



7th International Conference on Communication, Computing and Virtualization 2016

# Novel Content Based Image Classification Method Using LBG Vector Quantization Method with Bayes and Lazy Family Data Mining Classifiers

Neha R. Kasat<sup>a</sup>, Dr. Sudeep D. Thepade<sup>b</sup>

<sup>a</sup>Student, Department of information technology, Pimpri chinchwad college of engineering, Pune, 411044, India

<sup>b</sup>Professor, Department of information technology, Pimpri chinchwad college of engineering, Pune, 411044, India

---

## Abstract

Image Classification is important step in Image retrieval. Multimedia Classification gaining importance due to increase in number of application. Many approaches has been used for image classification. This paper proposes Image Classification approach using LBG vector quantization method with Bayes and Lazy family data mining classifiers. Here Experimentation is done using 6 data mining classifier and 7 codebook sizes. Total 42 variations are performed on a Test database. 500 images is used as an Image set to compute the classification accuracy. 74.77% performance is observed by Bayes Net classifier of Bayes Family with 128 Codebook size of LBG.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of ICCCV 2016

*Keywords:* Content Based Image Classification, Vector Quantization, LBG codebook algorithm, Classifiers Bayes, Lazy.

---

## 1. Introduction

Classification is nothing but dividing things into groups. So we can say that Categorization of Images with respect to its content is Content Based Image Classification. Trend of multimedia (Images, Videos) is increasing day by day and this creates the need of Image Classification. Image classification simplifies the work with Images by providing multiple images of same category. Image Classification is crucial when it comes to Decision making for statistical data (e.g. Presence of water or forest in wide geographic area). It is also important in the field of studying the variety of different types of life found on Earth and the variations within species (Bio-Diversity). Different Image classification techniques have been proposed in literature. Sine, Cosine, Walsh Transforms are used for

\* Corresponding author. E-mail address: [nehakasad2012@gmail.com](mailto:nehakasad2012@gmail.com)

Image classification [1]. Block Truncation Coding is performed on images for classification [3]. Kekre, Hartley, Slant and Haar Transforms are also used for classification [2]. Color spaces are used for classification [4].

## **2. LBG codebook in Vector Quantization**

Vector quantization has been used in various applications such as data compression, image segmentation, content based image retrieval and face recognition. Vector quantization deals with process of clustering. There is various method of clustering. Codebook is the set of values computed with the help of vector quantization algorithm. In this paper we are using Codebook for colour images. The image is divided in three colour plans. Each plan is further divided into fixed size blocks and that forms a training vector. Suppose for  $2 \times 2$  blocks, it will generated 12 values four of each colour plane. The centroid  $C_1$  is calculated to from Initial code vector. Two vectors  $v_1$  and  $v_2$  are generated by adding and subtracting error. Euclidean distance is calculated with  $v_1$  and  $v_2$ , two clusters are formed based on nearest of  $v_1$  and  $v_2$ . This procedure is repeated for these two clusters to generate four new clusters. And again repeated until the required codebook size is reached.

## **3. Lazy Family Data Mining Classifiers**

Lazy Family Data Mining Classifiers [2] supports incremental learning. It contains some classifiers such as IB1, IBK, K-Star, LWL, LBR. It takes less time for training and more time for predicting. For our data set we have used IB1 and IBK algorithms.

## **4. Bayes Family Data Mining Classifiers**

Bayes Family data mining classifiers are used for classification [1]. It is a statistical classifier. It contains some classifiers such as AODE, ADOEs, Naive Bayes, Bayesian Net, Naive Bayes Simple, Naive Bayes updatable, Naive Bayes Multinomial. It contains finite set of feature vector values. No single algorithm is there, for training the classifiers, there is a family of algorithms based on common principle. It uses numerical and categorical data. It is very effective for huge data and it is applicable to classification, spam and filtering.

