International Conference on Modeling Optimisation and Computing

Fuzzy based Multiple Dictionary Bag of Words for Image Classification

K.S.Sujatha\textsuperscript{a} P. Keerthana\textsuperscript{b} S. Suga Priya\textsuperscript{b} E.Kaavya\textsuperscript{b} B.Vinod\textsuperscript{c}

\textsuperscript{a}Assistant Professor, Department of Electronics and Communication Engineering  
PSG College of Technology, Coimbatore, India, 641004.  
E-mail: ksoorya@rediffmail.com

\textsuperscript{b}UG students, Department of Electronics and Communication Engineering  
PSG College of Technology, Coimbatore, India, 641004.

\textsuperscript{c}Head, Department of Robotics and Automation, PSG College of Technology, Coimbatore, India, 641004.  
E-mail: bvinod@rediffmail.com

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. An object recognition method based on the Bag-of-Words (BoW) model is implemented were descriptors are quantized to form a visual word dictionary called codebook with the help of soft clustering algorithm. To increase the recognition rate and accuracy of detection, the concept of Multiple Dictionary Bag of Words model (MDBoW) is implemented in which the dictionaries built using soft clustering algorithm from different subsets of the features are combined. The performances of existing BoW model with fuzzy codebook and the proposed MDBoW are evaluated in terms of macro precision, micro precision, accuracy and F1 measure. The proposed algorithm gives an increased recognition rate and accuracy of detection.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [name organizer]

Open access under CC BY-NC-ND license.

Key words: MDBoW; BoW; fuzzy; Codebook

1. Introduction

Object categorization through Bag of Words model is one of the most popular representation methods for object categorization. Bag of Words (BoW) approach has shown acceptable performance because of its fast run time and low storage requirements [14, 15, 16, 17, 18]. The key idea is to quantize each extracted key point into one of visual word, and then represent each image by a histogram of the visual words. For this purpose, a clustering algorithm like K-means is generally used for generating the visual words. Appropriate datasets are required at all stages of object recognition research, including learning visual
models of object and scene categories, detecting and localizing instances of these models in images, and evaluating the performance of recognition algorithms. Image databases are an essential element of object recognition research. They are required for learning visual object models and for testing the performance of classification, detection, and localization algorithms.

The process of object recognition using bag of words has the following stages: Firstly, it extracts local features from images by detectors or dense sampling and then calculates their descriptors. For local feature detection, classic detectors include Harris detector [1] and its extension [2], maximally stable extremal region detector [3], affine invariant salient region detector [4]. For local feature description, we usually use local descriptors such as Haar descriptor [5], scale-invariant feature transform (SIFT) descriptor [6], gradient location and orientation histogram (GLOH) descriptor [7], rotation-invariant feature transform (RIFT) descriptor [8], shape context [9], histogram of gradients (HOG) descriptor [10] and speeded up robust feature descriptor (SURF) [11].

In this paper Bag of Words model has been implemented for visual categorization of images using Harris corner detector for extracting features and Scale Invariant Feature descriptor (SIFT) for representing the extracted features. After obtaining local features called descriptors, a codebook is generated to represent them. The codebook is a group of codes usually obtained by clustering over all descriptors. 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. The performance of BoW depends on the dictionary generation method, dictionary size, histogram weighting, normalization, and distance function. In this paper the method of generation of the dictionary of visual words is being focused. A novel method, Multiple Dictionaries for BoW (MDBoW) [18] using soft clustering algorithm Fuzzy C-means, that 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. Unlike baseline method, more words are used from different independent dictionaries instead of adding more words to the same dictionary. 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. Base line method

In baseline method of Bag of Words model implemented in this paper, features are extracted using Harris corner detector and SIFT descriptor is used for representing the extracted features. The extracted features of the image should be distinctive. Features should be easily detected under changes in pose and lighting. There should be many features per object. Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters. The advantages of SIFT features are locality, distinctiveness, efficiency and extensibility.

