

Cost Effective Image Classification Using Distributions of Multiple Features

Vanitha Sivagami Sivasankaravel^{1*}

¹Mepco Schlenk Engineering College, Sivakasi
Tamilnadu, India

[e-mail: svanitha@mepcoeng.ac.in]

*Corresponding author: Vanitha Sivagami Sivasankaravel

*Received February 15, 2022; revised May 14, 2022; accepted June 25, 2022;
published July 31, 2022*

Abstract

Our work addresses the issues associated with usage of the semantic features by Bag of Words model, which requires construction of the dictionary. Extracting the relevant features and clustering them into code book or dictionary is computationally intensive and requires large storage area. Hence we propose to use a simple distribution of multiple shape based features, which is a mixture of gradients, radius and slope angles requiring very less computational cost and storage requirements but can serve as an equivalent image representative. The experimental work conducted on PASCAL VOC 2007 dataset exhibits marginally closer performance in terms of accuracy with the Bag of Word model using Self Organizing Map for clustering and very significant computational gain.

Keywords: Bag of Words Model, Image Classification, PASCAL VOC, Self Organizing Map, Gradients.

1. Introduction

In the digital era, tons of data are available free for access for everyone. Whenever we search for information in the Internet, huge volumes of data are retrieved where most of the data obtained are irrelevant to our search. So for the past three decades, tremendous research is going on for providing highly relevant information to the users.

In the Image processing domain, image classification and retrieval have many significant applications like remote sensing, robot navigation, intelligent transportation systems. The power of the image processing system depends upon the spectral and spatial components extracted from the image data, the algorithm used for classification, the timeframe required for completing the task and where it is applied. Considering all these factors instantaneously, designing an efficient image classification system is still a greatly challenging task.

Our human vision system has a good discriminating power and also has the support of the powerful nervous system which can spontaneously learn new things. Also our brain can store uncountable details in suitable form such that any information needed is retrieved within a fraction of a second. Thus it can easily classify different categories of images under variations of scale, orientation and illumination changes. An efficient machine vision system should try to imitate the human discrimination power for categorizing images.

For any image analysis purpose, the feature extraction is the crucial part. Many features based on color, shape and textures were reported in the literature. Thomas Deselaers et.al [1] gave a quantitative comparison of the image features for the retrieval. They explored many features including color, Gabor based texture features for applying on the Corel dataset. The texture features were used for image classification [2] and pixel classification leading to image segmentation [3].

Invariant features were helpful to analyze the images though the image under consideration are subjected to rotation and scale. Some of the work related to invariant features include Object Recognition from Local Scale-Invariant Features Transform (SIFT) by David G. Lowe [4], Texture classification with combined rotation and scale invariant wavelet features by Muneeswaran et.al [5]. Out of many invariant features reported in the literature, SIFT features were found most attractive for many researchers to proceed for further work.

Also human vision system classifies images by primarily looking at the color and shape structure of the object contained in it. In the proposed work we have the combination of features based on distribution of shape features which produce a signature holding edge details and its attributes. The rest of the paper is arranged as follows: section 2 discusses the related works in the literature, section 3 gives the proposed image classification system and its implementation, section 4 highlights the results and discussion

2. Related Work

Many works were reported in the literature on image classification and retrieval based on keywords. Afterwards researchers focused on Content Based Image Retrieval. However all the methods are used for syntax based image classification/retrieval. However, over the time there was a need for the retrieval based on semantics. Hence the model of word based classification of documents was considered by the researchers. Li Fei-Fei [6] proposed a Bayesian Hierarchical Model for learning the natural scenes. They introduced the concept of visual word and collection of visual words as code book to represent the set of images. Learning the

similar parts (region) across multiple images and abstracting them into visual words led to important paradigm shift in image analysis.

Pierre et.al [7] modeled the image as distribution of the visual words for image classification. Also they addressed the issues associated with the spatial relationship between key points and elimination of noisy words. Bag of features was also used by Hervé Jégou et.al [8] for describing the image. Each descriptor of the image is projected to a matrix obtained by the QR factorization of a random matrix to get the reduced dimension in the feature space. Each descriptor in the image is matched with a dictionary word and a binary encoding is done for the descriptor with the reduced dimension. The resulting binary vector for each descriptor is used to construct the term frequency for matching dictionary word in contrast to other descriptors in the image. Both computation time and storage requirement are very high. The experiments were conducted on Holidays, Oxford5k and University of Kentucky object recognition benchmark datasets. A survey of Bag of Features (BoF) was outlined by Stephen O'hara et.al [9] and they reported the techniques that mitigate quantization errors, improve feature detection, and speed up image retrieval. It provided a rich literature on the various techniques for constructing the BoF model. They also pointed out some of the important challenges of BoF model such as lack of spatial information and missing semantics. Also Veronika Cheplygina et al [10] gave a detailed survey of the different learning scenarios, which specify the methods for encoding the spatial relationship between the features.

