IJITR
Monica* et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.4, Issue No.3, April – May 2016, 3025 – 3032.

A Comparative Analysis of Object Recognition System Using SIFT, SURF and FAST Algorithms

MONICA
P.G Student
Department of Electronics & Communication Engineering
SKITM Engineering College
Bahadurgarh, Haryana, India

RAVIKANT KAUSHIK
Asst. Professor
Department of Electronics & Communication Engineering
SKITM Engineering College
Bahadurgarh, Haryana, India

RITU JANGRA
Asst. Professor
Department of Electronics & Communication Engineering
HMRITM Engineering College
Delhi, India

Abstract— **Object recognition has always been an area of interest for various researchers since decades. In this paper an attempt has been made to give a comparison between various techniques of object recognition mainly feature based approaches. In this paper an overview of the Famous and impressive technique by David Lowe, which is Scale Invariant Feature Transform (SIFT) has been given. Another very important technique called Speeded-Up Robust Feature Transform (SURF) has been used to conclude with certain interesting results. FAST is the third technique which has also been discussed in this paper. SIFT, SURF and FAST algorithms has been implemented on COIL dataset and a comparative analysis of these techniques has been given. The algorithms has been evaluated on two parameters i.e., number of features extracted and the time of execution. It has been seen that SIFT has detected more number of features as compared to SIFT and FAST. But the times of execution taken by SURF is comparatively less than SIFT and SURF.**

Keywords - **Object Recognition; Descriptors; Feature Extraction; SIFT; SURF; FAST Methods**

I. INTRODUCTION

Object recognition is a process of distinguish a particular object in an image or video. Basically object recognition algorithms based on matching, pattern recognition or feature-based techniques. Object recognition basically involves two processes: Identification and localization. It is useful in video stabilization, cell counting in bio-imaging and automated vehicle parking systems. Recently a lot of progress has been made in object categorization from images [1]. Object recognition is the subfield of computer vision which deals with recognizing the 3 D objects from image data. It also approximates the position and orientation of the recognized objects in the 3D world. Basically feature extraction is an important factor while object is required to be recognized. Feature extraction is a type in which dimension is reduced that efficiently represents interesting parts of an image as a compact feature vector. This term is also very important in image processing. This approach is useful when size of image is large and a reduced feature representation is required to complete tasks completely such as image matching and retrieval. In this paper we introduce three types of methods SIFT SURF and FAST to extract features of an image. Different results we will obtain in this approach. Object recognition in the field of computer vision describes the task to find and identifying objects in an image or video sequence form. Humans recognize a multitude of objects in images with little effort even when they are translated or rotated.

Dickinson et al [2] presents a system of object recognition which involves extraction of features and then grouping them. After this he performed object hypothesis generation and finally an object verification stage. Then, Shapiro and Stockman [3] gave a typical object recognition system which was composed of: Low-level image detection and localization of objects but also the recognition and understanding of the object/stimulus in the scene.

Object recognition involves three types of approaches:

(i) View-based
(ii) Feature- based
(iii) Model-based

View-based methods learn a model of object's appearance in two-dimensional image under different shapes and illumination condition. A number of view-based methods have been developed to recognize three dimensional objects. A full view of three dimensional structures can be drawn if enough two dimensional views of the object are provided.

Feature-based classifies images of object under variation by rotation, noise and scaling. It is robustly and efficiently recognizes a large database of objects. It is achieved by calculating a number of features and combining them into a feature vector.

In this approach, the set of edges extracted from image cannot be compared to the set of edges extracted from model object. The correspondence mapping from image edges to model edges is discovered before made a comparison. There are many ways to find the correspondence mapping but in every case computation is required.

Model-based recognition, the knowledge of an object's appearance is provided by model of its shape. The other type of knowledge cannot be considering that may be used to recognize an object. Image's content and object's model are needed for the model base object recognition.

Object recognition is that part of research which attracts the research community because of its various applications:

- automation, biometrics,
- medical diagnosis,
- defense,
- Content-based Image Retrieval (CBIR),
- surveillance and security systems,
- robotics and intelligent vehicle systems

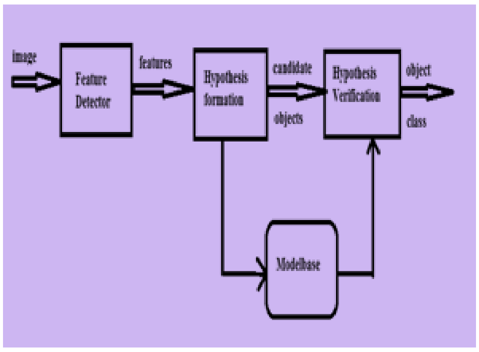

Figure 1 Object Recognition System

Feature – based approach involves the extraction of the features from the input image and then we compared these features with features that stored in database.