After feature extraction, clustering of the features is done by FCM clustering. Fuzzy C Mean (FCM) [12] is a data clustering technique in which a data set is grouped into clusters depending on the membership value. Fuzzy C-means is suited to identify clusters of the same geometry or the same order that is the clusters should have homogeneous order. After clustering a codebook with predefined number of visual words will be obtained. In training phase, the input vectors from the feature fool are assigned to one or more classes and any 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. It works well for large number of data and the distance metric used is good. The distance function used is Euclidean distance. Fig.1 shows the schematic of baseline method.
2.1. Fuzzy C-means Algorithm

Given the data set $X = \{x_1, x_2, x_3, \ldots, x_N\}$, 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 algorithm is based on the minimization of an objective function called C-means functional given by Equation (2.1).

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

$$D_{ik} = \|x_k - v_i\|^2$$

where $v_i$ is the cluster prototype or the cluster centre, $D_{ik}$ corresponds to the distance of the $k$th sample point from the $i$th cluster centre. The parameter $\mu_{ik}$ shall be interpreted as, the value of the membership function of the $i$th fuzzy subset for the $k$th datum. The value of $m$ varies from 1 to $\infty$ which is a real number which indicates the amount of fuzziness.

2.2. Steps for Fuzzy C-means Algorithm

The following are the steps to be followed for implementation of the algorithm. Initialize the partition matrix randomly, such that $U^{(0)} \in M_{fc}$. 

---

**Fig.1** Schematic for Base line method

**INPUT IMAGE**

**FEATURE EXTRACTION USING HARRIS DETECTOR AND SIFT DESCRIPTOR**

**FEATURE POOL**

**FUZZY CLUSTERING**

**TRAINING**

**CLASSIFICATION USING KNN**

**EVALUATION**
1. Compute the cluster prototypes (means)

\[ \mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^{N} (\mu_{i,k}^{(l-1)})^m \mathbf{x}_k}{\sum_{k=1}^{N} (\mu_{i,k}^{(l-1)})^m}, \quad 1 \leq i \leq c \]  

(2.3) for \( 1 = 1, 2, 3, \ldots \). Where \( \mathbf{v}_i \) is the cluster center calculated using the membership function.

2. Compute the distances:

\[ D_{ik}^2 = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A} (\mathbf{x}_k - \mathbf{v}_i), \quad 1 \leq i \leq c, 1 \leq k \leq N \]  

(2.4) where \( \mathbf{A} = \mathbf{I} \) for Euclidean Norm and is the distance matrix containing the square distances between data points and cluster centers.

3. Update the partition matrix:

\[ \mu_{i,k}^{(l)} = \frac{1}{\sum_{j=1}^{c} d_{ik}^{(l-1)}}^{2/m-1} \text{ Until } \| \mathbf{U}^{(0)} - \mathbf{U}^{(l-1)} \| < \epsilon \]  

(2.5)

The result of the partition is collected in structure arrays. \( \epsilon \) is the maximum termination tolerance and \( m \) is the fuzziness weighting exponent. Use of FCM algorithm requires determination of several parameters like \( c, m \), the inner product norm and the matrix norm. In addition, the set \( \mathbf{U}^{(0)} \in M_{lc} \) of initial cluster centers must be defined.

3. Multiple Dictionary Bag of Words model

Searching large scale collections of images has become an important application of machine vision. Multiple Dictionaries for BoW (MDBoW), that uses more visual words has significantly increased the performance for large scale classification of images. Multiple dictionaries can be implemented in two ways Unified and Separate. In single dictionary generation which is the baseline method a single dictionary of visual words is generated from the pool of features, which is used to generate the histogram for the image. In multiple dictionary generation each dictionary \( D_N \) is generated with a different subset of the image features. In Separate dictionary implementation the 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 Unified dictionary implementation a single unified dictionary is built from the concatenation of visual words from the dictionaries \( 1, \ldots, N \) and the image get a single histogram \( h \). Every feature gets only one entry in the histogram \( h \). 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 paper Separate dictionary implementation of Multiple Dictionaries for BoW (MDBoW) is implemented. Fig.2 shows the schematic of Separate dictionary implementation.

