Journal of Computer Science 2012, 8 (12), 2068-2074
ISSN 1549-3636

doi:10.3844/jcssp.2012.2068.2074 Published Online 8 (12) 2012 (http://www.thescipub.com/jcs.toc)

Soft Clustering Based Exposition to Multiple Dictionary Bag of Words

[1]K.S. Sujatha and [2]B. Vinod

[1]Department of Electronics and Communication Engineering,
[2]Department of Robotics and Automation,
P.S.G. College of Technology, Peelamedu, Coimbatore-641004, TamilNadu, India

Received 2012-07-02, Revised 2012-09-11; Accepted 2012-12-19

ABSTRACT

Object classification is a highly important area of computer vision and has many applications including robotics, searching images, face recognition, aiding visually impaired people, censoring images and many more. A new common method of classification that uses features is the Bag of Words approach. In this method a codebook of visual words is created using various clustering methods. For increasing the performance Multiple Dictionaries BoW (MDBoW) method that uses more visual words from different independent dictionaries instead of adding more words to the same dictionary was implemented using hard clustering method. Nearest-neighbor assignments are used in hard clustering of features. A given feature may be nearly the same distance from two cluster centers. For a typical hard clustering method, only the slightly nearer neighbor is selected to represent that feature. Thus, the ambiguous features are not well-represented by the visual vocabulary. To address this problem, soft clustering model based Multiple Dictionary Bag of Visual words for image classification is implemented with dictionary generated using modified Fuzzy C-means algorithm using R1 norm. A performance evaluation on images has been done by varying the dictionary size. The proposed method works better when the number of topics and the number of images per topics are more. The results obtained indicate that multiple dictionary bag of words model using fuzzy clustering increases the recognition performance than the baseline method.

Keywords: Bag of Words, Multiple Dictionaries BoW, Fuzzy C-Means

1. INTRODUCTION

One of the most important and challenging problem in machine vision is retrieving images from a large and highly varied image data set based on visual contents. In the present scenario there are currently several smart phone applications that allow the users to take a photo which has led to the rapid growth in the number of digital image collections. Automatic classification of images will be helpful in efficient search and management of these large collections of images. A new method of classification that uses features is the Bag of Words (Lazebnik *et al*., 2006) approach. This is an idea that solves the problem of recognition with an approach starting from visual features and not from segmentation. The first step in classifying images using Bag of Words is creating a codebook of visual words. For this features are extracted using detectors or dense sampling and descriptors are calculated at each and every local keypoints extracted. For local feature detection, classic detectors include Harris detector (Harris and Stephens, 1988) and its extension (Tuytelaars and Gool, 2004) and many more. For local feature description, local descriptors such as Haar descriptor (Viola and Jones, 2001), Scale-Invariant Feature Transform (SIFT) descriptor (Lowe, 2004), Histogram of Gradients (HOG) descriptor (Dalal and Triggs, 2005) and Speeded Up Robust Feature descriptor (SURF) (Bay *et al*., 2006) are commonly used.

In this study Bag of Words model has been implemented for visual categorization of images using Harris corner detector for extracting features and Scale

Corresponding Author: K.S. Sujatha, Department of Electronics and Communication Engineering, P.S.G College of Technology, Peelamedu, Coimbatore-641004, TamilNadu, India

Invariant Feature descriptor (SIFT) for representing the extracted features. After obtaining local features called descriptors, a codebook is generated to represent them. The overall performance of BoW depends mainly on the dictionary generation method and therefore in this implementation the method of generation of the dictionary of visual words is being focused. A novel method using Multiple Dictionaries for BoW (MDBoW) (Aly *et al*., 2011a,b) using soft clustering algorithm Fuzzy C-means with R1norm (Sujatha and Vinod, 2012) which uses more visual words is implemented. This method significantly increases the performance of the algorithm when compared to the baseline method for large scale collection of images which uses Bag of Words method. In baseline method, more words are added to the same dictionary whereas in MDBoW more words are taken from different independent dictionaries. The resulting distribution of descriptors is quantified by using vector quantization against the pre-specified codebook to convert it to a histogram of votes for codebook centers. K Nearest Neighbor algorithm (KNN) is used to classify images through the resulting global descriptor vector.

