State-of-the-Art: Transformation Invariant Descriptors

Asha S, Sreeraj M

Abstract— As the popularity of digital videos increases, a large number illegal videos are being generated and getting published. Video copies are generated by performing various sorts of transformations on the original video data. For effectively identifying such illegal videos, the image features that are invariant to various transformations must be extracted for performing similarity matching. An image feature can be its local feature or global feature. Among them, local features are powerful and have been applied in a wide variety of computer vision applications. This extensive use of local features is due to its efficient discriminative power in extracting image contents. In this paper, we focus on various recently proposed local detectors and descriptors that are invariant to a number of image transformations. This paper also compares the performance of various local feature descriptors under different transformation scenarios.

Index Terms— Visual Matching, Interest points, Global features, Local features, Feature Detectors, Feature Descriptor , Transformation

1 Introduction

NOWADAYS, the popularity of digital video is increasing in a faster rate. Each day, tens and thousands of video data are being generated. As the popularity increases, chances for the existence of pirated videos also increase. A video can undergo numerous changes so called transformations and get converted to pirated copies. So, an efficient mechanism for detecting such video copies becomes a necessity. Video copies are detected by performing visual matching between the pirated videos with its original video content.

Visual Matching is one of the major steps in various computer vision applications like image retrieval, image registration, object recognition, object categorization, texture classification, robot localization, wide baseline matching, and video shot retrieval. Most of the existing visual matching algorithms consider image contents for similarity checking. Image content can be represented using either local features or global features. Global features represent an image by considering the overall composition of the image. The most commonly used global features are color histogram, edges and texture. Local feature on the other hand, represents local patches in an image.

Any point, edge or a region segment that differs from its immediate neighbourhood is considered as a local patch. An ideal local feature would be a point having a location in space with no spatial extent. To localize the identified features in an image, the local neighbourhood of pixels needs to be analysed, thus giving all local features some implicit spatial extent. This local neighbourhood of pixels describes the interest points. Figure 1 illustrates some of identified interest points in an image.

A local feature vector corresponding to an image is obtained using both a feature detector and a feature descriptor. The local features are extracted and utilized in image processing applications through the following three stages: (i) feature detection (ii) feature description (iii) feature matching or clustering. Feature Detection is a method that detects or extracts features from the referenced image. These features are then described by considering a quantum of neighbourhood pixels, and finally they form the feature descriptor. During feature matching, the extracted feature vectors of images are compared for similarity checking. This stage also employs various distance metrics for similarity analysis.

A good local feature should possess properties like repeatability, distinctiveness, locality, quantity, accuracy and efficiency. The local features also provide a higher degree of invariance to changes in various viewing conditions, occlusions and cluttering.

This paper focuses on various local feature detectors and descriptors and is organized as follows. Various feature detectors are familiarized in Section 2. Section 3 describes the most promising feature descriptors. Section 4 evaluates the performance of descriptors under various transformation scenarios. Section 5 concludes the paper by briefly discussing the local features.

2 Feature Detection

In this section, we focus on how the features are detected. The initial step in describing a local feature is the feature detection stage. This stage identifies features like points, edges and regions for obtaining the feature descriptor. One of the desirable properties that a feature detector should hold is repeatability: deals with detecting same features in different images taken in the same scene. The existing detectors for extracting features can be classified into three types: (a) edge detectors (b) corner detectors (c) region detectors.

2.1 Edge Detectors

One of the most stable image features is the Edges. An Edge has characteristics that, at the edge, the image brightness changes sharply. Hence, even under viewpoints, scales and

- *Asha S is currently pursuing M.Tech in Computer Science and Information Systems from Federal Institute of Science and Technology (FISAT), M G University, India, PH-+91 99474 48628. E-mail:ashaabhilash123@gmail.com*
- *Sreeraj M is currently working as Assistant Professor Federal Institute of Science and Technology (FISAT)), M G University, India. E-mail: sreerajtkzy@gmail.com*

International Journal of Scientific & Engineering Research Volume 4, Issue [illegible]2013
ISSN 2229-5518

illumination changes, edges can be detected easily. Edge detectors applied on an image results in a set of connected curves. These connected curves can be boundaries of objects, surface markings or curves.

