International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Fusion of Image Feature Descriptors for Person Re-identification

Sathish P.K, S.Balaji

Abstract; Person re-identification has gained a lot of research interest in recent years. Extracting and matching features play an important role in this scenario. Past studies of image feature detectors and descriptors are more generic in nature. Different types of detectors and descriptors are used for person re-identification over the last few years. Most of these descriptors are a combination of two or more variants of descriptors. This research paper will focus on the comparative analysis and evaluation of various features detectors and descriptors used for image matching with relevance to person re-identification. We also explore how the combination of local and global descriptors can improve the re-identification rate. VIPeR dataset is used for the evaluation of descriptors.

Keywords : Person Re-identification; Feature Descriptors; Video Surveillance; Hybrid Descriptor

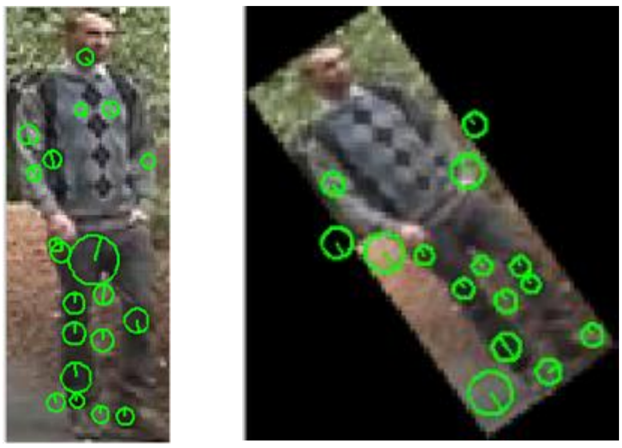
I. INTRODUCTION

The aim of any surveillance video application is to recognize objects such as vehicles and humans in these videos. Person re-identification plays an important **role** in surveillance. Person re-identification involves checking whether a given person has entered a surveillance network. To recognize an object or a person in an image, matching of various points across the query image and the target image is necessary. Image features are used to define image regions which have distinctive properties in an image. Feature detectors are used to detect these image regions. These descriptors provide information about the positions of the points in the image along with the details on the shape of their support regions. Feature detectors can be classified according to the extracted regions as corner detectors and blob detectors

The comparison is a difficult task because it is not a simple to define a quantitative measure to compare the descriptors and detectors. Some are good in particular situations compared to others. The data set in which they are applied plays an important role in their analysis. Different implementations can also lead to different results. Repeatability index is a widely used test for the detectors. This index measures how the features are repeated in an image in the different transformation of the image.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Sathish P K*, Assistant Professor, Computer Science and Engineering Department, Christ (Deemed to be University), Kengeri Campus – 560074.
S. Balaji, Centre for Incubation, Innovation, Research and Consultancy, Jyothi Institute of Technology, Bengaluru India.

Matching score is another measure used for comparison of feature descriptors and detectors. The performance of these descriptors and detectors also depend on the dataset used. In image recognition and classification problems, it is useful to extract the important features that are prominent in the image. Then the question arises that how can one determine the important features in an image? In simple words, a feature is something which is different from the surrounded pixels in an image. This difference could be the change of color, intensity or texture. Figure 1 (a). shows the interest points found in a surveillance image. The features extracted from the image can be used only if they are invariant to transformations such as scaling, rotation, and translation as in Figure 1 (b). Feature descriptor invariant to illumination is discussed in [1]. Multiple features are used form a complex descriptor in [2].



a. b.

Fig. 1 (a) Interest Points Detected. (b) Invariant Features After Rotation.