2. MATERIALS AND METHODS

2.1. Methods of Soft Clustering

In traditional bag of words that uses hard clustering a given feature may be nearly the same distance from two cluster centers and the slightly nearer neighbor is selected to represent that feature in the term vector. Thus, the ambiguous features are not well-represented by the visual vocabulary. To address this problem, in this study soft clustering methods are used to construct the codebook.

2.2. Fuzzy C-Means

Given the data set X= $\{x_1, x_2, x_3,\ldots\ldots,x_N\}$, choose the number of clusters $1< c < N$, the weighting exponent $m >1$, the termination tolerance $\epsilon >0$ and the norm-inducing matrix A. The fuzzy C-means clustering (Cannon *et al*., 1986) algorithm is based on the minimization of an objective function called C-means functional given by Equation (1)

$$J(x,u,v)=\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^{m}D_{ik}{}^{2} \quad (1)$$

Where Equation 2:

$$D_{ik}^{2}=\left\|X_k-v_i\right\|^{2} \quad (2)$$

Subject to the condition Equation 3:

$$\sum_{k=1}^{c}\mu_{ik}=1 \quad (3)$$

For all value of k.

2.3. Steps for Fuzzy C-means Algorithm:

The following are the steps to be followed for implementation of the algorithm. U is the fuzzy partition matrix. The i_{th} column of U contains values of the membership function of the i-th fuzzy subset of X. $U^{(0)}$ is the initial partition matrix. Initialize the partition matrix randomly, such that $U^{(0)} \in M_{fc}$. $X= \{x_1, x_2, x_3,\ldots\ldots,x_N\}$ is the given data set and $v = (v_1, v_2 \ldots\ldots v_c)$ are the vectors of centers. C is the number of clusters in X.

Compute the cluster prototypes (means) Equation 4:

$$V_i^{(l)}=\frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^{m}X_k}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^{m}},1\le i\le c \quad (4)$$

For l = 1, 2, 3, ….. .where v_i is the cluster centers calculated using the membership function.

Compute the distances Equation 5:

$$D_{ikA}^{2}=(X_k-V_i)^{T}A(X_k-V_i)1\le i\le c,1\le k\le N \quad (5)$$

where, A = I for Euclidean Norm and D_{ikA}^{2} is the distance matrix containing the square distances between data points and cluster centres.

Update the partition matrix Equation 6:

$$\mu_{i,k}^{(l)}=\frac{1}{\sum_{j=1}^{c}(D_{ikA}/D_{jkA})^{2/m-1}}$$
$$\text{Until}\left\|U^{(l)}-U^{(l-1)}\right\|<\epsilon \quad (6)$$

The result of the partition is collected in structure arrays. ϵ is the maximum termination tolerance and m is the fuzziness weighting exponent.

2.4. Modified Fuzzy C-Means

In the existing Fuzzy C means algorithm the objective function is defined in terms of mean squared error. In the proposed method instead of taking mean squared error the objective function is defined in terms of root mean squared error using R1 norm (Sujatha and Vinod, 2012). The root mean squared error is more

sensitive than other measures to the occasional large error and the squaring process gives disproportionate weight to very large errors. In matrix form X = (x_{ik}), index k sum over spatial dimensions, i = 1, · ·, c and index k sum over data points, k = 1, · · ·, N R1-norm is defined as Equation 7:

$$\|X\|_{R_1} = \sum_{i=1}^{c}\left(\sum_{k=1}^{N} x_{ik}^2\right)^{1/2} \quad (7)$$

It has been proved that R1-K-means performs slightly better than standard K-means (Ding *et al*., 2006).

The cost function to be minimised is given by Equation 8 and 9:

$$J(x,u,v) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^{m} D_{ik} \quad (8)$$

$$D_{ik} = (\|x_k - v_i\|^2)1^{1/2} \quad (9)$$

where, V = {v_1,v_2,…..v_c}, N is the number of classes and m is the smoothing parameter which controls fuzziness. When m = 1, μ_{ik}= 0 or 1 and it is hard partition as m increases the partition becomes more fuzzy.

