7th International Conference on Science, Technology & Management
Guru Gobind Singh Polytechnic, Nashik (ICSTM-17)
25th February 2017, www.conferenceworld.in ISBN: 978-93-86171-30-6

QUERY FUSION FOR IMAGE RETRIEVAL

Aarti S. Datir[1] , Dipak V. Patil[2]

[1]*Department of Computer Engineering, G.E.Ss R.H.Sapat C.O.E.M.S and R,Nashik (India)*

[2]*Department of Computer Engineering, G.E.Ss R.H.Sapat C.O.E.M.S and R,Nashik (India)*

ABSTRACT

Image retrieval system is a computer system for browsing, searching and retrieve images from a huge data set of digital images. Content-based image retrieval (CBIR) is the function of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large datasets. CBIR retrieves similar images from huge image dataset based on image features. Content-based means that the searching will evaluate the actual contents of the image. The content of image might refer colors, shapes, textures, or any other information that can be derived from the image itself. There are two major approaches to content-based image retrieval using local image descriptors. One of them is descriptor by descriptor matching and the other is based on comparison of global image representation that describes the set of local descriptors of each image. Image representation is one of the key issues for large-scale CBIR. Using MPEG-7 descriptors and local descriptors number of features are extracted from given query image, to reduce the feature size principle component analysis(PCA) is used. These features are embedded and aggregated into a compact vector to avoid indexing each feature individually. In the embedding step, each local descriptor is mapped into a high dimensional vector. The aggregation step integrates all the embedded vectors of an image into a single vector which obtains a compact representation for image retrieval. Subsequently, k-means algorithm is used, clusters are formed and images are trained. In re-ranking step, Euclidean distance is used to find similarity and provide efficient searching.

Keywords :CBIR, MPEG-7 Descriptors, PCA, Embedding, Aggregation, K-Means, Euclidean Distance

I INTRODUCTION

With the growth in social media, internet and multimedia technologies, a large amount of multimedia data in the form of images, video, audio has been used in many fields like satellite data, medical treatment, digital forensics, video and still images repositories and investigation system. This has created an ongoing demand of systems that can store and retrieve multimedia data in an effective manner. Many multimedia information storage and retrieval systems have been developed. The most common retrieval systems are Text Based Image Retrieval (TBIR) systems, where the search is based on automatic or manual annotation of images. A conventional TBIR searches the database

for the similar text nearby the image as given in the query string. The commonly used TBIR system is Google Images. The text based systems are fast as the string matching is computationally less time consuming process. However, sometimes it is difficult to express the entire visual content of images in words and TBIR may result in producing irrelevant results, also annotation of images is not always correct and consumes a lot of time[1][5]. To overcome these problems of TBIR systems content based image retrieval systems (CBIR) was developed[1]. A CBIR system uses visual contents of the images described in the form of low level features like shape, color, texture, and spatial locations to represent the images in the databases. The system retrieves similar images when an query image or sketch is presented as input to the system[1].

In a typical CBIR system image low level features like shape, texture, color and spatial locations are represented in the form of a multidimensional feature vector[2]. The collection of feature vectors of images in the database called as feature database. The retrieval process is started when a user query the system using an example image or sketch of the object. The query image is converted into the internal representation of feature vector using the same feature extraction method. In this Number of features are extracted from given input image using MPEG-7 and local descriptors. Features can be color, shape, edges, corners, regions etc. MPEG-7 is a multimedia content description standard which offers a comprehensive set of descriptors for the multimedia data description[3][4]. Following MPEG-7 descriptors are used for feature extraction:

1) Color Layout: This descriptor effectively represents the spatial distribution of color of visual signals in a very compact form. This compactness allows visual signal matching functionality with high retrieval efficiency at very small computational costs. It provides image-to-image matching as well as ultra high-speed sequence-to-sequence matching, which requires so many repetitions of similarity calculations[6].

2) Scalable Color: The Scalable Color Descriptor is a Color Histogram in HSV Color Space, which is encoded by a Haar transform. Its binary representation is scalable in terms of bin numbers and bit representation accuracy over a broad range of data rates. The Scalable Color Descriptor is useful for image-to-image matching and retrieval based on color feature. Retrieval accuracy increases with the number of bits used in the representation[6].