A. Person re-identification

In order to identify a person in a gallery of captured images, establishing a correspondence between the query image and gallery images are necessary. In real life, person re-identification is an open set problem where the number of images to which the query image has to be compared grows over time. To establish the performance of various detector and descriptors closed set person re-identification is considered in this paper. In a closed set person re-identification, the galleries are of fixed size. Figure 2 shows

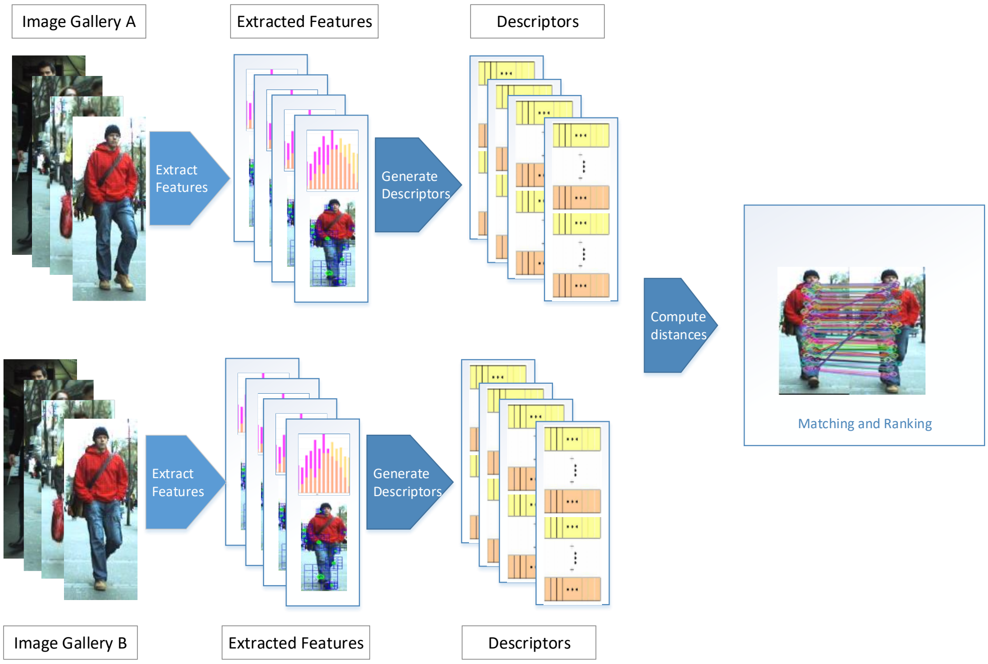


Figure 2. Person Re-identification

a general re-identification setup used in most of the literature. Extracting the correct set of features are important to obtain an accurate match between the query image and the target image.

This paper focuses on evaluating single feature detectors and descriptors for Person re-identification and also explores the possibilities of increasing the recognition rates by combining descriptors. Unlike most of the reviews on feature detectors and descriptors, we bring out the difference in different types of descriptor for person re-identification

II. IMAGE FEATURE DETECTORS

The image feature detectors are used to get the interest points in the image. These points can be used to obtain the feature descriptors. In this paper, we explore few of the most popular feature detectors. An image of the pedestrian from the ETHZ dataset is used to demonstrate the set of key points extracted by all the detectors (Figure. 3).

A. Harris Detector

Regions of the image with significant variation in intensity in all directions can detected using a corner detector. The Harris corner detector [11], Frostener detector [7] are good examples of corner detectors. Hessian matrix is also used as corner detector in [4].

The Harris detector uses auto-correlation matrix for feature detection. The gradient distribution of the local region is represented using the matrix Eq. (1).

$$M=\sigma_D^2 g(\sigma_I)*\begin{bmatrix} I_x^2(x,\sigma_D) & I_x(x,\sigma_D)I_y(x,\sigma_D) \\ I_x(x,\sigma_D)I_{y(x,\sigma_D)} & I_y^2(x,\sigma_D) \end{bmatrix} \quad (1)$$

where

$$I_x(x,\sigma_D) = \frac{\partial}{\partial x} g(\sigma_D) * I(x) \quad (2)$$

$$g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

σ_D is the scale Gaussian kernel computed over the image regions. The interest points detected in the ETHZ dataset [19] is shown in Figure 3 (b).

B. Hessian Detector

$$H = \begin{bmatrix} I_{xx}(x,\sigma_D) & I_{xy}(x,\sigma_D) \\ I_{xy}(x,\sigma_D) & I_{yy}(x,\sigma_D) \end{bmatrix} \quad (4)$$

Eq. (4) is an expansion of the Taylor expansion of the intensity function I(x) [13]. The result of key points extracted using Hessian detector is shown in Figure 3 (a).

