2011 International Conference on Advancements in Information Technology
With workshop of ICBMG 2011
IPCSIT vol.20 (2011) © (2011) IACSIT Press, Singapore

Fuzzy C-Means Clustering For Content Based Image Retrieval System

Saroj Shambharkar[1+] and Shubhangi Tirpude[2]

[1] Kavikulguru Institute of Technology & Science, Ramtek Department of Information Technology

[2] Ramdeobaba College of Engineering & Management, Department of Computer Science

Abstract. Due to the enormous increase in image database sizes, as well as its vast deployment in various applications, the need for CBIR development arose. The present Content Based Image Retrieval system is to describe the solution to the problem of retrieving the query image from the large image database using fuzzy C-Means clustering method. The CBIR can use the primitive features of an image such as texture, color, orientation and shape. These features are extracted and used as the basis for a similarity check between images stored in the database. In the proposed approach the color feature is used and uses the fuzzy c-means clustering algorithm. The database consists of 8 bit bmp format images of 256×256 size. The Hue Saturation Value color space is used. The color Histogram of each image in the database and the query image is obtained & then Median Filtering is applied to reduce the noise. The Fuzzy C-Means Clustering can used to obtain the more features of the images and to improve retrieval efficiency. The similarity between the query image and the images in the database is done using Quadratic Distance approach and the minimum distance image is retrieved from the database. The final result is obtained, that utilizes the features of the images as the basis for comparison and retrieval using different Matlab functions.

Keywords: Content-Based Image Retrieval, fuzzy C-Means Clustering, quadratic distance, Median Filtering, Matlab, HSV color space.

1. Introduction

Content based image retrieval (CBIR) is the application of Computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. The Content Based Image Retrieval tries to solve this problem as it provides the means to index, search and retrieve those images.

Content Based Image Retrieval is a task of searching images from a database and retrieval of an image, which are seems to be visually similar to a given example or query image. Content-based image retrieval uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. In typical content-based image retrieval systems, the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. These feature vectors can be computed by different methods available to the users. The CBIR system consists of following components:

- Query image: It is the image to be search in the image database whether the same image is present or not or how many are similar kind images are exist or not.
- Image Database: It consists of n number of images depends on the user choice.
- Feature extraction: It extracts visual information from the image and saves them as features vectors in a features database. The feature extraction finds the image description in the form of feature value (or a set of value called a feature vector) for each pixel. These feature vectors are used to compare the query with the other images and retrieval.
- Image matching: The information about each image is stored its feature vectors for computation

[+] Corresponding author. Tel.: + 9890938136.
E-mail address: Saroj.shambharkar@gmail.com.

process and these feature vectors are matched with the feature vectors of query image which helps in measuring the similarity.

- Resultant Retrieved images: It searches the previously maintained information to find the matched images from database. The output will be the similar images having same or very closest features as that of the query image.

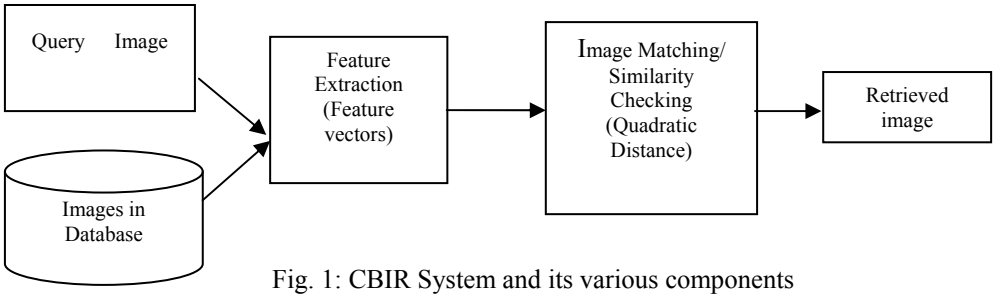

Fig. 1: CBIR System and its various components

1.1 Existing CBIR Systems

Some of the existing CBIR systems [3] are:

- QBIC or Query by Image Content
  It is the first commercial content based retrieval system. This system allows users to graphically pose and refine queries based on multiple visual properties such as color, texture and shape. It supports queries based on input images, user-constructed sketches, and selected colour and texture patterns.
- VisualSEEK and WebSEEK
  Virage is content based image search engine developed at Virage Inc.It supports color and spatial location matching as well as texture matching.
- NeTra
  This system uses color, shape, spatial layout and texture matching, as well as image segmentation.
- MARS or Multimedia Analysis and Retrieval System
  This system makes use of colour, spatial layout, texture and shape matching.
- Viper or Visual Information Processing for Enhanced Retrieval
  This system retrieves images based on color and texture matching.
- The img (Anaktisi**)** is a CBIR system on the web based on various descriptors which includes powerful color and texture features. The img (Anaktisi) provides different ways to search and retrieve them.