3) **Edge Histogram :**The edge histogram descriptor represents the spatial distribution of five types of edges, namely four directional edges and one non-directional edge. Since edges play an important role for image perception, it can retrieve images with similar semantic meaning. Thus, it primarily targets image-to-image matching (by example or by sketch), especially for natural images with non-uniform edge distribution. In this context, the image retrieval performance can be significantly improved if the edge histogram descriptor is combined with other Descriptors such as the color histogram descriptor. Besides, the best retrieval performances considering this descriptor alone are obtained by using the semi-global and the global histograms generated directly from the edge histogram descriptor as well as the local ones for the matching process[7].

When feature extraction is done on image using descriptors, the feature space is increased. To reduce the feature space there is a method called Principle Component Analysis(PCA). PCA is a useful statistical technique that has

 ISBN: 978-93-86171-30-6

found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. PCA is a method for finding these patterns in the data, and compress the data, ie. by reducing the number of dimensions, without much loss of information. This method perform dimension reduction in feature spaces. Principle component analysis(PCA) consist of image color reduction while 3 color component are reduced into one containing a major part of information and then object orientation is done. It mainly concerned with identifying correlation in the data. Using standard deviation, variance, co-variance feature size is reduced[8].

The k-means algorithm is used for clustering, by applying k-means algorithm the clusters of an images are formed and images are trained. The generalized k-means algorithm is explained in the following paragraph.

K-means clustering is a method commonly used to automatically partition a data set into k groups. It proceeds by selecting k initial cluster centers and then iteratively refining the results. The algorithm converges when there is no further change in assignment of instances to clusters[3][4].

1 Decide on a value for k.

2. Initialize the k cluster centers (randomly, if necessary).

3. Decide the class memberships of the N objects by assigning them to the nearest cluster center.

4. Re-estimate the k cluster centers, by assuming the memberships found above are correct.

5. If none of the N objects changed membership in the last iteration, exit. Otherwise go to 3.

The similarity measure like Euclidean distance is used to calculate the distance between the feature vectors of query image and target images which are in the feature database. Finally, the retrieval is performed using an indexing techniques which provide the efficient searching of the image database[4].

II LITERATURE REVIEW

R. Arandjelovic and A. Zisserman [9] proposedRootSIFT method instead of SIFT for improved retrieval performance. It provides a performance boost at no cost. It is very easy to implement, It does not increase storage requirements as SIFT. Using a square root (Hellinger) kernel instead of the standard Euclidean distance to measure the similarity between SIFT descriptors leads to a dramatic performance boost in all stages of the pipeline.

R. Arandjelovic and A. Zisserman introduced the Vector of Locally Aggregated Descriptors (VLAD) ,this image descriptor was designed to be very low dimensional(e.g. 16 bytes per image) so that all the descriptors for very large image datasets (e.g. 1 billion images) can fit into main memory. They presented three methods which improve standard VLAD descriptors over various aspects, namely cluster center adaptation, intra-normalization and MultiVLAD. Cluster center adaptation is a useful method for large scale retrieval tasks where image databases grow

 ISBN: 978-93-86171-30-6

with time as content gets added. Intra-normalization was introduced in order to fully suppress bursty visual elements and provide a better measure of similarity between VLAD descriptors. It was the best VLAD normalization scheme. Multi-VLAD method is used to improves retrieval performance for small objects[10].

A. Babenko and V. Lempitsky [11] introduced and evaluated a new data structure called the Inverted Multi-Index which is a new data structure for the large-scale retrieval in the datasets of high-dimensional vectors. The multi-indices produce much finer subdivisions of the search space without increasing the query time and the preprocessing time compared to inverted indices. Multi-indices provide faster and more accurate retrieval and approximate nearest neighbor search, especially when dealing with very large scale datasets. It replace the vector quantization inside inverted indices with the product quantization (PQ). PQ proceeds by splitting high-dimensional vectors into dimension groups , then effectively approximates each vector as a concatenation of several code words of smaller dimensionality.

L. Bo and C. Sminchisescu [12] propose efficient match kernels (EMK) that map local features to a low dimensional feature space and average the resulting vectors to form a set level feature. Efficient match kernels (EMK) combine the strengths of both bag of words and set kernels. It map local features to a low dimensional feature space and construct set-level features by averaging the resulting feature vectors. By illustrating the quantization limitations of such models and propose more sophisticated kernel approximations that preserve the computational efficiency of bag-of-words.