3.1 Steps for Separate dictionary generation

1. Generate N random possibly overlapping subsets of the image features \( \{S_n\}_{n=1}^{N} \)

2. Compute a dictionary \( D_n \) independently for each subset \( S_n \). Each dictionary has a set of \( K_n \) visual words.
3. Compute the histogram. Every image feature gets its visual word from every dictionary $D_n$. Accumulate these visual words as individual words into individual histograms $h_n$ for each dictionary. The final histogram is the concatenation of the individual histograms.

3.2 Multiple Dictionary Bag of Words model with FCM Clustering

In this paper, Separate dictionary concept has been implemented with Fuzzy C-means algorithm. Fuzzy clustering is the process of assigning membership levels and then using these member ship levels data elements are assigned to one or more clusters. The advantage of soft clustering is that it is insensitive to noise. In many real situations, fuzzy clustering is more natural than hard clustering, as objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial memberships. The schematic for Separate dictionary generation using fuzzy clustering is shown in Fig.3.

Features are extracted from the images using Harris corner detector and represented using SIFT descriptor. From the feature pool $N$ subsets of features are taken randomly and $N$ dictionaries are generated using Fuzzy C-means algorithm. For each of the dictionary generated histograms are generated for each image in the dataset and the final histogram is the concatenation of the individual histograms. This is done during the training phase of the algorithm. During the testing phase features are extracted from each image and histogram for the image is generated by the same process as stated above. The KNN classifier then finds the $k$ closest index and gives the classification result.

4. Experimental Result

Bag of words model for visual categorization of large scale images has been implemented using Harris corner detector for extracting features and 128 dimensional scale invariant feature descriptor (SIFT) for representing the extracted features. The features extracted are clustered using Fuzzy C-means algorithm and a code book is generated with each vector in it being a visual word which serves as the basis for indexing the images. Images are then represented as histogram counts of these visual words. K nearest neighbour algorithm (KNN) is used to classify images.

The performance of Bag of Words depends on dictionary generation method, dictionary size, histogram weighting, normalization, and distance function. In the proposed method the performance of Multiple
Dictionary Bag of Words model using Separate dictionary by varying the word per dictionary and also the number of dictionaries generated is analysed. Fuzzy C means soft clustering algorithm is used to generate dictionary. This paper work is based on the hypothesis that fuzziness in the codebook creation step as well as in the histogram creation process leads to more robust behaviour of the bag of visual words approach in terms of codebook size. The performance of the Multiple Dictionary Bag of Words model using Separate dictionary is compared with base line method by varying the word per dictionary and also by varying the number of individual dictionary generated by taking features randomly.

Fig. 3. Schematic for Separate dictionary implementation with FCM Clustering.

The parameters used for the evaluation of the different algorithms are:

I. Macro precision

\[
P_{\text{macro}} = \frac{1}{|c|} \sum_{i=1}^{|c|} \frac{TP_i}{TP_i + FP_i}
\]  

(4.1)

II. Micro precision

\[
P_{\text{micro}} = \frac{\sum_{i=1}^{|c|} TP_i}{\sum_{i=1}^{|c|} TP_i + FP_i}
\]

(4.2)
III. Accuracy

\[
\text{Accuracy} = \frac{\sum_{i=1}^{C} TP_i + \sum_{i=1}^{C} TN_i}{\sum_{i=1}^{C} (TP_i + FN_i + FP_i + TN_i)}
\] (4.3)

IV. Macro F1

\[
F = \frac{2 \cdot P_{\text{macro}} \cdot R_{\text{macro}}}{P_{\text{macro}} + R_{\text{macro}}}
\]
where

\[
R_{\text{macro}} = \frac{1}{|C| \sum_{i=1}^{C} \frac{TP_i}{TP_i + FN_i}}
\] (4.5)

V. Micro F1

\[
F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}
\] (4.6)

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. The notable difference between these two calculations is that micro-averaging gives equal weight to every document that is it is called a document-pivoted measure while macro-averaging gives equal weight to every category that is it is category-pivoted measure. 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.