The concept of Bag of Words, where the dictionary was created with the SIFT features by quantizing them using Over complete sparse coding was proposed by Chong Wang and Kaiqi Huang [11]. For the image representation, max pooling and average pooling were used and the experiments were conducted on the datasets such as Caltech 101, Caltech 256, Pascal VOC 2007 and ImageNet. Support Vector Machine was used to classify the images and the experiment was conducted for varying dictionary size. However no mention was there about the time taken to build the dictionary. But these Bag of Words model represent the local features of the image whereas they could not express the spatial relation between the local features. This was taken into consideration by Fernanda B.Silva et.al [12] who considered not only Bag of Visual Words but also the spatial relationship among the visual words in the form of visual graph. Dense SIFT features and k-means clustering were used to build the dictionary. The additional spatial information yielded improved results compared to normal Bag of Words and the experiments were conducted with Caltech-101 and Caltech-256 datasets. Scene categorization based contextual visual words was proposed by Jianzhao Qin et.al [13] where the contextual information surrounding the regions of interest is added to identify the appropriate visual word. The experiments were conducted on dataset with 8, 13 and 15 scene categories, respectively. They extracted SIFT features at different scales to make the context awareness. Bag of Local Binary Patterns from Three Orthogonal Planes were used to detect anomaly events in the visual scenes by Jingxin Xu [14]. Spatial Temporal patches are selected and subjected to LDA classifier to identify the region of interest, from which the visual texture features are extracted to construct the Bag of word patterns.

Avila et.al [15] proposed the pooling of descriptors towards the codeword. Instead of scalar value of the count of descriptors as used by the many researchers, they described a distance histogram at each codeword. This histogram of many bins if converted into one bin represents the traditional signature. However the distance histogram at each code word represents the consistent generalization of Bag of words pooling. They have applied these techniques to classify the videos into pornography and non-pornography. Rahat Khan et al [16] improved the performance of Bag of Visual Words model by augmenting spatial information between visual words. Also they introduced soft pair wise voting scheme for distance computation.

They conducted experiments on challenging data sets such as MSRC-2, 15Scene, Caltech101, Caltech256 and Pascal VOC 2007.

Bharath Ramesh et al [17] classified shapes having scale and rotation variations using Bag of Words model with the log polar transform features. Spatial co-occurrence matrices were constructed and bi-grams are selected to represent the spatial information efficiently. They used a special metric called weighted gain ratio for identifying the suitable dictionary size. They tested and proved their work on animal shapes dataset. Bag of Features model is used not only for image processing, but also used for text processing. In text processing, E. Khalifa et al [18] extracted grapheme features from handwritten documents. From these random grapheme features, they constructed multiple code book vectors. They also employed spectral regression for dimensionality reduction. They claim that their work performs better in matching a handwritten document with its author.

Lingxi Xie et al [19] used the bag of features model and slightly modified the pooling and normalization. After code book construction two types of quantization can be carried out: hard and soft quantization. They applied soft quantization where each descriptor is mapped as a sparse vector which denotes some of the closer visual words. Special pooling technique was made for constructing a better group of spatial bins. Their approach was tested for three major applications of image classification - scene recognition, general object recognition, and fine-grained object recognition with many ground truth datasets. Efficient Code book generation can be done by applying various nature inspired optimization algorithms [20-26]. All these works indicate the significance of Bag of Words model which has challenges such as lack of spatial information, high computation cost and large storage requirement. In our proposed work, some of the challenges such as high computations time, construction of the appropriate code words are addressed.

3. Proposed Work

In the semantic based image retrieval the image is represented by the visual word distribution. The visual word is mainly characterized by the interest points which are to be invariant features. SIFT features [4] are one of the dominant visual features used to construct the Bag of Visual Words model. From the SIFT features, the dictionary/code book creation is done using clustering techniques. Also the dictionary creation and bag of words formation is highly computational intensive. Hence to represent the shape of the objects, we have to use the edges in the form of gradient information (K), radius (R), and its orientation (B). Fig. 1 shows the image classification system using multiple dictionary of SIFT features whereas Fig. 2 shows our proposed image classification system using distribution of multiple shape features.