Object can be recognized by using a variety of models which includes:

1) Bag-of-words models with features such as SURF and MSER
2) Gradient-based and derivative-based matching approaches
3) Extracted features and boosted learning algorithms
4) Viola-Jones algorithm
5) Template matching
6) Image segmentation and blob analysis

In our paper, feature-based approaches of the object recognition techniques SIFT, SURF and FAST methods are discussed. Firstly, SIFT and its variants are discussed. SURF and its variants are discussed in next section. This paper is followed by FAST technique after SIFT and SURF. Dataset of images of objects with the results of SIFT, SURF. We have tried to conclude the survey of these various techniques. COIL dataset is used in this analytical approach of object processing. This paper compares the results of SIFT, SURF and FAST methods.

II. SCOPE AND PURPOSE

The main objective of this research paper is to enhance the already published research paper that was based on object reorganization by SIFT, SURF and FAST algorithms. The researcher has tried to study and to provide better knowledge of object recognition by using the three techniques that are named as SIFT, SURF and FAST. The researcher has tested on different dataset and have extracted the features of object by using SIFT, SURF and FAST. The result of each test methods is used to compare the precision of each feature has extracted and then the final conclusion will be as input for further research on mentioned subject. Besides, this is to be the initial stage towards the three-dimensional modeling and object reconstruction.

III. SCALE-INVARIANT FEATURE TRANSFORM (SIFT) METHOD IN OBJECT RECOGNITION SYSTEM

Scale-invariant feature transform (or SIFT) is a method/algorithm which is used for detecting and describing local features in images. The algorithm was proposed by David Lowe in 1999[4]. Lowe [4] proposed a scale invariant local descriptor, namely scale invariant feature transform (SIFT), it is basically a 3D histogram of gradient locations and orientations. It extracts large number of features as compared to SURF and FAST. In SIFT, to identify the particular object in image, interesting points on object that can be extracted to provide a feature description of the object. The extracted description from the training images can be used to identify the object when attempting to identify the object in test image containing many other objects also. To perform reliable recognition, the features extracted from testing image is to be detectable even certain changes in noise, image scale, etc. The important point of these features is that relative position between them in original image should not change from one image to another.

For example, the four corner co-ordinate point of window were used as features irrespective of window position but if considering points in frame were also used, in this case recognition would fail if the window is opened or closed. For flexible objects, it would not work fine, if any changes in its internal geometry between two images which are going to be processed. SIFT detects and uses a large number of features from images, which curtail the error caused by local variations. SIFT identify the objects in robustly way even among clutter because SIFT feature descriptor is invariant to scaling, orientation and illumination changes. SHIFT keypoints are extracted from dataset images and stored in a database. An object is recognition in particular image when comparison is made individually each feature from the image to features that are already stored in database.

Applications include:

1) Object recognition,
2) Robotic mapping and navigation and image stitching,
3) Modeling in 3D,
4) Gesture recognition, video tracking,
5) Individual identification of wildlife and match moving.

Other properties of SIFT are:

- They are highly distinctive.
- They are relatively easy to extract.
- SHIFT allows correct object identification with low probability of mismatch.

They are easy to match against a (large) database of local features however the high dimensionality can be an issue. Basically SIFT is expensive, mainly when the number of objects and learning data increase significantly. Ke and Sukthankar [5] proposed PCA-SIFT, which applies principal component analysis on normalized gradient patches so as to reduce the size of the original SIFT descriptor. SIFT provides various features to be extracted but the speed of execution in case of SURF is more as compared.

Mikolajczyk and Schmid [6] proposed gradient location and orientation histogram (GLOH) descriptor similar to SIFT and replaces the rectangular location grid used by SIFT with a log-polar one. It is similar to both SIFT and GLOH, because it uses both rectangular and log-polar location grids. Belongie et al. [7] introduce shape context, which is also similar to SIFT, but it is based on edges. Table 1 shows the various descriptor classifications. SIFT contain various real image feature extract in image processing.