2.5. Steps for Modified Fuzzy C-means Algorithm

The following are the steps to be followed for implementation of the algorithm .Given the data set X, choose the number of clusters 1< c < N, the weighting exponent m >1, the termination tolerance ϵ >0 and the norm-inducing matrix A. U is the fuzzy partition matrix. The i_{th} column of U contains values of the membership function of the i-th fuzzy subset of X. $U^{(0)}$ is the initial partition matrix. Initialize the partition matrix randomly, such that $U^{(0)} \in M_{fc}$.

X= {x_1, x_2, x_3,……, x_N} is the given data set and v = (v_1, v_2 . . . . . v_c) are the vectors of centers. c is the number of clusters in X. The objective function J is to be minimised such that the root mean squared error between the original vectors and the reallocated centers is minimised.

Compute the cluster prototypes (means) Equation 10:

$$V_i^{(l)} = \frac{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^{m} X_k}{\sum_{k=1}^{N}\left(\mu_{ik}^{(l-1)}\right)^{m}}, 1 \le i \le c \quad (10)$$

For l = 1,2…. where vi is the cluster centres calculated using the membership function.

Compute the distances Equation 11:

$$D_{ikA}^2 = (X_k - V_i)^T A(X_k - V_i) \quad 1 \le i \le c, 1 \le k \le N \quad (11)$$

where, A = I for Euclidean Norm and D_{ikA}^2 is the distance matrix containing the square distances between data points and cluster centres.

Update the partition matrix Equation 12:

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^{c}(D_{ikA} / D_{jkA})^{2/m-1}} \quad \text{Until } \left\|U^{(l)} - U^{(l-1)}\right\| < \in \quad (12)$$

The result of the partition is collected in structure arrays. ϵ is the maximum termination tolerance and m is the fuzziness weighting exponent.

2.6. Baseline Method

In baseline method, features are extracted using Harris corner detector and SIFT descriptor is used for representing the extracted features. The extracted feature pool is then clustered using the modified FCM to get a codebook with predefined number of visual words. Features extracted from training images are assigned to the nearest code in the codebook. The image is reduced to the set of codes it contains, represented as a histogram. The normalized histogram of codes is exactly the same as the normalized histogram of visual words. The k closest points from training data is found in testing phase, for the test data point and classification is done using KNN classifier.

2.7. Multiple Dictionary Bag of Words Model

Multiple Dictionaries for BoW (MDBoW), which uses more visual words, have significantly increased the performance of classification of images from a large and highly varied image data set. In MDBoW model implemented in this study, features are extracted using Harris corner detector and SIFT descriptor is used for representing the extracted features. In multiple dictionary generation from each dictionary D_N which is generated with a different subset of the image features each training image gets a histogram h_N from every dictionary D_N which is concatenated to form a single histogram h. Every feature gets N entries in the histogram h, one from every dictionary. In this approach, more words are taken from different independent dictionaries where as in base line method more words will be taken from same dictionary. Thus multiple dictionary method has less storage than baseline approach. In this study Separate dictionary implementation of Multiple Dictionaries for BoW (MDBoW) is implemented using modified FCM.