## **5. Proposed Content Based Image Classification Technique Using LBG and Data Mining Classifiers.**

Similar to multiple methods of image classification, the classification can also be performed using LBG and Data mining Classifiers. In the proposed work LBG Codebook generation algorithm is applied on colour images. A initial vector of colour image is formed. Clusters are formed of size  $2^n$  Calculating centroid of each clusters and form a Codebook.

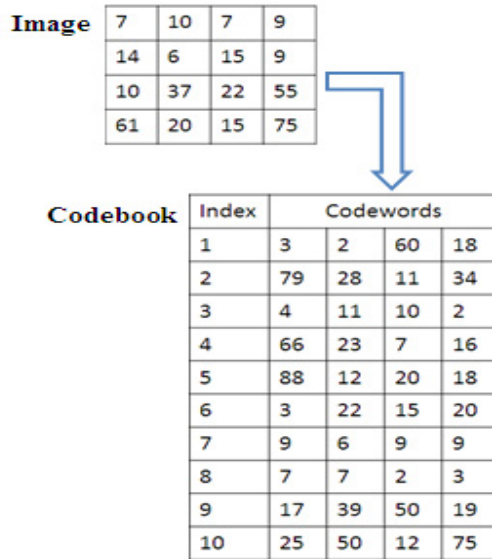


Fig. 1. Feature vector using codebook

Image Classification is performed in two steps feature extraction and classification.

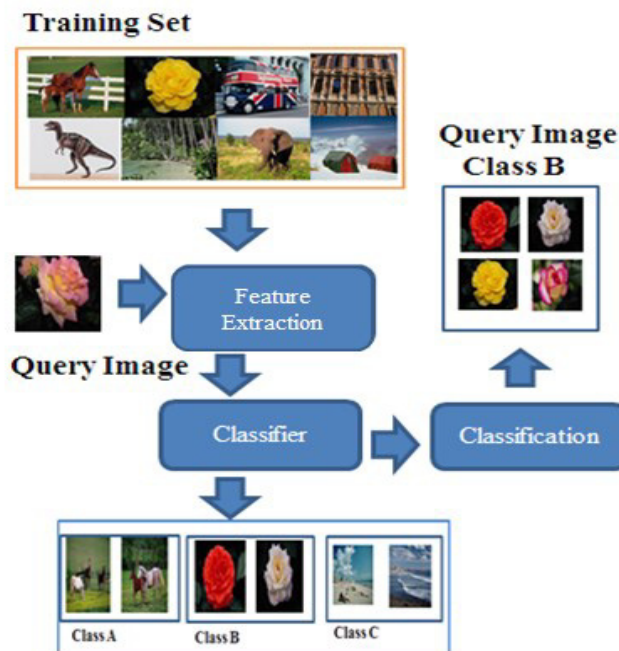


Fig. 2. Proposed CBIC process

Training dataset is supplied to Feature Vector, for Feature extraction using LBG algorithm. Feature vector is given to the classifier for Training. Then query image is provided to Feature vector, features are extracted and given to trained

classifier for classification. Classification, classifies the image as per its belonging class.

## 6. Experimentation Environment

In proposed method Wang Image database is used which is split in training and testing dataset each of 320 and 40 respectively, with 8 different classes. Platform used for feature extraction was Matlab on computer with Intel core i5 processor and 4 GB RAM. For classification Weka Tool is used.

Accuracy is the only performance measure to compare the techniques. Seven different codebook sizes and six classifiers are used. Total 7\*6 combinations are carried out for finding result

## 7. Result and Discussion

In Proposed technique 42 variations are Experimented on 500 Images with 320 Training set Images and 180 Testing set Images. Training and Testing images are spread across 8 categories namely Buses, Roses, Dino, Beachs, Horses, Elephants, Mountains & etc.

Bayes Net gives best over all result 74.77% for all codebook sizes. Codebook size 128 performs better with all the classifiers, it gives 72.5% result.

Table 1. The Table shows Average accuracy with respect to classifiers.

