International Journal of Electronics Signals and Systems

Volume 2 | Issue 4 Article 1

April 2013

# Evaluation of Bag of Visual Words for Category Level Object Recognition

K. S. Sujatha
*Dept. of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, India, 641004.*, bsoorya@rediffmail.com

G. M. Karthiga
*Dept. of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, India, 641004.*, karthiga.gmk@gmail.com

B. Vinod
*Dept. of Robotics and Automation, PSG College of Technology, Coimbatore, India, 641004.*, bvinod@rediffmail.com

# Evaluation of Bag of Visual Words for Category Level Object Recognition

**K. S. Sujatha[1], G. M. Karthiga[2] & B. Vinod[3]**

[1&2]Dept. of Electronics and Communication Engineering, [3]Dept. of Robotics and Automation
PSG College of Technology, Coimbatore, India, 641004.
E-mail : bsoorya@rediffmail.com[1], karthiga.gmk@gmail.com[2], bvinod@rediffmail.com[3].

***Abstract*** - Object recognition in a large scale collection of images has become an important application in machine vision. The recent advances in the object or image recognition for classification of objects shows that Bag-of-visual words approach is a better method for image classification problems. In this work, the effect of different possible parameters and performance evaluation of Bag of visual words approach in terms of their recognition performance such as Accuracy rate, Precision and F1 measure using 8 different classes of real world datasets that are commonly used in restaurant applications is explored. The system presented here is based on visual vocabulary. Features are extracted, clustered, trained and evaluated on an image database of 1600 images of different categories. To validate the obtained results,a performance evaluation on vehicle datasetsunder SURF and SIFT descriptors with K-means and K-medoid clustering and KNN classifier has been made. Among these SURF K-means performs better.

***Keywords*** - *Bag-of-visual words; SURF; SIFT; K-means; K-medoid; KNN classifier.*

## I. INTRODUCTION

Object recognition in computer vision is the task of finding a given object in an image or video sequence. Every day a multitude of familiar and novel objects are recognized, though these objects may vary somewhat in form, colour and texture and so on. Theultimate goal of object recognition is validation, detection, category recognition, scene or context recognition and also activity recognition.

Object Recognition using Bag of Features (BoF) has gained enormous popularity in image classification techniques. It has been shown that Bag of Words (BoW)[1, 2, 3, 4, 5] approach provides several advantages with acceptable recognition performance, faster run time and reduced storage.Similar techniques are implemented in text information retrieval and text categorization.The main idea in this algorithm is that the descriptors are quantized to form a visual word dictionary called codebook with the help of different clustering algorithms. The BoW model allows a dictionary-based modelling. Computer vision researchers use a similar idea for image representation which is called the Bag of Feature (BoF) model. Each vector in the codebook being a visual word serves as the basis for indexing the images. The main task ofBag of Feature (BoF) is to classify a given image in to one of the pre-determined objects based on the trained classes of objects and to increase detection rate of images by adapting different descriptors, clustering algorithms and classifiers. Here hard clustering algorithms like K-means and K-medoidare used. The performance of the algorithms implemented is compared by varying the dictionary size.