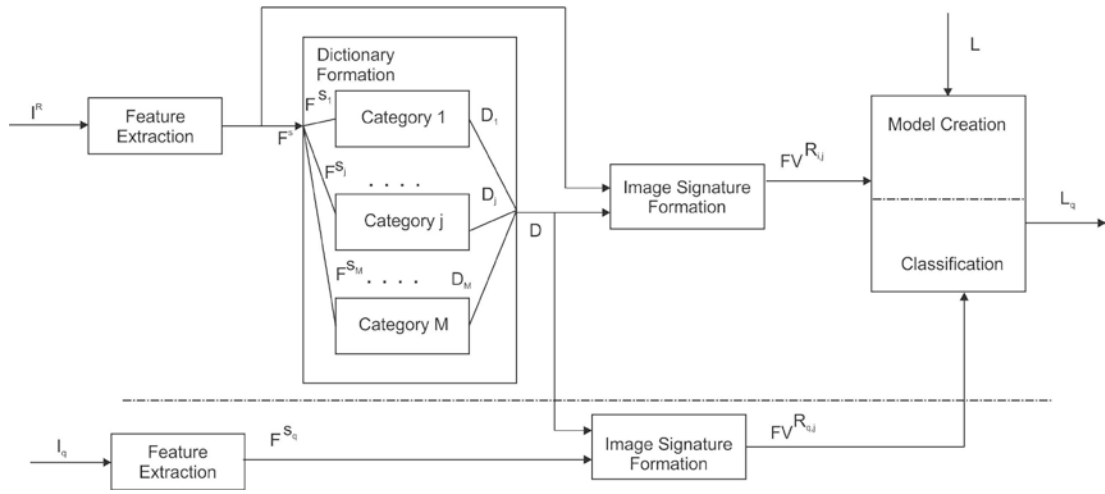


Fig. 1. Classification using single / multiple dictionary of SIFT features

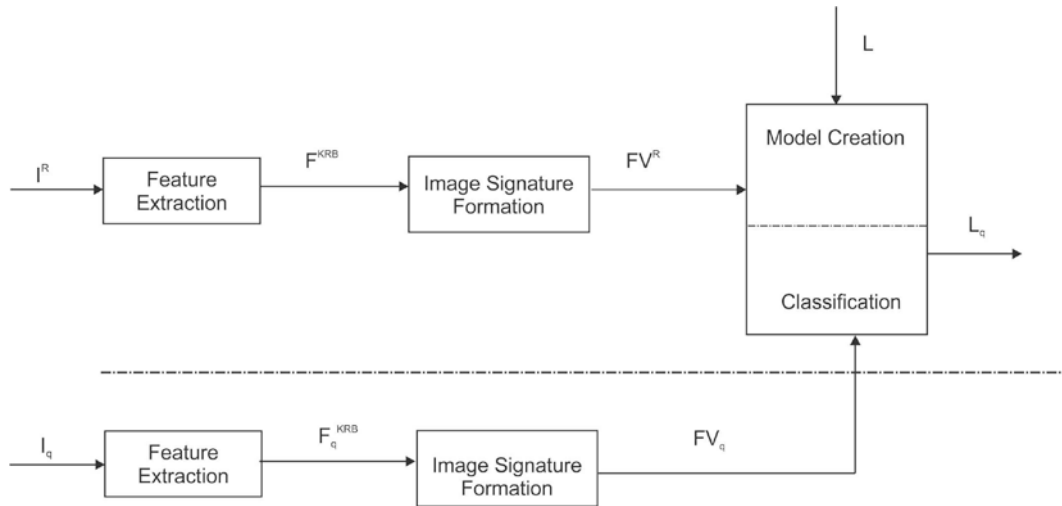


Fig. 2. Classification using using KRB features

Notations Used

N - Number of training images

I^R - Set of training images $\{I_1^R, I_2^R \dots I_i^R \dots, I_N^R\}$

I_q - Query image

F^S - Sift features of all images

F_j^S - Sift features for images in j^{th} category

F^{KRB} - KRB features

D - Single Dictionary of all training images

D_j - Multiple Dictionary for j^{th} category images

$FV_{i,j}^R$ - j^{th} category signature of training Image i

$FV_{q,j}$ - j^{th} category signature of query Image

L - Image labels of training images $\{L_1, L_i, \dots, L_N\}$

L_q - Classified label of query image

3.1 Feature Extraction

For each image in the training set I_i^R in I^R , SIFT features are extracted and combined to give F^S . Each SIFT key point is 128 dimensional feature vector. The number of SIFT key points for each image is not uniform. The large number of resulting SIFT features are clustered to form code book. Clustering very large number of SIFT key points will lead to intolerable computational overhead. Hence the proposed work shows the alternate way of representing the shape of the object. Hence KRB features are extracted where K denotes the distribution of image gradients, R denotes the distribution of radius and B denotes the distribution of radii angle.

Fig. 3 shows an arbitrary shaped object. Initially the image gradients along the horizontal and vertical direction are computed for each pixel as G_x and G_y . The magnitude (Mag) of gradients and angle of orientation are calculated as in (1):

$$G_x = \frac{\partial I_i}{\partial x}, \quad G_y = \frac{\partial I_i}{\partial y}$$

$$Mag = \sqrt{G_x^2 + G_y^2} \quad (1)$$

$$\theta = \tan^{-1} \left(\frac{G_y}{G_x} \right)$$

The gradients determine the edges of the image which also includes many noisy pixels along with the edge pixels. Noise is eliminated by using the simplest technique called thresholding. The filtered edge pixels contribute highly to describe the shape / contour of the object of interest. A centroid (x_c, y_c) around the edges of the object is found. From each boundary point (x_e, y_e) of the object, radii (r) to the centroid is computed. Since the edge points can be present to the left and right of the centroid, the direction of the edge pixel is indicated by representing the radii as a signed number. Hence the radius and sign of the radius are represented as in (2) and (3) :

$$r = \sqrt{(x_e - x_c)^2 + (y_e - y_c)^2} \quad (2)$$

$$\text{sgn}(r) = \begin{cases} +1, & \text{if } y_c < y_e \\ -1, & \text{if } y_c > y_e \end{cases} \quad (3)$$

where y_c and y_e are the y co-ordinates of the centroid and the edge pixel respectively.

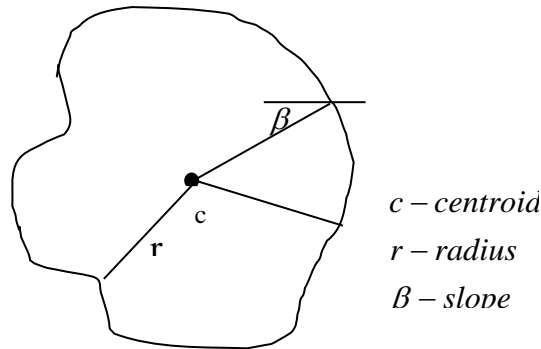


Fig. 3. Arbitrary shape Representation

The radial line makes an angle incident on the boundary of the object, which is considered as another feature representing the orientation of the object denoted as β and computed as in (4):

$$\beta = \tan^{-1} \left(\frac{y_c - y_e}{x_c - x_e} \right) \quad (4)$$

These resulting features from the object such as gradient orientation (θ), radii (r) and object orientation (β) are made use for the construction of the feature vector of the object and in turn the image feature.