TABLE I. DIFFERENT TECHNIQUE IN FEATURE EXTRACTION WITH THEIR USE

Given by	Technique in Object Recognition	Use
David Lowe	SIFT	Uses Rectangular location grid
Mikolajczyk and Schmid	GLOH	Applies PCA and uses Log-Polar location grid
Dalal and Triggs	HOG	Uses both rectangular and log-polar location grids

IV. OBJECT RECOGNITION USING SPEEDED UP ROBUST FEATURES (SURF)

SURF is acronym of Speeded Up Robust Features (SURF) is a local feature detector that can be used for object recognition or three dimensional reconstruction. SURF method motivated by SIFT descriptor was first presented by Herbert Bay et al. [8] for object recognition. It is faster than SIFT. SURF algorithm is based on the principles as SIFT.

SURF is basically several times faster than SIFT and claimed more robust than SIFT. SURF uses an integer approximation for detects the interest points. For points of interest, SURF is a detector and a descriptor in images where the image is transformed into coordinates using the technique of multi-resolution pyramid, Laplacian Pyramid shape is used to obtain an image with the same size but with reduced bandwidth. So a special blurring effect on the original image, called Scale-Space, is achieved. SURF technique ensures that the points of interest are scale invariant. New indexing step based on the sign of the Laplacian is introduced, which not only increases the robustness of the descriptor but also the matching speed. SURF method motivated by SIFT descriptor was first presented by Herbert Bay et al. [9] for object recognition. It is faster than SIFT. SURF algorithm is based on the principles as SIFT.

SURF is mainly based on:

- Interest point detection.
- Matching.
- Local neighborhood detection.

The speed has improved due to the use of integral images, which reduced the number of operations for single box convolution. Dalal and Triggs [9] provides the histogram of oriented gradient (HOG) descriptor, which is a histogram of gradient locations and orientations in 3D. But the performance of SURF can be improved by using technique FAST in recognition of image. An application of the algorithm is patented in the United States. Rani et al [10] gave the comparative analysis of SIFT and SURF and concluded that the number of features extracted using SIFT is more than the SURF but the run time in case of SURF is less than the SIFT. Here in this paper light is thrown on the FAST technique with SIFT and SURF. Disregarding discretization effects the real image descriptor in SIFT is much better as compared to the pure image descriptor in SURF. FAST algorithm proposed by Rosten and Drummond for finding interest points in an image. Chincha et al[11] introduced a paper on finding objects for blind people based on SURF features. It is a good approach for human. But like SIFT, SURF is presented for gray scale images. Many efforts are done for the improvement of SIFT and SURF using color information recently. SURF shows similar performance to SIFT, but much faster. By the Study it is concluded that when speed is not important, SIFT outperforms SURF. The performance of image matching by SIFT descriptors can be improved to achieve higher efficiency scores and lower 1-precision scores. Jalilvand et al [12] proposed a color local invariant descriptor based on SURF (CWSURF) in that using a color invariant model gives higher accuracy about 5% than SIFT and SURF. SURF descriptor is more distinctive and not much slower to compute, but it is slower to match due to its higher dimensionality.

Various features of SURF is given below:

- It is a part-based approach.
- Object is decomposed into the parts and based on the parts the object is classified.
- Effective even under occlusion.
- Extracts a set of features, determines a matching of these features to parts in the model.

V. FEATURES FROM ACCELERATED SEGMENT TEST (FAST)

4.1 Fast Indexing For Matching

During the matching stage for fast indexing, the sign of the Laplacian for the underlying interest point is included. Rosten and Drummon [13] proposed an algorithm for identifying interest points in an image. Interest points are found at blob-type structures. Sign of the Laplacian recognize the bright blobs on dark backgrounds from the reverse situation in object recognition. This feature is acquirable at no extra computational cost. In the stage of matching, we only compare features. Hence, this minimal information permits for faster matching, without reducing the performance of descriptor. It is to be noted that this advantage is for more advanced indexing methods. Interest point in an image is a pixel which has a well-specified position and can be detected in a robust manner. Interest points have high local information contents and they should be ideally repeatable between various images. Interest point detection has many of the applications in object recognition, image matching, tracking etc. The idea of interest point detection or we can say corner detection is not a new idea.

There are several well established algorithms for the same such as:

- Moravec detection algorithm used for corner detection,
- Corner detection algorithm introduced by Harris & Stephens,
- SUSAN corner detector.

The working principle behind the work of the FAST algorithm was to evolve an interest point detector to use in real time applications like SLAM on a mobile robot, which have computational resources [14] in limit.

