriptor vectors are mapped to Hamming space. The BRIEF descriptor is developed based on intensity difference tests. The binarized descriptors are compared using Hamming distance metric. Rublee et al. [18] developed the Oriented Fast and Rotated BRIEF(ORB), which is invariant to rotation and robust to noise. The ORB is an extension of BRIEF descriptor and the keypoints are found using FAST algorithm. Strecha et al. [19] suggested a method based on similarity-sensitive hashing algorithms applied to local SIFT descriptors to improve the performance of feature matching in wide-baseline stereo problems. Alahi et al. [5] proposed a descriptor called *Fast Retina Keypoint*(FREAK) motivated by the way in which the images are processed in the human visual system. It computes binary strings by comparing image intensities over a retinal sampling pattern.

We have seen that there are several techniques available for developing a local descriptor. However, the robustness to various challenges such as geometric and photometric transformations could not be achieved completely. So, we propose an efficient method of designing a local descriptor based on Kirsch's directional features to address the above challenges. The proposed model shows better results for several object datasets and face datasets, and hence suitable for real-time feature matching applications. The details of the proposed model are given below.

### 3 Proposed Model

In the proposed descriptor, the keypoints are extracted using the well known FAST detector as it is proved to be very efficient and robust in real time feature detection.

In order to achieve rotation invariance, the dominant orientation for each keypoint needs to be calculated. To obtain the orientation, we adopt the same method as applied in SURF descriptor [6] since the convolution operation becomes faster with the use of integral image. The Haar wavelet responses for  $x$  and  $y$  directions within the circular neighbourhood having a fixed radius around each keypoint are calculated. The dominant orientation is calculated by obtaining the sum of all responses within a sliding window of size  $\pi/3$ . The  $x$  and  $y$  responses within the window are added to get a local orientation vector. The longest such vector gives the orientation for the keypoint.

To extract the descriptor for each keypoint, a square window of size  $16 \times 16$  pixels centred on the keypoint is considered. This descriptor window coordinates are aligned in the direction of dominant orientation of the keypoint. To describe the features in this local region we use Kirsch’s features [11] as it gives the directional features. Also, Kirsch edge detector accurately detects the four directional edges since all the eight neighbours are considered in the operation. The directional gradient proposed by Kirsch is given by [15]:

$$G(i, j) = \max_{k=0}^7 \{|5P_k - 3Q_k|\} \tag{1}$$

where,

$$\begin{aligned} P_k &= A_k + A_{k+1} + A_{k+2} \\ Q_k &= A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7} \end{aligned} \tag{2}$$

In Eqn(1),  $G(i, j)$  represents the gradient of the pixel at  $(i, j)$ . The subscripts of  $A$  are evaluated modulo 8,  $A_k$  denotes the 8-neighbours of the pixel at  $(i, j)$  as shown in fig 1,  $i = 0, 1, \dots, 7$ .

$A_0$	$A_1$	$A_2$
$A_7$	$(i, j)$	$A_3$
$A_6$	$A_5$	$A_4$

**Fig. 1.** 8-neighbours of a pixel  $(i, j)$

Now, the four directional features are calculated as follows:

$$\begin{aligned} G(i, j)_H &= \max(|5P_0 - 3Q_0|, |5P_4 - 3Q_4|) \\ G(i, j)_V &= \max(|5P_2 - 3Q_2|, |5P_6 - 3Q_6|) \\ G(i, j)_{RD} &= \max(|5P_1 - 3Q_1|, |5P_5 - 3Q_5|) \\ G(i, j)_{LD} &= \max(|5P_3 - 3Q_3|, |5P_7 - 3Q_7|) \end{aligned} \tag{3}$$

where  $G(i, j)_H$ ,  $G(i, j)_V$ ,  $G(i, j)_{RD}$  and  $G(i, j)_{LD}$  represent the directional edge features for horizontal, vertical, right-diagonal and left diagonal directions respectively. The filter masks to be used for calculating directional feature vectors are shown in fig 2.

In the above approach, the global characteristics of images are not considered. So, the normalized image intensities within the local region is used as global feature in addition to the local directional features. Each of the  $16 \times 16$  features within the local region are compressed to  $4 \times 4$  using bicubic interpolation technique where the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighbourhood and then the descriptor is normalized. Thus we have a descriptor of length  $4 \times 4 \times 5 = 80$ . The fig 3 summarizes the different stages of descriptor building.



The descriptors are compared using City block distance measure. The nearest neighbour distance ratio (NNDR) [13] is used to match the keypoints. The two regions are matched by comparing the distance ratio between the first and second nearest neighbours against a threshold. If  $D_B$  and  $D_C$  are the first and second nearest neighbours of  $D_A$ , then the region is matched if  $\frac{\|D_A - D_B\|}{\|D_A - D_C\|} < T$ , we fixed the threshold value  $T = 0.75$  empirically.

