Performance comparison of various feature descriptors in object category detection application using SVM classifier

Kavitha B R , Ramya G, Priya G

***Abstract:* Feature extraction involves feature detection, description and matching which is the baseline of many computer vision applications like content based image retrieval, image classification, image recognition, object detection etc. Features detected should have greater repeatability and should be able to derive descriptors out of it that are highly distinctive and robust to changes in scale, orientation, rotation, illumination etc. This paper provides an insight about the performance comparison of the long existing SIFT and SURF descriptors. The evaluation is carried out in an experimental setup of object category detection which uses a SVM classifier to detect the category.**

***Index Terms*: Feature detectors, descriptors, SIFT, SURF, ORB, BRISK, Bag-of-features.**

I. INTRODUCTION

One of the major tasks in computer vision applications like object recognition, image retrieval, image registration etc. is to identify the exact match of the features in two images of same scene or object. The features should match irrespective of their geometric and photometric transformations such as translation, rotation, blur, view point etc. Features of an image can be anything that is associated with the property of image such as colour, intensity, texture or shape. Features can be classified in different perspectives. The features can be low level features such as colour, shape, texture or spatial dependency. High level features represent the interpretation of the image there by reducing the semantic gap. Features can be global features or local features. Global features describe the image as a whole rather than a particular object. For example, colour histogram represent the whole image as a distribution of colours but the drawback here is there can be two different images having same colour histogram. The other way, there can be two similar objects which differ only in colour, possessing two different colour histograms. Features can be local such as interest points, edges, corners and regions or blobs. Good Features should possess the following properties: repeatability, distinctiveness, locality, quantity, accuracy and efficiency[1]Features should be invariant to geometric and photometric transformations. In many computer vision applications, two images are matched to identify the point to point correspondences using these features that are determined using a detector and described using a descriptor. The descriptors are matched using any similarity measure. Several detectors and descriptors are in state of the art over a decade. Feature extraction algorithms generally follow the three steps viz. detection, description and matching. Feature detectors are used to detect the features such as interest points, blobs, corners etc using which descriptors are computed. Harris and Stephens[2] developed a detector which is used to detect the corners in the image. This detector is invariant to rotation and intensity but not scale. Lindeberg [3] developed a blob detector which is scale invariant. This detector used Hessian matrix and Laplacian to detect the blob features. Mikolajczyk and Schmid[4] coined the Harris-Laplace and Hessian-Laplace detector which is scale and rotation invariant. David Lowe [5] developed a robust and scale invariant detector which uses Laplacian of Gaussian(LoG) and Difference of Gaussian(DoG) to detect the features. This paper discusses about some of the detectors and descriptors such as SIFT, SURF, ORB, FAST, BRIEF, BRISK and some of their performance is compared. In Section II we discuss some of the related work and some of the methods proposed in the literature and used in state of the art methods. In Section III and IV we discuss about the experimental setup and the performance comparison results respectively.

Revised Manuscript Received on March 10, 2019.

Kavitha B.R,School of Information Technology and Engineering(SITE), Vellore Institute of Technology,Vellore ,India.

Ramya G,School of Information Technology and Engineering(SITE), Vellore Institute of Technology,Vellore ,India.

Priya G,School of Computing Science and Engineering(SCOPE), Vellore Institute of Technology,Vellore,India.

II. BACKGROUND

This section presents an overview of various feature detectors and descriptors

SIFT (Scale Invariant Feature Transform)

SIFT descriptor is the most robust and is highly distinctive[5]. The high dimensionality of the descriptor (128-vector) is the only drawback as it takes more computational time which is difficult in low power devices like mobile phones and tablets.

SURF (Speeded up Robust Features)

SURF[6] is one method which preserves the distinctiveness of SIFT still producing results faster than SIFT. This method uses Hessian matrix approximation to detect the interest points and scale space interpolation to localize the interest points.

The descriptor is computed by constructing square regions around the interest points and Haar wavelet responses are computed for that sub region. The integration of gradient information within each subregion makes this descriptor faster in computational speed. Also, the dimensionality here is reduced to half when compared to SIFT. This descriptor is highly invariant to rotation, scale, intensity and camera calibration.