L. Chu, et al [14] propose a rotation invariant PDIR method, which improves the image retrieval performances by exploiting the group spatial consistency of visual word matches. It propose the Combined-Orientation-Position (COP) consistency to softly quantize the relative spatial relationship between visual word matches in a rotation invariant way, then embed the COP consistency into a simple consistency graph model to efficiently find the group of most consistent visual words. The proposed PDIR system has only one system parameter, which improves its robustness while dealing with different data. Proposed method is also effective in retrieving the near duplicate images with large area of duplicate regions.

O. Chum and J. Matas [15] proposed an efficient algorithm, based on min-Hash for discovery of dependencies in sparse high dimensional data. The dependencies are represented by co-ocsets, i.e. sets of features co-occurring with high probability. These structures dominate the computed similarity, completely ruining the results of standard retrieval. Two methods for managing co-ocsets in such case have been proposed. Both methods significantly outperform the state-of-the-art.

O. Chum, et al [16] done query expansion into the visual domain via two novel contributions. Firstly, strong spatial constraints between the query image and each result allow to accurately verify each return, suppressing the false positives which typically ruin text-based query expansion. Secondly, the verified images can be used to learn a latent feature model to enable the controlled construction of expanded queries.

H. Jegou, M. Douze, and C. Schmid [17] proposed IDF like weighting method to address burstiness problem, for this three strategies are used namely multiple match removal (MMR), Intra-image Burstiness, Inter image burstiness. Herv´eJ´egou, et al [18] addressed the problem of large-scale image search. By considering search accuracy, efficiency, and memory usage, evaluate different ways of aggregating local image descriptors into a vector and show that the Fisher kernel achieves better performance than the reference bag-of-visual words(BOW) approach for any given vector dimension. Then optimize dimensionality reduction using PCA and indexing using Approximate nearest neighbor (ANN) in order to obtain a compact representation.

H. J´egou and A. Zisserman [19] proposed a method to reduce the interference between local descriptors when combining them to produce a vector representation of an image. In this they uses two methods namely T-embedding to reduces the impact of unrelated matches on the image similarity and Democratic aggregation to explicitly limits the interference between descriptors when aggregating them.

H. J´egou and O. Chum [20] proposed Different techniques are introduces to improve dimensionality reduction by PCA for large scale image retrieval, they proposes an effective way to alleviate the quantization artifacts through a joint dimensionality reduction of multiple vocabularies.

J. Philbin, et al [21] proposed descriptor-space soft-assignment. It improves the state of the art performance on standard datasets by collecting information about image patches that is lost in the quantization step of previously published methods.

G. Tolias and H. J´egou proposes a query expansion technique for image search, representation of the query is obtained by exploiting the binary representation provided by Hamming Embedding image matching approach.[22]

R. G. Cinbis, J. Verbeek, and C. Schmid [23] introduced latent variable models for local image descriptors, which avoid the common but unrealistic iid assumption. The Fisher vectors of non iid models are functions computed from the same sufficient statistics as those used to compute Fisher vectors of the corresponding iid models. Using the Fisher kernel we encode an image by the gradient of the data log-likelihood w.r.t. hyper-parameters that control priors on the model parameters. This system involves discounting transformations similar to taking square-roots, by including Gaussian mixtures over local descriptors, and latent topic models to capture the co-occurrence structure of visual words, both improve performance.

Takashi Takahashi and Takio Kurita [24] introduced the mixture of subspaces image representation, which obtains both high accuracy and low memory cost in large-scale image retrieval. This representation outperforms the state-of-the-art FV-based approach in both plain and encoded representation. The distribution of local descriptors is modeled for each database image, and the likelihood function of each model is used for matching a query to the database images.

Mohammed Alkhawlani, et al [25] Proposed a system for CBIR, which uses local feature descriptors to produce image signatures that are invariant to rotation and scale. This system combines the robust techniques, such as SIFT(Scale Invariant Feature Transform), SURF (Speeded Up Robust Features), and BoVW(Bag-of-Visual Word),

 ISBN: 978-93-86171-30-6

to enhance the retrieval process , Using k-means algorithm cluster the feature descriptors in order build a visual vocabulary and SVM(Support Vector Machine) is used as a classifier model to retrieve much more images relevant to the query efficiently in the features space.

III SYSTEM ARCHITECTURE

The proposed system can be roughly divided into following three blocks:

1. Feature Extraction using Local and MPEG-7 Descriptors
2. Feature Space Reduction using PCA
3. Cluster Formation Using K-means and Similarity Measure

The details of steps performed in proposed system are as follows:

Step 1: Get feature vectors of Training images in the dataset

The collection of feature vectors of images in the database called as feature database.