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

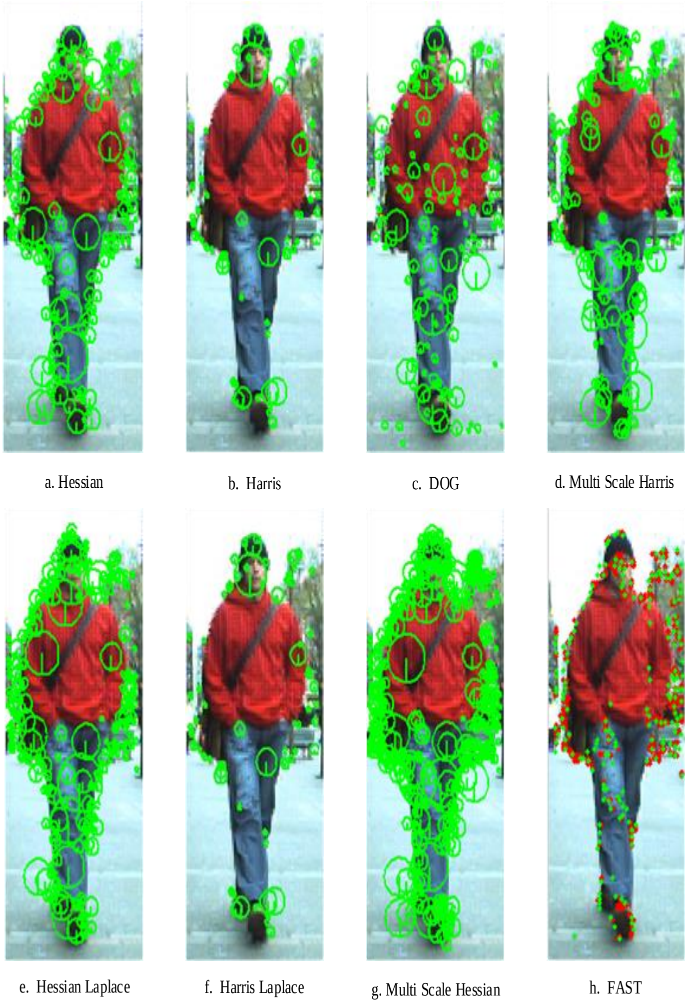


Figure. 3. Image feature detectors

C. Harris Laplace Detector

Harris- Laplace regions is proposed by Mikolajczyk and Schmid [15]. The interest points are detected by the Harris functions and are selected in scale-space using Laplacian of Gaussian operator. This detector is invariant to rotation and scaling. Fig. 3 (f) shows the interest points extracted using this detector.

D. Difference of Gaussians (DoG)

The Difference of Gaussians is a scale-invariant image feature detector. This is discussed in [6, 8, 10, 13, 14] . The $D(x,\sigma)$ maps are obtained by combining the pairwise sets of Gaussian smoothened versions of the image (Eq. 5). Local Maximum is considered over the scale and space of the image. The DoG detector based extraction is depicted in Fig. 3 (c).

$$D(x,\sigma) = \big(G(x,k\sigma) - G(x,\sigma)\big) * I(x) \qquad (5)$$

E. Hessian Laplace Detector

Hessian Laplace detector works similar to Harris Laplace detector. The result of this detector is shown in Figure. 3
(e). It can be observed that it returns more interest points than Harris Laplace detector.

F. Multi-Scale Detectors

Mutli-Scale detectors extract key points at different scales and all the extracted points are combined together. This mechanism helps in dealing with the scale changes in images. Fig. 3(d) and 3(g) shows two multiscale detectors such as Multiscale Harris detector and Multiscale Hessian detector.

G. Features from Accelerated Segment Test (FAST) detector

Rosten and Drummond have proposed FAST detector in [17,18]. The detector compares pixels on a circle of fixed radius around the candidate pixel and classifies it as a corner if there are n pixels with a higher or lower intensity than the selected pixel. Initially, pixel p selected and is compared with neighboring pixels 1,5, 9 and 13 as shown in Figure 4. If the intensity of p is greater or lesser than any three of the pixels compared then p can be classified as a corner or else the test is continued with the other pixels. The application of FAST on EHTZ dataset is shown in Figure 3 (h).

