Clustering based CBIR System

Anushri Aglawe [1] Supriya Bothara [2] Shraddha Burhade [3] Deepa Chandak [4]

[1] Department of Computer Engineering, K.K.W.I.E.E.R, Maharashtra, India
[2] Department of Computer Engineering, K.K.W.I.E.E.R, Maharashtra, India
[3] Department of Computer Engineering, K.K.W.I.E.E.R, Maharashtra, India
[4] Department of Computer Engineering, K.K.W.I.E.E.R, Maharashtra, India

Abstract—*The CBIR (Content Based Image Retrieval) implementation in searching images into image database requires usually for a sufficiently prolonged time because such image searching process is performed with comparison between searched images and individually records in an image database. In this work, it is proposed a K-Means clustering algorithm aiming to develop clusters from each image database records, it can later be used for optimizing image searching access period. Results of the clustering process in form of cluster table would be made as indexing in early image searching for cluster position determination from searched images to image records.*

Keywords— K-Means Clustering, CBIR, Image Database

1. INTRODUCTION

Early in the 1990, CBIR, which conducted a retrieval process based on a visual content in a form of the compositions of image colors, began to be developed. Currently, retrieval systems have also involved the user feedbacks, irrespective of whether or not an image of retrieval results was relevant (relevance feedback) which was used as a reference in modifying a retrieval process to obtain more accurate results.

Clustering is a method of grouping data objects into different groups, such that similar data objects belong to the same group and dissimilar data objects to different clusters . Image clustering consists of two steps the former is feature extraction and second part is grouping. For each image in a database, a feature vector capturing certain essential properties of the image is computed and stored in a feature base. Clustering algorithm is applied over this extracted feature to form the group.

The authors propose to use clustering techniques to allow for efficient access to large image databases. More efficient access is important, since due to the size of large image databases, querying becomes expensive even if the images are represented in a compact manner. With clustering, the task of retrieval is decomposed into a two stage process. In the first step an appropriate cluster is selected and in the second step the best matches from this cluster are returned. They compare a clustering technique which uses relative entropy to techniques using the Euclidean norm. We propose to use image clustering techniques to allow for faster searching in image databases. They compare different clustering techniques to find out which suits the task of clustering images best.

The authors propose to use image clustering to give a good overview of an image database to help a user find a sought image faster. To cluster this images, they estimate the distribution of image categories and search the best representative for each cluster . They represent images by a high-dimensional feature vector and propose a new clustering algorithm which they compare to other clustering techniques. We give general information about clustering of data and the evaluation of results.

In second section we will discuss about existing system and its drawbacks. After that in third section, architecture of proposed system is discussed followed by working of proposed system and then finally concluding with conclusion and future scope.

2. EXISTING SYSTEM

During Current Market survey, we found that there is no efficient system for retrieval of images, which are stored in database and also there is no such system which will give a better accuracy than our proposed system.
Due to the tremendous demands of image retrieval form databases, there arises need of efficient system which will give the result with more accuracy and within limited amount of time. In this project, we are implementing clustering based CBIR system using pattern matching. The systems purpose is to use clustering techniques to allow efficient image retrieval from large image databases .Due to large size of databases image retrieval becomes more difficult and querying becomes expensive even if the images are represented in a compact manner. The implemented system was designed using CBIR system with color model (RGB model) which was inefficient. Now, in this project we are using CBIR system with Pattern matching technique which is more efficient than CBIR system with RGB model. Which eventually leads to higher accuracy and speed enhancement.With clustering, the task of retrieval is decomposed. We propose to use image clustering techniques to allow for faster retrieval of images from image databases.

3. PROPOSED SYSTEM

In this section, we describe the architecture of the proposed system.

