Content based Image Retrieval using Gaussian Mixture Model based Subspaces Representation

Jadhav Shweta D.
M. E. Computer (Student)
Department of Computer Engg.
K. K. W. I. E. E. R., Nasik
S. P. P. U.

Shahane Nitin M.
Associate Professor
Department of Computer Engg.
K. K. W. I. E. E. R., Nasik
S. P. P. U.

ABSTRACT
Content Based Image Retrieval (CBIR) plays a significant role in case of image processing. Generally, in case of large scale dataset the two problems which are common viz. lower memory cost and higher retrieval accuracy. To solve the problem of the large scale retrieval the mixture of subspaces image representation is used. In this approach the group of the local descriptors of every individual image is used for global image representation. The Principal Component Analysis (PCA) is used for the dimensionality reduction. So that large number of images can be retrieved easily. Accuracy of the proposed system is measured in terms of mean average precision. Through the experiment it shows that the proposed system gives better result than earlier system.

Keywords
Content Based Image Retrieval (CBIR); Image Retrieval; Subspaces.

1. INTRODUCTION
Today as there is advancement in the technology lots of images are being downloaded from the internet. So the gathering of the images is increased. In this paper the problem of the content based image retrieval (CBIR) which is based on local image descriptor. Even by using the most commonly used local image descriptor i.e. SIFT each sample descriptor-by-descriptor matching can achieve high retrieval accuracy. But the faces the two difficulties by applying this approach to large scale image retrieval [17]. The first is the computational cost which is required for mapping every local descriptor in the database with those features which are extracted from the query image. The second is memory cost required for storing the large number of the images in the memory. To solve the problem different methods are carried out. Categorization and indexing is done by using bag of features. It uses the powerful descriptor but it fails in case of large scale image dataset. This paper focuses on the high performance of the image retrieval methods that utilizes an efficient and compact image feature representation.

Bag-of-features (BOF) are used for image representation. It counts only the number of features. Because of this the other statistics method are used for calculating the mean, standard deviation and weight of the local descriptor by using Gaussian Mixture Model (GMM). To overcome the drawback of the bag-of-features the fisher vector is used.

For better understanding of the concepts it is divided into sections: Section 1 describes the introduction of the system and motivation of the proposed system. Section 2 describes the related work in which motivational survey, efficiency and drawbacks of previous system are discussed. Section 3 describes the detailed design of the proposed system. Section 4 presents the experimental results for demonstrating the validity of the proposed system for the large scale datasets and Section 5 describes the conclusion.

2. LITERATURE SURVEY
Section 2.1 explains the methods of feature extraction. Section 2.2 explains the representation of the image features.

2.1 Feature Extraction
In CBIR system the most commonly used is the color features. The extracted features are stored in the form of feature vector. There are different local image features for representing an image. Among the local descriptor the most common is the SIFT [12] which gives better accuracy but it fails in case of large scale image retrieval.

Thomas Deselaers et. al [13] in this MPEG-7 descriptor is a Multimedia content descriptor interface. It contains different descriptors viz. edge histogram descriptor [1], Scalable color descriptor, Color Layout Descriptor and color edge directivity descriptor. All the features which are extracted are then combined into single vector.

2.2 Image Representation
F. Perronin et. al [2] fisher kernel is the dominant structure which combines strengths of the discriminative and generative approaches. The structure is applied to image categorization in which the input is the images and generative model is the visual vocabulary. The approximation of the low level features in the images is done by using a Gaussian Mixture Model (GMM). The computational needs are less for training and testing time. The vocabularies are trained on the set of categories and it can be applied on another set without any loss in performance.

H. Jegou et. al [6] [10]the three constraints are considered the effectiveness, searching accuracy and the storage of the memory for representation. In this the aggregating of the local image descriptor into the vector representation then performs the dimensionality reduction of these vectors. Finally the indexing algorithm is applied. As compared to the other it gives better performance.

T. Takashi et. al [12] the two constraints are considered the low memory cost and the better retrieval accuracy. In this asymmetric approach is applied in which the probability distribution of the local descriptor is modeled for every image in the database and query image of the local descriptor is used. The likelihood function is used for matching the score between the query image and the image stored in the database.

Through the results it shows that the every database image in less than hundred bytes can achieve higher accuracy than state of art method using Fisher Vector.

C. Chen et. al [15] subspace modeling structure is applied which can retrieve the images from the highly imbalanced data. It builds the positive subspace model on the sets of the positive training dataset. Each of which is generated by a Gaussian Mixture Model (GMM) that partition data instances of the target into subsets. Merging of each subset with original positive data instances. Lastly, the joint scoring method is applied for final ranking score.