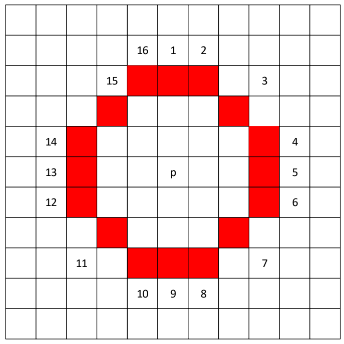


Figure 4 [17][18] FAST detector.

III. DESCRIPTORS

A. Scale Invariant Feature Transform (SIFT) descriptor

Scale Invariant Feature Transform is an algorithm proposed by David Lowe in 2004 [13]. This descriptor is designed to extract features from grayscale images. The key points obtained are invariant to scaling and rotation. They are extracted using Difference of Gaussian method from a scale-space represented the image. A set of scale space images are obtained by convolving with Gaussians and the Difference of Gaussians are found by subtracting adjacent images. The key points are calculated by the maximum or minimum of the DoG (x,y,σ) The magnitude and orientation of the gradient are represented as in Eq. (6).

$$m(x,y) = \sqrt{\big(L(x+1,y) - L(x-1,y)\big)^2 + \big(L(x,y+1) - L(x,y-1)\big)^2} \qquad (6)$$

$$\theta(x,y) = tan^{-1}\Big(\big(L(x,y+1) - L(x,y-1)\big) / \big(L(x+1,y) - L(x-1,y)\big)\Big)$$

where m(x, y) is the gradient magnitude and θ(x, y) is the orientation.A histogram of the gradient orientations of 36 bins is formed around the region covering the key point. The gradient and magnitude around these points are calculated. The histogram of the gradients is concatenated as a vector of 128 entries to form the descriptor. The Figure 5 shows the SIFT descriptors obtained for the EHTZ dataset image showing different orientations.

B. Dense SIFT descriptor

Improved classification results are observed by computing SIFT descriptors over dense grids. Calculating a larger set of image descriptors provide more information about the image. This algorithm was proposed by Bosch et al.[5]. The fast implementation of dense SIFT is discussed in [21] Figure 5 (b). shows the dense SIFT key points extracted.

C. Pyramidal Histogram of Visual Words (PHOW) Descriptor

Pyramidal Histogram of Visual Words (PHOW) descriptor is a SIFT-based descriptor extracted at different scales. This feature makes it invariant to scaling. Figure 5 (c). shows the interest points extracted for PHOW descriptors.

D. Speeded-up robust features (SURF) Descriptor

The SURF descriptor [3] uses the similar approach SIFT. The Scale-space levels and octaves are used in SURF as in SIFT. Interest points are obtained using Hessian matrix based detector (Eq. 7).

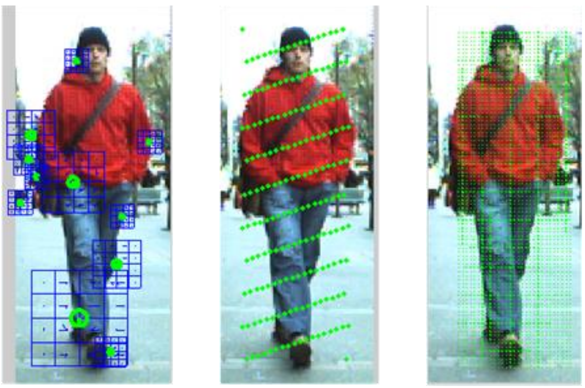
$$H(x,\sigma) = \begin{bmatrix} Lxx(x,\sigma) & Lxy(x,\sigma) \\ Lxy(x,\sigma) & Lyy(x,\sigma) \end{bmatrix} \tag{7}$$

where

$$L_{xx}(x,\sigma) = I(x) * \frac{\partial^2}{\partial x^2} g(\sigma) \tag{8}$$

$$L_{xy}(x,\sigma) = I(x) * \frac{\partial^2}{\partial xy} g(\sigma) \tag{9}$$

Lxy(x, σ) in the Eq. (8) is the convolution of the given image using Gaussian second derivate. The computational cost of Gaussian derivatives of second order is reduced using integral images. SURF employs upscaling of filters.



a. b. c.