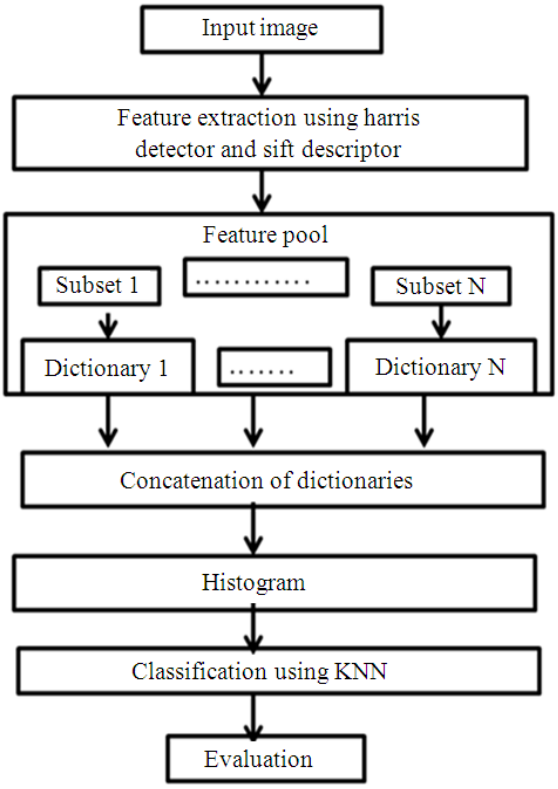

Fig. 1. Schematic for Separate dictionary implementation with FCM Clustering

2.8. Steps for Separate Dictionary Generation

Figure1 shows the schematic of Separate dictionary implementation:

- Generate N random possibly overlapping subsets of the image features
- Compute a dictionary D_N independently for each subset S_N using the modified FCM. Each dictionary has a set of K_N visual words
- Compute the histogram. Every image feature gets its visual word from every dictionary D_N. Accumulate these visual words as individual words into individual histogram h_N for each dictionary. The final histogram is the concatenation of the individual histograms

This process of histogram construction is done during the training and the testing phase of the algorithm. The KNN classifier then finds the k closest index and gives the classification result.

3. RESULTS AND DISCUSSION

The effect of variation of different parameters and performance evaluation of MDBoW approach for image classification is done in terms of Micro Precision, Macro Precision, MicroF1-measure, MacroF1-measure and Accuracy rate (Al-Salemi and Ab Aziz, 2011) for eight different topics namely burger, spaghetti, egg, spoon, bottle, can, coffee pot and mug from dataset created from Google images. The dataset is created for real time application for visual recognition of objects for a humanoid used in restaurant environment. The images in the dataset used can be categorised as tiny images. The performance measures used in this evaluation are Equation 13-18:

- Macro Precision

$$P_{macro}\frac{1}{|c|}\sum_{i=1}^{|c|}\frac{TP_i}{TP_i + FP_i} \tag{13}$$

- Micro Precision

$$P_{micro} = \frac{\sum_{i=1}^{|c|} TP_i}{\sum_{i=1}^{|c|} TP_i + FP_i} \tag{14}$$

- Macro F1- measure

$$F = \frac{2 * P_{macro} * R_{macro}}{P_{macro} + R_{macro}} \tag{15}$$

Where:

$$P_{macro} = \frac{1}{|c|}\sum_{i=1}^{|c|}\frac{TP_i}{TP_i + FN_i} \tag{16}$$

- Micro F1- measure

$$F = \frac{2 * precision * recall}{precision + recall} \tag{17}$$

- Accuracy rate

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

Table 1. Macro Precision for different words per dictionary

Words per dictionary	Base line method	MDBoWFCM					MDBoWMODFCM				
		Dic 1	Dic 2	Dic3	Dic 4	Dic 5	Dic 1	Dic 2	Dic3	Dic 4	Dic 5
80	0.5547	0.5908	0.6003	0.5596	0.6109	0.5711	0.5896	0.5877	0.5805	0.57789	0.5615
120	0.5976	0.6488	0.6331	0.6255	0.5864	0.6023	0.5758	0.5949	0.5862	0.61010	0.6047
160	0.5748	0.6197	0.6374	0.6531	0.6285	0.6767	0.6473	0.6475	0.6776	0.60920	0.6354
200	0.5361	0.6378	0.6021	0.5919	0.6019	0.5881	0.5680	0.5964	0.6174	0.61070	0.6276

Table 2. Accuracy rate for word per dictionary 160 for various numbers of dictionaries

No: of dictionary	1	2	3	4	5
Accuracy rate (MDBoWFCM)	0.9137	0.9100	0.914	0.9075	0.92
Accuracy Rate (MDBoWM ODFCM)	0.9125	0.9128	0.920	0.9031	0.91

