Multiresolution Gray-Scale And Rotation Invariant Feature Extraction For An Object Recognition System In A Cluttered Scene

Priya M.S [1].
St.Anne's F.G.C,
Bangalore, India;
priwah@gmail.com

Dr. G.M. Kadhar Nawaz [2].
Sona College of Technology,
Salem, India,

Received January 2018

Abstract

The local image features used in an object recognition system should be invariant to image scaling, translation, rotation and also to the illumination changes on the image. Thus, these features should be efficiently detected through a staged filtering approach which can identify the stable points in a scale space. In this paper we discuss the application of SIFT in feature extraction which is the fundamental concept in object recognition. An important aspect of this approach is that it generates large numbers of features that densely cover the image over the full range of scales and locations. The key points are detected using a cascade filtering approach which enables the correct match for a key point to be selected from a large database of other key points. With visual reference maps that consists of more or less organized images there is a compromise between the density of reference data stored and the capacity to identify, when it is not exactly in the same position as one of the reference views. Thus, we have also proposed SURF to improve the performance of appearance-based localization methods that perform image retrieval in large data sets.

Keywords: keypoints, feature extraction, PCA-SIFT, SIFT, SURF

I. INTRODUCTION

Images taken from different viewpoints may suffer transformations such as noise, rotation, scaling, translation, which leads to the two images of the same scent to appear different. Thus it is challenging task in vision applications to find similarity correspondences between two images of the same scene or object. To provide reliable matching between different viewpoints of the same image, extraction of prominent features is required. Feature detection occurs within an image and seeks to describe only those parts of that image where we can get unique feature descriptors. During the training session, the feature descriptors are extracted from sample images and stored. In classification, feature descriptors of an image will be matched with all trained image features of the trained images giving maximum correspondence and the best match.

Feature detection algorithms [14][8] are proposed in the literature to compute reliable descriptors [7][2] for image matching [5][4]. SIFT and SURF descriptors [3] are concluded as the best in their performance and have now been used in many applications[15]. A thorough comparison of many feature descriptor algorithms is reported in [9] which concluded that overall SIFT outperforms other detectors. SURF was not included in the comparisons[10] and although it has been claimed to be superior to SIFT[6] by the proposers of SURF [5].

II. METHODS

A. SIFT

Scale Invariant Feature Transform (SIFT) features are features extracted from images to help in reliable matching between different views of the same object [16]. The extracted features are invariant to scale and orientation, and are highly distinctive of the image. They are extracted in four steps. The first step computes the locations of potential interest points in the image by detecting the maxima and minima of a set of Difference of Gaussian (DoG) filters applied at different scales all over the image. Then, these locations are refined by discarding points of low contrast. An orientation is then assigned to each key point based on local image features. Finally, a

local feature descriptor is computed at each key point. This descriptor is based on the local image gradient, transformed according to the orientation of the key point to provide orientation invariance. Every feature is a vector of dimension 128 distinctively identifying the neighborhood around the key point.

The following steps are involved in SIFT algorithm:

a. Scale-space Extrema Detection: The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

b. Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

c. Orientation assignment: One or more orientation is assigned to each keypoint to achieve invariance to image rotation. A neigbourhood is taken around the keypoint location depending on the scale, and the gradient magnitude and direction is calculated in that region.

d. Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination. The first stage used difference-of-Gaussian (DOG) function to identify potential interest points, which were invariant to scale and orientation. DOG was used instead of Gaussian to improve the computation speed.

$$D(x,y,\sigma)=(G(x,y,k\,\sigma)\,G(x,y,\sigma)) * I(x,y) \longrightarrow (1)$$

$$= L(x,y,k\sigma)\,L(x,y,\sigma) \longrightarrow (2)$$

where * is the convolution operator, $G(x,y,\sigma)$ is a variable scale Gaussian, $I(x,y)$ is the input image $D(x,y,\sigma)$ is Difference of Gaussians with scale k times.

For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted[11] from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

SIFT keypoints[1] of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches. The determination of consistent clusters is performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

e. Keypoint Matching: Keypoints between two images are matched by identifying their nearest neighbours. But in some cases, the second closest-match may be very near to the first. It may happen due to noise or some other reasons.

Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. For example, if only the four corners of a door were used as features, they would work regardless of the door's position; but if points in the frame were also used, the recognition would fail if the door is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. However, in practice SIFT detects and uses a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors.