For improving the performance of the large scale image dataset, it is key to find out the real and well organized method for image retrieval. To improve the accuracy the MPEG-7 descriptor is used.

3. IMPLEMENTATION DETAILS

The images are trained in the training phase. In the testing phase the query image is given to the system and the relevant images are retrieved.

3.1 Feature Extraction

In the proposed system the visual features are extracted and stored in the MPEG-7 descriptor. Edge Histogram Descriptor (EHD), Color Edge Directivity Descriptor (CEDD), Color Structure Descriptor (CSD), and Scalable Color Descriptor (SCD) are used as dominant color features. In the proposed system we are using EHD, SCD, and CEDD.

3.1.1 Edge Histogram Descriptor (EHD)

The input image is given of the size M × N and it is divided into 16 blocks. Again each image block is divided into sub-image of size 2×2. The different edges of the image are represented with the 5 edge masks viz. horizontal, vertical, 45 degree, non directional and 135 degree. Each sub-image contains 5 bins of histogram. At last 80 bins are created for 16 sub-image block.

$$x_{ehd} = \{x_{0,}\, x_1, \dots, x_N\}$$

3.1.2 Color Edge Directivity Descriptor (CEDD)

In this both the texture and color features are extracted from an image. The image is divided into 22 sub image which contains 4040 blocks. The CEDD contains the six texture regions and twenty four color regions. EHD represents the rotation invariance of 45 degree because of quantized color space that it uses; it represents the tolerance to change in lighting condition and hue variance. Total it contains 144 bins.

$$x_{cedd} = \{x_{0,}\, x_1, \dots, x_{143}\}$$

3.1.3 Scalable Color Descriptor (SCD)

It is represented in the HSV color space. Encoding of the color features is done using Haar transform. The rang of bin is from 16 to 256.

$$x_{SCD} = \{x_{0,}\, x_1, \dots, x_{255}\}$$

The features which are extracted are stored in the form of single vector.

3.2 Subspace

The dimensions are reduced by using Principal Component Analysis (PCA). Generally it is used for the dimensionality reduction for the large scale dataset. The input is the single vector which is given and the mean and covariance matrix is found out. Then calculating the eigenvalues and eigenvectors. All these features are combined to form a single vector.

3.3 Mixture Model

It is generative and probabilistic model. GMM has the individual mean and covariance matrix. The pdf corresponds to a GMM defined in D-dimensional space is given as

$$p(x) = \sum_{i=1}^{k} w_i N(x|\theta_i)$$

where $\theta = (w_1, \theta_1, w_{2,} \theta_2, \dots, w_k, \theta_k)$

$$N(x|\theta_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{D/2}} exp^{\left(\frac{-1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right)}$$

Where k= Gaussian component

w_i = mixing of i^{th} component and $0 \le w_i \le 1$ and probability of mixing is given by $\sum_{i=1}^{k} w_i = 1$

μ_i= mean of the i^{th} component

Σ_i = covariance matrix of the i^{th} component

$N(x|\theta_i)$= Normal density function

3.4 Process Block Diagram

The system is divided into two phases i.e. training and testing. In the training phase the images are trained. The large scale of features are extracted and the features dimensions are reduced. In the testing phase, the query image is given and the color features are extracted and stored. While the PCA is applied to features the size of features are reduced. The k-means algorithm is applied and the GMM is used to find the mean and covariance matrix. Mapping is done between the query image and the image stored in the dataset. Atlast the indexing and top-k relevant image is found out.

