Is ORB Efficient Over SURF for Object Recognition?

Mohan Ramakrishna, Shylaja S S

Abstract— **Machine vision systems have fascinated humans since the emergence of Computing. Technological advancements, both in terms of hardware and software have led to development and deployment of numerous Machine Vision systems. In the modern era of computing, object recognition form a deeply entrenched and omnipresent component of intelligent social systems. Extensive research carried out by the community from past five decades have paved path for advancements in factories, offices, industrial inspection and defect identification systems, medical imaging, defence and biometrics. Accuracy of these applications depends on the robustness and efficiency of the feature detection, feature extraction and matching techniques deployed. In this direction, a wide range of algorithms for object recognition have been devised. Two buzz words of the day are SURF (Speed Up Robust Feature) which rely on Integral images and Hessian Matrix for feature detection and ORB (Oriented Fast and Rotated BREIF) which is a combination of two major techniques: FAST (Features from Accelerated Segment Test) and BRIEF (Binary Robust Independent Elementary Features). Keeping, this as a focal point, the work proposed aims at investigating and performing a comparative study of both the approaches for object recognition.**

Index Terms— **Object recognition, SURF,ORB,BRIEF, FAST**

I. INTRODUCTION

Humans have a capability to recognize a wide range of familiar and novel objects with little effort even though they may vary in attributes like form, colour, texture, etc. Typically, objects are recognized in many different places, from many different vantage points, of different sizes and sometimes, they may even be partially obstructed from view point. Thus, object recognition in humans is invariant with respect to changes in the size, translation and rotation of the object. In simpler terms, humans have the ability to recognize objects effortlessly and instantaneously. On other hand machine recognition of object is a trivial Computer Vision problem that has not been completely solved from past five decades. Given a database of objects and an image, automated object recognition system aims at determining if any of the objects are present in the image. However, the problems associated with this process are images may be big; viewing conditions can be infinite where as computers are finite; objects can be surrounded by many similar or dissimilar objects. Put in other words, machine recognition of objects has remained challenging due to the significant variations exhibited by images including varying illuminations, partial occlusions, cluttered backgrounds, viewpoint changes, intra-category appearance variations, etc. Hence a good object recognition system must have ability to detect objects in a given image. This requires deployment of good model-detector, feature extractor, feature-model matching techniques and object verification.

The search of interest point correspondence can be divided into three main steps. In the first step, the key-points are selected at distinctive locations in image for corners, blobs, T-junctions and the key-point detector must be repeatable. As a second step, the neighbourhood of the key-points are represented by feature vector and are known as descriptors. The descriptors must be robust to noise and distinctive in nature. Final step is to match these descriptor vectors across the images. In this direction, this paper aims at comparative analysis of two newer techniques namely ORB and SURF that have found immense application in the field of object recognition. ORB is an acronym for Oriented FAST and Rotated BRIEF which is a fusion of FAST key-point detector and BRIEF descriptor. On other hand, SURF (Speeded up Robust Features) is a robust local feature detector.

The work is presented as per the following layout, a review into literature on object recognition techniques is discussed in section 2, section 3 provides insight into ORB highlighting on how the two techniques namely FAST and BRIEF have been utilized to devise this new approach and this section also describes the SURF with focus on the steps involved in recognizing object. Finally, the result of applying the aforementioned techniques on a standard caltech dataset with a comparative study is presented in section 4. Finally, section 5 concludes the work carried out with a focus on scope for future work that has been planned.

II. LITERATURE SURVEY

Appearance based object recognition can be classified into one of the two main philosophies namely generative and discriminative models. These two techniques can be described as: "Given an input x and a label y then a generative classifier learns a model of the joint probability p(x; y) and classifier using p(y|x), which is obtained by using Bayes' rule. In contrast, a discriminative classifier models the posterior p(y|x) directly from the data or learns a map from input to labels: y = f(x)". Typical examples of generative models are Principal Component Analysis

Manuscript received Aug, 2014.

Mohan Ramakrishna, Information Science, VTU/ PESIT/ Bangalore, India ,9590576184

Dr. Shylaja SS, Information Science, VTU/ PESIT/ Bangalore, India ,9449867804

(PCA) [8], Independent Component Analysis (ICA) [9], Non-negative Matrix Factorization (NMF) [10]. These models aim at finding the most apt representation of the original data. They achieve this through approximating the original data and retaining as much information as possible. On other hand, discriminant models like Linear Discriminant Analysis (LDA) [11], Support Vector Machines (SVM) [12], Ada boost aim at performing classification. Given, the training set and set of corresponding labels, discriminant models try to find the most optimal decision boundaries.