1.2 Objectives

There are two major issues concerned while designing a CBIR system:

- Every image in the image data base is to be represented efficiently by extracting significant features.
- Relevant images are to be retrieved using similarity measure between query and every image in the image data base.

2. Related Work

Fuzzy c means clustering method with thresholding for underwater image segmentation is explained. The paper focuses on comparison of fuzzy c means clustering algorithms with proposed method for underwater images. The paper focuses on comparison of fuzzy c means clustering algorithms with proposed method for underwater images. To evaluate the nonlinear image region segmentation, quantitative statistical measures have been used, such as the gray level energy, discrete entropy, relative entropy, mutual information and information redundancy [1].

A system based on the fuzzy c-means clustering algorithm, the CBIR system fuses color and texture features in image segmentation. A technique to form compound queries based on the combined features of different images is devised. This technique allows users to have a better control on the search criteria, thus a higher retrieval performance can be achieved [2].

Surveyed almost 300 key theoretical and empirical contributions in the current decade related to image retrieval and automatic image annotation, and in the process discuss the spawning of related subfields and discuss significant challenges involved in the adaptation of existing image retrieval techniques to build systems that can be useful in the real world [3].

The image is represented by a Fuzzy Attributed Relational Graph (FARG) that describes each object in the image, its attributes and spatial relation. The texture and color attributes are computed in a way that model the Human Vision System (HSV). A new approach for graph matching that resembles the human thinking process is proposed [4].

3. Proposed System Approach And System Architecture

3.1 System Architecture

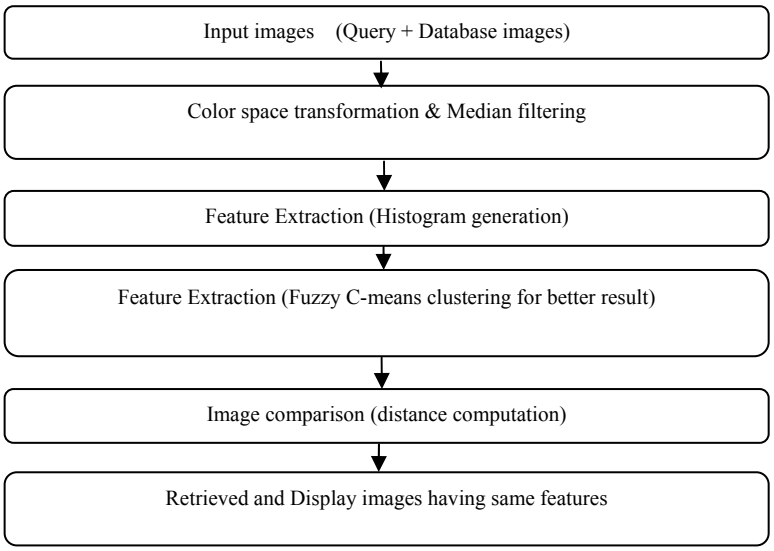

Fig. 2: CBIR using proposed approach

3.2 Image Feature Extraction

This process computes the image feature vectors which are then used by Quadratic distance calculation and helps in retrieval process .The HSV color model is used. To extract the color feature, color median filtering is applied recursively to deemphasize noises for preprocessing and color histogram are calculated. The median filter is to run through the signal entry by entry, replacing each entry with median of neighboring entries. After feature extraction, the pixel features are clustered into groups using the fuzzy c-means clustering algorithm if the result is to be improved. The query images are processed and transform to different color space for better performance, then similarity matrix can be calculated using fuzzy clustering of images. The distance between the two images is thus found and image having minimum distance that is the images similar to query image is retrieved and displayed.

The use of color features has become increasingly important. With the assistance of color features, objects in an image can be distinguished easily. For example, in skin cancer imagery, color is a useful feature to determine different types of skin cancer. A color model is specified in terms of 3-D coordinate system and a subspace within that system where each color is represented by a single point. The more commonly used color models are RGB (red, green, blue), HSV (hue, saturation, value), HIS, NTSC, YCbCr. Thus the color content is characterized by 3-channels from some color model. Color histograms are used in extracting the color features of images.

3.3 Proposed Fuzzy C-Means Clustering

The Fuzzy C-Means (FCM) is a clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. The FCM objective function and its generalizations are the most heavily studied fuzzy model in Pattern Recognition. There is an infinite range of possible fuzzy partitions. Therefore, an optimization model or objective function must be devised to search

for the optimal partition according to the chosen objective function. The way that most researchers have solved the optimization problem has been through an iterative locally optimal technique, called the FCM algorithm. The FCM objective function weighted the distance between a given data point and a given prototype by the corresponding degree of membership between the two. Thus, partitions that minimize this function are those that weight small distances by high membership values and large distances by low membership values.