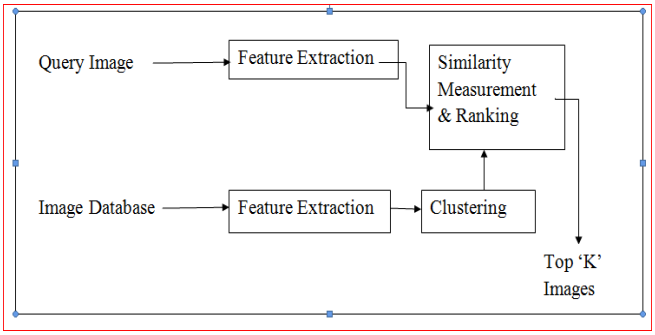


Fig -1: Clustering Based CBIR System Architecture

Clustering Based CBIR System is designed so that the task of retrieval becomes efficient. More efficient access is important, since due to the size of large image databases, querying becomes expensive even if the images are represented in a compact manner. The System is decomposed into a two stage process. In the first step an appropriate cluster is selected and in the second step the best matches from this cluster are returned.

The architecture consist of this blocks:

- Feature Extraction
- Clustering
- Similarity Measurement and Ranking

The first block, called the Feature Extraction, extracts the features of Query image and Database Images, while the second block called Clustering, forms the group of database images having similar extracted features. After Clustering, Similarity is measured and accordingly top 'K' ranking images are Displayed. User get top 'K' images as the output. Thus we make the system efficient.

4. WORKING OF PROPOSED SYSTEM

Clustering Based CBIR System is used to retrieve similar images as of query image from large image database. For retrieving similar images, we have to perform following functions –

- Image Upload
- Feature Extraction
- Clustering
- Pattern Matching
- Display Top 'K' Images

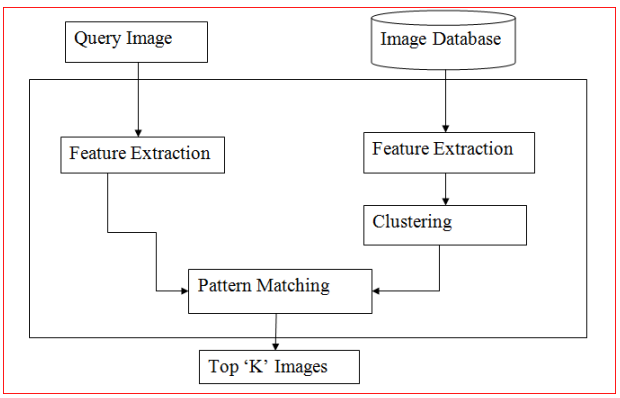


Fig -2: Working of Proposed System

The foremost efforts in building this system is to get accurate and efficient similar images from image database. The user gives the Query image to the system. The system performs the following functions to display top 'K' images using K-Means Clustering Technique.

4.1. Image Upload Process

The first task of the system is to get Query Image from the user. So, Image uploading must be easy and efficient. Suppose if we want to upload the image to the database then there should be no difficulty in doing that, for this we need to use efficient image uploading algorithm which will be efficient.

4.2. Feature Extraction Process

In image processing, feature extraction is a special form of dimensional reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract therelevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature Extraction will be done based on Color . After performing Feature Extraction on Query image and Database Images we will perform Clustering.

4.3. Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). Clustering can be done using different algorithms like k-mean, fuzzy-c, etc. But the efficient way is to use K-means for clustering ,which will give us more accurate result .

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n-observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The accuracy depends on number of clusters formed, so more the number of clusters, more accurate will be the result.

4.3.1 K-Means Algorithm

It is assumed that there were n clustered objects to get a sample set, that is, $X=\{x_1, x_2, x_3, \ldots\ldots, x_n\}$. By using K-means

algorithm, n sample objects are grouped into K clusters to ensure the similarities among samples in the same cluster and the differences among samples in different clusters. Specific procedure is as follows:

1) Randomly select K objects as initial cluster centers as following: $c_1, c_2, c_3, \ldots\ldots, c_k$.

2) According to the minimum distance principle, that is,

$$D_j = \sqrt{\sum_{j=1}^{k} (X - c_j)^2}$$

$X = \{x_1, x_2, \ldots, x_n\}$, $j = 1,2,\ldots,k$

each sample object is assigned to one of K clusters.