## 4 Experimental Results

In this section, we present the experimental results to reveal the performance of the proposed descriptor. We conducted experiments on several standard image datasets such as COIL-20 object dataset [2], 53-objects dataset [3], YALE face database [4], AT & T face database [1] etc. The experiments have been performed with various combinations of training and testing sets. Also, we have provided a comparative analysis with the other standard descriptors such as SIFT, SURF and ORB by performing the similar experiments for these descriptors on the same datasets. All the descriptors were tested with their default parameters settings (i.e. default threshold parameters for interest point detection and matching). The recognition rate is defined as the number of correctly recognized test objects divided by the total number of test objects. Some of the results are given in Table 1. The figure 4 demonstrates the result obtained by the proposed descriptor for matching the corresponding keypoints of two images affected by some transformations such as rotation, scale change, viewpoint change etc.

The COIL-20 dataset consists of gray scale images of 20 objects taken at a pose interval of 5 degree resulting in 72 images per each object, and it contains a total of 1440 images. In the first training set, we considered the images taken at a pose interval of 15 degree, resulting in 24 training images for each subject and the remaining images were used for testing purpose. The images taken at every 30 degree interval were chosen for the second training set, resulting in 12 training images for each object and the remaining 60 images were used for testing. In the third phase, the objects captured at a pose interval of 60 degree were selected for training, which leads to 6 training images for each subject and the remaining 66 images were used for testing purpose. The proposed model achieved the recognition rates: 98.75%, 96.67% and 86.97% respectively for the three different training sets. From Table 1, it can be noticed that the proposed descriptor outperforms the other descriptors on each test combinations. The SURF descriptor (with default parameters) produces weak results since the number of detected interest points are too less to be matched. The YALE face database consists of 165 face images with 15 subjects that include variation in both facial expression and lighting. In the first training set, the first four face images of each person were considered and the remaining images have been chosen for testing. The proposed model achieved a recognition rate of 95.23% for this test images. In the second phase, the first three face images of each subject were chosen for training and the remaining eight images were used for testing. The recognition

rate for this test set is 93.33%. Objects-53 dataset consists of images of 53 objects with five different views for each subject, containing a total of 265 images. The images are captured at different camera view points. The first, third and fifth images of each subject were chosen for the training set and the remaining images were selected for testing purpose. ORB descriptor produced good results for this dataset, followed by the proposed model, achieving a recognition rate of 82.69%. The AT & T face database contains 10 different images of 40 subjects. Some images were captured at different times. The dataset contains images with changes in lighting and facial expressions. In our experiment, the odd numbered samples were used for training set and the remaining images were used for testing purpose. The SIFT produced better accuracy, followed by the proposed model (98%). From the experimental results it can be noticed that the proposed

**Table 1.** Descriptor matching performance on different datasets

Dataset + (No. of TrainSamples per sub.)	Recognition rate(%)			
	proposed model	SIFT	SURF	ORB
COIL-20(24)	98.75	96.56	83.33	95.73
COIL-20(12)	96.67	93.42	72.00	92.25
COIL-20(6)	86.97	84.47	48.94	81.81
YALE(4)	95.23	95.23	85.71	79.05
YALE(3)	93.33	93.33	83.33	71.66
53-objects(3)	82.69	77.88	74.04	86.53
AT & T(5)	98.00	99.50	92.50	86.50
Average Results(%)	<b>93.09</b>	<b>91.48</b>	<b>77.12</b>	<b>84.79</b>

descriptor gives consistently better accuracy for all the databases tested when compared to other local descriptors. Other approaches produced better results for object datasets, however, the performance decreased when tested on face databases.

## 5 Conclusion

We have presented a robust and efficient local descriptor based on Kirsch's directional features for representing the texture feature of images. The interest point detection is carried out efficiently using FAST detector. The experimental results show that the proposed local descriptor is on a par with the standard descriptors such as SIFT, SURF and ORB. The proposed method is computationally efficient, simple to implement and hence suitable for many computer vision and video processing applications.

## References

1. AT & T face database, <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
2. COIL-20 object dataset, <http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>
3. object dataset, <http://www.vision.ee.ethz.ch/showroom/zubud/>
4. YALE face database, <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>
5. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: fast retina keypoint. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 510–517 (June 2012)
6. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
7. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
8. Gauglitz, S., Höllerer, T., Turk, M.: Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision* 94, 335–360 (2011)
9. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of Fourth Alvey Vision Conference, pp. 147–151 (1988)
10. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, pp. II-506–II-513 (2004)
11. Lee, S.W.: Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(6), 648–652 (1996)
12. Li, J., Allinson, N.M.: A comprehensive review of current local features for computer vision. *Neurocomputing* 71(10–12), 1771–1787 (2008)
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10), 1615–1630 (2005)
15. Pratt, W.: *Digital image processing*. John Wiley & Sons (2007)
16. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
17. Rosten, E., Porter, R., Drummond, T.: Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(1), 105–119 (2010)
18. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (November 2011)
19. Strecha, C., Bronstein, A., Bronstein, M., Fua, P.: LDAHash: Improved matching with smaller descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(1), 66–78 (2012)
20. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. *Found. Trends. Comput. Graph. Vis.* 3(3), 177–280 (2008)