3.2 Dictionary Formation

Out of the features extracted, varying and large number of SIFT features are extracted from each image. A visual word under the context of SIFT features is the cluster center of group of SIFT features. Hence a Bag of Visual Words (BoVW) are formed from the extracted SIFT features. The BoVW is also called the dictionary (D) or Code Book (CB) resulting from the images under consideration. The process of constructing the CB involves the SIFT features extraction for the entire set of images used to build knowledge model of the proposed system and clustering them. Most of the existing works use KMeans clustering for forming the dictionary. But KMeans clustering technique updates the weights of the cluster centers alone and is highly sensitive to noise. Also in order to have better clustering, each cluster requires to have relatively equal number of observations. Hence alternate clustering process such as Kohonen Self Organizing Map [27] SOM is used for the clustering.

SOM is a neural network which not only updates the weight of the cluster centers (winning node) but also updates the weights of the neighboring nodes of the winning node thereby SOM explores the state space in a better way compared to KMeans clustering, which is also an iterative process. However when using SOM, the learning time is high and hence the computation cost is very high for constructing the dictionary. Another advantage of using SOM is to have different topological structures for arriving at the cluster center mimicking the human brain in learning. To overcome the high learning time, multiple dictionaries $[D_1, D_2, D_3, \dots, D_M]$ are created, one for each of the M categories as proposed by Umit Lutfu et al [28]. This process of multiple dictionary construction can be performed in parallel and thereby the computation cost could be reduced. Each dictionary includes the features of all trained images of a particular category. Initially, the weights associated with each neuron in the SOM are

randomly assigned. Hence the clustering obtained by both KMeans and SOM is mainly dependent on the initialization.

The weights (w) associated with the winning and neighboring neurons during the iteration $t+1$ are updated as in (5):

$$w_{t+1} = w_t + h_c(t)(x - w_t) \quad (5)$$

where $h_c(t)$ is the neighborhood function of the winning neuron which is computed normally as a Gaussian function as in (6):

$$h_c(t) = e^{-\frac{[d(r_c, r_i)]^2}{2\sigma^2(t)}} \quad (6)$$

where the distance $d(r_c, r_i)$ is the distance between the winning neuron (r_c) and neighboring neuron (r_i).

3.3 Image Signature Formation

Using the BoVW and the SIFT features extracted from each image the signature of the image is obtained. The signature of the image I_i ($i=1\dots N$) with dictionary of size NC is represented as $FV_{i,j,k}^R$, which is initialized to 0.

Here $i=1\dots N$ (number of training images), $j = 1\dots M$ (number of categories), $k=1\dots NC$ (number of clusters - Dictionary size) and R denotes the training images and the signature is found as in (7):

$$FV_{i,j,k}^R = FV_{i,j,k}^R + 1 \quad \text{for } k = \arg \min_k (\text{dist}(VW_k, F_{i,l}^S)) \quad (7)$$

where $F_{i,l}^S$ is the l^{th} SIFT feature associated with the i^{th} image. The signature of the images will be represented as in Fig. 4.

$$\begin{matrix} & 1 & & j & & M \\ \begin{bmatrix} FV_{1,1,1}^R & \dots & FV_{1,1,k}^R & \dots & FV_{1,1,NC}^R \\ \vdots & & \vdots & & \vdots \\ FV_{i,1,1}^R & \dots & FV_{i,1,k}^R & \dots & FV_{i,1,NC}^R \\ \vdots & & \vdots & & \vdots \\ FV_{N,1,1}^R & \dots & FV_{N,1,k}^R & \dots & FV_{N,1,NC}^R \end{bmatrix} & \begin{bmatrix} FV_{1,j,1}^R & \dots & FV_{1,j,k}^R & \dots & FV_{1,j,NC}^R \\ \vdots & & \vdots & & \vdots \\ FV_{i,j,1}^R & \dots & FV_{i,j,k}^R & \dots & FV_{i,j,NC}^R \\ \vdots & & \vdots & & \vdots \\ FV_{N,j,1}^R & \dots & FV_{N,j,k}^R & \dots & FV_{N,j,NC}^R \end{bmatrix} & \begin{bmatrix} FV_{1,M,1}^R & \dots & FV_{1,M,k}^R & \dots & FV_{1,M,NC}^R \\ \vdots & & \vdots & & \vdots \\ FV_{i,M,1}^R & \dots & FV_{i,M,k}^R & \dots & FV_{i,M,NC}^R \\ \vdots & & \vdots & & \vdots \\ FV_{N,M,1}^R & \dots & FV_{N,M,k}^R & \dots & FV_{N,M,NC}^R \end{bmatrix} \end{matrix}$$

Fig. 4. Formation of Image Signature

In case of single dictionary, j is set to 1. The signature for each image is obtained by computing the frequency of occurrence of the SIFT features matching (minimum distance) with the appropriate Visual Word (VW).

The process of finding the image signature using SIFT features, clustering them into dictionary of visual words and distributing the SIFT features into various Visual Word Bins is a time consuming process and hence our proposed work is motivated.

In the proposed work, the image signature using KRB features are obtained by distributing the gradient orientation (θ) into K bins, radii (r) into R bins and object orientation (β) into B bins and combining them. The size of the image signature varies depending on the contents of the image. Experiments were conducted with varying bin sizes of each type (K, R and B) from 1 to 20 and it is optimized empirically with 27 (K=10, R=8 and B=9) and hence the storage

requirement is very less compared with the dictionary based work.