## II. BAG- OF- VISUAL WORDS ALGORITHM

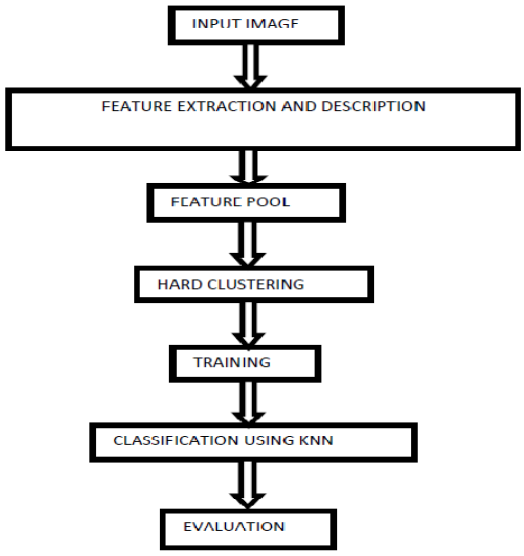


Fig. 1 : Schematic of Bag- of- Visual Words algorithm

In Bag of Words model features are extracted using detectors or dense sampling and descriptors are calculated at each and every local feature extracted. Thefeaturesextracted from the images should be easily detected under changes in pose,illumination and should be distinctive. There should be many features per object. After feature extraction,features are hard clustered bydata clustering techniques like K-means and K-medoid. After clustering a visual vocabulary is obtained with predefined number of visual words. In training phase, the input vectors from the feature pool are assigned to one or more classes. The decision rule divides input space into decision regions separated by decision boundaries and histogram is built up. In testing phase, for the test data point, the k closest points from training data is found and classification is done using KNN classifier. Figure.1 shows the schematic of Bag-of- Visual Words algorithm.

## III. FEATURE EXTRACTION

Recognition is a basis issue for robots, and featureextraction is the very important part of this process. Highquality of feature extraction will play a crucial role on the results of recognition.Features are generally a random collection of points in an image or it may be a collection of distinctive points in an image like blobs and corners which are likely to be repeatable. For local feature detection, classic detectors include Harris detector [6] and its extension [7], maximally stable external region detector [8], affine invariant salient region detector [9]. For local feature description, local descriptors such as Haar descriptor [10], scale-invariant feature transform (SIFT) descriptor [11], gradient location and orientation histogram (GLOH) descriptor [12], rotation-invariant feature transform (RIFT) descriptor [13], shape context [14], histogram of gradients (HOG) descriptor [15] and speeded up robust feature descriptor (SURF) [16] are usually used.Inthis work Harris detector with SIFT (128 D) descriptor and Hessian detector with SURF (64 D) descriptor are used.

### *A. SIFT descriptor using Harris detector*

The major stages in computations include four steps. (1) Identifying key points (Harris detector) (2) key point localisation (3) orientation assignment (4) key point descriptor computation (SIFT).

The Harris detector is a scale, rotation, and translation invariant interest point detection algorithm that has also shown to be robust to illumination variations, camera noise and viewpoint changes.The Harris detector is based on the second moment matrix. The second moment matrix given in equation (1), also called the auto-correlation matrix,defined for each pointp is often used for feature detection or for describing local image structure.

$$C = \begin{bmatrix} \sum L_x^2(x,y) & \sum L_x L_y(x,y) \\ \sum Lx\ Ly\ (x,y) & \sum L_y^2(x,y) \end{bmatrix} \quad (1)$$

Here, Lx and Ly denote the respective x and y gradient magnitude images, and the summation is taken over all points in a neighbourhood of point p. Thematrix C can be seen as a covariance matrix of the Gradient magnitude within a neighbourhood. It is real and symmetric, and so it can be decomposed into its principal components, with Eigenvalues $\lambda_1$ and $\lambda_2$. Harris noted the following in regions of constant intensity, $\lambda 1 = \lambda 2 = 0$.In regions with very little intensity variation, both $\lambda 1$and $\lambda 2$ will be small.For an edge region, the variation in one direction will be strong, and the variation in the otherdirection will be weak. Therefore, $\lambda 1$ will be large, and $\lambda 2$will be small. For a corner region, there will be strong variations in both directions and so both $\lambda 1$and $\lambda 2$willexceed some threshold value.

Key point localisation attempts to remove unstable key points from the final list by finding those that are poorly localised on an edge or corner.Orientation assignment aims to assign a consistent orientation to the key points based on local image properties. The key point descriptor can then be represented relative to this orientation, achieving invariance to rotation. The gradient magnitude, *m,* orientation(x,y) are given by the equations (2) and (3).

$$m(x,y) = \sqrt{\big(L(x+1,y) - L(x-1,y)\big)^2 + \big(L(x,y+1) - L(x,y-1)\big)^2} \quad (2)$$

$$\mu(x,y) = \big((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))\big) \quad (3)$$

Further the local image gradientsare measured in the region around each key point. These aretransformed into a representation that allows for significant levels of local shape distortion and change in illumination. The local gradient data is also used to create key point descriptors. The gradient information is rotated to line up with the orientation of the key point. These data are then used to create a set of histograms over a window centered on the key point. Key point descriptors typically use a set of 16 histograms, aligned in a 4×4 grid, each with 8 orientation bins, one for each of the main compass directions and one for each of the mid-points of these directions. This process results in a feature vector, containing 128 elements.