FAST (Features from Accelerated Segment Test)

This detector[7] is basically based on the SUSAN corner detection technique. A circular mask of N pixels is applied surrounding the selected pixel. Then the intensity of the selected pixel is compared with surrounding pixels. A machine learning approach is used to classify whether the detected corners are real corners. The keypoints detected using this detector can be used to determine the descriptors using methods like BRIEF, ORB etc. This method works faster even in low power devices which has limited computational ability. Drawback of FAST is it is not robust to noise.

BRIEF(Binary Robust Independent Elementary Features):

BRIEF descriptor developed by Calonder, Michael, et al [8] computes binary strings from image patches by comparing pixel intensities which are pre-determined and randomly selected. This descriptor is a simple and efficient one which performs similar to SIFT. This descriptor can be combined with detectors like SURF or FAST. The speed of this descriptor is very high when compared to SIFT or SURF. Also it works efficient even in low power devices with limited computational power. But the drawback here is it is not scale invariant as well as orientation invariant.

ORB(Oriented FAST and Rotated BRIEF):

Rublee et.al [9] developed this binary descriptor based on BRIEF which is faster than SIFT or SURF. This descriptor is built on FAST detector and BRIEF descriptor. By default FAST is variant to orientation and BRIEF is variant to in-plane rotation. This method adds the orientation component to FAST detector (oFAST) using the intensity centroid technique to measure the orientation of corners and rotation component to BRIEF descriptor (rBRIEF) by developing a learning method for choosing a good subset of binary tests. It is faster than SIFT in two orders of magnitude. It is rotation invariant and resistant to noise. The drawback of ORB is it is not scale invariant.

BRISK (Binary Robust Invaraint Scale Keypoints)

This descriptor developed by Leutenegger et al[10] is yet binary descriptor that is faster than SIFT or SURF, provides high quality description at a less computational cost. It involves basically two steps: the scalespace keypoint detection and keypoint description.The keypoints are detected in and between the octave layers of the image pyramid and an oriented sampling pattern is used to match the brightness between image pairs. However BRISK can be combined with other descriptor and vice versa

Visual Vocabulary(Bag of Features):

BoF(Bag of Features)[11] is a concept taken from Bag of words of text retrieval. This can be used for categorization of objects by developing a large vocabulary of many visual words(i.e features) .Each image is represented as a histogram of the number of visual words that are in the image.

SVM Classifier:

Support Vector Machine is a type of classifier which determines a hyperplane to separate two classes in a multidimensional space. The points nearest to the margin on both the sides are known as support vectors. The objective of SVM is to determine a hyperplane which maximizes the width of the margin between the support vectors. SVM handles linear separation with a line between two classes. For non linear separation SVM uses kernel functions to convert the low dimensional input to high dimensional space so that the hyperplane separates the two classes. Different kernel functions are available such as linear,RBF(Radial Basis Function) , polynomial and sigmoid. Generally, SVMs are used for binary classification. For multi class classification, One vs one or One vs rest SVM classifiers are used.

III. EXPERIMENTAL SETUP

The performance of the descriptors is compared with the object category detection experiments in which the category of the given object should be determined. In order to facilitate this, the setup made by Csurka, Gabriella, et al[12] is used. It involves three steps:

- First step is to create the vocabulary for the given set of images. This is done by detecting the features using a detector and computing the descriptor for each image. For the number of bins cluster the feature descriptors which would create the visual vocabulary
- Second step is to determine the features for the given image and obtain the feature descriptors. These can be matched with the feature descriptors in the vocabulary that is created earlier and the histogram is built which gives the Bag of Features descriptor for that image.
- Third step is to train the classifier with the BoF descriptors of positive and negative images. Once the classifier is trained any image can be fed as input and its respective category will be determined.

The dataset used for experimentation is the original COREL database used in [13] which contains 1000 images. The dataset has been divided into 10 categories each 100 images. The images consisted of dinosaurs, elephants, roses , buses etc. In each category 80 images are used for training and 20 images are used for testing. System is Core i5 with 1.6GHZ, 8GB RAM with windows 8.1 operating system. The system is implemented in Python2.7 using opencv2.4.9.

