A Review on Content Based Image Retrieval using Mixture Model for Image Representation

Jadhav Shweta D¹ Prof. Shahane Nitin M.²
¹M. E. ²Professor
1,2Department of Computer Engineering
1,2K. K. W. I. E. E. R., Nasik(Maharashtra), India

Abstract— Content Based Image Retrieval has drawn attention in last few years. CBIR is used for retrieving images from the large scale. Without the use of compressed code for image retrieval large scale of images cannot be searched expeditiously. Bag-of-Words BoW is not realistic for huge image database because it is concentrated in memory. The main objective is to achieve high performance and lower memory cost. The new framework Mixture of Subspace Image Representation is used. In this local descriptor is modelled for each image present in the database. A query image is given to the system is matched with the image present in the database and the likelihood function is computed. It is found out through the review of papers that the new framework represents every database image in a small amount of number of bytes which achieves good performance than the state-of –art. The memory required to store the data is more so the new framework is proposed. So that they require the less memory and improved performance of retrieval. The retrieval performance is improved because of the reduction of the large scale database.

Key words: Image Retrieval, Image Search, Content Based Image Retrieval, Image representation

I. INTRODUCTION

Today with the advancement of the Internet, and the ease of use of image capturing devices such as elevated resolution camera, the collection of the digital images is increasing. So the images should be stored well and the retrieval should be done. The main steps in CBIR system is feature extraction, indexing algorithm and similarity matching. Large scale of image cannot be searched efficiently from the database. Typically to search the relevant images from database the Bag of Words was used. BoW representation groups local descriptor together. The codebook [3] of k-centroids is obtained by k-means clustering [3]. BoW is represented in the form of histogram of number of image descriptor. BOW only counts the features which are extracted. Fisher Vector (FV) for image representation is introduced to overcome the drawbacks of Bag-of-words. Fisher Vector act as an extension of the bag-of-words [3][14].

In image classification, feature coding act as solution. Generally, image classification is used to assign the labels to an image. It uses global image descriptor using fisher coding. For estimating the distribution of local descriptor Gaussian Mixture Model (GMM) is employed over a training set [6]. Large scale of data retrieval is one of the problems. So to achieve high search accuracy, efficiency and low memory usage a mixture model is used. Mixture of subspace image representation is used instead of the image feature representation. For representing the image, Fisher Vector is used for getting the higher retrieval accuracy. Gaussian Mixture Model is modelled by grouping the local descriptor. GMM is a probabilistic model. Gaussian Mixture is conceptually a soft k-means which can better model data distribution. For learning the mixture model, the GMM is the fastest algorithm which maximises the log-likelihood function. Disadvantages are when a system has inadequately many points per mixture then the estimate of the covariance matrices become difficult.

Representation of Subspace image is done in n-dimensional feature space. Dimensionality reduction is achieved by extracting the important features from dataset and then performs the learning. The projection of the original high-dimensional feature space to a low-dimensional space is done by using subspace. There are various methods for the dimensionality reduction. Generally dimension reduction is performed for the large scale database. So that they can fit into that area [12]. The lots of memory size will be reduced. The free memory can be used for the other purpose also.

For the better understanding we divide this paper in three sections, section I discusses the techniques of CBIR with their advantages. Section II gives focus on Related Work. Section III Background: aggregating local descriptors Section IV Performance measures. V Conclusion and References.

II. RELATED WORK

The CBIR process includes the following structure
1) Query image
   System is given a query image in the form of image.
2) Feature extraction
   The features which are used to represent an image such as texture, color etc. are extracted.
3) Visual dictionary
   For creating the visual dictionary the different methods are used Gaussian Mixture Model, k-means etc.
4) Image Representation
   For representing an image the Bag-of-Features, Fisher Vector is used.
5) Similarity Matching
   It involves matching the feature to get the
6) Similar visual.
   The similarity is calculated by using the distance formula.
7) Retrieval
   Finally the output is the relevant images from the database i.e. decreasing order.

A. Feature Extraction
There are different methods of the feature extraction. The features which are used to represent an image such as texture, color etc. are extracted. Brief details are given below of the features.
1) Generally the color feature include the color histogram, color covariance matrix etc. The features which are color are mostly used features in image retrieval. The MPEG-7 is a standard, which improves content based image retrieval by providing a set of descriptors. MPEG-7 has included main color, scalable color, color structure, color layout as color feature.

2) Texture based is the another feature and the useful information for the image retrieval. There are various approaches to extract and represent texture. They are the wavelet and Gabor filter are used to extract the features.

3) Shapes is the another feature extraction methods. The natural scene object are generally recognized by the shape. They are local and global feature.

B. Image Representation

Different methods are used for image representation as it is specified above. So the related word is given below

- F. Perronin [1] Fisher kernel framework is used to cope up with the problem of large scale data. In this Fisher Kernel is used to solve the problem of identical scene image retrieval. Generic framework Fisher kernel is introduced in for classification purposes to combine the best of the generative and discriminative worlds. The Fisher kernel has not been any thorough evaluation for retrieval. The drawback of Fisher vector is that it is high dimensional and as it is dense. Fisher representation is used for image retrieval and normalization procedure is used to improve the retrieval performance for huge amount of database. As the Fisher vector is dense they applied the standard binary encoding technique. The results are shown on the dataset which shows that the Fisher vector performs very well using a little as a few hundred bits per image and significantly better than a very recent compressed BOV approach.