3.3 Model Creation and Classification

The proposed work attempted to classify the images with multiple objects using multi label classification. So M separate binary SVM classifiers are used for each category of images. Initially the feature vectors and the label pair (FV^R , L) are used for classifier model construction using multiple binary SVM classifiers as given by Yi Liu et al [29]. When multiple dictionaries are employed, the images belonging to that category is given a positive label whereas all the other images are labeled as negative samples.

During the process of classification, the output of the multiple classifiers is subjected to the frequency based voting and decision is taken based on the majority voting. All these processes discussed for SIFT features are of high computation cost theoretically. But the KRB features describe the image by its signature with lesser number of bits and less computation cost.

The dictionary based knowledge building and the proposal of our overall work and are shown in the algorithms in Fig. 5 and Fig. 6 respectively.

```

Algorithm model_Building_UsingBoW( $I^R$ ,  $L$ )
//Input :  $I^R$  – Collection of training images  $\{I_1^R, I_2^R \dots I_N^R\}$ 
            $L = \{L_j \mid j = 1 \dots M\}$  set of labels associated with each image in  $I^R$ 
// Output :  $\Gamma = \{\Gamma_j \mid j = 1 \dots M\}$  – Collection of classifier models
{
  for each image  $I_i$  in  $I^R$ 
     $F_i^S = \text{extractSIFT}(I_i)$ 
     $F^S = F^S \cup F_i^S$ 
  end
  case 1 : // single dictionary
     $D_j = \text{som}(F^S, NC)$ 
    //  $D_j = \{VW_1, VW_2 \dots VW_{NC}\}$ 
    //  $j = 1$  for single dictionary
    //  $j = 1 \dots M$  for multiple dictionary
  case 2 : // Multiple dictionaries
    for each class  $j = 1 \dots M$ 
       $D_j = \text{som}(F^{s,j}, NC)$ 
    end
   $FV_{i,j} = \text{formSignature}(D_j, F_i^S)$  // Eq. 6 is used
   $\Gamma_j = \text{buildSVM}(FV_{i,j}, L_j)$ 
  return  $\Gamma$ 
}

```

Fig. 5. Dictionary based Model Building Algorithm

```

Algorithm model_Building_Using_KRB( $I^R, L$ )
//Input :  $I^R$  – Collection of training images  $\{I_1^R, I_2^R \dots I_N^R\}$ 
            $L = \{L_j \mid j = 1 \dots M\}$  set of labels associated with each image in  $I^R$ 
// Output :  $\Gamma = \{\Gamma_j \mid j = 1 \dots M\}$  – Collection of classifier models
{
  for each image  $I_i$  in  $I^R$ 
     $I_e = \text{findStrongEdges}(I_i)$  // use gradients
     $\theta = \text{extractOrientation}(I_e)$  // Eq.1 is used
     $r = \text{extractRadius}(I_e)$  // Eq.2 is used
     $\beta = \text{extractSlopeAngle}(I_e)$  // Eq.3 is used
     $fv^K = \text{distribute}(\theta, K)$ 
     $fv^R = \text{distribute}(r, R)$ 
     $fv^B = \text{distribute}(\beta, B)$ 
     $FV_i = \text{combineDistributions}(fv^K, fv^R, fv^B)$ 
  end
   $\Gamma_j = \text{buildSVM}(FV_i, L_j)$ 
return  $\Gamma$ 
}

```

Fig. 6. Proposed work based Model Building Algorithm

Once the model is built, the images are classified as shown in **Fig. 7.**

```

Algorithm classify_Image( $I_q, D, \Gamma$ )
//Input :  $I_q$  – Query image
            $D = \{D_j \mid j = 1 \dots M\}$  collection of dictionaries
            $\Gamma = \{\Gamma_j \mid j = 1 \dots M\}$  – Collection of classifier models
// Output :  $L_q$  – Classified Label of query image
{
  case1: // using BoW features
     $F_q^S = \text{extractSIFT}(I_q)$ 
     $FV_{q,j} = \text{formSignature}(D_j, F_q^S)$ 
  case2: // using KRB features
     $I_e = \text{findStrongEdges}(I_q)$  // use gradient
     $\theta = \text{extractOrientation}(I_e)$  // Eq.1 is used
     $r = \text{extractRadius}(I_e)$  // Eq.2 is used
     $\beta = \text{extractSlopeAngle}(I_e)$  // Eq.3 is used
     $fv^K = \text{distribute}(\theta, K)$ 
     $fv^R = \text{distribute}(r, R)$ 
     $fv^B = \text{distribute}(\beta, B)$ 
     $FV_q = \text{combineDistributions}(fv^K, fv^R, fv^B)$ 
   $L_q = \text{classifyImage}(FV_q, \Gamma)$  // Use SVM
return  $L_q$ 
}

```

Fig. 7. Algorithm for Image Classification

3. Results and Discussion

The experiments were conducted using PASCAL VOC 2007 dataset which consists of 9963 images divided into 5011 training images and 4952 testing images. These images are grouped under 20 categories such as aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, diningtable, dog, horse, motorbike, person, pottedplant, sheep, sofa, train, tvmonitor. The experiments were conducted in the computing environment of Intel Xeon E5-2609 (Hexa Core) v3 1.9 GHz Processor and 48GB RAM. In the feature extraction phase, the Bag of Visual Words are formed from single and multiple dictionaries resulting from the SIFT features. Also the distribution of shape information with gradient (K), Radius (R) and Angle beta (B) are made use of for the construction of shape features.