*B. SURF descriptor using Hessian detector*

SURF [16] outperforms or approximates previously proposed schemes with respect to repeatability, distinctiveness and robustness and it can be computed and compared faster. It is achieved by relying on integral images for image convolutions and by building strengths on leading detectors (Hessian detector) and descriptors.

The major stages in computations include:

(1) Interest point detection which involves computing integral images and Hessian matrix- based interest points.

(2) Scale space representation.

(3) Interest point localization.

Integral images allows for fast computation of box type convolution filters. The entry of an integral image $I_{\epsilon(x)}$ at a location x = $(x,y)^T$ given by equation (4) represents the sum of all pixels in the input image I within a rectangular region formed by the originand *x*.

$$I_{\epsilon(x)} = \sum_{i=0}^{i\leq x} \sum_{j=0}^{j\leq y} I(i,j) \quad (4)$$

Hessian matrix- based interest points are used for its good performance and accuracy.Given a point *x=(x,y)* in an image I, the Hessian matrix H*(x,ρ)* in *x* and scale *ρ* is given in equation (5).

Where *Lxx(x, y, σ)* is the convolution of the second order derivative of Gaussian with image *I* at *(x,y)*. This also applies to *Lxy(x,y,σ)* and *Lyy(x,y,σ)*.Scale spaces are usually implemented as an image pyramid. The images are repeatedly smoothed with a Gaussian and then sub-sampled in order to achieve a higher level of the pyramid.

Gaussians are optimal for scale-space analysis. In real applications, Gaussians have to be discretized and cropped. The SURF approximates the second order Gaussian derivate with box filters, which is calculated fast through integral images. The localization of interest point is determined by the determinant of Hessian matrix. So, interest points are finally localized in scale space and image space by using non-maximum suppression in their 3×3×3 neighbourhood. In the construction of descriptor of an interest point, a circular region around a detected interest point is first constructed. Then, a dominant orientation based on this circular region is calculated and assigned to this region, which enable the descriptor invariance to image rotations. The dominant orientation is calculated by the response of Haar wavelet. This process is also very fast by integral images. After the estimation of the dominant orientation, a square patch around an interest point is extracted to construct the SURF descriptor. The square patch is divided into a 4×4 sub-blocks. The gradients of each sub-block are used to construct the final descriptor vector.

## IV. CLUSTERING

Clustering is the process of assigning a set of objects into groups so that the objects of similar type will be in one cluster. Clustering can be classified as hard clustering and soft clustering. Hard clustering and soft clustering differ in assigning the value to the partition matrix. Hard clustering assigns the value of either 0 or 1 whereas soft clustering allows the value to be in more than one cluster. The object will be assigned to a cluster that has the highest value. The general algorithms available for hard clustering are K-means [17] and K-medoid [18]. In this work hard clustering algorithms like K-means and K-medoid algorithms are used for clustering with Euclidean measure as distance metric learning.

*A. K means Algorithm*

K-meansis one of the simplest unsupervised algorithm that partition extracted features $(x_1 x_2, .., x_N)$ from images into *k* number of clusters. Given an image set *X*, initial random cluster centers are chosen.

*B. Steps for K means Algorithm*

The procedure follows in a simple way:

1) Compute the distances $D_{ik}^2$ between each data point $x_k$and cluster centers $v_i$.

$$D_{ik}^2 = (x_k - v_i)^T (x_k - v_i), \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (6)$$

2) Select the points for the cluster, with minimal distances, they belong to that cluster.

3) Recalculate the cluster centers $v_i^{(l)}$, by finding the mean of the data points$x_i$ of corresponding clusters.

$$v_i^{(l)} = \frac{\sum_{j=1}^{N_i} x_i}{N_i} \quad (7)$$

until,

$$\prod_{k=1}^{n} max \left| v^{(l)} - v^{(l-1)} \right| \neq 0 \quad (8)$$

4) Repeat the process until *k* distinct clusters are obtained. Finally calculate the partition matrix.