2.2 Corner Detectors

A corner in an image is the region where more than one edges intersects. Hence, to detect corners, consider the local neighborhood of the corners. Those corners are selected as interest points that have their edges with dominant and different directions. From the detected interest points, real corners are determined by performing additional local analysis. Corner Detectors [1]; detect points that correspond to points in an image with high curvature.

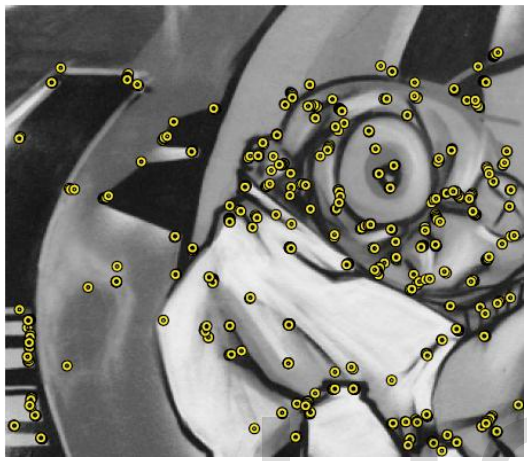

Fig. 1. Interest Points identified using corner detector [1]

2.3 Region Detectors

Region Detection also called Blob Detection, deals with detecting regions of interest in a digital image. These detected regions possess properties like brightness, color, sharpness different from its surrounding regions.

One of the earliest feature detectors is the Moravec's corner detector [4] that tests the intensity value associated with each pixel in the image and finds the local maximum of minimum intensity changes. Moravec's detector is modified by Harris and Stephens in [2] by considering the differential of the corner score with respect to direction and named it as the Harris corner detector. However, it cannot deal with scaling changes. A simple and efficient detector is proposed in [3] called the SUSAN detector. It computes the fraction of pixels within a neighborhood which have similar intensity to the center pixel. Corners can then be localized by thresholding this measure and selecting local minima. A similar idea was explored in [5] where pixels on a circle are considered and compared to the center of a patch.

Lowe [9] introduced a new local feature, called SIFT (Scale Invariant Feature Transform) which is invariant to scaling changes. It uses a detector to find the local maxima from the difference of Gaussian (DoG) values taken at successive scales. Another scale invariant detector is proposed by Mikolajczyk and Schmid [10] is the Harris–Laplace / Affine detector that could deal with both scaling and affine transformations. It combines the Harris corner detector and the Laplace function for characteristic scale selection. Hessian-Laplace /Affine detector selects interest points based on Hessian matrix at a point. An intensity based feature detector was pioneered by Tuytelaars and Van Gool in [23]. A method developed by Matas et al. [11] to find image correspondence is the Maximally Stable Extremal Region (MSER) detector. Table 1 shows the classification of various feature detectors.

TABLE 1
CLASSIFICATION OF FEATURE DETECTORS

Feature Detector	Edge	Corner	Region
Moravec's corner detector [4]		X	
Harris corner detector [2]	X	X	
SUSAN [3]	X	X	
Trajkovic operator [5]		X	
High-speed corner detector [22]		X	
Förstner corner detector [6]		X	
Canny[7]	X		
Sobel	X		
Harris–Laplace / Affine	X	X	
Hessian–Laplace / Affine	X	X	
Laplacian of Gaussian		X	X
Difference of Gaussians[9]		X	X
Determinant of Hessian [19]		X	X
MSER [11]			X
PCBR			X
Grey-level blobs			X

3 Feature Description

Feature description follows feature detection. For describing the identified features, a local patch around the detected features is extracted. This extracted vector is referred as the feature vector or the commonly called feature descriptor. Feature descriptors are mainly classified into the following four types.