The Fig. 8 shows the resulting accuracy of the proposed method and the existing methods. Visual Dictionary was constructed with Bag of Visual words. Experiments were conducted with single dictionary and multiple dictionary (for each category). The performance of the single dictionary is marginally better compared with the case of multiple dictionaries. The advantages of multiple dictionary is that the code book creation could be done in offline in parallel minimizing dictionary creation time. However, in both single and multiple dictionary creation, the computation cost is very high in the order of hours, but when the KRB based method is deployed the resulting accuracy is marginally less at the same time, the computational cost is very negligible. The average precision and recall for the KRB based method and dictionary based methods are (0.92, 0.64) and (0.99, 0.69) respectively. It is observed the proposed method shows better performance considering both the computation time and the resulting accuracy.

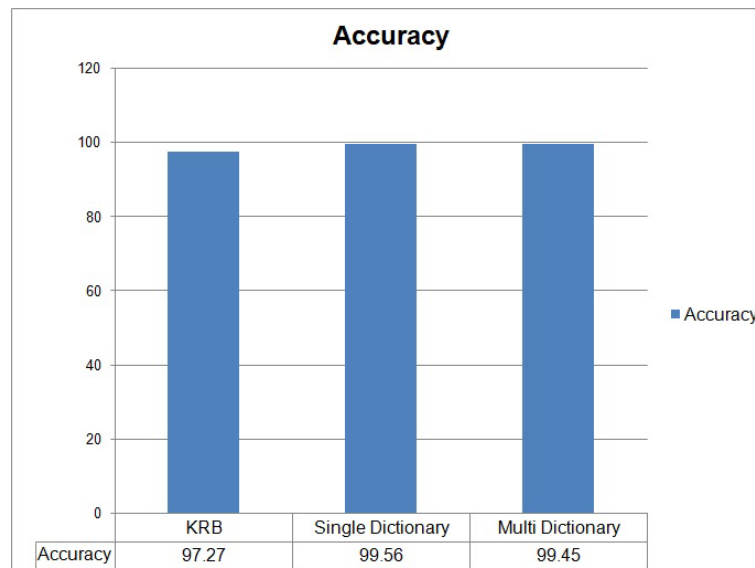


Fig. 8. Accuracy for the methods: KRB, Single Dictionary and Multiple Dictionary

Fig. 9 shows both the numerical and graphical view of the time taken for the feature extraction, training. The testing time is insignificantly less.

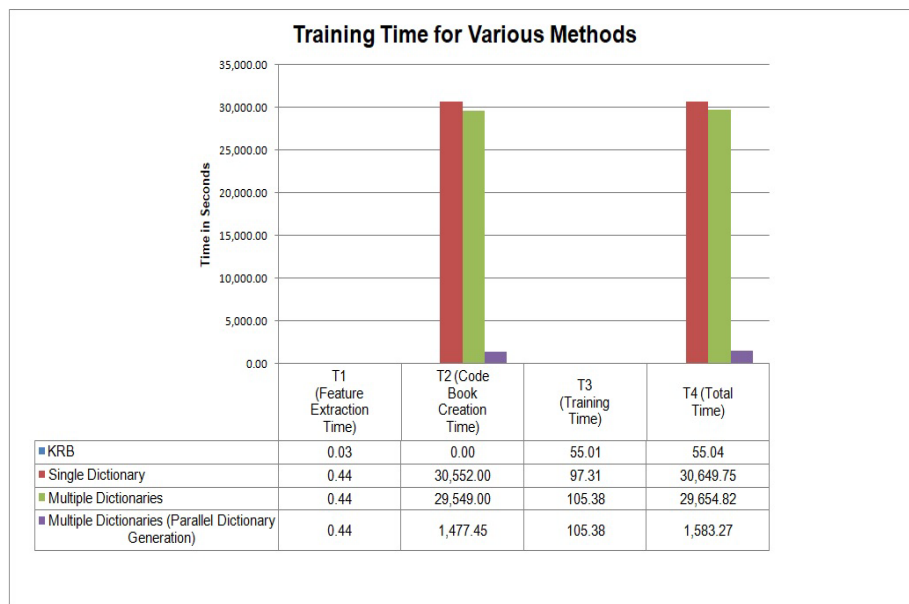


Fig. 9. Training Time for the proposed methods and other existing methods

Since multiple binary classifiers are used, voting based decision is proposed in case of the multi label classification. The accuracy shown in the previous figure is for voting threshold of 0.5. For perfect match by all the classifiers, the threshold is to be at 1.0. However majority voting is considered in the proposed work. **Fig. 10** shows the accuracy when the voting threshold of multi label classifier is varied from 0.5 to 1.0. We can observe the decline in the accuracy for increasing threshold.