FAST algorithm was offered by Edward Rosten and Tom Drummond in their paper "Machine learning for high-speed corner detection" in 2006 (later, revised in 2010). A basic summary of the algorithm is given below:

Feature Detection using FAST:

- Select a pixel denoted as p in the image which is to be identified as an interest point or not and assume its intensity I.
- Select an appropriate threshold value denoted as t.
- Consider a circle of 16 pixels around the pixel is to be under test.
- Now the pixel p is a corner if there exists a set of n immediate pixels in the circle of value 16 pixels which are all brighter than I p + t and all darker than I p – t, shown as white dash lines in the above image. The value n was chosen to be 12.

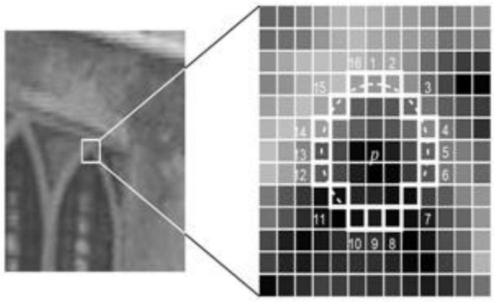

Fig. 2 This figure is for feature detection using FAST.

- A high-speed test was aimed to exclude a large number of non-corners. This study tests only the four pixels at 1, 9, 5 and 13 points (First 1 and 9 are tested if they are too darker or brighter. If it is so then checks 5 and 13). If p is a corner point then at least three of these must all be brighter than I (p + t) or darker than I (p – t). If the case is neither of these, then p cannot be a corner. Then, the full segment test criterion can be applied to the passed candidates by exploring all pixels in the circle.

This detector is itself displays high performance, but there are several limitations:

1. It does not reject as many candidates for the value n < 12.
2. The choice of pixels is not optimal as its efficiency depends on the order of the questions and distribution of the corner appearances.
3. The results of high-speed tests are discarded.

4. Various features are detected next to one another.

First 3 points are to deal with a machine learning approach. In 1997, almost a decade later Harris Detector was published. It was a new corner detector algorithm called FAST, was presented by Trajkovic and Hedley[15] in 1998. With FAST, the detection of corners was prioritized over the edges as they claimed that corners are one of the most unlogical types of features that show a strong two dimensional intensity change. Therefore they are well distinguished from the neighboring 8 points (by Trajkovic and Hedley, 1998).

Hedley and Trajkovic (1998) stated that to enable feature point matching from a detected corner, the corner detector should fulfil the following criteria:

1) ***Consistency:*** The detected positions should be insensitive to the variation of noise and it is more important that they should not move when multiple images are acquired of the same scene,
2) ***Accuracy:*** Corners should be detected as near as possible to the correct positions,
3) ***Speed:*** Even the best corner detector is useless if it is not fast enough.

According to a comparative study of the existing corner detectors based on the above criteria (given by Trajkovic and Hedley, 1998), found that most of these detectors satisfied the first two criterions but failed in the third.

The main function of FAST was the increment of the computational speed required in the detection of corners. Corner detector uses a CRF i.e. corner response function that gives a numerical value for the corner strength based on the image intensity in local neighborhood. CRF was computed over the image and the corners which were treated as local maxima of the CRF. With CRF a multi-grid technique is employed which was responsible for the improvement in the computational speed of the algorithm and it was also responsible for the suppression of false corners being detected. The main contribution of FAST contains a new algorithm to get over some limitations of currently used corner detectors. But FAST also modified the Harris detector to decrease the computational time of the algorithm without compromising the results (Trajkovic and Hedley, 1998)

In this section we will discuss about SIFT, SURF and FAST techniques. We have tried to perform an experiment using COIL (Columbia Object Image Library) dataset. The dataset includes four classes of images for our experiment as shown in figure 3.

VI. EXPERIMENTS AND RESULTS

DATABASE IMAGES: COIL database images are used to analyse the comparative result. Each class comprises of ten color images acquired under different environment i.e. scale, illumination and orientation. The color images are (128×128 pixels).

Fig. 3 Some images (128×128 pixels) from the COIL dataset

SIFT descriptor extracts features and recognise the image by comparing with their keypoints. The result is shown in figures of SIFT, SURF and FAST.

USING SIFT:

The features extracted using SIFT descriptor have been shown below in Figure 4 for both object 1 and 2. cup and box are the two objects used as dataset.