Fig. 2. Sample images from dataset

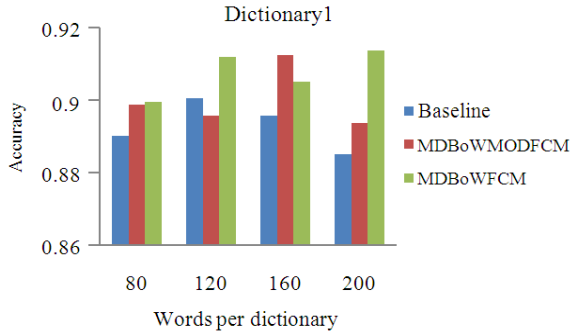

Fig. 3. Accuracy vs. words per dictionary for Dictionary1

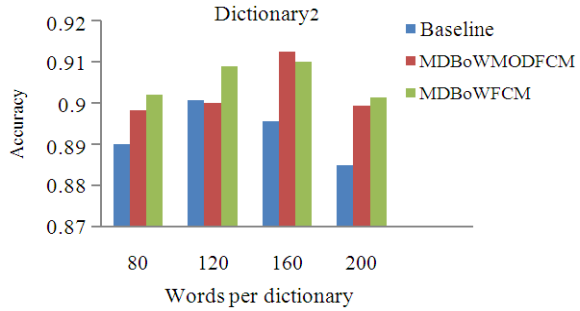

Fig. 4. Accuracy vs. words per dictionary for Dictionary2

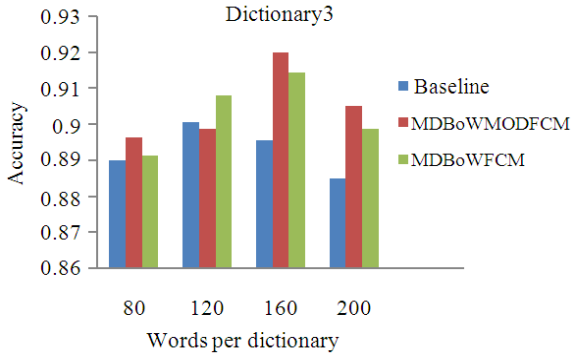

Fig. 5. Accuracy vs. words per dictionary for Dictionary3

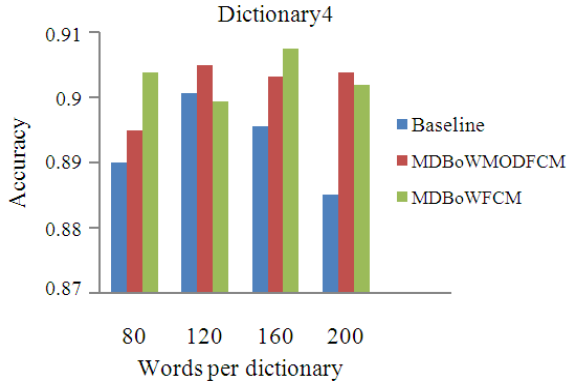

Fig. 6. Accuracy vs. words per dictionary for Dictionary4

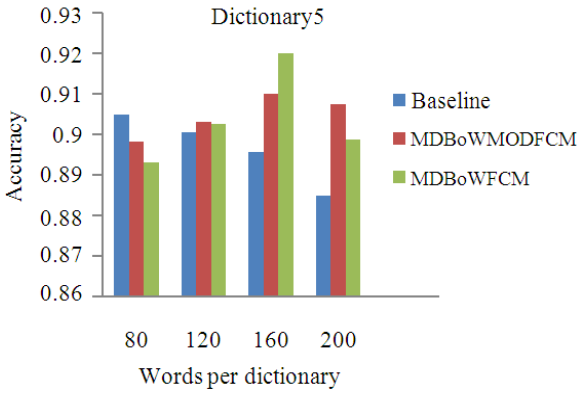

Fig. 7. Accuracy vs. words per dictionary for Dictionary5

Table 3. Micro Precision for different words per dictionary

		MDBoWFCM					MDBoWMODFCM				
Words per dictionary	Base line method	Dic 1	Dic 2	Dic3	Dic 4	Dic 5	Dic 1	Dic 2	Dic3	Dic 4	Dic 5
80	0.5600	0.5975	0.6075	0.5650	0.6150	0.5725	0.5950	0.5925	0.5850	0.5810	0.5675
120	0.6025	0.6475	0.6350	0.6325	0.5965	0.6100	0.5825	0.6120	0.5950	0.6200	0.6125
160	0.5825	0.6200	0.6400	0.6570	0.6300	0.6800	0.6500	0.6530	0.6850	0.6115	0.6430
200	0.5400	0.645	0.6050	0.5950	0.6075	0.5950	0.5750	0.5975	0.6193	0.6150	0.6290

Table 4. Macro F1 for different words per dictionary

		MDBoWFCM					MDBoWMODFCM				
Words per dictionary	Base line method	Dic 1	Dic 2	Dic3	Dic 4	Dic 5	Dic 1	Dic 2	Dic3	Dic 4	Dic 5
80	0.572	0.6047	0.6143	0.5868	0.6307	0.5884	0.5981	0.6024	0.6022	0.5937	0.5811
120	0.610	0.6708	0.6479	0.6432	0.6100	0.6278	0.5909	0.6185	0.6072	0.6265	0.6230
160	0.587	0.6289	0.6439	0.6635	0.6366	0.6821	0.6575	0.6576	0.6841	0.6184	0.6413
200	0.546	0.6558	0.6121	0.6037	0.6109	0.6015	0.5846	0.6088	0.6304	0.6209	0.6419

Table 5. Micro F1 for different words per dictionary

		MDBoWFCM					MDBoWMODFCM				
Words per dictionary	Base line method	Dic 1	Dic 2	Dic3	Dic 4	Dic 5	Dic 1	Dic 2	Dic3	Dic 4	Dic 5
80	0.5547	0.5908	0.6003	0.5596	0.6109	0.5711	0.5896	0.588	0.5805	0.57789	0.5615
120	0.5976	0.6488	0.6331	0.6255	0.5864	0.6023	0.5758	0.595	0.5862	0.61010	0.6047