Apart from appearance based models, learning of hierarchical visual dictionaries have also been deployed for object recognition. These methods operate by learning meaningful visual parts from various image features, like image primitives [9], segmented image regions , interest points [1],[2], and histogram of gradients. Different learning algorithms have been used: discriminative criterion, data mining heuristics [9], maximum likelihood learning where a hierarchical latent Dirichlet process is assumed, deep neural networks and Grammar-Markov models. These methods have demonstrated the usefulness of the learned structures mainly through good classification performance.

III. An Overview of Methods

A. SURF Algorithm

Herbert Bay et al. [1] in 2006 came up with interest point detector and descriptor called Speed up Robust Features (SURF). SURF algorithm is invariant to rotation, scale, illumination and change in the view point which makes it appropriate for object recognition. Typically, SURF uses integral images which contain the sum of gray scale pixel values of the image; this technique reduces the computation time. The key-point detector method is based on Hessian matrix to make use of its good performance. The key stages involved in SURF are discussed in the subsequent sections.

1) Integral Images

The integral images also called as Summed Area Table (SAT) was introduced by Viola and Jones. The integral images are used as quick and effective way of calculating the sum of pixel values in a given image. One of the applications of integral images is for calculating average intensity within a given image. The value of integral image I(x) at x= (x, y) represents sum of all pixel values in Input image I within the rectangular region formed by x and origin.

$$H(x,y) = \sum_{i<x,j<y} I(i,j)$$

(1)

Fig.1.(a) Original Image 1

Fig.1.(b) Original Image 2

2) Interest Point Based on Hessian Matrix

SURF uses Hessian matrix because of its good performance and accuracy. Blobs and corners are detected at locations, if the determinant of the matrix is maximum. For a given point x = (x, y) in an image I, the Hessian matrix H(x,σ) in x at scale σ is defined as:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (2)$$

3) Interest Point Description

To construct the scale-space pyramid, SURF increases the dimension of the Gaussian filter rather than reducing the size of the image. After constructing the scale-space pyramid, it finds the extreme of the Hessian matrix determinant values at different scales by comparing a point with its 26 neighbors in the pyramid similar to that of the SIFT. This gives the SURF key-points with their scales.

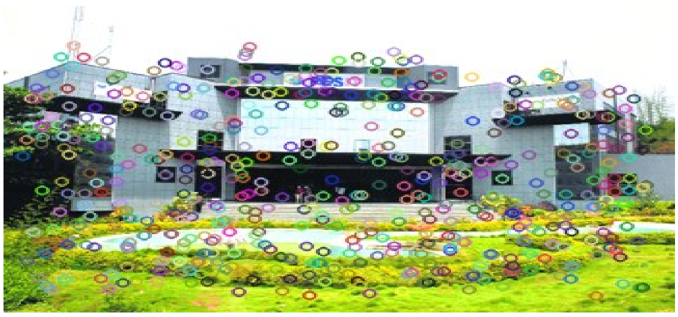

Fig.2. (a) Keypoints Detection using SURF

Fig.2. (b) Keypoints Detection using SURF

This is followed by constructing a square region centred on the interest point which is further oriented along the orientation. Then the point of interest is spilt in to 4x4 square sub regions. Haar wavelet responses are calculated along vertical dx and horizontal dy directions. These responses are weighted around the interest point with the Gaussian Kernel.

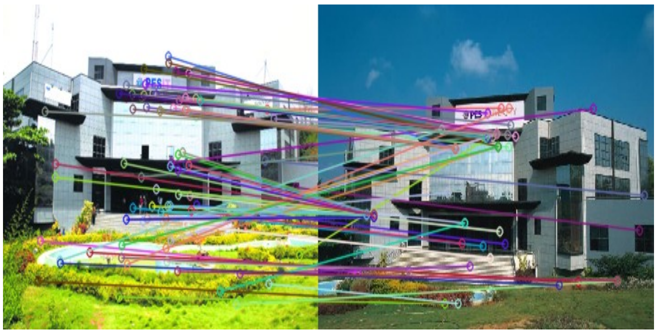

Fig.3. Matched Image using SURF algorithm

B. ORB Algorithm

ORB is Key-point detector and descriptor technique which is much efficient and faster than SURF algorithm. It is scale and rotation invariant, more robust to noise and affine transformation. The algorithm is combination of two popular techniques namely, FAST (Features for Accelerated Segment Test) proposed by Edward Rostenet. al. [13] for key point detection and BRIEF (Binary Robust Independent Elementary Features) proposed by Michael Colander [14] for description of the key point.