3) Take the average values of objects of each cluster as new clustering centers, average values can be get by

$$C_j = \frac{1}{n_j} \sum^{n_j} x_i, \quad j=1,2, \ldots..,K$$

nj is the number of objects in the cluster j

4) If the cluster centers have changed, repeat 2), 3) steps until the cluster centers do not change. As a result, clustering criterion function can be converged

$$I_c = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \| X_i^{(j)} - c_j \|^2 , X_i^{(j)} \in S_j$$

Cj is the clustering center of cluster Sj .

4.4. Pattern Matching

In this, We perform pattern matching on the extracted features of Query Image and extracted feature of the database image. It is the comparison of the extracted features to get top 'K' images.

4.5. Display Top K Images

After Performing Pattern matching, Top 'k' images are displayed to the user. The output should be accurate and efficient.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The COREL photo database of ten thousand images is used to evaluate the performance of CBIR system. We have selected a database of almost 'Thousand images' and then divide the database into number of concepts of different complexities to verify the results. Here we have chosen the Beach concept and worked on that by applying the target search methods on that database we have got the most relevant images.

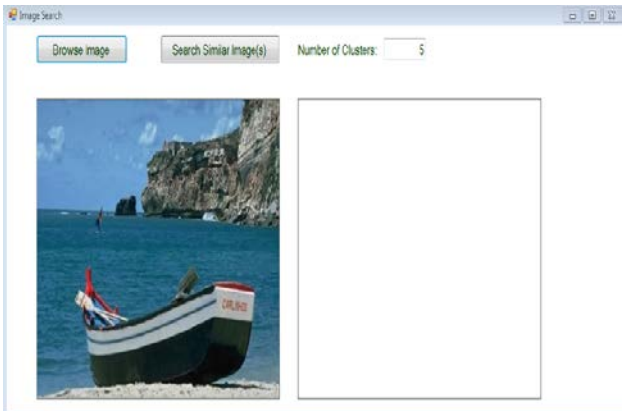


Fig 5.1 Query image

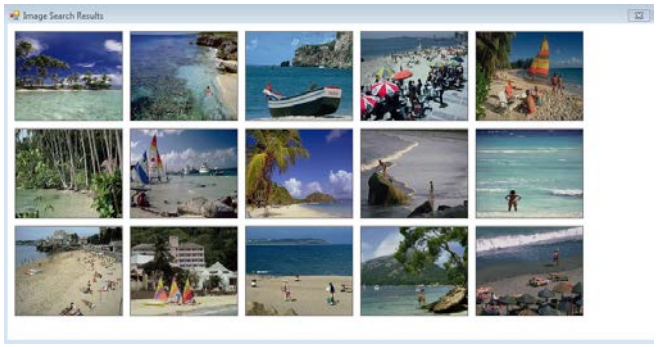
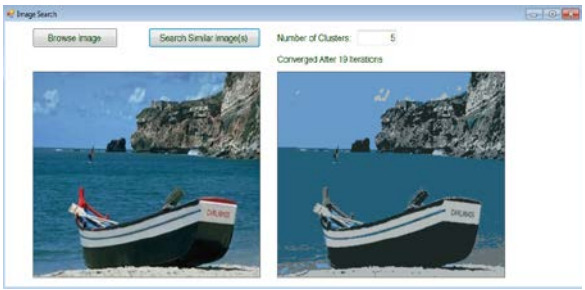