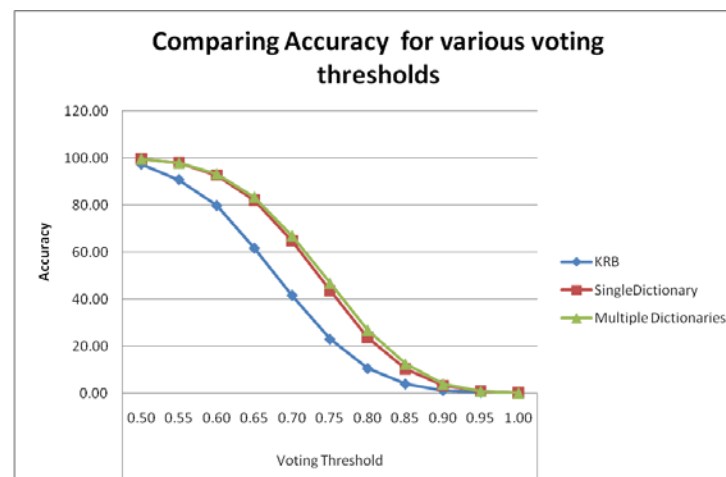


Fig. 10. Accuracy for increasing voting threshold in the classifiers

Also the size of the dictionary has an impact on the classification. **Fig. 11** shows the accuracy for the varying dictionary size. It is observed that the increase in code book size results good accuracy. But in the literature [28] Umit Lutfu et al selected the code book size as five times of square root of the number of features, which leads to large sized dictionary resulting in high storage capacity requirements and computation time. However the proposed

work has made use of the optimal code book size of 50, 100 and 200 empirically selected achieving the comparable good results (accuracy nearer to 100). However, the code book size varies depending the contents of the images under consideration.

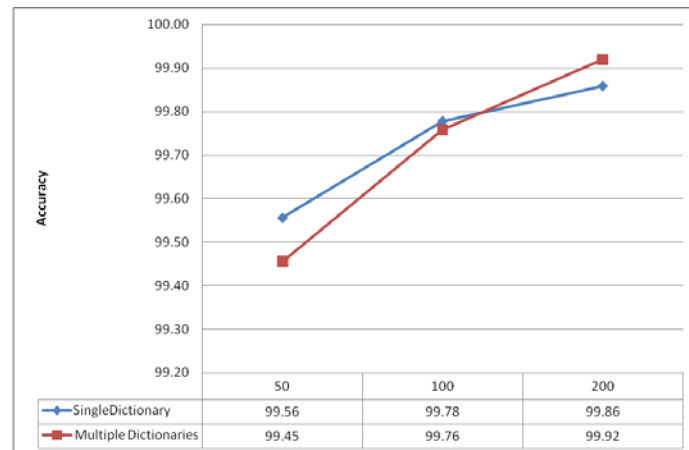


Fig. 11. Accuracy for varying dictionary size

All of the above discussions shows that the proposed work with KRB features performs well equally well similar to the multiple code book based features. The image representation using these KRB features are suitable not only for Natural Image Classification and retrieval but also more suitable for Scene Classification, Image Compression, real applications like Image based search, Unknown Object Detection in Video Surveillance, Satellite Imaging based Landscape Classification, Medical Image diagnosis such as Brain Tumor Detection, Lung Cancer Detection etc.

5. Conclusion

Multi label classification was done using multiple SVM classifiers which are trained with the BoW features and KRB features. When using the BoW features models, the image signature construction needs heavy computations due to the clustering of large number of SIFT features. In the proposed work, KRB features were used and found that the results are comparable with BoW model. The novelty of the work is in the huge reduction (hours to seconds) of the computational cost using the proposed KRB features. The experiments were conducted on the PASCAL VOC 2007 dataset. The KRB features are shape oriented, so the technique works well for classifying objects of different shapes. But since KRB features are histogram of various shape oriented features, they cannot handle the spatial information more accurately similar to BoW features. In the future, this work can be improved by segregating the salient object of interest and then classifying them.

References

- [1] Thomas Deselaers, Daniel Keysers, Hermann Ney, "Features for Image Retrieval: A Quantitative Comparison," *Pattern Recognition, Springer Publications*, pp 228-236, 2004.
[Article \(CrossRef Link\)](#)