(i) Distribution based descriptors: these descriptor uses histogram representation to characterize the recognized interest points. In image processing applications, the performance of these descriptors are found to be better compared to other descriptors.

(ii) Spatial-Frequency based descriptors: Here the detected

International Journal of Scientific & Engineering Research Volume 4, Issue [illegible]2013
ISSN 2229-5518

interest points are described by considering the frequency contents of the image.

(iii) Differential descriptors: In these descriptors, the local neighbourhood of a point is approximated by computing the image derivatives up to a given order.

(iv) Others: This descriptor classification includes descriptors other than the above mentioned three types. This category includes Moment based feature description, Phase based feature description, and Color- based description.

The various commonly used local feature descriptors for computer vision applications are explained briefly in the following subsections

3.1 SIFT

David Lowe in [9] proposed the SIFT (Scale Invariant Feature Transform) descriptor. SIFT extracts distinctive invariant image features. These features are invariant to various image transformations like scaling, rotation, affine- distortion, noise and illumination changes.

To extract the SIFT feature descriptor for an image, the reference image is initially convoluted with Gaussian filters at different scales. Then DoG values at successive scales are taken and the maximum and minimum DoG values are found. The points having maximum and minimum DoG values are referred as key points. From the identified candidate key points, points with low contrast are removed. For each of the remaining key points, orientation and gradient magnitude of the location are computed. Finally, the feature descriptor is obtained by considering the image gradients of the local neighbourhood of key points. These are then transformed into a 128 dimensional vector representation.

SIFT feature descriptors are powerful with high distinctiveness. It is relatively easy to extract and allows efficient object identification with low probability of mismatch. The high dimensionality of this local feature descriptor is an issue.

3.2 PCA-SIFT

This is an improved SIFT descriptor proposed by Ke and Sukthankar in [12]. The PCA-SIFT descriptor is more compact, highly distinctive and as robust as SIFT. PCA-SIFT, like SIFT encode the characteristics of image by considering the feature point's neighbourhood. Here, SIFT descriptor is improved by applying Principal Component Analysis for dimensionality reduction.

In PCA-SIFT, for a given image, the key points are detected using SIFT detector. Centered on these key points a 41 x 41 patch at the given scale and dominant orientation is extracted. From these patches the feature vector is computed by concatenating both the horizontal and vertical gradient maps. Thus the feature vector makes up to a total of 2x39x39=3042 elements. Then it is further normalized to unit magnitude. PCA is applied to the feature vector and a 36 dimension compact feature description is generated.

Like SIFT, PCA-SIFT descriptor generation is simple; the descriptor is compact, faster and more accurate than the standard SIFT descriptor. It requires less storage space and minimal retrieval time. Due to dimensionality reduction, it loses some discriminative information.

3.3 GLOH

Gradient Location Orientation Histogram (GLOH) [13] is proposed as an extension of the standard SIFT descriptor. It is a 64 dimensional descriptor. Unlike SIFT, in GLOH histogram representation considers more spatial regions. The GLOH feature descriptor is constructed using Histogram of location and Orientation of pixels in a window around the interest point. In GLOH, SIFT descriptor is computed in log polar co-ordinate system with three bins in radial directions (using three different radii) and three in angular direction. Thus a total of 17 location bins are considered. It also takes into account gradient orientations that are quantized in 16 bins. Finally provides with a 102 dimensional descriptor. Like PCA-SIFT, GLOH also employs PCA for performing dimensionality reduction.

GLOH is more distinct, robust and faster than SIFT. This descriptor also results in information loss due to dimensionality reduction.

3.4 MI-SIFT

Mirror and Inversion invariant SIFT (MI-SIFT) descriptor was proposed by R.Ma in [14]. This descriptor improves the SIFT descriptor by enhancing the invariance to mirror reflections and grayscale inversions. Mirror reflection and Inversion invariance is achieved by combining the SIFT histogram bins of both the image and its mirror reflected image. This approach provides a unified descriptor for the original, mirror reflected and grayscale inverted images with an additional cost of computation.