Figure 5. SIFT and SIFT-based descriptors computed for EHTZ dataset image.

E. Binary Descriptors

- Binary Robust Independent Elementary Features (BRIEF) descriptor

BRIEF classifies the image patches by using intensity comparisons between pairs of pixels [6]. The resultant bits are stored as a string. A test τ on patch p of S x S size is given by Eq. (10).

$$\tau(p;x,y) := \begin{cases} 1 & if\ I(p,x) < I(p,y) \\ 0 & otherwise \end{cases} \tag{10}$$

the pixel intensity is given by I(p,x) of p at x = (u,v). which is a bit string given by Eq. (11).

$$\sum_{1\le i\le n_d} 2^{i-1}\,\tau(p;x_i,y_i) \tag{11}$$

- Binary Robust Invariant Scalable Keypoints (BRISK) descriptor

The BRISK descriptor is formed as a binary string which is derived from comparing the pixel intensities between pairs of pixels [12]. The pixel pairs are selected according to a predefined pattern, unlike BRIEF. The local gradient g (pi, pj) can be found by the following equation (Eq. 12).

$$g(p_i,p_j) = (p_j - p_i).\frac{I(p_j,\sigma_j)-I(p_i,\sigma_i)}{\|p_j - p_i\|^2} \tag{12}$$

Where I(pi-σi) and I(pj-σj) are smoothed intensity values of the sampling pairs(pi, pj). Each bit b in the descriptor is derived as follows

$$b = \begin{cases} 1 & I(p_j^{\bar{\alpha}},\sigma_j) > I(p_i^{\alpha},\sigma_i) \\ 0 & otherwise \end{cases} \qquad \forall (p_i^{\alpha}, p_j^{\alpha} \in S \tag{13}$$

Where α =arctan2(gx,gy)through which the sampling pattern is rotated around the keypoint k.

- Oriented FAST and Rotated BRIEF (ORB) descriptor

ORB (Oriented FAST and Rotated BRIEF)[34] is a BRIEF based descriptor. ORB uses a modified version of the FAST detector to detect the key points called as FAST-9 using a circular radius of 9.FAST is detected at various levels of the scale pyramid using Harris filter.

ORB uses intensity centroid [16] to find the corner orientation measure. The moments of patch is found by using the equation

$$m_{pq} = \sum_{x,y} x^p\, y^q I(x,y) \tag{14}$$

These moments can be used to find the centroid.

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right) \tag{15}$$

A vector is constructed from the center of the corner to the centroid. The orientation of the patch is given by θ = atan2(m01,m10). ORB descriptor is obtained by using learned sampling pairs.

- Fast Retina Keypoint (FREAK) descriptor

In Fast Retina Keypoint (FREAK) descriptor a retinal sampling pattern is used to compute the binary descriptor by the comparing the image intensities [1]. The sampling pairs are optimized using learning methods similar to ORB. The orientation is taken care similar to

BRISK. Matching of descriptors uses a coarse-to-fine approach in order to speed up the matching process. Figure 6 shows the matching of various binary descriptors.

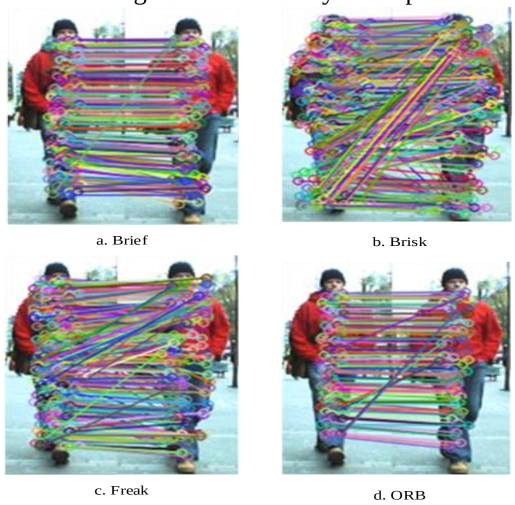


Figure 6. Binary descriptors and matching images

III. PERFORMANCE EVALUATION

A. Datasets

VIPeR Dataset[9] and ETHZ dataset is used for performance evaluation of the detector and descriptors. Sample images of VIPeR Dataset are shown in Figure 7 and ETHZ dataset is shown in Fig. 8.The VIPeR dataset contains 632 image pairs of pedestrians in the size of 128 x48 pixels. The images are stored in separate folders CAM A and CAM B. The folders contain images of the same person with different viewpoints, poses, and change in illumination.

