iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 Issue: V Month of publication: May 2014

DOI:

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

A Survey on State-of-Art Flip-Invariant Descriptors

Aysha Mol K S
M. Tech Student, Department of Computer Science
Viswa jyothi College of Engineering & Technology, Vazhakulam, Kerala

Abstract - In image processing, flip operation creates the mirror of an image. Flip is common copyright infringement technique used in creating near-duplicate videos, since it will not change the video content. Hence in a video copy detection system, flip operation need to be identified. Most of the salient region descriptors do not take flip into consideration. This paper compares the performance of existing salient region descriptors over flip operation.

Keywords— Flip, Video copy detection, copyright infringement, descriptor, key frame

I. Introduction

With the advance in multimedia technology, it has become easier to access and store video data of huge volume. Many of the videos stored in a video database are near-duplicate copies of an original video. Near-duplicate videos are approximately identical videos with similar appearance, but vary in terms of rotation, scale, photometric variation etc. Hence to avoid copyright infringement of videos, a video copy detection system is needed.

In the existing video copy detection system [10], salient regions from the key frames of videos stored in a database are identified and the descriptors of these salient regions are extracted. These descriptors are invariant to various transformations like rotation, scaling, lighting changes etc. When a new video (reference video) arrives, its key frames are identified and the salient regions are detected and feature descriptors are extracted. Visual matching of the feature-invariant descriptors of reference video and the database videos are performed to detect the copies.

Flip is a common operation used in creating near-duplicate videos. Flip produces the mirror of an image. Flip operation are of two types: horizontal and vertical (Fig. 1). Horizontal flip performs flipping around vertical axis and vertical flip performs flipping around horizontal axis. The main advantage of this operation is that it will not cause a change in the video content, only the direction of information flow will get changed. Hence it is easy to create the copy of a video without much change in content. Hence to identify flip, the feature-invariant descriptor used in a video copy detection system must be invariant to flip transformation. The flip-invariance property of a descriptor depends mainly on its partitioning scheme.

This paper focuses on the partitioning scheme of various flip-invariant descriptors and their performance over flip operation. Various feature descriptors and their performance over flip is explained in session II.

II. Feature descriptors

Salient regions in an image are represented by using a vector, which describes the characteristics of a region. Such a vector representation is called a feature descriptor. Feature descriptors [1] are mainly divided into 4 types:

(a) Distribution-Based Descriptors: Such descriptors use histograms to represent the salient region descriptor. Scale Invariant Feature Transform (SIFT) descriptor is a distribution based descriptor.

(b) Spatial-Frequency Based Descriptors: Such descriptors describe the interest points using frequency contents of an image. Gabor filter based descriptors are examples of spatial-frequency based descriptors.

(c) Differential Descriptors: In such descriptors, the local neighbourhood of a point is approximated by computing the image derivatives up to a given order. Steerable filters are examples of differential filters.

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

(d) Others: This session includes descriptors other than those mentioned above. Various moment-based and color-based descriptors are included in this session.

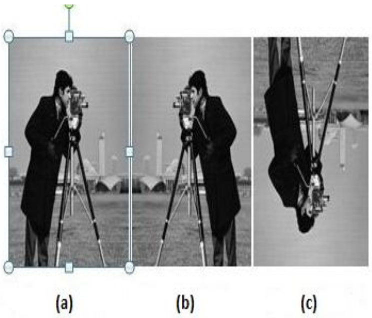

Fig. 1 Flipped images of Camera Man (a) Original Image (b) Horizontal Flip (c) Vertical Flip

Various commonly used flip-invariant descriptors are:

A. SIFT

David Lowe in [7] proposed the SIFT (Scale Invariant Feature Transform) descriptor. SIFT extracts distinctive invariant image features. These features are invariant to various transformations like scaling, rotation, illumination changes etc. To extract SIFT descriptors from an image, reference image is initially convoluted with Gaussian filters at different scales. The 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 considered as key points. From the identified key points, points with low contrast are removed. From the remaining key points, orientation and gradient magnitude of the location are computed. Finally feature descriptor is obtained by considering the image gradients of the local neighbourhood of the key points. The descriptor is having a dimension of 128.

This descriptor is highly powerful and distinctive. Partitioning scheme of SIFT [6] is as follows: SIFT divides a region into 4 × 4 blocks. It then describes each grid with an 8 directional gradient histogram (fig 2). Histograms are concatenated in row major order from left to right. Histogram bins are arranged in clockwise manner. Flip transformation will disorder the placement of blocks and bins. This results in a different version of descriptor due to the predefined order of feature scanning. Since the descriptor is derived from directionally sensitive gradient fields, SIFT is not flip invariant. The SIFT descriptor still seems to be the most appealing descriptor for practical uses, and hence also the most widely used nowadays. It is distinctive and relatively fast, which is crucial for on-line applications.