Classifiers	Average Accuracy
Naive Bayes	67.32
Bayes Net	74.77
Naive Bayes Multinomial	72.85
Naive Bayes Updateable	66.071
IB1	71.78
IBK	71.78

Above table shows the Average accuracy for all codebook sizes with all Classifiers. Bayes Net Classifiers gives best result.

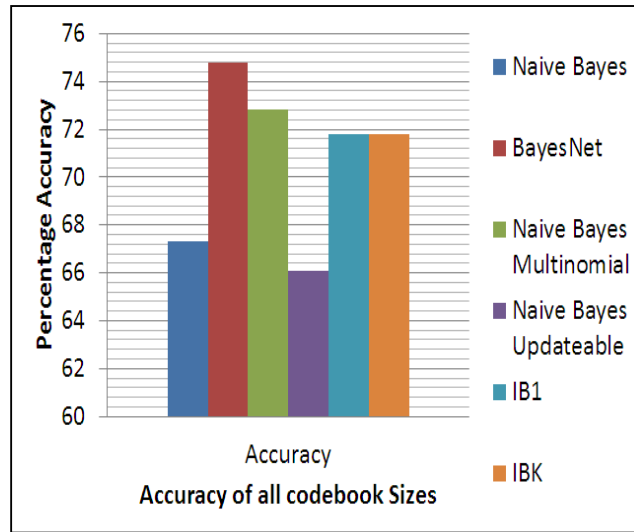


Fig. 3. Average Accuracy with respect to Classifiers.

Table 2. The Table shows Average Accuracy with respect to Codebook size.

Codebook size	Average Accuracy
16	71.25
32	70
64	70
128	72.5
256	70
512	71.88
1024	61.25

Above Table shows Average Accuracy for all the codebook size. Codebook size 128 gives highest accuracy of all other codebook sizes.

Table 3. The Table contains all the Experimentation results with all classifiers and codebook sizes.

Classifiers	Codebook size						
	16	32	64	128	256	512	1024
Naive Bayes	72.5	70	70	67.5	67.5	68.75	55
Bayes Net	72.5	85.93	72.5	72.5	72.5	75	72.5
Naive Bayes Multinomial	72.5	72.5	72.5	75	75	75	67.5
Naive Bayes updateable	72.5	70	70	67.5	67.5	60	55
IB1	70	70	70	77.5	72.5	75	67.5
IBK	70	70	70	77.5	72.5	75	67.5

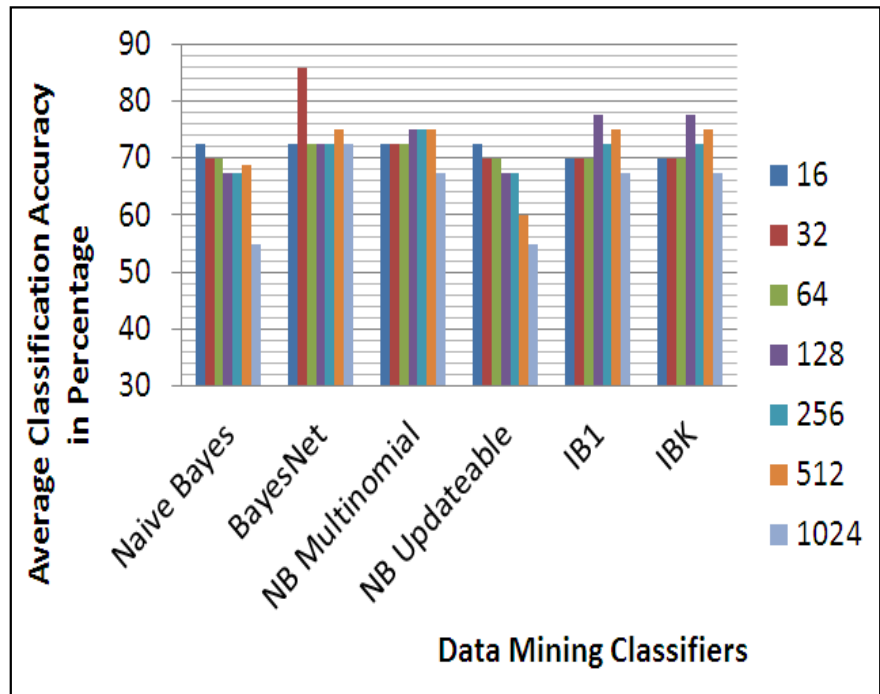


Fig 4. Average Classification Accuracy in Percentage with respect to Data Mining Classifiers.