3.3.1 Fuzzy C-Means Algorithm

Fuzzy C-Means is a method of clustering which allows one piece of data to belong to two or more clusters [11]. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^{m} \| x_i - c_j \|^2, 1 \le m < \infty$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i^{th} of d-dimensional measured data, is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\| x_i - c_j \|}{\| x_i - c_j \|}\right)^{\frac{2}{m-1}}} \qquad c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This iteration will stop when $\max_{ij}\{| u_{ij}^{k+1} - u_{ij}^{(k)} |\} < \varepsilon$,

Where ε a termination criterion between 0 and 1, and k is is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

The algorithm is composed of the following steps:

- Initialize U=[u_{ij}] matrix, U(0)
- At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

- Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\| x_i - c_j \|}{\| x_i - c_j \|}\right)^{\frac{2}{m-1}}}$$

- If $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$ then STOP; otherwise return to step 2.

4. Result of CBIR System

The CBIR system is working on color transformation, filtering, feature extraction, calculating distance between the feature vectors, similarity matrix .Using the feature extraction described above ,the feature vectors of query and images in the database or file are obtained and compared using the Quadratic distance Metric .The result is retrieved images matched on the basis on the feature vector values .

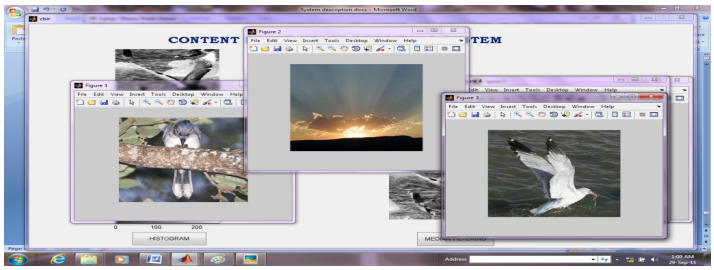

Fig. 3: Resultant retrieved images by CBIR system

5. Conclusion & Future Work

Content based image retrieval system is using the existing inbuilt function of MATLAB software is easiest way to implement. It is not necessary that image having same color is of same domain, so there is a need of comparing texture and shape also to improve results. As image collections grow in size the system may take a lot of time, and eventually reduce the query-retrieval process. To increase the speed and the user's interaction with image retrieval systems, the images to be access from the web/Internet sources and the CBIR system can be implemented over the World Wide Web and applying proposed fuzzy C-means algorithm in a more efficient manner.

6. References

[1] http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html, 27 July 2011, 6. 30 pm

[2] www.scribd.com/doc/59845513/91/Fuzzy-c-Means-clustering,15 July 2011, 5. 10 pm

[3] http://en.wikipedia.org/wiki/Content-based_image_retrieval.html, 5 September 2011, 3.54 pm

[4] www.ijcaonline.org/volume4/number2/pxc3871139.pdf, 2 July 2011, 2.10 am

[5] spict.utar.edu.my/WAIC-09CD/contents/pdf/WAIC09-1.3.pdf,6 Sept2011,3.24 pm

[6] www.mathworks.com/help/toolbox/fuzzy/fp310.html, 6 July 2011, 4.10 pm

[7] Ahmed Ismail sihab (Department of Computing, Imperial College of Science, Technology and Medicine, University of London), "FUZZY CLUSTERING ALGORITHMS AND THEIR APPLICATION TO MEDICAL IMAGE ANALYSIS".

[8] Dr.G.Padmavathi (Department of Computer Science, Avianshilingam Deemed University for Women Coimbatore) and Mr.Muthukumar (Department of Computer Science, Avinashilingam Deemed University, Coimbatore, "Image Segmentation using fuzzy c means clustering method with thresholding for underwater images"

[9] F. D. Donna, L. Maddalena, and A. Petrosino, About the," Embedding of Color Uncertainty in CBIR Systems, WILF, 4578:394-403, 2007".

[10] Heba Aboulmagd Ahmed-Cairo Technology Development Centre (CTDC) – IBM, Neamat El Gayar, Hoda Onsi (Faculty of Computers and Information – Cairo University), "A New Approach in Content-Based Image Retrieval Using Fuzzy Logic".

[11] Reza Toushmalani (Department of Physics, Hamedan branch, Islamic Azad University, Hamedan), "APPLICATION OF FUZZY C-MEANS ALGORITHM IN EULER DECONVOLUTION DEPTH OPTIMIZATION".

[12] Savvas A. Chatzichristofis and Yiannis S. Boutalis, Mathias Lux, "Img (Rummager), "An Interactive Content Based Image Retrieval System".

[13] W. S. Ooi and C.P. Lim, School of Electrical & Electronic Engineering, University of Science Malaysia," A Fuzzy Clustering Approach to Content-Based Image Retrieval" -Workshop on Advances in Intelligent Computing 2009.