*C. K- Medoid Algorithm*

The K-Medoid algorithm is a clustering algorithm related to the K-means algorithm and the medoid shift algorithm. Both the K-means and K-medoid algorithms

break the dataset up into groups and both attempt to minimize squared error, the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the K-means algorithm, K-medoid chooses data points as centers. K-medoid is also a partitioning technique of clustering that clusters the data set of n objects into k clusters with k known a priori. It is more robust to noise and outliers as compared to K-means because it minimizes a sum of general pairwise dissimilarities instead of a sum of squared Euclidean distances. The possible choice of the dissimilarity function is very rich but in this implementation the squared Euclidean distance is used.

*D. Steps for K- Medoid Algorithm*

A medoid of a finite dataset is a data point from this set, whose average dissimilarity to all the data points is minimal that is it is the most centrally located point in the set. Given the data set $X$, choose the number of clusters $1 < c < N$. Initialize with random cluster centers $v_i$ chosen from the data set $X$.

1) Compute the distances $D_{ik}^2$ between each data point $x_k$and cluster centers $v_i$.

$$D_{ik}^2 = (x_k - v_i)^T (x_k - v_i), \quad 1 \le i \le c, \quad 1 \le k \le N \quad (9)$$

2) Associate each data point to the closest medoid.

3) Recalculate the cluster centers $v_i^{(l)*}$, by finding the mean of the data points $x_i$of corresponding clusters.

$$\boldsymbol{v}_i^{(l)*} = \frac{\sum_{j=1}^{N_i} x_i}{N_i} \quad (10)$$

4) Choose the nearest data points to be the cluster center

$$D_{ik}^{i*} = (x_k - v_i^*)^T (x_k - v_i^*) \quad (11)$$

and

$$x_i^* = argmin_i \left(D_{ik}^{2*}\right);\ v_i^{(l)} = x_i^* \quad (12)$$

until

$$\prod_{k=1}^{n} max \left| v^{(l)} - v^{(l-1)} \right| \neq 0 \quad (13)$$

5) Repeat the process for l = 1, 2,.,..until *k* distinct clusters are obtained. Finally calculate the partition matrix.

K-medoidruns similar to K-means algorithm and this algorithm takes reduced computation time. This algorithmtests several methods for selecting initial medoid and calculates the distance matrix once.

## V. CLASSIFIER

In pattern recognition, the k-nearest neighbor algorithm [19] (k-NN) is a method for classifying objects based on closest training examples in the feature space. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors where k is a positive integer, typically small. If k= 1, then the object is simply assigned to class of its nearest neighbor. The nearest-neighbor method is perhaps the simplest of all algorithms for predicting the class of a test example.

The training phase is simple, that is to store every training example, with its label. To make a prediction for a test example, first compute its distance to every training example. Then, keep the k closest training examples, where k ≥ 1 is a fixed integer. This basic method is called the k-NN algorithm. The most common distance function is Euclidean distance neighbors (k is a positive integer, typically small). If k= 1, then

$$d(x,y) = \|x - y\| = \sqrt{(x-y).(x-y)} = \sum_{j=1}^{m} (x_i - y_i)^2)^{1/2} \quad (14)$$

K-Nearest Neighbor algorithm (KNN) is a part ofsupervised learning that has been used in many applications inthe field of data mining, statistical pattern recognition andmany others. KNN is a method for classifying objects based on closest training examples in the feature vector. An object is classified by a majority vote of its neighbors. K is always a positive integer. The neighbors are taken from a set of objects for which the correct classification is known. It is usual to use the Euclidean distance, though other distance measures such as the Manhattan distance can be used.

## VI. RESULTS AND DISCUSSION

The effect of different possible parameters and performance evaluation of Bag of visual words approach is done in terms of Precision, F1- measure and Accuracy rate for four different topics from dataset1 namely airplanes, cars, motorbikes and faces. Performance evaluation is also done for eight different topics namely burger, spaghetti, egg, spoon, bottle, can, coffee pot and mug of real world datasets that has been used in restaurant applications. Dataset1 is taken from Caltech database and since dataset2 is taken for real time application for visual recognition of objects used in restaurant,it is created from Google images. When compared to the images in dataset1 the size of images in dataset2 is comparatively small and can be categorized as tiny images when compared to dataset1.