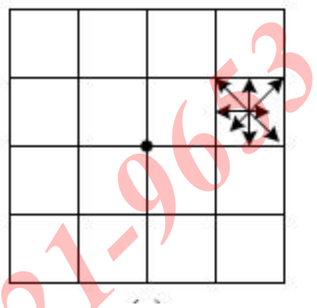

Fig. 2 Partitioning scheme of SIFT

B. SPIN

SPIN [2] make use of a two dimensional histogram. The 2-D histogram encodes the distribution of image brightness values in the neighbourhood of center point. The two dimensions of the histogram are d, the distance from the center point and i, the intensity value. The dimension of SPIN (Fig. 3) descriptor is 50.

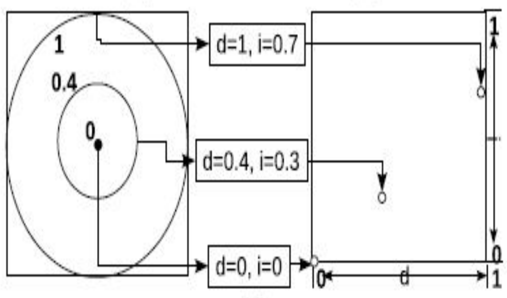

Fig. 3 Partitioning scheme of SPIN

SPIN preserves the flip invariance property by enforcing spatial information. But the empirical evaluation [1] shows that SPIN is outperformed by SIFT.

C. RIFT

It is the rotation invariant generalization of SIFT. Rotation-Invariant Feature Transform (RIFT) [2] will first detect a circular normalized patch within each interest point in an image. Each circular patch is then divided into concentric

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

rings of equal width. A gradient orientation histogram is computed within each ring. The concatenation of the gradient orientation histogram about all the identified key points yields the RIFT descriptor. To maintain rotation invariance, at each key point, this orientation is measured relative to the direction pointing outward from the center.

RIFT (Fig. 4) uses four rings and eight histogram orientations, yielding a 32-dimensional descriptor. The RIFT descriptor is not invariant to flip [2], since flip reverses the order of directions in the orientation histogram. The spatially loose representation [6] also results in RIFT a descriptor not as distinctive as SIFT.

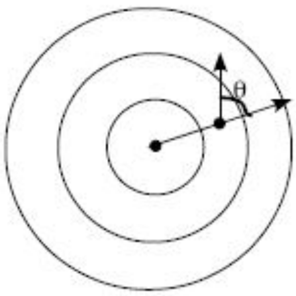

Fig. 4 Partitioning scheme of RIFT

D. GLOH

Gradient Location Orientation Histogram (GLOH) [1] is an extension of SIFT descriptor. It is designed to increase SIFT's robustness and distinctiveness.The GLOH (Fig. 5) 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 direction and three in angular direction. Thus a total of 17 bins are considered. It also takes into account gradient orientations that are quantized into 16 bins. This gives a 272 bin histogram.

GLOH make use of Principal Component Analysis (PCA) for dimensionality reduction. Unlike SIFT, In GLOH, histogram representation considers more spatial regions. The invariance property no longer exists [6] after strengthening the spatial constraint. Also GLOH is computationally more expensive [5].

Fig. 5 Partitioning scheme of GLOH

E. FIND

Flip Invariant Descriptor (FIND) [4] introduced by X.Guo is a novel flip-Invariant descriptor. FIND introduced a new cell ordering scheme. FIND (Fig. 6) make use of a new splitting strategy called Overlap-Extension (O-E) strategy to obtain the feature descriptor. For each detected key point it reads the 8 directional gradient histograms in 'S' order. As a result the descriptor obtained before and after flip operation are mirror of each other.

By making use of this structure, FIND explicitly makes the structure invariant to flip. Comparing to SIFT, FIND reduces 35.94% of a descriptor. FIND utilizes a concise structure with less storage space. But the partitioning scheme of FIND doesn't produce a descriptor as distinctive as SIFT [6].

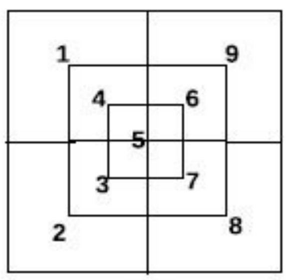

Fig. 6 Partitioning scheme of FIND

F. SURF

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

SURF (Speeded Up Robust Features) [5] is a robust local feature descriptor. SURF is based on sums of 2D Haar wavelet responses. It makes an efficient use of integral images. It uses an integer approximation to the determinant of Hessian blob detector, which can be computed extremely quickly with an integral image. For features, it uses the sum of the Haar wavelet response around the point of interest. Again, these can be computed with the aid of the integral image. The descriptor is 64-dimensional.

SURF is shown to have similar performance as SIFT, while at the same time being much faster. The standard version of SURF is more robust against different image transformations than SIFT. The SURF descriptor outperforms the other descriptors in a systematic and significant way, with sometimes more than 10% improvement in recall for the same level of precision. At the same time, it is fast to compute. But the accurate version of SURF (SURF-128) [5], showed better results than the regular SURF.