Step 2: Extract Features like CLD, EHD, SCD of images

Number of features are extracted from an image using descriptors.

Step 3: Aggregate / Combine all these features

Combine all extracted features of image such as color, shape, texture and other information into a vector.

Step 4: Apply PCA on these features

This method perform dimension reduction in feature spaces. Principle component analysis(PCA) consist of image color reduction while 3 color component are reduced into one containing a major part of information and then object orientation is done. It mainly concerned with identifying correlation in the data. Using standard deviation, variance, co-variance feature size is reduced.

Step 5: Calculate trained Matrix from PCA output

By applying PCA on features of an image the feature size is reduced, then calculate rained matrix from generated PCA output.

Step 6: Train K-means for clustering

K-Means is a least-squares partitioning method that divide a collection of objects into K groups. The algorithm iterates over two steps: Compute the mean of each cluster.

Compute the distance of each point from each cluster by computing its distance from the corresponding cluster mean. Assign each point to the cluster it is nearest to. Iterate over the above two steps till the sum of squared within group errors cannot be lowered any more.

Step 7: Query Image

User provides image as an input to the system.

Step 8: Extract feature of query images

ISBN: 978-93-86171-30-6

Feature extraction done on given input image.

Step 9: Classify cluster of image

Using k-means algorithm classify images in the database as compared to query image

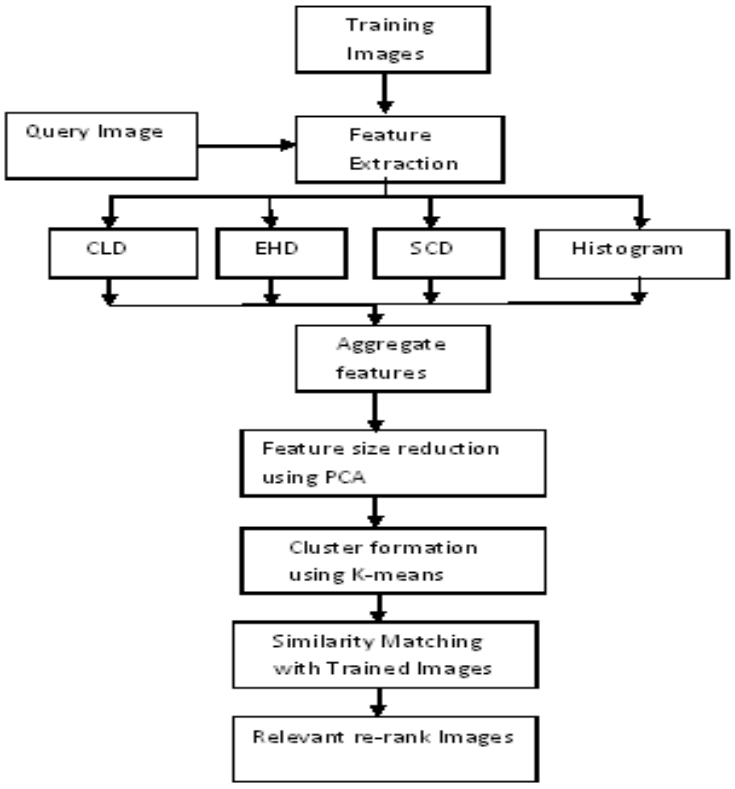
Step 10: Re-rank images.

Using similarity measure ranking of images is done.

Step 11: Retrieve relevant images

Provide relevant result as a output.

Following figure shows flow of the proposed system:



IV PERFORMANCE EVALUATION

1) Precision : Precision is the ratio of the number of relevant images retrieved to the total number of irrelevant and relevant images retrieved.

$$Precision = \frac{\text{Number of relevant images retrieved}}{\text{Total numer of images retrieved}}$$

 ISBN: 978-93-86171-30-6

2) Recall : Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. Recall is also known as sensitivity or true positive rate. It's the ratio of correctly predicted positive events

$$Recall = \frac{\text{Number of relevant images retrieved}}{\text{Total numer of images in database}}$$

V CONCLUSION

The proposed system is expected to provide compact image representation for large scale CBIR. Firstly we extract the multiple features from an image using MPEG-7 descriptors, reduce the feature space using principle component analysis (PCA).By using embedding and aggregation we reduces the impact of unrelated matches on the image similarity and limit interference between descriptors. K-means algorithm is used for image clustering which can improve the performance of image retrieval. Using euclidean distance we calculate the similarity distances between visual features and provide efficient searching.