This setup consists of two phases – training and testing. In training phase, using the images a dictionary is built by detecting the features of each image and defining a descriptor for the same. The size of the dictionary is varied as 200,400,600,800,1000 and 1200 and the best results were seen with size=1200. Using k-means clustering dictionary is built. Then, the BoF descriptor of each image is determined using this dictionary.

These features are used to train the multiclass SVM classifier with the BoF descriptors. In this step, 80 images of any one category is taken as positive images and the rest of the images are taken negative images, thereby creating a model of One-vs-Rest classifier. Once the classifier is trained, 20 images from all the categories is given as input for testing, wherein the same procedure is followed to detect and determine the BoF descriptor of test images and

fed to the SVM classifier to classify the categories of the images. This setup is done for all the four descriptors namely, SIFT, SURF, ORB and BRISK.

EVALUATION MEASURES:

The performance is measured using three metrics: Confusion matrix, ROC curve and accuracy.

Confusion matrix

Confusion matrix is used to find the accuracy of the classifier which is determined using the following terms:

True Positives: Number of samples that are predicted positively and it is actually positive

False Positives: Number of samples that are predicted positively but, it is actually negative

True Negatives: Number of samples that are predicted negatively and it is actually negative

False Positives: Number of samples that are predicted negatively but, it is actually positive

ROC curve ROC curve determines the area under curve metric which is the probability of a classifier to rank a positive example than a negative example, both randomly chosen. The area under the curve is plotted between True positive rate(TPR) and False positive rate(FPR) and gives a value in the range 0-1. Accuracy It is the ratio of the number of correct predictions to the total number of examples.

IV. RESULTS

This section presents the results of experiments done with SIFT and SURF descriptors. First we have tried various values for codebook size. We tried values from {200,400,600,800,1000,1200} and identified that k=1200 produce better results.The histogram presented in Fig.1 shows that almost all the classes perform better for codebook size 1200.

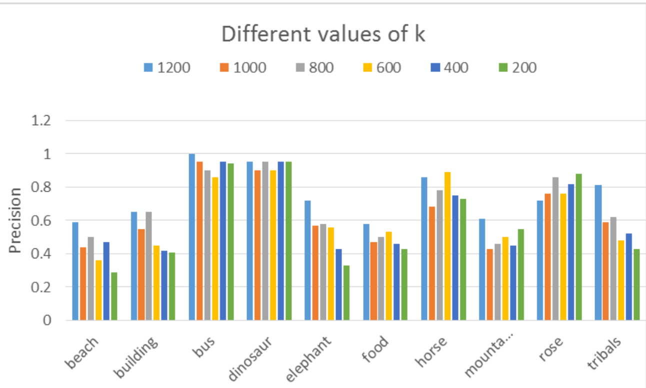

Figure 1 Histogram of precision for different values of k for 10 classes. Almost all the classes perform better for k=1200.

The other evaluation measure is confusion matrix which is plotted for both SIFT and SURF descriptors with codebook size k=1200.The precision for class bus and class dinosaur is the highest. Also ,we have plotted the ROC curves for SIFT and SURF descriptors for value k=1200(Fig 2 and Fig 3).