For the Fuzzy C means the parameter \( m = 1.7 \) and stop condition \( \varepsilon = 0.001 \). The test data set includes eight different topics 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 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. Since dataset is taken for real time application for visual recognition of objects for a humanoid used in restaurant, it is created from Google images. The images in the dataset used can be categorised as tiny images. The sample images from dataset are as shown in Fig. 4.

![Sample images from dataset](image)

Fig 5 to 9 shows the variation of accuracy rate with words per dictionary by varying the number of dictionary generated randomly from the feature pool from 1 to 5 which is named as dictionary1, dictionary2, dictionary3, dictionary4 and dictionary5. The results obtained are compared with the baseline method implemented in the paper. In both baseline method and Multiple Dictionary Bag of Words model the clustering of words are done using Fuzzy C means soft clustering algorithm. The algorithm was also implemented for Dataset taken from Caltech database which includes four different topics each topic...
containing 200 images. It was found that the Multiple Dictionary Bag of Words model works for large scale image search where the number of topics and the number of images per topics are more.

**Table 1.** Accuracy rate for word per dictionary 160 for various numbers of dictionaries

<table>
<thead>
<tr>
<th>No: of Dictionary</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy Rate</td>
<td>0.9137</td>
<td>0.9144</td>
<td>0.9075</td>
<td>0.92</td>
<td></td>
</tr>
</tbody>
</table>
**Fig. 8** Accuracy vs. words per dictionary for Dictionary 4

**Fig. 9** Accuracy vs. words per dictionary for Dictionary 5

**Table 2.** Macro Precision for different words per dictionary for Base line method and Separate Dictionary (MDBoW)

<table>
<thead>
<tr>
<th>No: Of Words Per Dictionary</th>
<th>Base Line Method</th>
<th>Separate Dice 1</th>
<th>Separate Dice 2</th>
<th>Separate Dice 3</th>
<th>Separate Dice 4</th>
<th>Separate Dice 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.5714</td>
<td>0.6122</td>
<td>0.6213</td>
<td>0.6103</td>
<td>0.6472</td>
<td>0.6051</td>
</tr>
<tr>
<td>120</td>
<td>0.6332</td>
<td>0.6958</td>
<td>0.6613</td>
<td>0.6542</td>
<td>0.6236</td>
<td>0.6467</td>
</tr>
<tr>
<td><strong>160</strong></td>
<td><strong>0.6225</strong></td>
<td><strong>0.6381</strong></td>
<td><strong>0.6478</strong></td>
<td><strong>0.6702</strong></td>
<td><strong>0.6433</strong></td>
<td><strong>0.6842</strong></td>
</tr>
<tr>
<td>200</td>
<td>0.5739</td>
<td>0.6666</td>
<td>0.6193</td>
<td>0.6127</td>
<td>0.6144</td>
<td>0.6082</td>
</tr>
</tbody>
</table>