Fig.4.(a) Original Image 1

Fig.4.(b) Original Image 2

1) Oriented FAST-Keypoints

FAST key-point detector is used to detect key-points in real-time system such as parallel tracking and mapping. It is more efficient and reasonable corner detection algorithm. FAST algorithm has good computation properties. Detection of key-points in FAST involves consideration of intensity threshold between centre pixel and circular ring about the centre. FAST in ORB detects corners at multiple scales by making a scale pyramid of the image and add orientation to the corners.Orientation to key-points is assigned using a technique known as intensity centroid. In the intensity centroid it assumes that the corners intensity is offset from its centre, and vector is used to impute orientation. Rosin [13] defines moments of patch as:

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{3}$$

The intensity centroid of patch of a pixel is:

$$C = \left(\frac{m10}{m00} - \frac{m01}{m00}\right) \tag{4}$$

Where,

$$mpq = \sum_{x,y} x^p y^q I(x,y) \tag{5}$$

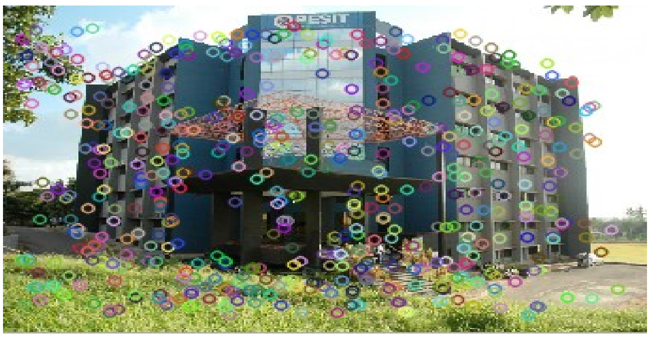

Fig.5. (a) Key-points Detection using ORB

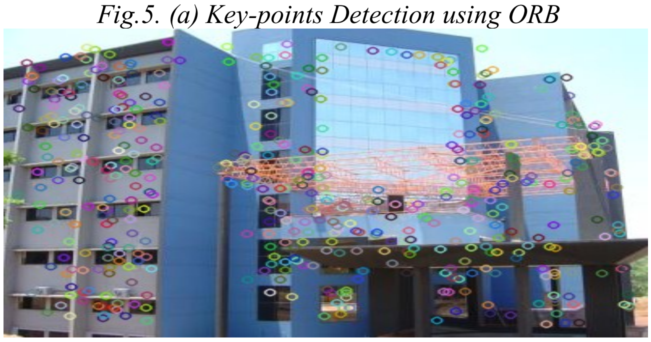

Fig.5. (b) Key-points Detection using ORB

ISSN: 2278 – 1323

2) BRIEF Key-point Descriptor

The BRIEF descriptor uses a shortcut method to find the descriptor bit string. A binary test is performed on the pixels. This is carried out by selecting $_d$ pixel location pairs randomly. Assume the first location pair to be p and q and binary test on them will yield:

$$T(p;x,y) := f(x) = \begin{cases} 1, & p(x) < p(y) \\ 0, & p(x) \geq p(y) \end{cases} \qquad (6)$$

If intensity of I(p) < I(q) then the result is 1 or else the result is 0. This process is applied to all n_d locations. The length of the n_d can be 128, 256 or 512 and the hamming distance is applied on the bit string to match the descriptors.

Fig.6.Matched Image using ORB algorithm

IV. RESULTS AND DISCUSSION

Four classes of clatech-110 dataset namely tiger, airplane, bike and side view car have been used for experimentations. A sample set of images have been depicted in Fig. 7. The images were grouped into gallery and test sets with 50 known and 150 unknown images respectively.