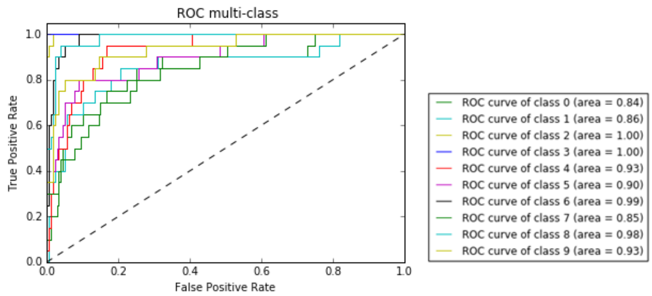

Figure 2 ROC curve using SIFT descriptor for 10 classes

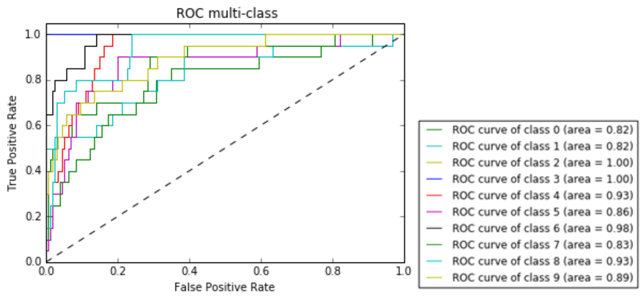

Figure 3 ROC curve for SURF descriptor for 10 classes

Actual/ Predicted	Beach	Buildin	Bus	Dinosa	Elepha	Food	Horse	Mounta	Rose	Tribals
Beach	13	2	0	0	1	0	0	2	2	0
Building	1	11	0	0	2	3	1	2	0	0
Bus	0	1	19	0	0	0	0	0	0	0
Dinosaur	1	0	0	19	0	0	0	0	0	0
Elephant	2	0	0	0	13	1	0	1	2	1
Food	1	1	0	1	0	14	1	1	0	1
Horse	1	0	0	0	0	1	18	0	0	0
Mountain	3	2	0	0	1	1	1	11	1	0
Rose	0	0	0	0	0	1	0	0	18	1
Tribals	0	0	0	0	1	3	0	1	2	13

Figure 1 Confusion matrix of SVM classifier using SIFT descriptor

This section discusses about the performance comparison of the four descriptors SIFT,SURF,ORB and BRISK applied to object category detection. Table 1 shows the comparison results. The average results of 10 repetitions of train/test images for each category are reported. In terms of computational speed, ORB is the fastest when compared to all the other three descriptors.

Table 1 Classification accuracy(%) for SIFT,SURF,ORB and BRISK descriptors

Images	SIFT	SURF	ORB	BRISK
Elephants	94	94	90	86
Dinosaurs	100	98	100	100

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Buses	100	100	98	98
Buildings	86	94	88	84
Roses	98	92	88	84
Horses	99	82	90	96
Mountains	90	90	88	86
Beach	86	66	78	90

V. CONCLUSION

In this paper, the performance of feature descriptors is compared using object detection application. The performance of SIFT is consistent and robust for all the categories but takes longer time for computation. ORB and BRISK are the fastest descriptors. The descriptors should be chosen based on the application that is used. Further, the performance comparison should be made based on geometric and photometric transformations done on images of each category. .

REFERENCES

1. T. Tuytelaars and K. Mikolajczyk, Local Invariant Feature Detectors: A Survey, Foundations and Trends in Computer Graphics and Vision, vol. 3, no. 3, pp. 177-280, 2008.
2. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference. (1988) 147-151
3. Lindeberg, T., 1998. Feature detection with automatic scale selection. Int. J. Comput.Vision 30 (2), 79–116.
4. Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proc. Fourth Alvey Vision Conf., Manchester, UK, pp. 147–151.
5. Lowe, DavidG. "Distinctive image features from scale-invariant keypoints."International journal of computer vision 60.2 (2004): 91-110.
6. Bay, Herbert, et al. "Speeded-up robust features (SURF)." Computer vision and image understanding 110.3 (2008): 346-359.
7. Rosten, Edward, and Tom Drummond. "Machine learning for high-speed corner detection." Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006. 430-443.
8. Calonder, Michael, et al. "Brief: Binary robust independent elementary features." Computer Vision–ECCV 2010 (2010): 778-792.
9. Rublee, Ethan, et al. "ORB: an efficient alternative to SIFT or SURF."Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
10. Leutenegger, Stefan, Margarita Chli, and Roland Y. Siegwart. "BRISK: Binary robust invariant scalable keypoints." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
11. Sivic, Josef, and Andrew Zisserman. "Video Google: A text retrieval approach to object matching in videos." Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003.
12. Csurka, Gabriella, et al. "Visual categorization with bags of keypoints."Workshop on statistical learning in computer vision, ECCV. Vol. 1. 2004.
13. http://wang.ist.psu.edu/docs/related/
14. X. Yuan, et al., "A sift-lbp image retrieval model based on bag-of-features, " in International conference on image processing, 2011.