MI-SIFT perform better than SIFT in finding transformations including mirror reflection and inversions. It also achieves comparable performance in image matching with ordinary images.

3.5 F-SIFT

Flip invariant SIFT (F-SIFT) introduced by Zhao and Ngo in [15] is a SIFT descriptor tolerant to image flips, while preserving the properties of SIFT. They considered flip as a decomposition of flip along a predefined axis followed by a rotation. F-SIFT perform a selective flipping on the image regions based on the dominant curl associated with the regions. Flip invariant descriptor is obtained by first rotating the region patches to its dominant orientation and estimating the Curl associated with the regions. The direction of the curl indicates the direction of rotation. If the curl is negative it denotes that the region patch is flipped and is needed to be rotated anti clockwise. Then SIFT descriptors are extracted from these normalized regions. In F-SIFT, the regions are normalized geometrically by flipping horizontally or vertically and complementing their dominant orientations.

This descriptor shows similar performance as SIFT. For flipped images F-SIFT out performs SIFT. F-SIFT descriptor computation involves high computational overhead. The extraction of F-SIFT descriptors is approximately one third slower than SIFT. Performance degrades if there are errors during finding the curl.

3.6 FIND

X. Guo [16] introduced a novel flip invariant descriptor through a new cell ordering scheme. As like SIFT, it also

International Journal of Scientific & Engineering Research Volume 4, Issue [illegible]2013
ISSN 2229-5518

adopts the DoG detector. FIND employs an overlap – extension strategy to obtain the feature descriptor. For each detected key points, it reads the 8 directional gradient histograms by following an 'S 'order. Here, for a given image, the descriptor obtained before and after a flip operation are mirror of each other.

FIND exhibits stable performance in both flip and non-flip case. FIND is also tolerant to scaling, rotation and affine transformations.

3.7 RIFT

RIFT (Rotation invariant SIFT) descriptor is a rotation-invariant generalization of SIFT proposed by S. Lazebnik in [17]. One of the earliest flip invariant SIFT descriptor. This descriptor is somewhat sensitive to scale changes and is less discriminative than original SIFT.A circular normalized patch is considered around each detected interest points. This circular patch is divided into concentric rings of equal width. And within each ring, from each division a 8 directional gradient orientation histogram are computed. The concatenation of gradient orientation histogram about all the identified key points yields the RIFT descriptor. To preserve the rotation invariance property, at each key point, the orientation is measured relative to the direction pointing outward from the center. RIFT uses four rings and eight histogram orientations, yielding 32-dimensional descriptors.

3.8 SPIN

SPIN is also proposed by S. Lazebnik in [17]. It is also a flip invariant descriptor. SPIN preserves flip invariance property by taking into account the spatial information. Here, it encodes a region using 2D histogram of pixel intensity and also considers the distance from region center.

3.9 MIFT

MIFT (Mirror reflection Invariant Feature Transform) is a framework for providing feature descriptor which is robust to transformations including mirror reflections. MIFT [18] is a local feature descriptor providing mirror reflection invariance while preserving existing merits of SIFT descriptor. A Mirror reflected version of an image can be obtained by reversing the axis of the image, hence, in a horizontally reflected image the row order of pixels remains the same, but the column order changes. So in MIFT mirror invariance is achieved by simple descriptor reorganization. This descriptor reorganization first organises the arrangement of cell order around the interest points. This is done by checking the values of total left pointing (ml) and right pointing (mr) orientations. Based on the winning orientation, column order may change (ml>mr) or not. Second, for each cell, it checks whether the order of orientation bins to follow clockwise or anticlockwise direction. Thus MIFT provides a descriptor which is identical in all cases of mirror reflections.