Fig.7.Sample Images from Clatech-110 dataset

Different statistical parameters like True Positive, False Positive, True Negative, and False Negative have been used to compute Specificity, Sensitivity, True positive Rate, False positive Rate and Accuracy. For the purpose of evaluation, the aforementioned parameters have been considered and based on the values of various parameters; accuracy and precision are computed and tabulated in Table 1.

$$\textit{Accuracy (ACC)} = \frac{\textit{(True Positive + True Negative)}}{\textit{(P + N)}} \qquad (7)$$

$$P = \textit{(True Positive + False Negative)}$$
$$N = \textit{(False Positive + True Negative)}$$

$$\textit{Precision} = \frac{\textit{(True Positive)}}{\textit{(True Positive + False Positive)}} \qquad (8)$$

Table 1: A comparison between SURF and ORB for various scenarios

Parameters	***SURF***	***ORB***
True Positive Image (Hit Ratio)	**40 Images**	**42 Images**
True Negative (Correct Rejection)	**120 Images**	**127 Images**
False Positive (False Alarm)	**10 Images**	**8 Images**
False Negative (Miss)	**30 Images**	**23 Images**
Sensitivity or True Positive Rate	**57%**	**64%**
Specificity or True Negative Rate	**92%**	**98%**
Precision or Positive Predictive value	**80%**	**84%**
Negative Predictive Value	**80%**	**85%**
False Positive Rate or Fall Out	**8%**	**6%**
False Discovery Rate	**2%**	**2%**
Accuracy	**80%**	**84%**

Further, we plot the Receiver Operating Characteristics (ROC) which is graph indicating the ratio of TPR (True Positive Rate) over FPR (False Positive Rate). It is also known as graph between Sensitivity and (1-Specificity). A sample ROC plot for SURF and ORB is illustrated in Fig. 8(a) and (b).

$$\textit{Sensitivity or True Positive Rate} = \frac{\textit{(True Positive)}}{\textit{(True Positive + False Negative)}} \qquad (9)$$

$$\textit{False Positive Rate} = \frac{\textit{(False Positive)}}{\textit{(False Positive + True Negative)}} \qquad (10)$$

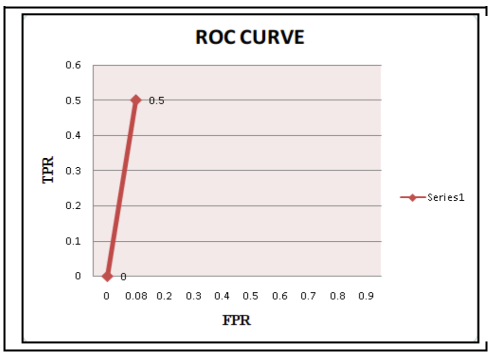

Fig.8. (a) ROC curve for SURF

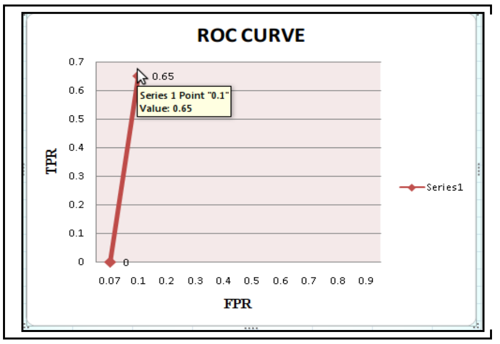

Fig.8. (b) ROC curve for ORB

Similarly, the Cumulative Match Characteristics (CMC) is measured by varying the threshold value used in the algorithm; we rank the accuracy of the algorithm by varying the threshold value and plot a graph between the accuracy and the ranks. Accuracy for first four ranks has been depicted in table 2 and in Fig. 9 (a) and(b) illustrates sample CMC curves.

Table 2: A comparison between SURF and ORB for different threshold values

Rank	***Parameter***	***SURF***	***ORB***
Rank 1	***True Positive***	**40**	**42**
	True Negative	**120**	**127**
	Total dataset Size	**200**	**200**
	Accuracy	**80%**	**84%**
Rank 2	***True Positive***	**42**	**44**
	True Negative	**125**	**130**
	Total dataset Size	**200**	**200**
	Accuracy	**83%**	**87%**
Rank 3	***True Positive***	**43**	**45**
	True Negative	**128**	**132**
	Total dataset Size	**200**	**200**
	Accuracy	**86%**	**89%**
Rank 4	***True Positive***	**44**	**45**
	True Negative	**130**	**130**
	Total dataset Size	**200**	**200**
	Accuracy	**87%**	**89%**