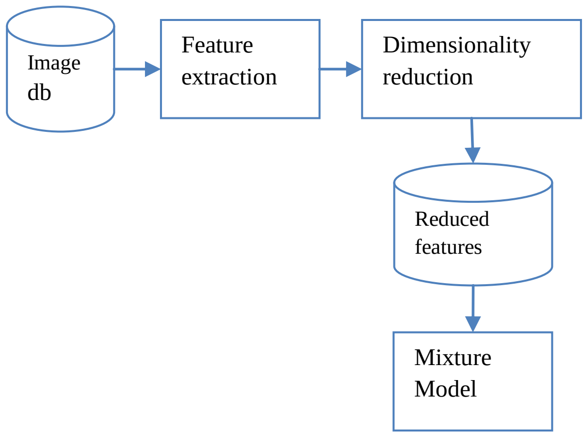

Fig. 1: Training phase

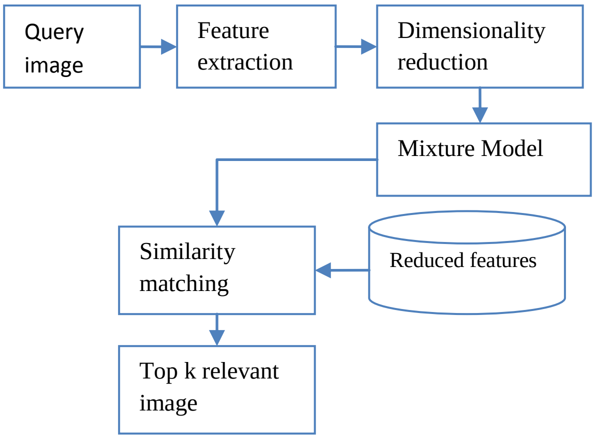

Fig. 2: Testing Phase

4. RESULTS AND DISCUSSION

4.1 Performance Measures

The accuracy of Holiday dataset is measured using Mean Average Precision(mAP). For each query image q, mean is taken over precision at every relevant image is called mean average precision. It is given by formula

$$AP(q) = \frac{1}{n}\sum_{i=1}^{n_q} P_q(R_i)$$

Where, R_i = recall after i^{th} relevant image retrieved

Mean of the average precision taken over all query image q is called Mean Average Precision (mAP). It is given as

$$MAP = \frac{1}{Q}\sum_{q\in Q} AP(q)$$

Q = queries q is a set.

4.2 Experimental Setup

In order to assess the performance of the proposed system, an image data set containing1,491 images from INRIA Holiday data set of outdoor scene jpg images is used. Resizing of the image is done so that the maximum length of the side is less than or equal to 1024 pixels [17]. The ground truth of all databases is known so that we can test each image from the database and evaluate the performance of the proposed system. The images are classified manually into categories and their ground truth will be the categorization of the CBIR system. The accuracy is measured using Mean Average Precision.

Corel dataset contains 1000 images which are in the jpg format. It contains the 10 categories each containing the 100 images. The images are of the size 256 X 384.

4.3 Result Table

Input query is given to the system dimensionality reduction using PCA is performed on the feature set including the features extracted from the query image and then k-means clustering algorithm is being applied on the reduced feature set. For each query image the nearest cluster is retrieved and precision is calculated. The procedure is repeated for every query image. Average precision is calculated for all 1000 images present in the dataset. The experiment is conducted for number of clusters k = 128.

Table 1 shows the average precision for Corel data set with clustering. The table shows the number of images from all the ten categories present in Corel Dataset associated in each cluster.

Table 1. Average Precision with Clustering (Corel Dataset)

Sr. No.	Original feature vector	Reduced feature vector	Number of Clusters	Average Precision
1.	480	416	128	68.85

When the query is submitted to the system dimensionality reduction using PCA is performed on the feature set including the features extracted from the query image and then k-means clustering algorithm is being applied on the reduced feature set. For the query image the nearest cluster is retrieved and precision is calculated based on the ground truth of the dataset. The procedure is repeated for every query image. Average precision is calculated for all 1491 images present in the dataset. The experiment is repeated for number of clusters k = 32, 64 and 128. Table 2 shows the average precision for Holiday dataset at k = 32, 64 and 128.

Table 2. Average Precision with Clustering (Holiday Dataset)

Sr. No.	Original feature vector	Reduced feature vector	Number of Clusters	Average Precision
1.	480	416	32	64.32
2.	480	416	64	67.07
3.	480	416	128	68.36

When the clustering is performed on reduce feature set and number of clusters is set to 128 the average precision obtained over the entire data base, 68.85% for Corel Dataset and 68.38% for Holiday Dataset, is almost same which is evident from Table 1 and 2. Also, from Table 2 it can be observed that as number of cluster are increased from 32 to 128 average precision increases.

Precision for Top 100 Retrievals Each image in the Corel Data base was given as a query image, top 100 images were retrieved each time and precision is calculated. Top 100 retrievals where precision and recall values are same.

Table 3. Top 100 Retrievals (Corel data set)

Category	Precision
African	40.09
Beach	23.17
Building	42.90
Bus	50.72
Dinosaur	88.90
Elephant	26.73

Rose	37.29
Horse	59.22
Mountain	29.86
Food	38.48

5. CONCLUSION

Content based image retrieval is a big challenge in case of large scale database for getting the relevant images. From the last few decades the researchers are working for getting the better retrieval accuracy. Whatever the size of the database the human can easily identify the image. The researchers are working on the color and the texture features. So that the accuracy can be increased.

The goal is to obtain high image retrieval accuracy by increasing the number of cluster size. The color features using the MPEG-7 Color descriptor are used. The feature space is reduced to less number of bytes by using the PCA. All the reduced features are merged to form one single vector. The memory size required to store the features is reduced.