160	0.5748	0.6197	0.6374	0.6531	0.6285	0.6767	0.6473	0.648	0.6776	0.60920	0.6354
200	0.5361	0.6378	0.6021	0.5919	0.6019	0.5881	0.5680	0.596	0.6174	0.61070	0.6276

In these equations TP indicates true positive, FP false positive, FN false negative and TN true negative of the classification result. For the modified Fuzzy C means and FCM the parameter m = 1.7 and stop condition ϵ = 0.001. The test data set includes eight different topics each containing 50 images. 200 images per concept were used to build the codebooks. The classifier is trained for another 200 images from each topic. The number of dictionaries formed randomly is varied from 1 to 5 and the word per dictionary is varied from 80 to 200. The distance measure used is Euclidean distance. The sample images from dataset are as shown in **Fig. 2**.

Figure 3-7 shows the variation of accuracy rate with words per dictionary by varying the number of dictionary generated randomly from 1 to 5 which is named as Dictionary1, Dictionary2, Dictionary3, Dictionary4 and Dictionary5. In both baseline method and Multiple Dictionary Bag of Words model the clustering of words are done using modified Fuzzy C means soft clustering algorithm using R1 norm. The results obtained are compared with the Multiple Dictionary Bag of Words model with FCM (Sujatha *et al*., 2012).

As the number of words per dictionary is increased from 80 to 200, accuracy increases and reaches a maximum for a particular value of word per dictionary and then reduces. The results obtained shows that Multiple Dictionary Bag of Words model using modified Fuzzy C means soft clustering algorithm using R1 norm gives the maximum accuracy rate for words per dictionary of 160 and it is more than baseline and MDBoW with dictionary formed using FCM. As the number of dictionaries generated increases the classification accuracy rate increases and then for a given number of dictionary the method gives maximum measure and then reduces.