- F. Perronin [2] Fisher kernel exted the bag-of-words. The important limitation is the scalability to large quantities of training images. Fisher kernels are the good way to exploit your generative model. A Fisher kernel is based on GMMs. Fisher kernel framework is used to show that they can increase the accuracy of image classification. For learning the classifier ImageNet and Flickr groups, they compare the two plentiful sources of training image [2]. Main advantage of Fisher Kernel over Bag-of-Words is having smaller dictionaries.

- H. Jegou [3] an efficient way of aggregating local image [3] descriptor is proposed. applying a vector of dimension which is the simplifies the Fisher Kernel representation. Finally optimizing dimension and then indexing is done using algorithm. The contribution provide the better search accuracy with a reasonable dimensionality and the advantage is to combine the optimizing of trade-off between dimensionality reduction and the indexing algorithm[3].

- M. Jain [4] In this paper the image classification framework based on patch matching is proposed. Hamming-Embedding system consist of similarity-based matching method and similarity space encoding. It allows the user to use a linear Support Vector Machine.

- H. Jegou [5] this paper tries to overcome the problem of performance of the retrieval. The limitations which are taken into consideration: accuracy, efficiency and memory cost. Firstly how the local descriptor aggregation is performed using Fisher kernel framework. Other is to reduce the dimension so that it can provide search accuracy. Fisher Kernel is used for image representation which transforms variablesize set into a fixed size representation. Product quantization is used for vector compression which jointly minimizes the dimensionality reduction and compression. The high accuracy is preserved by image representing which is reduced to less number of byte.

- J. Linyz [6] fisher kernel is used for improving the performance of retrieval and smaller memory. Fisher kernel is a most commanding device, which combine the both generative and discriminative models. Fisher vector representation of fixed length for a given image is formed by aggregating the local descriptor. Robust fisher vector representation is used to degrade negative impact of noise on Fisher Vector. For large scale data the performance of retrieval has been measured. The proposed system gives the benefits of good performance and lower memory complexity.

- M. Reddy [7][16] without the use of dense image representation the big size of data cannot be retrieved. To make the process capable the image need to be represented in the form of dense codes. The dense representation should occupy a smaller amount of memory and also separate the classes present in dataset. Sparse coding is done for representation of the local features which are extracted from an image and is gained by k-Singular Value Decomposition which is used for training the dictionary. VLAD aggregates the residuals to obtain a dense image representation. Dense image representation is improved and so it gives improved performance of retrieval.

- T. Takashi [8] to achieve the good performance and lower memory charge. They investigate the distribution of local descriptor is modeled for every individual image in the database and the query is used from the local descriptor. Feature representation is done by using Gaussian mixture model of subspace for representing an image. In a learning stage the PCA is applied for dimension reduction and then the mixture matrix is optimized and the optimized covariance matrix is calculated. For image retrieval, we further need to reduce the memory charge. So for this the compact coding is used.

- To achieve all the low memory cost and high accuracy Fisher vector is used. Fisher Vector is the method for image representation. Fisher Vector and Vector of Locally Aggregated Descriptor gives better performance than BOF.

III. BACKGROUND: AGGREGATING LOCAL DESCRIPTORS

A. Bag-Of-Features

In Bag-of-features representation grouping of the local descriptor is done and the system is learned the visual vocabulary by using visual vocabulary [3][7]. Lastly image
representation is done by using the frequencies of visual words.

B. Fisher Vector

In order to overcome the drawback of the BoW the fisher kernel is used. The fisher kernel is the most commanding instrument to change an received variable volume set off free samples into rigid volume vector representation [3][7]. Independent images are considered in this feature extraction. Representation of an image is done by using the likelihood function with respect to its parameters. Mapping the distribution of the features around key is captured by this approach for representing an image. Fisher vector are compressed to lower dimensional index vector of the image. Maximum Likelihood estimation is used for matching the similarity.

C. Vector of Locally Aggregated Descriptors

It is the simplification of the Fisher Kernel. In the learning phase, the vector quantifier is used by k-means [3][7]. The codebook \( \omega_1, \ldots, \omega_k \) is learned by using k-means. For each codeword \( \omega_i \) is given by

\[
V \cdot \omega := v_i + (x - \omega_i)
\]

The local descriptor to be d-dimensional, then the dimension is kd. And finally the L2 is normalized.

D. Discussion

BOF versus FV [6] both generate a dense image representation by aggregating local descriptor. Fisher vector employs top order statistics of local descriptor. BOF only considers the count of features which are extracted and fisher vector considers the mean and covariance.

IV. PERFORMANCE MEASURE

The performance is evaluated using the mean average precision. The mean average precision for number of class is the average of the average precision of each class.

\[
\text{Average Precision} = \frac{\sum_{i=1}^{n_q} p(i) \text{relevant}(i)}{n_q}
\]

Where,

\[
n_q = \text{number of images of query class present in database}
\]

\[
p(i) = \text{precision at cut-off i in received record}
\]

V. CONCLUSION

Content Based Image Retrieval is a challenging method for the feature extraction from the large scale image dataset. Though lots of method is used to increase the accuracy but they have not achieved. In CBIR system, texture and color is considered to be a primitive visual cue like color and shape of an image. But there is lots of scope to enhance the retrieval speed with a proper technique. The mixture of the subspace image representation is introduced for achieving high accuracy and efficiency and the compact coding is done for the large scale image representation. Fisher vectors are the state-of-the-art model-dependent but class image representation. Hence, the FV can be successfully used in much application: Image classification, semantic picture segmentation, mono and multi-model image. It is found out through survey that the survey that its representation outperforms the sate-of-art FV based approach. The accuracy is increased and less memory is required through approach. However it is found through the survey that it give the better accuracy by using the Support Vector Machine denoted as SVM.

VI. ACKNOWLEDGEMENT

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.

REFERENCES