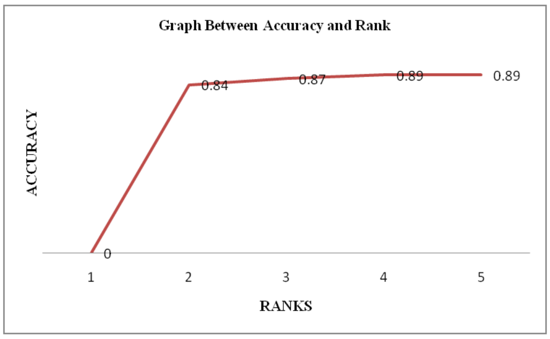

Fig.9 (a) CMC curve for SURF

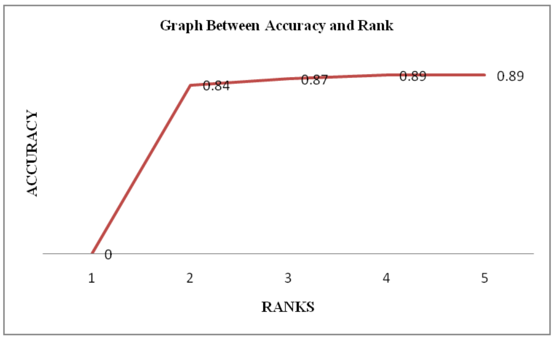

Fig.9. (b) CMC curve for ORB

V. Conclusion and Future Enhancement

This paper aimed at investigating the efficiency of ORB over SURF. During the experimentation stage, the ORB algorithm exhibited good stability, accuracy and better Hit ratio when compared to SURF. Thus, ORB algorithm is computationally better than SURF for object recognition. The Key-point detection in ORB is much faster than SURF, this is due to the fact that it doesn't use orientation component unlike SURF which has computation overhead. Rather, it makes use of Rosin method i.e. Intensity Centroid. Apart from this, the feature descriptors in ORB uses binary test between the pixels, which is much faster than other feature descriptor algorithms. Thus, based on the work carried out, we conclude that "ORB is an efficient alternative to SURF". One future direction would be to replace the feature matching algorithm FLANN with other techniques like BFMatcher, Hamming distance, etc.

Acknowledgment

We acknowledge the support from Mr. Dhannajay Rathod (Student, ISE Dept, PESIT) for extending help in worked carried out.

REFERENCES

[1] Herbert Bay, Andreas Ess , TinneTuytelaars , and Luc Van Gool , "Speeded-Up Robust Features (SURF)", IEEE Conference on Computer Vision, pp. 532-539, 2006

[2] Ethan Rublee Vincent Rabaud Kurt Konolige Gary Bradski, "ORB: An Efficient Alternative to SIFT or SURF", In European Conference on Computer Vision, Vol.23, No.6, pp.681-685, 2011.

[3] M. Brown and D. Lowe, "Invariant Features from Interest Point Groups", British Machine Vision Conference, pp. 1-8, 2002

[4] David G. Lowe "Object Recognition from Local Scale-Invariant Features", IEEE International Conference on Computer Vision Workshops (ICCVW), pp.354-361, 1999

[5] Marius Muja, David G. Lowe, "Fast approximate Nearest Neighbours with Automatic Algorithm Configuration", IEEE Conference on Computer Vision Workshop, pp. 3476-3483,2013.

[6] Chris Harris & Mike Stephens, "Harris Corner Detection", Pattern Recognition, 41, pp. 3054-3067, 1988

[7] Shi and C. Tomasi, "Shi-Tomasi Corner Detector", IEEE International Conference onEuropean Conference on Computer Vision, pp. 265-271, 1994.

[8] MackiewiczAndrzej,RatajczakWaldemar"PrincipalComponet Analysis(PCA)", Computers and Geosciences, Volume 19, Issue 3, p. 303-342.

[9] Pierre Comon, "Independent Component Analysis (ICA)", Signal ProcessingVolume 36, Issue 3, April 1994, Pages 287–314.

[10] Patrik O. Hoyer, "Non-negative Matrix Factorization (NMF)", The Journal of Machine Learning, Research Volume 5, 12/1/2004, Pages 1457-1469

[11] Baozong Yuan, "Linear Discriminant Analysis (LDA)"Pattern Recognition LettersVolume 26, Issue 5, April 2005, Pages 527–532.

[12] Chih-chungchang" Support Vector Machines (SVM)", Volume 2 Issue 3, April 2011 Article No. 27. The journal of Machine Learning.
[13] Edward Rosten, Gregory " Adaptive and Generic Corner Detection Method Based on Accelerated Segment Test" In Computer Vision Springer Volume 6312, 2010 pp 183-196
[14] Micheal Colander, Vincent Lepetit and Pascal Fau, " BRIEF : Binary Robust Independent Elementary Features", at CVLab, EPFL, Lausanne, Switerland.