G. MI-SIFT

Mirror and Inversion invariant SIFT (MI-SIFT) [3] improves the SIFT descriptor by enhancing the invariance to mirror reflections and grayscale inversions. Mirror reflections and inversion invariance is achieved by combining the SIFT histogram bins of both the image and its mirror-reflected image. This provides a single descriptor for original image, mirror-reflected image and greyscale inverted image with an additional cost of computation.

MI-SIFT is invariant to mirror-like images which are derived from flip operation. MI-SIFT is also invariant to grayscale-inverted images which are derived from invert operation. MI-SIFT descriptor which is based on moment [6] is not discriminative. MI-SIFT also results in more than 10% of matching performance degradation [6] than SIFT during no-flip transformations. It also requires more computational time [3] when compared to SIFT.

H. PCA-SIFT

PCA-SIFT [8] is an improved SIFT descriptor proposed by Ke and Sukhthankar. It encodes the characteristics of an image by considering the feature point's neighbourhood. PCA-SIFT improves SIFT descriptor by applying Principal Component Analysis (PCA) 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. The feature vector thus makes a total of 3042 elements. It is then normalized to unit magnitude. PCA is then applied to feature vector and a 36-dimension feature descriptor is generated.

As reported in [8] PCA-SIFT descriptor is more compact, accurate and faster than SIFT. It requires less storage space. But when error is introduced into the orientation assignment phase or in the scale estimation, PCA-SIFT's accuracy [8] begins to degrade. The standard SIFT representation can handle such errors.

I. MIFT

MIFT (Mirror reflection Invariant Feature Transform) [9] is a framework for providing feature descriptor that is robust to transformations including mirror reflections. It enhances SIFT by adding mirror reflection invariance property to SIFT. 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. Hence in MIFT, mirror invariance is achieved by simple descriptor reorganization. This descriptor reorganization first organizes the arrangement of cell order around the interest points. This is done by checking the values of total left pointing and right pointing orientations. Based on the winning orientations, the column order may change or not. Second, for each cell, it checks whether the order of orientation bins to follow clockwise or anti-clockwise direction.

MIFT is robust to mirror-reflections. But it requires longer computational time compared to SIFT. MIFT [9] detects the symmetry axis of a planar symmetric object using RANSAC.

J. F-SIFT

F-SIFT [6] is a new descriptor that incorporates the flip invariance property to SIFT, while preserving its original properties. F-SIFT generates descriptors as follows: Given a region rotated to its dominant orientation, Curl is computed to estimate the flow direction of image i.e. either clockwise or anti-clockwise. F-SIFT ensures flip invariance property by enforcing that the flows of all regions should follow a predefined direction indicated by the sign of Curl. For regions whose flows are opposite of the predefined direction, flipping the regions along the horizontal (or vertical) axis and complementing their dominant orientations are performed to

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

geometrically normalize the regions. SIFT descriptors are then extracted from the normalized regions. In other words, F-SIFT operates directly on SIFT and preserves its original property.

During flip, F-SIFT exhibits significantly stronger performance than SIFT [6]. The number of matching pairs recovered by F-SIFT is much more than SIFT. For transformation involving no flip, F-SIFT shows similar performance as SIFT. But fewer matching pairs are found by F-SIFT due to estimation error during Curl computation. This error comes from regions lacking of texture pattern. The extraction of F-SIFT descriptors from an image is approximately one third slower than SIFT.

III. Conclusions

Feature invariant descriptors have gained research attention now-a-days because of their wide range of applications. A number of feature invariant descriptors have been proposed in literature based on the needs of various applications. Among them SIFT is the most widely used descriptor. A number of enhancements on SIFT is also proposed. Attempts were made to reduce the dimensionality of SIFT, there by speeding up feature descriptions. This work performs a survey of some of the already existing flip-invariant descriptors. A brief overview of different descriptors and their advantages and disadvantages are included as part of the survey.

Acknowledgment

The author would like to thank the Department Head, Project Guide and Group Tutor and for their constructive comments and informative suggestions that have helped her to improve this paper.

References

[1] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no.10, Oct. 2005.

[2] 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.

[3] R. Ma, J. Chen, and Z. Su, “MI-SIFT: Mirror and inversion invariant generalization for SIFT descriptor,” in *Proc. Int. Conf. Image Video Retr.*, 2010, pp. 228–236.

[4] X. Guo and X. Cao, “FIND: A neat flip invariant descriptor,” in *Proc.Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 515–518.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded up robust features,” *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[6] W.L Zhao and C. Ngo, “*Flip-Invariant SIFT for Copy and Object Detection*,” IEEE.Trans. Image Processing, Vol.22, no. 3, MARC H 2013, pp. 980-991.

[7] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

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

[9] 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.

[10] M. Douze, A. Gaidon, H. Jégou, M. Marszatke, and C. Schmid, “INRIA-LEAR’s video copy detection system,” in *Proc. NIST TREVCIDWorkshop*, 2008, pp. 1–8.

[11] W.-L. Zhao, X. Wu, and C.-W. Ngo, “On the annotation of web videos by efficient near-duplicate search,” *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 448–461, Aug. 2010.