Figure 7. VIPeR dataset. The top row shows the images in CAMA folder and the bottom row shows the images of CAMB folder.

ETHZ dataset contains 8,580 images of 146 pedestrians organized as SEQ1, SEQ2, and SEQ3. Each sequence contains images with a change in viewpoint, illumination, pose and occlusion. SEQ3 is used in our evaluation experiments.

Figure 8. ETHZ dataset. Sample images from SEQ1 sequence.

B. Experimental Setup

Performance evaluation is carried out on the VIPeR dataset. Two random sets of images are taken. Feature detection is done using the FAST detector for the evaluation of descriptors. Descriptors are computed for each image on the gallery set and query set. A similarity matrix is obtained by computing the Euclidian distance between the descriptor vectors. The experiments are repeated for different descriptors and results are evaluated using Cumulative Matching Characteristic curve (CMC) as mentioned in [9][20]. CMC gives a better comparison of descriptors for person re-identification. The graphs show the CMC computed for a dataset of 30 image pairs selected from the gallery set and the query dataset.

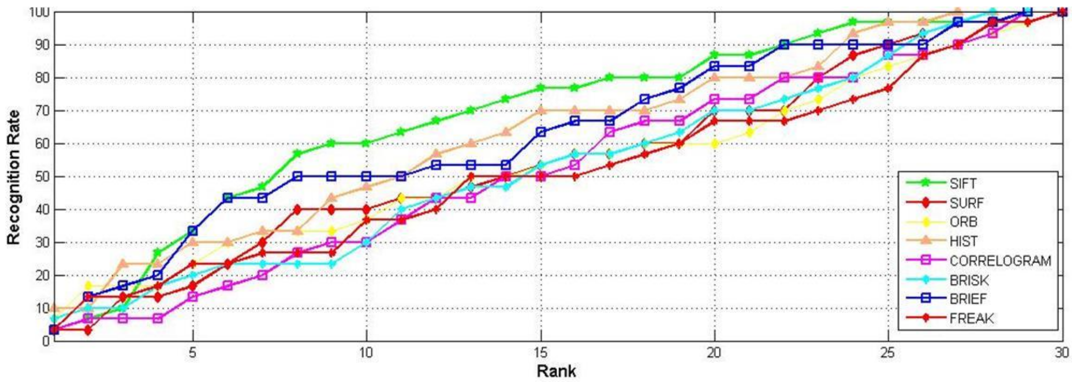


Figure 9. CMC on VIPeR dataset

The CMC shows that SIFT descriptor performs slightly better than the rest of the descriptors for a small random set of images of the above-mentioned dataset (Figure 9). The Figure 10 shows the performance of the descriptors when combined with histograms of the images. The hybrid descriptors which use local and global features perform better than single image descriptors in re-identification.

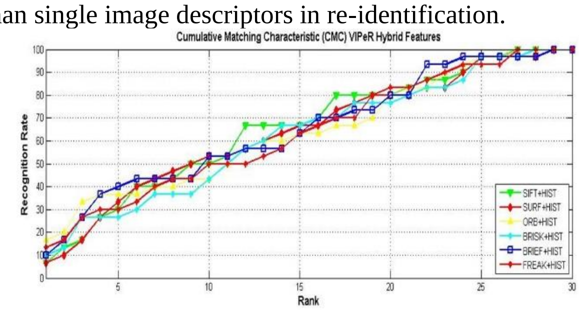


Figure 10 CMC curve VIPeR dataset for Hybrid descriptors

V. CONCLUSION

This work analyses different keypoint detectors and feature descriptors for the person re-identification datasets for video surveillance. The CMC graph shows that single feature descriptors recognition rate at rank 1 is comparatively less. To improve the accuracy of the matching different descriptors can be combined together which shows considerable improvement. This work does not cover the various types of color descriptors which can be considered an extension of this work. The nearest neighbor classifier used in this work uses only Euclidian distance metric for computing the similarity matrix. Other distance metrics such as cosine similarity, Mahalanobis, Bhattacharya distances can also be studied.