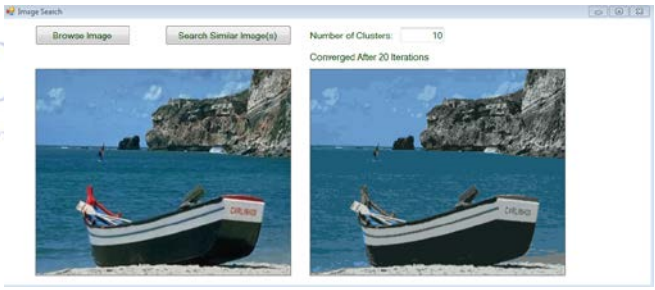
Fig 5.2 Image Retrieved

5.1 COMPARING ITERATIONS DEPENDING ON NUMBER OF CLUSTERS

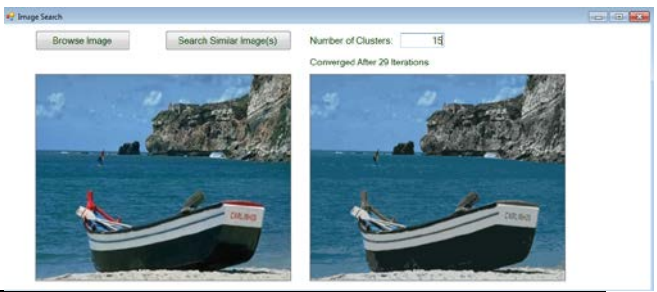
1. If number of clusters are 5, we required 19 iterations to get the result.



2. If number of clusters are 10, we required 20 iterations to get the result.



3. If number of clusters are 15, we required 29 iterations to get the result.



Sr. No.	Number of Clusters	Number of Iterations Required
1	01	02
2	02	05
3	03	07

4	04	13
5	05	19

Table I – Number of iterations required for Number of Cluster on same image.

6. CONCLUSIONS AND FUTURE SCOPE

The proposed system is designed to operate the content based image retrieval system. It has been verified with the photos of places of interest in dataset. Our experimental results demonstrate that our CBIR system architecture not only works well for image retrieval, but also improves its precision. In our knowledge, this paper first combines feature extraction module, K-means clustering and pattern matching module to build the CBIR system. The experimental results confirm that the proposed CBIR system architecture attains better solution for image retrieval. Our model represents the first time in which combine new modules and techniques proposed in the paper have been integrated with CBIR system. Images can be retrieved correctly through the proposed CBIR system. For those images which are contained in the dataset, all of them can be searched as the most similar result. Also for general images selected randomly, the query results are similar to the input data. Since the CBIR system is based on the color feature, the retrieval results are directly and easy to tell the performances. In the future work, we hope to build a generalized query method which increase the system searching ability and provide more accurate content descriptions of places of interest places by performing color feature analysis and CCH image extraction simultaneously. As a result, the CBIR system will be able to suggest more relevant annotations and descriptions. Furthermore, we hope to optimize the system architecture and modules proposed in this paper. There exists some detail setting can be discussed and optimized with the images retrieval issues.

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley, 1999.

[2] K. Barnard and D. Forsyth, “Learning the Semantics of Words and Pictures,” Proc. 8th Int’l Conf. on Computer Vision, vol. 2, pp. 408–415, 2001.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image Segmentation Using Expectation-Maximization and its Application to Image Querying,” IEEE Trans. Pattern Anal. Machine Intell., vol. 24, no. 8, pp. 1026–1038, 2002.

[4] Y. Chen and J. Z. Wang, “A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval,”

[5] Remco, C.V., Mirela, T., “Content-based image retrieval systems: a survey”. Tech. Report, Department of Computing Science, Utrecht University, 2000.

[6] Long, F., Zhang, H., and Feng, D., “Fundamentals of content-based image retrieval”. In Feng, D., Siu, W. C., and Zhang, H. J., (eds.), Multimedia Information Retrieval and Management–Technological Fundamentals and Applications. Springer, 2002.

[7] Han, J., Kamber M., “Data Mining: Concepts and Techniques”. Morgan Kaufmann Publishers, 2nd Edn., New Delhi, ISBN: 978-81-312-0535-8, 2006.

[8] Iyengar G., Lippman A. “Clustering Images Using Relative Entropy for Efficient Retrieval”. Proc. Workshop on Very Low bitrate Video Coding, Urbana, IL. 1998.