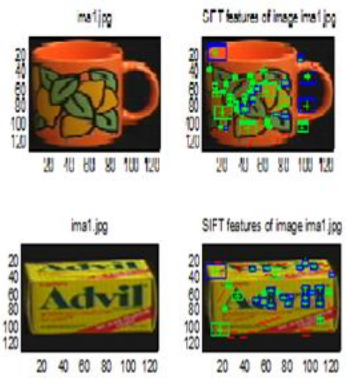

Fig.4 Feature extraction using SIFT technique.

USING SURF:

Feature extracted using SURF is given by below figure. Features extracted, in case of SURF is less as compared to SIFT. The values of extraction time and extracted features are shown in comparison table.

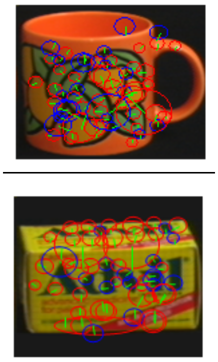

Fig. 5, Feature extraction using SURF.

USING FAST:

Using FAST technique the experiment is done on the database object images cup and box i.e. object 1 and object 2. Here we conclude that the extracted feature in case of fast for object 1 is 69 X 2 and extraction time is 10.847503 seconds. Object 2 is also taken under analytical operation using FAST technique.

Fig. 6 Feature point extracted using FAST technique.

TABLE 2: TO SHOW COMPARISON BETWEEN SIFT, SURF AND FAST

SIFT	Object 1	128x128	483	1.039472 seconds
	Object 2	128x128	392	5.137656 seconds
SURF	Object 1	128x128	58	2.890023 seconds
	Object 2	128x128	54	0.958410 seconds
FAST	Object 1	128x128	138	10.847503 seconds
	Object 2	128x128	210	5.184382 seconds

The feature extraction and extraction time is different for different technique in object recognition. Object 1 and object 2 both have different value of extraction time. Different extracted features of SIFT, SURF and FAST for object 1 and 2 is graphically represented by below graphs. Extraction time in case of FAST is highest for object 1 and object 2. This is clearly shown by the graphs.

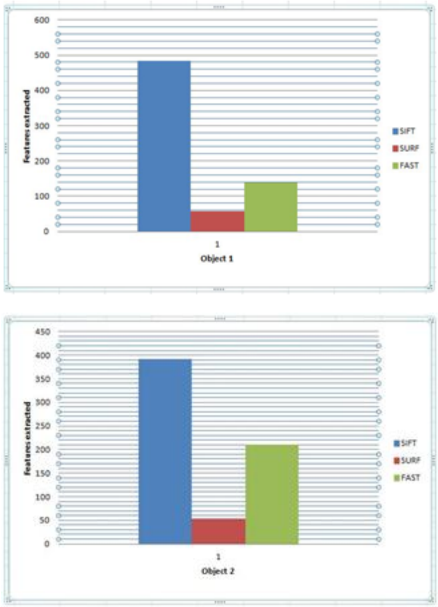

Fig. 7 this graph shows feature extraction in case of SIFT is highest as compared to both SURF and FAST for object 1.

For both object 1 and object2 SURF have lowest feature extraction.

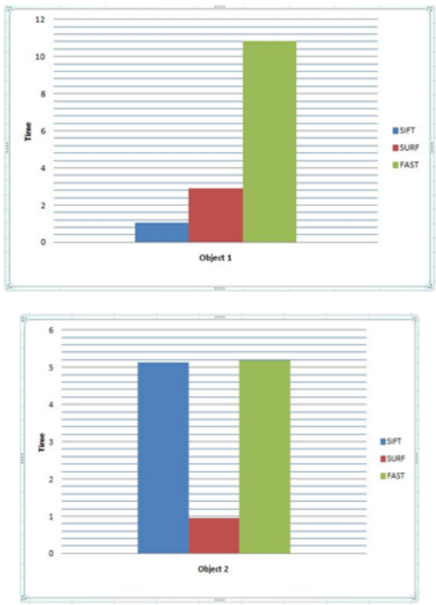

Fig. 8 Above graph gives the representation of extraction time for all of the three techniques used in object recognition i.e. SIFT, SURF and FAST. Extraction time in case of FAST is highest but in the case of SURF it is lowest. This is the result with object 2.