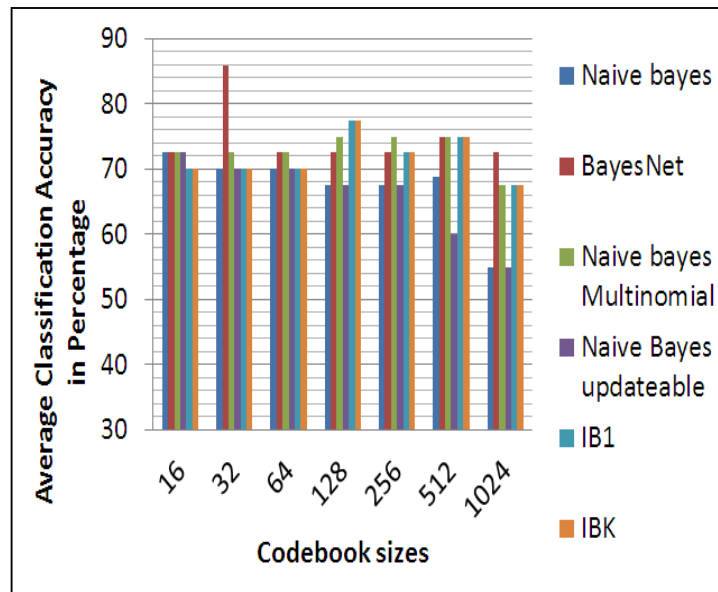


Fig 5. Average Classification Accuracy in Percentage with respect to Codebook Sizes.

## 8. Conclusion

Content Based Image Classification is important for improving the Accuracy of Image Retrieval. Conventional Data Mining Classifiers are Experimented here on 500 images using 7 different Codebook (16,32, 64, 128 ,256 , 512 ,1024 ) of LBG vector quantization method to proposed Novel image classification technique.

Overall best codebook size is 128. Best family is Lazy with 71.78%. Best Data Mining Classifiers is Bayes Net. Overall best performance is observed by the proposed Image Classification technique is Bayes Net classifier with 128 codebook size.

## References

1. Dr. Sudeep D. Thepade, Madhura M. Kalbhor, "Novel Data Mining based Image Classification with Bayes, Tree, Rule, Lazy and Function Classifiers using Fractional Row Mean of Cosine, Sine and Walsh Column Transformed Images." 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Jan. 16-17, Mumbai, India
2. Dr. Sudeep D. Thepade, Madhura M. Kalbhor, "Extended Performance Appraise of Bayes, Function, Lazy, Rule, Tree Data Mining Classifier in Novel Transformed Fractional Content Based Image Classification." 2015 International Conference on Pervasive Computing (ICPC), Jan. 9-10. Pune, India.
3. Dr. Sudeep D. Thepade, Rik Das, "Performance Comparison of Feature Vector Extraction Techniques in RGB Color Space using Block Truncation Coding for Content Based Image Classification with Discrete classifiers." INDICON 2014.
4. Ms. Sandhya R. Shinde, Ms. Sonali Sabale, Mr Siddhant Kulkarni, Ms. Deepti Bhatia, "Experiments on Content Based Image Classification using Color Feature Extraction", 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Jan. 16-17, Mumbai, India
5. Singh, V.P., Srivastava, R., "Design & performance Analysis of CBIR system Based on Image Classification using Feature Sets", Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015 .