REFERENCES

1. Alahi, Alexandre, Raphael Ortiz, and Pierre Vandergheynst: “Freak, Fast retina keypoint”, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
2. Bąk, S., Corvee, E., Brémond, F., & Thonnat, M, “Person re-identification using Haar-based and DCD-based signature”, Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010, 1–8. http://doi.org/10.1109/AVSS.2010.68. 2010
3. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF). Computer Vision and Image Understanding”, 110(3):346–359, 2008
4. P.R. Beaudet, “Rotationally invariant image operators”, In International Joint Conference on Pattern Recognition, pages 579–583, 1978.
5. Bosch, Anna; Zisserman, Andrew, and Munoz, Xavier, “Scene Classification via pLSA.Proc”, 9th European Conference on Computer Vision (ECCV'06), Springer Lecture Notes in Computer Science 3954: 517~530, 2006.
6. J. L. Crowley and A. C. Parker, “A representation for shape based on peaks and ridges in the difference of low pass transform”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 2, pp. 156–170, 1984.
7. M. A. Föstner and E. Gülch, “A fast operator for detection and precise location of distinct points, corners and centers of circular feature”, In ISPRS Intercommission Workshop, 1987.
8. P. Gaussier and J. P. Cocquerez, “Neural networks for complex scene recognition: Simulation of a visual system with several cortical areas”, in Proceedings of the International Joint Conference on Neural Networks, vol. 3, pp. 233–259,1992.
9. D. Gray, S. Brennan, and H. Tao, “Evaluating appearance models for recognition, reacquisition, and tracking”,In PETS, 2007.
10. S. Grossberg, E. Mingolla, and D. Todorovic, “A neural network architecture for preattentive vision”, IEEE Transactions on Biomedical Engineering, vol. 36, pp. 65–84, 1989.
11. C. Harris and M. Stephens, “A combined corner and edge detector”, In Alvey Vision Conference, pages 147–151, 1988.
12. Leutenegger, Stefan, Margarita Chli, and Roland Y. Siegwart, “BRISK: Binary robust invariant scalable keypoints”, Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
13. D.G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, 60(2):91–110, 2004.
14. D. G. Lowe, “Object recognition from local scale-invariant features”, in Proceedings of the International Conference on Computer Vision, pp. 1150–1157, 1999.
15. K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors”, International Journal of Computer Vision, 63–86, 2004.
16. Rosin, Paul L. “Measuring corner properties. Computer Vision and Image Understanding”, 73.2): 291-307, 1999.
17. E. Rosten and T. Drummond, “Machine learning for high-speed corner detection”, in Proceedings of the European Conference on Computer Vision, pp. 430–443, 2006.
18. E. Rosten and T. Drummond, “Fusing points and lines for high-performance tracking”, in Proceedings of the International Conference on Computer Vision, pp. 1508–1511, 2005.
19. chwartz, W., Davis, L.: Learning discriminative appearance-based models using partial least squares. In: Proc. XXII SIBGRAPI , 2009
20. Štruc V., Pavešic, N., “Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition”, Informatica (Vilnius), vol. 20, no. 1, pp. 115-138, 2009.
21. A. Vedaldi, B. Fulkerson, “Vlfeat — an open and portable library of computer vision algorithms”, in: Proceedings of the 18th Annual ACM International Conference on Multimedia, Firenze, Italy,25–29 October 2010, pages 1469–1472, 2010.

AUTHORS PROFILE

Sathish P K has completed his Ph.D. in Computer Science and Engineering from Visvesvaraya Technological University in 2019. He is working as Assistant Professor in Christ (Deemed to be University),Bengaluru,India. His areas of interest are pattern recognition, computer vision and machine learning.

S Balaji received his Ph.D. in Computer Science and Engineering from Department of Electrical Engineering, Indian Institute of Science, Bengaluru in 1993. Currently he is working as Professor of Computer Science and Engineering, Centre for Incubation, Innovation, Research and Consultancy, Jyothi Insititute of Techhnology, Thataguni, Begaluru Rural-560082, India.