The results projected in **Table 1-5** shows that Multiple Dictionary Bag of Words model using Separate dictionary and dictionary generated using modified FCM with R1 norm shows better performance than baseline method and MDBoW using FCM. The results obtained shows that the method gives maximum accuracy rate for word per dictionary of 160 and number of dictionary 3. **Table 2** shows the variation of accuracy rate for word per dictionary 160 for various numbers of dictionaries. The accuracy rate increases as the number of dictionary is increased from 1 to 5. The parameters Macro

Precision, Micro Precision, Micro F1 and Macro F1 shows better values for Multiple Dictionary Bag of Words with modified FCM. The results obtained validate that MDBoW performs better for datasets having large number of classes and more number of images per topics. Macro-averaging gives an equal weight to each category and is often dominated by the systems performance on rare categories. Micro-average is a useful measure when dataset varies in size and gives an equal weight to each document and is often dominated by the system's performance on most common categories. Macro-average method can be used to analyse how the system performs overall across the sets of data.

4. CONCLUSION

In this study, the performance of Multiple Dictionary Bag of Words model with code book generated using modified FCM with R1 norm in the objective function is investigated. The analysis is done by varying the words per dictionary and also the number of dictionaries generated. It is compared with the base line method and MDBoW with FCM for dictionary generation. In base line method more words will be taken from same dictionary where as in this approach, more words are taken from different independent dictionaries. It is seen that the method works better when the number of topics and the number of images per topics are more.

5. REFERENCES

Al-Salemi, B. and M.J. Ab Aziz, 2011. Statistical bayesian learning for automatic arabic text categorization. J. Comput. Sci., 7: 39-45. DOI: 10.3844/jcssp.2011.39.45

Aly, M., M. Munich and P. Perona, 2011a. Multiple dictionaries for bag of words large scale image search.

Aly, M., M.E. Munich and P. Perona, 2011b. Bag of words for large scale object recognition-properties and benchmark. Proceedings of the 6th International Conference on Computer Vision Theory and Applications, Mar. 5-7, Vilamoura, Algarve Portugal.

Bay, H., T. Tuytelaars and L.V. Gool, 2006. SURF: Speeded up robust features. Comput. Vis., 3951: 404-417. DOI: 10.1007/11744023_32

Cannon, R.L., J.V. Dave and J.C. Bezdek, 1986. Efficient implementation of the fuzzy c-means clustering algorithms. IEEE Trans. Patt. Anal. Mach. Intell., 8: 248-255. DOI: 10.1109/TPAMI.1986.4767778

Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 25-25, IEEE Xplore Press, San Diego, CA, USA., pp: 886-893. DOI: 10.1109/CVPR.2005.177

Ding, C., D. Zhou, X. He and H. Zha, 2006. R1-pca: Rotational invariant l1-norm principal component analysis for robust subspace factorization. Proceedings of the 23rd International Conference on Machine Learning, (ICML' 06), ACM Press, USA., pp: 281-288. DOI: 10.1145/1143844.1143880

Harris, C. and M. Stephens, 1988. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference, (AVC' 88), pp: 147-151.

Lazebnik, S., C. Schmid and J. Ponce, 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17-22, IEEE Xplore Press, pp: 2169-2178. DOI: 10.1109/CVPR.2006.68

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis., 2: 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94

Sujatha, K.S. and B. Vinod, 2012. Performance evaluation of different soft clustering algorithms for bag of words model. Eur. J. Sci. Res., 70: 228-239.

Sujatha, K.S., P. Keerthana, S. SugaPriya, E. Kaavya and B. Vinod, 2012. Fuzzy based multiple dictionary bag of words for image. Classification Proc. Eng., 38: 2196-2206. DOI: 10.1016/j.proeng.2012.06.264

Tuytelaars, T. and L.V. Gool, 2004. Matching widely separated views based on affine invariant regions. Int. J. Comput. Vis., 59: 61-85. DOI: 10.1023/B:VISI.0000020671.28016.e8

Viola, P. and M. Jones, 2001. Robust real-time object detection. Proceedings of the 2nd International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing and Sampling, Jul. 13-13, Vancouver, Canada, pp: 1-25.