**Table 3.** Micro Precision for different words per dictionary for Base line method and Separate Dictionary (MDBoW)

<table>
<thead>
<tr>
<th>No: of Words Per Dictionary</th>
<th>Base Line Method</th>
<th>Separate Dice 1</th>
<th>Separate Dice 2</th>
<th>Separate Dice 3</th>
<th>Separate Dice 4</th>
<th>Separate Dice 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.555</td>
<td>0.5975</td>
<td>0.6075</td>
<td>0.565</td>
<td>0.615</td>
<td>0.5725</td>
</tr>
<tr>
<td>120</td>
<td>0.6</td>
<td>0.6475</td>
<td>0.635</td>
<td>0.6325</td>
<td>0.5965</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>160</strong></td>
<td><strong>0.6075</strong></td>
<td><strong>0.62</strong></td>
<td><strong>0.64</strong></td>
<td><strong>0.6566</strong></td>
<td><strong>0.63</strong></td>
<td><strong>0.68</strong></td>
</tr>
<tr>
<td>200</td>
<td>0.555</td>
<td>0.645</td>
<td>0.605</td>
<td>0.595</td>
<td>0.6075</td>
<td>0.595</td>
</tr>
</tbody>
</table>
The results projected in Tables 2 to 5 shows that Multiple Dictionary Bag of Words model using Separate dictionary shows better performance than baseline method. It can be seen from the results that on an average the method gives maximum accuracy rate for word per dictionary of 160 and the accuracy rate increases as the number of dictionary increases from 1 to 5. The tabulation of this result is given in Table 1. The parameters Macro Precision, Micro Precision, Micro F1 and Macro F1 have better values for Multiple Dictionary Bag of Words than baseline method. For word per dictionary of 160 all these parameters increase as the number of dictionary increases.

**Table 4. Micro F1 for different words per dictionary for Base line method and Separate Dictionary (MDBoW)**

<table>
<thead>
<tr>
<th>No: of Words per Dictionary</th>
<th>Base Line Method</th>
<th>Separate Dic 1</th>
<th>Separate Dic 2</th>
<th>Separate Dic3</th>
<th>Separate Dic 4</th>
<th>Separate Dic 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.5457</td>
<td>0.5908</td>
<td>0.6003</td>
<td>0.5596</td>
<td>0.6109</td>
<td>0.5711</td>
</tr>
<tr>
<td>120</td>
<td>0.5949</td>
<td>0.6488</td>
<td>0.6331</td>
<td>0.6255</td>
<td>0.5864</td>
<td>0.6023</td>
</tr>
<tr>
<td>160</td>
<td>0.6038</td>
<td>0.6197</td>
<td>0.6374</td>
<td>0.6531</td>
<td>0.6285</td>
<td>0.6767</td>
</tr>
<tr>
<td>200</td>
<td>0.5493</td>
<td>0.6378</td>
<td>0.6021</td>
<td>0.5919</td>
<td>0.6019</td>
<td>0.5881</td>
</tr>
</tbody>
</table>

**Table 5. Macro F1 for different words per dictionary for Base line method and Separate Dictionary (MDBoW)**

<table>
<thead>
<tr>
<th>No: of Words per Dictionary</th>
<th>Base Line Method</th>
<th>Separate Dic 1</th>
<th>Separate Dic 2</th>
<th>Separate Dic3</th>
<th>Separate Dic 4</th>
<th>Separate Dic 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>0.5631</td>
<td>0.6047</td>
<td>0.6143</td>
<td>0.5868</td>
<td>0.6307</td>
<td>0.5884</td>
</tr>
<tr>
<td>120</td>
<td>0.6161</td>
<td>0.6708</td>
<td>0.6479</td>
<td>0.6432</td>
<td>0.61</td>
<td>0.6278</td>
</tr>
<tr>
<td>160</td>
<td>0.6149</td>
<td>0.6289</td>
<td>0.6439</td>
<td>0.6635</td>
<td>0.6366</td>
<td>0.6821</td>
</tr>
<tr>
<td>200</td>
<td>0.5643</td>
<td>0.6558</td>
<td>0.6121</td>
<td>0.6037</td>
<td>0.6109</td>
<td>0.6015</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, the performance of fuzzy clustering Multiple Dictionary Bag of Words model using Separate dictionary used for image classification is investigated by varying the words per dictionary and also the number of dictionaries generated and it is compared with the base line method. 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. It is seen that the 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 which uses fuzzy codebook in
Bag of Words method. The performance measures used for evaluation increases as the number of dictionary is increased for a particular value of word per dictionary.

References