The performance measures used in this evaluation are

1) Precision

$$P = \frac{1}{|c|}\sum_{i=1}^{|c|}\frac{TP_i}{TP_i+FP_i} \quad (15)$$

2) F1- measure

$$F = \frac{2 * P * R}{P + R} \quad (16)$$

where

$$R = \frac{1}{|c|}\sum_{i=1}^{|c|}\frac{TP_i}{TP_i+FN_i} \quad (17)$$

3) Accuracy rate

$$Accuracy = \frac{\sum_{i=1}^{c} TP_i+\sum_{i=1}^{c} TN_i}{\sum_{i=1}^{c}(TP_i+FN_i+FP_i+TN_i)} \quad (18)$$

In these equations TP indicates true positive, FP false positive, FN false negative and TN true negative of the classification result. Precision and recall are the most common measures for evaluating an information retrieval system. F1 score is a measure of test's accuracy. It considers both the precision P and recall R of the test to compute the score.The sample images of two different datasets are shown in Figure 2.

A performance evaluation of Bag of Words (BoW) is done using dataset1 and dataset2.The test data set includes four different topics in dataset1and eight topics in dataset2 each containing 50 images. 200 images per concept were used during the training phase to build the codebooks. The classifier is trained for another 200 images from each topic. The sizes of visual vocabularies are varied from 25 to 500 and evaluation performed. The distance measure used is Euclidean distance.

The parameters that affect the performance ofBoW are extraction of features and its description methods, dictionary generation methods, dictionary size, and distance function. A performance evaluation of four different combinations of parameters has been doneby varying the feature descriptor and dictionary generation method. The four combinations are (1) SIFT descriptor with K-means clustering (SIFTK-means) (2) SIFT descriptor with K-medoid clustering (SIFT K-medoid) (3) SURF descriptor with K-means clustering (SURF K-means) and (4) SURF descriptor with K-medoid clustering (SURF K-medoid). In the first two combinations Harris corner detector is used for feature extraction and in the next combinations Hessian detector is used. The distance function used in all implementation is Euclidean distance. KNN classifier is used for classifying the images. Evaluation is done by varying codebook size.

(a)

(b)

Figure.2 (a) sample images from dataset1 (b) sample images from dataset2

Among all, SURF K-means performs better for a codebook size of 100 for dataset1 and 500 for dataset2.In terms of run time K-medoid serves advantageous with reduced computation time. The results which give the effect of different possible parameters by varying codebook size for dataset1are shown in Table 1, 2 and Figure 3 and for dataset2 in Table 3, 4 and Figure 4. Sincedataset2 can be categorized as tiny image set when compared to dataset1 the performance measures of these image set is slightly less when compared to dataset1 as the number of features extracted depends on the size of the images. The performance measures can be further increased for Dataset2 by increasing the number of images from which features are extracted and trained.

TABLE I. Precision vs. Codebook Size for Dataset1

| *Codebook Size* | *SIFT K-means* | *SIFT K-medoid* | *SURF K-means* | *SURF K-medoid* |
|---|---|---|---|---|
| 25 | 0.75365 | 0.74215 | **0.8571** | 0.37364 |
| 50 | 0.8181 | 0.7715 | **0.8828** | 0.58628 |
| 75 | 0.85275 | 0.78945 | **0.9123** | 0.76951 |
| 100 | 0.8577 | 0.81935 | **0.9271** | 0.46384 |
| 200 | 0.9183 | 0.79725 | **0.9056** | 0.61672 |
| 300 | 0.9171 | 0.7675 | **0.9132** | 0.5873 |

TABLE II. F1 Measure vs. Codebook Size for Dataset1

| *Codebook Size* | *SIFT K-means* | *SIFT K-medoid* | *SURF K-means* | *SURF K-medoid* |
|---|---|---|---|---|
| 25 | 0.74676 | 0.7257 | **0.8484** | 0.37181 |
| 50 | 0.8064 | 0.7606 | **0.8764** | 0.58312 |
| 75 | 0.8438 | 0.7796 | **0.9111** | 0.75963 |
| 100 | 0.8513 | 0.80699 | **0.926** | 0.44896 |
| 200 | 0.9166 | 0.775 | **0.9003** | 0.6108 |
| 300 | 0.9135 | 0.769 | **0.9066** | 0.609 |