MIFT is one of the promising local feature detector based on SIFT, that is robust to mirror reflection also. Since MIFT is based on SIFT, high dimensionality of the descriptor is a curse. It also requires longer computational time compared to SIFT.

3.10 SURF

H. Bay et al. in 2006 pioneered a robust local feature descriptor called SURF [19] (Speeded Up Robust Features). SURF has proven its efficiency in various computer vision tasks. SURF descriptors are more commonly used for applications like object recognition and 3D reconstruction. The extraction of such a robust descriptor is inspired from the efficient SIFT [9] descriptor. SURF is proven to be several times faster than SIFT. The SURF feature descriptor is 64 dimensional and is robust against various image transformations as compared to SIFT. SURF makes efficient use of integral images. This integral representation aids in calculating the integer approximation to the determinant of Hessian blob detector. SURF is based on sums of 2D Haar wavelet responses. For features, it uses the sum of the Haar wavelet response around the point of interest; the integral image computation eases this step.

SURF is more robust to image transformations like rotation, scaling and noise. SURF is faster than SIFT.

3.11 BRIEF

BRIEF [21] is a general-purpose feature point descriptor that can be combined with arbitrary detectors. It is robust to typical classes of photometric and geometric image transformations. BRIEF is targeting real-time applications. It is highly discriminative even when using relatively few bits and can be computed using simple intensity difference tests.

3.12 ORB

ORB (Oriented and Rotated BRIEF) is an efficient, fast binary descriptor based on BRIEF descriptor [21]. It is introduced by E. Rublee, V. R Rabaud, K. Konolige, G. Bradski in [24]. ORB feature descriptor achieves rotation invariance and noise resistance. ORB is a 32 bit binary feature descriptor. In ORB, for detecting key points, it uses the FAST key point detector [20]. ORB algorithm uses FAST in pyramids to detect stable key points, selects the strongest features using FAST or Harris response, then computes the descriptors using BRIEF.

Binary descriptors are a very efficient feature descriptor for time-constrained applications. ORB also provides good matching accuracy. Another advantage of binary descriptors are very fast extraction times and very low memory requirements

4 Performance Comparison

This section tentatively concludes the performance of various feature descriptors based on several studies. The descriptors are evaluated under various image transformation scenarios.

4.1 Rotation

Almost all descriptors show similar performance under various image rotations. None of the descriptors performed well when textured images are considered. Among all, GLOH and SIFT obtained the best results. SURF finds the least matches and gets the least repeatability as compared to SIFT. PCA-SIFT found to be better than SURF under this scenario. ORB found to efficient than SIFT and SURF.

4.2 Viewpoint

Invariance in viewpoint changes can be achieved effectively in

International Journal of Scientific & Engineering Research Volume 4, Issue 7, July-2013
ISSN 2229-5518

textured images than structured images. The SURF descriptor obtained best results for structured images and SIFT performs best in the case of textured images. When the view point change area increases PCA-SIFT performs better.

4.3 Viewpoint

Under various scale changes, SIFT, SURF and GLOH performs best for both textured and structured images as compared to other descriptors. PCA-SIFT detects only few matches and is not as stable as SIFT and SURF to scale invariance. ORB is not so stable during scale changes.

4.4 Image blur

As the image blur increases, the performance of all descriptors is degraded significantly. Among them, GLOH and PCA-SIFT obtain the best results; SURF also shows good performance.

4.5 Illumination

Under illumination changes, SURF provides better repeatability. GLOH performs the best, followed by SIFT, for illumination normalized regions. They achieve the most stable and efficient results.

4.6 Processing Time

The factors that influence the processing time are the size, quality of the image and image types. Time is counted for the complete processing of image, which includes feature detection and matching. SURF is the fastest one, SIFT is the slowest but it finds most matches. ORB being a binary descriptor is faster than SURF.

4.7 Mirror Reflections

MIFT outperforms the various mirror invariant version of SIFT algorithms. SURF could only find least match points.