- [2] Robert M. Haralick, K. Shanmugam, Its'hak Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, 3(6), pp. 610 - 621, 1973. [Article \(CrossRef Link\)](#)
- [3] Muneeswaran K, Ganesan L, Arumugam S and Soundar K R, "Texture image segmentation using combined features from spatial and spectral distribution," *Pattern Recognition Letters*, 27(7), pp. 755–764, 2006. [Article \(CrossRef Link\)](#)
- [4] David G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proc. of the International Conference on Computer Vision*, Corfu, Sept. 1999. [Article \(CrossRef Link\)](#)
- [5] K. Muneeswaran, L. Ganesan, S. Arumugam, K. Ruba Soundar, "Texture classification with combined rotation and scale invariant Wavelet Features," *Pattern Recognition*, 38(10), pp. 1495-1506, 2005. [Article \(CrossRef Link\)](#)
- [6] Fei-Fei Li, P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. [Article \(CrossRef Link\)](#)
- [7] Pierre Tirilly, Vincent Claveau and Patrick Gros, "Language modeling for bag-of-visual words image categorization," in *Proc. of the International Conference on Content-based image and video retrieval*, pp. 249 - 258, 2008, [Article \(CrossRef Link\)](#)
- [8] Hervé Jégou, Matthijs Douze, Cordelia Schmid, "Improving Bag-of-Features for Large Scale Image Search," *International Journal of Computer Vision*, Springer, 87, pp. 316–336, 2010. [Article \(CrossRef Link\)](#)
- [9] Stephen O'hara and Bruce A. Draper, "Introduction to the Bag of Features Paradigm for Image Classification and Retrieval," *arXiv:1101.3354*, Cornell University, Jan 2011. [Article \(CrossRef Link\)](#)
- [10] Veronika Cheplygina, David M.J. Tax, Marco Looga, "On classification with bags groups and sets," *Pattern Recognition Letters*, 59, pp. 11–17, 2015. [Article \(CrossRef Link\)](#)
- [11] Chong Wang, Kaiqi Huang, "How to use Bag-of-Words model better for image classification," *Journal of Image and Vision Computing*, vol. 38, pp. 65-74, 2015. [Article \(CrossRef Link\)](#)
- [12] Fernanda B. Silva et.al, "Image classifications based on Bag of Visual Graphs," *IEEE proceedings, ICIP 2013*, 2013. [Article \(CrossRef Link\)](#)
- [13] Jianzhao Qin, Nelson H.C. Yung, "Scene categorization via contextual visual words," *Pattern Recognition*, 43(5), pp. 1874–1888, 2010, [Article \(CrossRef Link\)](#)
- [14] Xu Jingxin, Denman, Simon, Fookes, Clinton B., & Sridharan, Sridha, "Unusual Event Detection in Crowded Scenes Using Bag of LBPs in Spatio-temporal Patches," in *Proc. of DICTA 2011, IEEE Computer Society*, pp. 549-554, 2011. [Article \(CrossRef Link\)](#)
- [15] Sandra Avila, Nicolas Thome, Matthieu Cord, Eduardo Valle, Arnaldo de A. Araújo, "Pooling Image Representation: The visual codeword point of view," *Computer Vision and Image Understanding*, 117(5), pp. 453–465, 2013. [Article \(CrossRef Link\)](#)
- [16] Rahat Khan, Cécile Barat, Damien Muselet, Christophe Ducottet, "Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model," *Journal of Computer Vision and Image Understanding*, vol. 132, no. C, pp. 102-112, 2015. [Article \(CrossRef Link\)](#)
- [17] Bharath Ramesh, Cheng Xiang, Tong Heng Lee, "Shape classification using invariant features and contextual information in the bag-of-words model," *Pattern Recognition, IET publications*, 48 (3), pp. 894-906, March 2015. [Article \(CrossRef Link\)](#)
- [18] E. Khalifa, S. Al-maadeed, M.A. Tahir, A. Bouridane, A. Jamshed, "Off-line writer identification using an ensemble of grapheme codebook features," *Pattern Recognition Letters*, 59, pp. 18–25, 2015. [Article \(CrossRef Link\)](#)
- [19] Lingxi Xie, Qi Tian, Bo Zhang, "Simple Techniques Make Sense: Feature Pooling and Normalization for Image Classification," *IEEE Transactions on Circuits and Systems for Video Technology*, 26 (7), pp. 1251-1264, July 2016. [Article \(CrossRef Link\)](#)
- [20] Dhiman, G., and Kumar, V., "Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications," *Advances in Engineering Software*, 114, 48-70, 2017. [Article \(CrossRef Link\)](#)

- [21] Dhiman, G., & Kumar, V., "Emperor penguin optimizer: A bio-inspired algorithm for engineering problems," *Knowledge-Based Systems*, 159, 20-50, 2018. [Article \(CrossRef Link\)](#)
- [22] Kaur, S., Awasthi, L. K., Sangal, A. L., and Dhiman, G., "Tunicate Swarm Algorithm: A new bio-inspired based metaheuristic paradigm for global optimization," *Engineering Applications of Artificial Intelligence*, 90, 103541, 2020. [Article \(CrossRef Link\)](#)
- [23] Dhiman, G., and Kaur, A., "STOA: a bio-inspired based optimization algorithm for industrial engineering problems," *Engineering Applications of Artificial Intelligence*, 82, 148-174, 2019. [Article \(CrossRef Link\)](#)
- [24] Gupta, V. K., Shukla, S. K., and Rawat, R. S., "Crime tracking system and people's safety in India using machine learning approaches," *International Journal of Modern Research*, 2(1), 1-7, 2022. [Article \(CrossRef Link\)](#)
- [25] Kumar, R., and Dhiman, G., "A Comparative Study of Fuzzy Optimization through Fuzzy Number," *International Journal of Modern Research*, 1, 1-14, 2021. [Article \(CrossRef Link\)](#)
- [26] M. Laxmi Prasanna Rani, Gottapu Sasibhushana Rao and B. Prabhakara Rao, "An efficient codebook generation using firefly algorithm for optimum medical image compression," *Journal of Ambient Intelligence and Humanized Computing*, 12, 4067–4079, 2021. [Article \(CrossRef Link\)](#)
- [27] Kohonen Teuvo, "Self-Organized Formation of Topologically Correct Feature Maps," *Biological Cybernetics*, 43 (1), pp. 59–69, 1982. [Article \(CrossRef Link\)](#)
- [28] Umit Lutfu, Altintakan, Adnan Yazici, "Towards effective image classification using Class - Specific Codebooks and Distinctive local features," *IEEE Transactions on Multimedia*, 17(3), pp. 323-332, March 2015. [Article \(CrossRef Link\)](#)
- [29] Yi Liu, and Yuan F. Zheng, "One-Against-All Multi-Class SVM Classification Using Reliability Measures," in *Proc. of International Joint Conference on Neural Networks*, Montreal, Canada, 2005. [Article \(CrossRef Link\)](#)



Vanitha Sivagami Sivasankaravel received her Ph.D.degree in Computer Science and Engineering from Anna University, Chennai, India in 2017 She is currently working as Associate Professor in the Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India. Her areas of interest are Image Processing, Internet of Things, Automata Theory, System Programming.