SIFT can robustly identify objects even among clutter and under partial occlusion, because the SIFT feature descriptor [2] is invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes. This section summarizes Lowe's object recognition method and mentions a few competing techniques available for object recognition under clutter and partial occlusion. In the keypoint localization step, they are rejected the low contrast points and eliminated the edge response. Hessian matrix was used to compute the principal curvatures and eliminate the keypoints that have a ratio between the principal curvatures greater than the ratio. An orientation histogram was formed from the gradient orientations of sample points within a region around the keypoint in order to get an orientation assignment. According to experiments, the best results were achieved with a 4x4 array of histograms with 8 orientation bins in each. So the descriptor of SIFT that was used is 4x4x8= 128 dimensions.

B. PCA-SIFT

PCA-SIFT Principal Component Analysis (PCA) is a standard technique for dimensionality reduction and has been applied to a broad class of computer vision problems, including feature selection. PCA-SIFT can be summarized in the following steps:

1. Pre-compute an eigenspace to express the gradient images of local patches

2. Given a patch, compute its local image gradient

3. Project the gradient image vector using the eigenspace to derive a compact feature vector.

The input vector is created by concatenating the horizontal and vertical gradient maps for the 41x41 patch centered at the keypoint. Thus, the input vector has 2x39x39=3042 elements. Then normalize this vector to unit magnitude to minimize the impact of variations in illumination. Projecting the gradient patch onto the low-dimensional space appears to retain the identity related variation while discarding the distortions induced by other effects. Eigenspace can be build by running the first three stages of the SIFT algorithm on a diverse collection of images and collected 21,000 patches. Each was processed as described above to create a 3042-element vector, and PCA was applied to the covariance matrix of these vectors. The matrix consisting of the top n eigenvectors was stored on disk and used as the projection matrix for PCA-SIFT. The images used in building the eigenspace were discarded and not used in any of the matching experiments [2].

C. SURF

A basic second order Hessian matrix approximation is used for feature point detection. The approximation with box filters is pushed to take place of second-order Gaussian filter [11]. And a very low computational cost is obtained by using integral images. The Hessian-matrix approximation lends itself to the use of integral images, which is a very useful technique. Hence, computation time is reduced drastically [5]. In the construction of scale image pyramid in SURF algorithm, the scale space is divided into octaves, and there are 4 scale levels in each octave. Each octave represents a series of filter response maps obtained by convolving the same in put image with a filter of increasing size. And the minimum scale difference between subsequent scales depends on the length of the positive or negative lobes of the partial second order derivative in the direction of derivation [2].

III. EXPERIMENTAL RESULTS

Scale Invariant Feature Transform is used to extract features from the face. The implementation is done using MATLAB. Feature extraction enables you to derive a set of feature vectors, also called descriptors, from a set of detected features. Computer Vision System Toolbox offers capabilities for feature detection and extraction.

SURF is used to detect blobs and regions. The SURF local feature detector function is used to find the corresponding points between two images that are rotated and scaled with respect to each other.

Fig 1. Detect the facial features in an image using SURF.

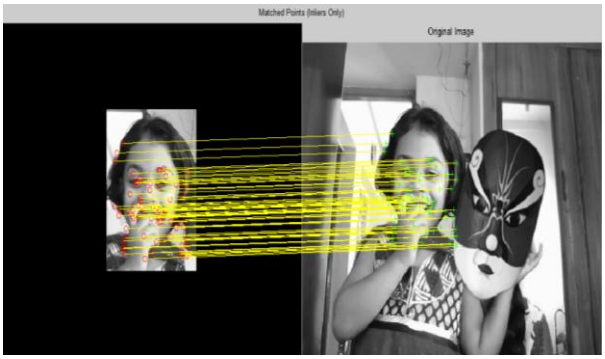

Fig.2 Matching features in an image using SURF.

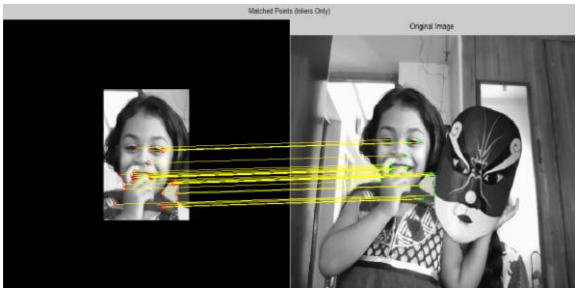

Fig.3 Matching features in an image using SIFT

SIFT starts by detecting edges and corners in the image. On the resulted image, SIFT tries to find Region of Interest points that are differentiating that image from the others. Then, out of each ROI, it extracts a histogram where each of the bins is count of particular edge or corner orientation. These histograms can be concatenated or quantized into some smaller number of groups with a clustering method like K-means.

The key points are the interest points which are the spatial locations, or points in the image that define what is **interesting** or what **stand out** in the image. The key points should remain same even when the image changes with rotation, shrinking or distortion.