Hence it can be concluded that the system gives better results because of the use of MPEG-7 descriptors, dimensionality reduction and dynamic clustering.

Though, significant improvement is achieved using MPEG-7 descriptors dimensionality reduction and dynamic clustering still there is a scope to go better outcomes. Some experimentation is also required so as to reduce the feature space further. In the system the features are reduced from 480 to 416 using PCA. Hence it is to be investigates that how it can be reduced further.

6. ACKNOWLEDGMENTS

I am heartily thankful and have a deep sense of gratitude towards my guide Prof. N. M. Shahane, for his fraternal guidance, encouragement and astute advice from the preliminary to concluding level which enabled me to complete this work. Besides illuminating technical guidance, all time support and his friendly attitude has been the main driving force of success in completion of paper. I am also thankful to Prof. Dr. S. S. Sane, Head Department of Computer Engineering, K.K.W.I.E.E.R., Nashik.

I would also like to express my thanks to Prof. Dr. K. N. Nandurkar, Principal, K.K.W.I.E.E.R., Nashik. Finally, this work would not have been possible without the confidence, endurance and support of my family. My family has always been a source of inspiration and encouragement. I am quite fortunate to have support and help of my family members.

7. REFERENCES

[1] Chee Sun Won, Dong Kwon Park, and Soo-Jun Park , ”Efficient Use of MPEG-7 Edge Histogram Descriptor”, ETRI Journal, volume 24, Number 1, Feburary 2002.

[2] F. Perronnin and C. Dance , ”Fisher kernels on visual vocabularies for image categorization”, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2007, pp. 18.

[3] F. Perronnin, Y. Liu, J. Sanchez, and H. Poirier , ”Large Scale Image Retrieval with Compressed Fisher Vectors”, Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2010.

[4] F. Perronnin, J. Sanchezz, and T. Mensink , ”Improving the Fisher kernel for largescale image classification”, in Proc.Eur. Conf. Comput. Vis., 2010, pp. 143156.

[5] H. Jegou, M. Douze, and C. Schmid, ”Improving bag-of features for large scale image search,Int. J. Comput. Vis., vol.87, no. 3, pp. 316336, 2010.

[6] H. Jegou, M. Douze, C. Schmid, and P. Perez, ”Aggregating Local Descriptors into a Compact Image Representation ”,Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2010.

[7] T. Tuytelaars, M. Fritz, K. Saenko, and T. Darrell, ”The nbnn kernel”,in Proc. IEEE Int. Conf. Comput. Vis., 2011, pp.824-1831.

[8] M. Jain, R. Benmokhtar, P. Gros, and H. Jegou, Hamming embedding similaritybased image classification”, in Proc. 2nd ACM Int. Conf. Multimedia Retrieval, 2012, article 19, DOI:10.1145/2324796.2324820.

[9] M. Young, The Technical Writers Handbook. Mill Valley, CA: University Science, 1989.

[10] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid, ”Aggregating local image descriptors into compact codes”, IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no. 9,pp. 1704-1716, Sep. 2012.

[11] Jie Linyz, Ling-Yu Duany, Tiejun Huangy, Wen Gaoy, ”Robust Fisher Codes for Large Scale Image Retrieval”, IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP), May 2013.

[12] T. Takahashi, T. Kurita, ”Mixture of Subspaces Image Representation and Compact Coding for Large Scale Image”, in IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 37, No.7, July 2015.

[13] Thomas Deselaers and Daniel Keysers and Hermann Ney,”Features for Image Retrieval: An Experimental”, Journal Information Retrieval archive, Volume 11 Issue 2, April 2008 Pages 77 - 107 .

[14] Qilong Wang, Peihua Li, Lei Zhang, Wangmeng Zuo, ”Towards Effective Codebookless Model for Image Classification”, Computer Vision and Pattern Recognition, July 2015.

[15] Chao Chen, Coral Gables, Mei-Ling Shyu, Shu-Ching Chen , ”Gaussian Mixture Model-Based Subspace Modeling for Semantic Concept Retrieval”, Information Reuse and Integration (IRI), 2015 IEEE International Conference,13-15 Aug. 2015.

[16] Jadhav Shweta D and Prof. Shahane Nitin M.,“ A Review on Content Based Image Retrieval using Mixture Model for Image Representation “, IJSRD - International Journal for Scientific Research & Development| Vol. 3, Issue 11, 2016 | ISSN (online): 2321-0613.

[17] Jadhav Shweta D and Prof. Shahane Nitin M, “Content Based Image Retrieval using Mixture of Subspaces Image Representation”,cPGCON-2016.