VI ACKNOWLEDGMENT

I would like to express gratefulness to Dept. of Computer Engineering, GES's R.H. Sapat C.O.E.M.S and R, Nasik.

REFERENCES

[1] ZhanningGao, JianruXue, Wengang Zhou, Shanmin Pang, and Qi Tian, "Democratic Diffusion Aggregation for Image Retrieval" , IEEE Trans. Multimedia, , vol. 18, no. 8, pp. 16611674, Aug. 2016

[2] NidhiSinghai et al,"A Survey On: Content Based Image Retrieval Systems", International Journal of Computer Applications (0975 8887) Volume 4 No.2, July 2010

[3] Monika Jain and Dr. S.K.Singh,"A Survey On: Content Based Image Retrieval Systems Using Clustering Techniques For Large Data sets", International Journal of Managing Information Technology (IJMIT) Vol.3,No.4, November 2011

[4] Aarti Datir and D. V. Patil," Survey on Different Techniques of Content Based Image Retrieval", International Journal of Science Technology Management and Research, Volume 1, Issue 8, November 2016

[5] P.S. Malge and Pasnur M.A. ," Performance Evaluation of Texture based Image retrieval", International Journal of Computer Applications (09758887) Volume 72 No.2, May 2013

[6] Hamid A. Jalab,"Image Retrieval System Based on Color Layout Descriptor and Gabor Filters",IEEE conf. on open systems,sep 2011

[7] Dong Kwon Park et al. Efficient Use of Local Edge Histogram Descriptor.

[8] Rafael do EspritoSanto,"Principal Component Analysis applied to digital image compression",Study carried out at Institute of Cerebro,jun 2012.

[9] R. Arandjelovic and A. Zisserman,"Three things everyone should know to improve object retrieval," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2012, pp. 29112918.

[10] R. Arandjelovic and A. Zisserman,"All about VLAD," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2013, pp. 15781585

[11] A. Babenko and V. Lempitsky,"The inverted multi-index," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2012, pp. 30693076

[12] L. Bo and C. Sminchisescu,"Efficient match kernel between sets of features for visual recognition," in Proc. Adv. Neural Inf. Process. Syst., 2009, pp. 135143

[13] Y.-L. Boureau, J. Ponce, and Y. LeCun,"A theoretical analysis of feature pooling in visual recognition," in Proc. Int. Conf. Mach. Learn., 2010, pp.111118.

[14] L. Chu, S. Jiang, S. Wang, Y. Zhang, and Q. Huang,"Robust spatial consistency graph model for partial duplicate image retrieval," IEEE Trans. Multimedia, vol. 15, no. 8, pp. 19821996, Jun. 2013

[15] O. Chum and J. Matas,"Unsupervised discovery of co-occurrence in sparse high dimensional data," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2010, pp. 34163423

[16] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman,"Totalrecall:Automatic query expansion with a generative feature model for objectretrieval," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2007, pp. 18. 91

[17] H. Jegou, M. Douze, and C. Schmid,"On the burstiness of visual elements," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2009,pp. 11691176

[18] H. Jegouetal.,"Aggregating local image descriptors into compact codes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 17041716, Sep. 2012

[19] H. Jegou and A. Zisserman,"Triangulation embedding and democratic aggregation for image search", in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2014, pp. 33103317

[20] H. Jegou and O. Chum,"Negative evidences and co-occurences in image retrieval: The benefit of PCA and whitening," in Proc. Eur. Conf. Comput. Vis., 2012, pp. 774787

[21] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman,"Lost in quantization: Improving particular object retrieval in large scale image databases," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008, pp. 18

[22] G. Tolias and H. Jegou,"Visual query expansion with or without geometry: Refining local descriptors by feature aggregation," Pattern Recog., vol. 47, no. 10, pp. 34663476, 2014

[23] R. G. Cinbis, J. Verbeek, and C. Schmid,"Image categorization using fisher kernels of non-IID image models," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2012, pp. 21842191

ISBN: 978-93-86171-30-6

[24] Takashi Takahashi, Takio Kurita," Mixture of Subspaces Image Representation and Compact Coding for Large-Scale Image Retrieval", IEEE Trans. on Pattern Anal. and Mach. Intell. , JULY 2015, VOL. 37, NO. 7,PP .1469-1479

[25] Mohammed Alkhawlani, Mohammed Elmogy, HazemElbakry, "Content-Based Image Retrieval using Local Features Descriptors and Bag-of- Visual Words, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9, 2015