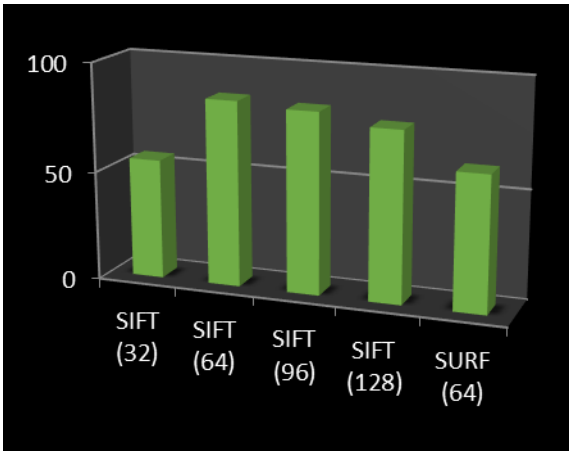

Fig.4 Performance of SIFT Vs SURF

Scaling	32D SIFT	64D SIFT	96D SIFT	128D SIFT	SURF
Time	12 sec	20 sec	35 sec	55 sec	75 sec

Table.1 Average matching time by all descriptors

4. CONCLUSION

SIFT is invariant to rotation, illumination and affine transformation changes, and shows good performance in most of the cases but has slow execution time. SURF is superior with execution time. The evaluations carried out suggests strongly that SIFT-based descriptors, which are region-based, are the most robust and distinctive, and are therefore best suited for feature matching. SURF has similar performance to SIFT, while at the same time being much faster. Another experiment concludes that when speed is not critical, SIFT outperforms SURF. SIFT, SURF and PCA-SIFT are descriptor based algorithms and have advantages over different conditions. PCA-SIFT reduces the execution time of SIFT matching but was proved to be less effective than SIFT in extracting the feature points.

REFERENCES:

[1] A. Annis Fathima, R. Karthik, V. Vaidehi, *"Image Stitching with Combined Moment Invariants and SIFT Features",* Procedia Computer Science 19 (2013) pp.420 – 427.

[2] David G. Lowe, "*Distinctive Image features from scale invariant keypoints*", International journal of Computer Vision, Vol. 60, pp. 91-110, 2004.[3] F. Schaffalitzky, A. Zisserman, "*Multi view image matching for unordered image sets*", Proceedings European Conference on Computer Vision, Vol. 1,pp. 414-431, 2002.

[4] G. Carneiro, A.D. Jepson, "*Multi scale phase based local features"*, Proceedings International Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 736-743, 2003.

[5] H. Bay, T. Tuytelaars, L. Van Gool, "*SURF: Speeded Up Robust Features"*, Proceedings European Conference on Computer Vision, Vol. 110, pp. 407-417, 2006.

[6] J. Bauer, N. Sunderhauf, P. Protzel, *"Comparing several implementations of two recently published feature detectors"*, Proceedings International Conference on Intelligent and Autonomous Systems, 2007.

[7] J. Matas, O. Chum, M. Urban, T. Padjla, *"Robust wide baseline stereo from maximally stable external regions"*, Proceedings British Machine Vision Conference, Vol. 1, pp. 384-393, 2002.

[8] K. Mikolajczyk , C. Schmid, *"An affine invariant interest point detector",* Proceedings European Conference on Computer Vision, pp. 128-142, 2002.

[9] K. Mikolajczyk, T. Tuytelaars, C Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, *"A Comparison of Affine Region Detectors, International journal of Computer Vision",* Vol. 65, pp. 43-72, 2005.

[10] L. Juan, O. Gwun, *" A Comparison of SIFT , PCA-SIFT and SURF",* International Journal of Image Processing, Vol. 65, pp. 143-152, 2009.

[11] M.M. El-gayar, H. Soliman, N. meky, *"A comparative study of image low level feature extraction algorithms",* Egyptian Informatics Journal (2013) 14, pp. 175–181.

[12] P.A. Viola, M.J. Jones, *"Rapid Object Detection using a boosted cascade of simple features",* Proceedings International Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511-518, 2001.

[13] Simranjeet Kaur, Gagandeep Singh Saini, Sandeep Kaur, *"Performance Evaluation of Fuzzy Sift and Canny Feature Extraction",* IJARCSSE, Vol 5, Issue 7, July 2015 pp.810-818.

[14] T. Linderberg,*"Feature Detection with automatic scale selection",* International journal of Computer Vision, Vol. 30, pp. 79-116, 1998.

[15] T. Tuytelaar, L. Van Gool, *"Wide baseline stereo based on local, affinely invariant regions",* Proceedings British Machine Vision Conference, pp. 412-425, 2000.

[16] Y. Ke, R. Sukthankar, *" PCA-SIFT a more distinctive representation for local image descriptors",* Proceedings International Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 506-513, 2004