5 Conclusion

Local feature descriptors are found to have several promising properties and capabilities compared to global feature descriptors. Hence, they are widely used in applications like object recognition, object class recognition, texture classification, image retrieval, robust matching and video data mining.

Through this paper we reviewed several detectors like edge, corner, region and intensity based detectors. The key part of feature extraction is the feature description process. Detectors identify the features. These features are described using any of the suitable feature descriptor. Feature descriptors are broadly classified into distribution-based descriptors, differential descriptors, spatial-frequency based descriptors and other descriptors. There are a large number of feature descriptors that are invariant to various image transformations.

However, a large number of feature detectors and descriptors are constantly evolving and they might override the performance of currently efficient descriptors.

References

[1] T. Tuytelaars, K. Mikolajczyk, "Local invariant feature detectors: a survey," Found.Trends Comput. Graph. Vis. 3 (3) (2007) 177–280.

[2] C. Harris and M. Stephens, "A combined corner and edge detector," in Alvey Vision Conference, pp. 147–151, 1988.

[3] S. M. Smith and J. M. Brady, "SUSAN — A new approach to low level image processing," International Journal of Computer Vision, vol. 23, no. 34, pp. 45–78, 1997.

[4] H. Moravec, "Towards automatic visual obstacle avoidance," in: Proceedings of the International Joint Conference on Artificial Intelligence, 1977, p. 584.

[5] M. Trajkovic, M. Hedley, "Fast corner detection," Image Vis. Comput. 16 (2) (1998) 75–87.

[6] Forstner, W. Guich, " A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular features", ISPRS, 1987

[7] J. Canny, "A Computational Approach To Edge Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986

[8] M. Trajkovic and M. Hedley, "Fast corner detection," Image and Vision Computing, vol. 16, no. 2, pp. 75–87, 1998.

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints, "Proc. In Int.J. Comput. Vis. 60 (2) (2004) 91–110.

[10] K. Mikolajczyk, C. Schmid, "Scale & affine invariant interest point detectors," Int.J. Comput. Vis. 60 (1) (2004) 63–86.

[11] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," Proc. in: Proceedings of the British Machine Vision Conference, 2002, pp. 384–393.

[12] Y. Ke, R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," Proc. in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 506–513.

[13] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.

[14] R. Ma, J. Chen, Z. Su, "MI-SIFT: mirror and inversion invariant generalization for SIFT descriptor," Proc. of ACM Int. Conf. on Image and Video Retrieval, 2010, pp. 228–235.

[15] W. L. Zhao and C. Ngo, "Flip-Invariant SIFT for Copy and Object Detection," IEEE Trans. Image Processing, vol. 22, no. 3, MARCH 2013, pp. 980–991.

[16] X. Guo, X. Cao, "FIND: A Neat Flip Invariant Descriptor," Proc. in Conf. on Pattern Recognition, 2010.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[18] X. Guo, X. Cao, J. Zhang, and X. Li, "MIFT: A Mirror Reflection Invariant Feature Descriptor", Springer, ACCV 2009, Part II, LNCS 5995, pp. 536–545, 2010.

[19] H. Bay, T. Tuytelaars, L.V. Gool, " SURF: speeded up robust features," Computer Vision and Image Understanding 2008;110(3): pp 346-359

[20] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection." IEEE Trans. Pattern Analysis and Machine Intelligence, 32:105–119, 2010.

[21] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," Proc. in European Conference on Computer Vision, 2010.

[22] E. Rosten, T. Drummond, "Machine learning for high-speed corner detection," Proc. in European Conference on Computer Vision, vol. 1(1), 2006, pp. 430–443.

[23] T. Tuytelaars, L. Van Gool, "Matching widely separated views based on affine invariant regions," J. Comput. Vis. 59 (1) (2004) 61–85.

[24] E. Rublee, V. R Rabaud, K. Konolige, G. Bradski, "ORB: an efficient alternative to SIFT or SURF," Proc. in European Conference on computer Vision, 2011.