VII. DISCUSSION AND CONCLUSSION

In this paper we discussed about some of the state of art techniques and some recent techniques trying to overcome the shortcomings of the contemporary techniques. First we discussed about object recognition. Then we discussed about SIFT, SURF and FAST techniques or methods by which different features of image of an object extracted. Khaleel et al [16] combined SURF and BoW(Bag of Words) feature descriptor and used the K- means clustering to form different way to place or to focalize the license plate's region in an image and it concludes that this method is more effective than the SURF. Light is thrown on the object recognition discussion which is followed by the comparative study of SIFT, SURF and FAST. We mainly focused on the feature based approaches. We observed that the performance of object recognition systems has improved with descriptors like ORB, PCA-SIFT, GLOH and HOG, COLOUR-SURF, FAST. We also reviewed the performance of systems using wavelet functions as feature extractors. Also, we observed the comparison of traditional feature extraction methods like SIFT, SURF and FAST. Graphically representation clearly describes the comparison among these three techniques by measuring under the parameter of feature extraction and their time taken to extract these features. We concluded that SIFT contain highest number of feature extraction but FAST provide highest extraction time. Performance in object recognition can be improved by using FAST over SIFT and SURF. Object recognition using FAST is more accurate than using SIFT and SURF. But FAST extract less number of features.

VIII. ACKNOWLEDGMENT

The author is grateful to Mr. Ravikant Kaushik sir and Mrs. Ritu Rani for enlightening discussions and for help in developing the knowledge for this paper.

IX. REFERENCES

[1]. J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors. “Toward Category-level Object Recognition”. Springer-Verlag, 2006.

[2]. S. Dickinson,in: E. Lepore and Z. Pylyshyn (eds.), What is Cognitive Science?, Basil Blackwell publishers, 1999, pp 172--207.

[3]. Linda G. Shapiro and George C. Stockman (2001). Computer Vision. Prentice Hall. ISBN 0-13-030796-3.

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

[5]. Yan Ke; Sukthankar, R., "PCA-SIFT: a more distinctive representation for local image descriptors," in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on , vol.2, no., pp.II-506-II-513 Vol.2, 27 June-2 July 2004 .

[6]. Mikolajczyk, K.; Schmid, C., "A performance evaluation of local descriptors," in Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on , vol.2, no., pp.II-257-II-263 vol.2, 18-20 June 2003

[7]. Belongie, S.; Malik, J.; Puzicha, J., "Shape matching and object recognition using shape contexts," in Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.24,no.4, pp.509-522, Apr 2002

[8]. Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008.

[9]. Dalal, N.; Triggs, B., "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol.1, no., pp.886-893 vol. 1, 25-25 June 2005.

[10]. Ritu Rani, Surender Kumar Grewal, Object Recognition: Performance evaluation using SIFT and SURF, International Journal of Computer Applications (0975 – 8887) Volume 75– No.3, August 2013.

[11]. Chincha, R.; YingLi Tian, "Finding objects for blind people based on SURF features," in Bioinformatics and Biomedicine Workshops (BIBMW), 2011 IEEE International Conference on , vol., no., pp.526-527, 12-15 Nov. 2011.

[12]. Jalilvand, A.; Boroujeni, H.S.; Charkari, N.M., "CH-SIFT: A local kernel color histogram SIFT based descriptor," in Multimedia Technology (ICMT), 2011 International Conference on , vol., no., pp.6269-6272, 26-28 July 2011.

[13]. E. Rosten and T. Drummond, “Fusing points and lines for high performance tracking,” in 10th IEEE International Conference on Computer Vision, vol. 2, 2005, pp. 1508–1515.

[14]. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, “A solution to the simultaneous localization and map

building (SLAM) problem," IEEE Trans. Robot. Autom., vol. 17, no. 3, pp. 229– 241, Jun. 2001.

[15]. Trajkovic, M. and M. Hedley. 1998. FAST corner detector. Image and Vision Computing 16: 75-87.

[16]. Firas Mahmood Khaleel, Siti Norul Huda Sheikh Abdullah, Muhamad Khuzaifah Bin Ismail , "License Plate Detection Based on Speeded Up Robust Features and Bag of Words Model" Proc. of the IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA) 26-27 November 2013, Kuala Lumpur, Malaysia.

[17]. Chitaliya, N.G.; Trivedi, A.I., "Feature Extraction Using Wavelet-PCA and Neural Network for Application of Object Classification & Face Recognition," in Computer Engineering and Applications (ICCEA), 2010 Second International Conference on , vol.1, no., pp.510-514, 19-21 March 2010.