TABLE III. Precision vs. Codebook Size for Dataset2

| *Codebook Size* | *SIFT K-means* | *SIFT K-medoid* | *SURF K-means* | *SURF K-medoid* |
|---|---|---|---|---|
| 25 | 0.6872 | 0.68146 | **0.66866** | 0.48835 |
| 50 | 0.66456 | 0.6148 | **0.71706** | 0.5987 |
| 75 | 0.69277 | 0.64067 | **0.73018** | 0.56731 |
| 100 | 0.69863 | 0.68483 | **0.74239** | 0.52716 |
| 200 | 0.7048 | 0.68744 | **0.71092** | 0.575 |
| 300 | 0.6899 | 0.63899 | **0.73874** | 0.55026 |
| 400 | 0.67612 | 0.68638 | **0.7577** | 0.64355 |
| 500 | 0.6412 | 0.64744 | **0.76873** | 0.6479 |

TABLE IV. F1 Measure vs. Codebook Size for Dataset2

| *Codebook Size* | *SIFT k-means* | *SIFT K-medoid* | *SURF K-means* | *SURF K-medoid* |
|---|---|---|---|---|
| 25 | 0.67849 | 0.67568 | **0.66176** | 0.40945 |
| 50 | 0.65848 | 0.60463 | **0.70327** | 0.59305 |
| 75 | 0.68988 | 0.63529 | **0.72123** | 0.56236 |
| 100 | 0.68271 | 0.67219 | **0.73359** | 0.45621 |
| 200 | 0.69347 | 0.66687 | **0.70287** | 0.58109 |
| 300 | 0.68618 | 0.62935 | **0.73308** | 0.54252 |
| 400 | 0.67052 | 0.67552 | **0.74875** | 0.62235 |
| 500 | 0.633 | 0.62816 | **0.75796** | 0.6334 |

The maximum performance measures of four different combinations for dataset1 (Vehicular datasets) and dataset2 (Restaurant datasets) are shown in Table 5.

## V. CONCLUSION

In this paper, the performance of Bag of visual words approach is investigated in terms of its degree of recognition using performance measures like accuracy rate, Precision and F1 measure. Dataset1 includes four different topics namely airplanes, cars, motorbikes and faces and dataset2 includes eight different topics namely burger, spaghetti, egg, spoon, bottle, can, coffee pot and mug of real world datasets that are commonly used in restaurant applications. The system presented here is based on visual vocabulary. The parameters that affect performance of Bag of Words (BoW) are, extracted features, description methods, dictionary generation methods, dictionary size, and distance function. A performance evaluation of four different combinations has been done by varying the feature descriptor and dictionary generation method. The four combinations are (1) SIFT descriptor with K-means clustering (SIFT K-means) (2) SIFT descriptor with K-medoid clustering (SIFT K-medoid) (3) SURF descriptor with K-means clustering (SURF K-means) and (4) SURF descriptor with K-medoid clustering (SURF K-medoid). Among all these combinations SURF K-means performs better for both the datasets.

TABLE V. Maximum performance measures for different methods for a given codebook size

| **Datasets** | **Method** | **Codebook Size** | **Max Accuracy Rate** | **Max. Precision** | **Max. F1 Measure** |
|---|---|---|---|---|---|
| *Dataset 1* | *SIFT K-means* | 200 | 0.9587 | 0.9183 | 0.9166 |
| | *SIFT K-medoid* | 100 | 0.8975 | 0.81935 | 0.80699 |
| | *SURF K-means* | **100** | **0.9625** | **0.9271** | **0.926** |
| | *SURF K-medoid* | 200 | 0.8025 | 0.61672 | 0.6108 |
| *Dataset 2* | *SIFT K-means* | 200 | 0.92063 | 0.7048 | 0.69347 |
| | *SIFT K-medoid* | 100 | 0.915 | 0.68483 | 0.67219 |
| | *SURF K-means* | **500** | **0.93688** | **0.76873** | **0.75796** |
| | *SURF K-medoid* | 400 | 0.90063 | 0.64355 | 0.62335 |

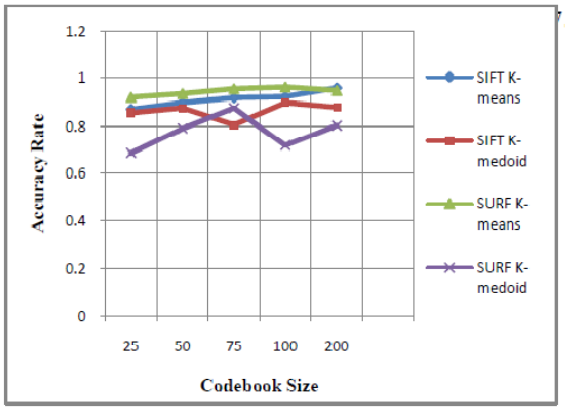


Figure.3 Accuracy Rate Vs Codebook Size for Dataset 1

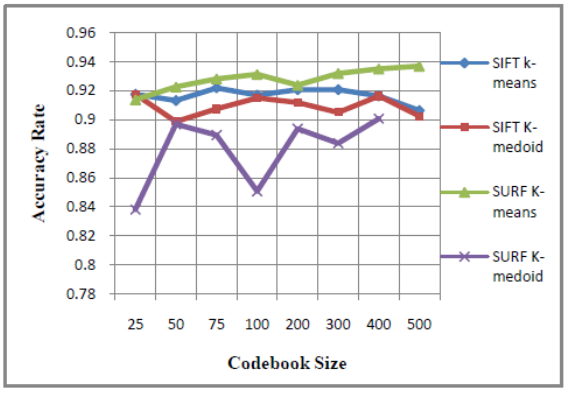


Figure.3 Accuracy Rate Vs Codebook Size for Dataset 2

## V. REFERENCES


1. D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree.CVPR, 2006.
2. M. Muja and D. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In VISAPP, 2009.
3. Mohamed Aly, Mario Munich and Pietro Perona, "Bag of Words for Large scale object recognition," in computational vision lab, Caltech, Pasadena, CA, USA.
4. O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In ICCV, 2007.
5. H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search.In ECCV, 2008.
6. C. Harris and M. Stephens. A combined corner andedge detector.Proceedings of the Fourth Alvey Vision Conference, pages 147–151, 1988.
7. T. Tuytelaars and L. V. Gool. Matching widely separated views based on affine invariant regions. International Journal of Computer Vision, 59(1):61–85, 2004.
8. J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. Image and Vision Computing, 22(10):761–767, 2004.
9. K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. International Journal of Computer Vision, 60(1):63–86, 2004.
10. P. Viola and M. Jones. Robust real-time object detection .Proc.of IEEE Workshop on Statistical and Computational Theories of Vision, 2001.
11. D.G. Lowe. Distinctive image features from Scale-invariant key-points. International Journal of Computer Vision, 2(60):91–110, 2004.
12. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 27(10):1615–1630, 2005.
13. S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. Technical Report, Beckman Institute, University of Illinois, 2004.
14. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. IEEE Trans. On Pattern Analysis and Machine Intelligence, 24(4):509–522 2002.
15. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection.CVPR, 2005.
16. Bay.H, Tuytelaars.T, &Van Gool, L.(2006),SURF: Speeded Up Robust Features.9th European Conference on Computer Vision.
17. TapasKanungo, SenoirMember, IEEE, David.M.Mount, Member, IEEE, Nathan S. Nethanyahu,Member, IEEE, ChritineD.Piatko, Ruth Silverman, and Angela Y. Wu.Senior Member, IEEE, An efficient K-means clustering Algorithm :Analysis and Implementation. IEEE Trans. On Pattern analysis and Machine Intelligence, Vol.24, No. 7 July 2002.
18. Park, H.S., J.S. Lee and C.H. Jun, 2006. A K-means like algorithm for K-medoid clustering and its performance.http://citeseerx.ist.psu.edu/viewdoc/download doi=10.1.1.90.7981&rep= rep1& type=pdf.
19. R.Muralidhara, Dr.C.Chandrasekar, Object Recognition using SVM-KNN based on Geometric Moment Invariant. International Journal of Computer Trends and Technology July-August Issue 2011.


❑❑❑