International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Survey on Content Based Image Retrieval

Mehak Garg[1], Jyoti Rani[2]

[1, 2]Computer Science Engineering Department, Giani Zail Singh PTU Campus, Bathinda, India

Abstract: ***In the present scenario image retrieval plays an essential role. CBIR aims at finding image databases for specific images that are similar to a given query image based on its features. Users can query example images based on these features such as texture, color, shape, region and others. Target or close Images can be retrieved in a little fast if it is clustered in a right manner. For clustering, we use fuzzy- c mean. In this way relevant images will be retrieved from database.***

Keywords: content based image retrieval, RGB components, Texture, Auto-correlation.

1. Introduction

Image retrieval is the processing of searching and retrieving images from a huge database. As the images grow complex, retrieve the right images become a difficult problem. Content Based Image Retrieval is a task of searching images from a database and retrieval of an image, which are looking 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, texture, and spatial layout to represent the image. The CBIR system consists of the following components.

1.1 Query image

It is the image to be found in the image database, whether the similar image is present or not. And how many are similar kinds images are existing or not.

1.2 Image database

It exists of n number of images depends on the user choice.

1.3 Feature extraction

It separates visual information from the image and saves them as features vectors in a features database. The feature extraction finds the image detail 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 image with the other images and retrieval.

1.4 Image matching

The information about each image is stored its feature vectors for computation process and these feature vectors are matched with the feature vectors of query image which helps in measuring the similarity.

1.5 Resultant retrieved images

It finds the previously maintained information to find the matched images from database. The output will be the similar images having same or closest features as that of the query image.

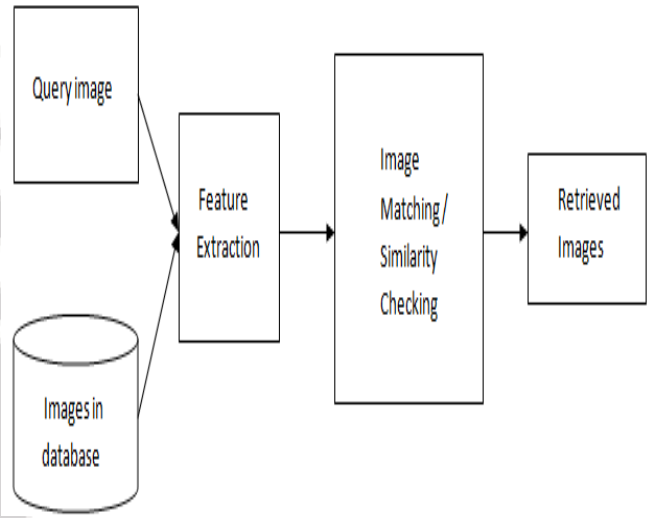

Figure 1: CBIR system and its various components

2. Application of CBIR Systems

- Architectural and engineering design
- Art collections
- Crime prevention
- Medical diagnosis
- Military
- Photograph archives
- Retail catalogs
- Face Finding

3. Problem Definition

3.1 Image database

First we have to create an image database. For this we need to search RGB components of images

3.2 RGB components

In the color based image retrieval the RGB Color model is used. Color images normally are in three dimensional.RGB color components are taken from each and every image. The mean value of Red, Green, and Blue components of target images are calculated and stored in the database. Based on RGB component mean values, images are clustered as Red, Green and blue component classification. These three mean values for each image are deposited and considered as features.

3.3 Feature extraction

The top ranked images are re-grouped according to their texture features. In the texture feature-based approach the parameters are gathered on the basis of statistical approach. Statistical features of gray levels are one of the systematic methods to classify texture. The various texture parameters like entropy, contrast, Auto-correlation, dissimilarity, standard deviation, mean, and variance of both query image and target images are calculated. From the calculated values required image from the database is extract.

The Gray Level Co-occurrence Matrix (GLCM) is used to extract second order statistics from an image. GLCMs have been used successfully for texture calculations.

3.4 Fuzzy-c mean

In fuzzy-c means clustering method, every point has a degree of belonging to the clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, the points on the edge of a cluster may be in the cluster to a smaller degree than points in the center of the cluster.

4. Image Retrieval Process

Image Retrieval from the image collections involves the following steps:

- Pre-processing
- Image Classification based on a true factor
- RGB Components processing
- Preclustering
- Texture feature extraction
- Similarity Comparison
- Target image selection

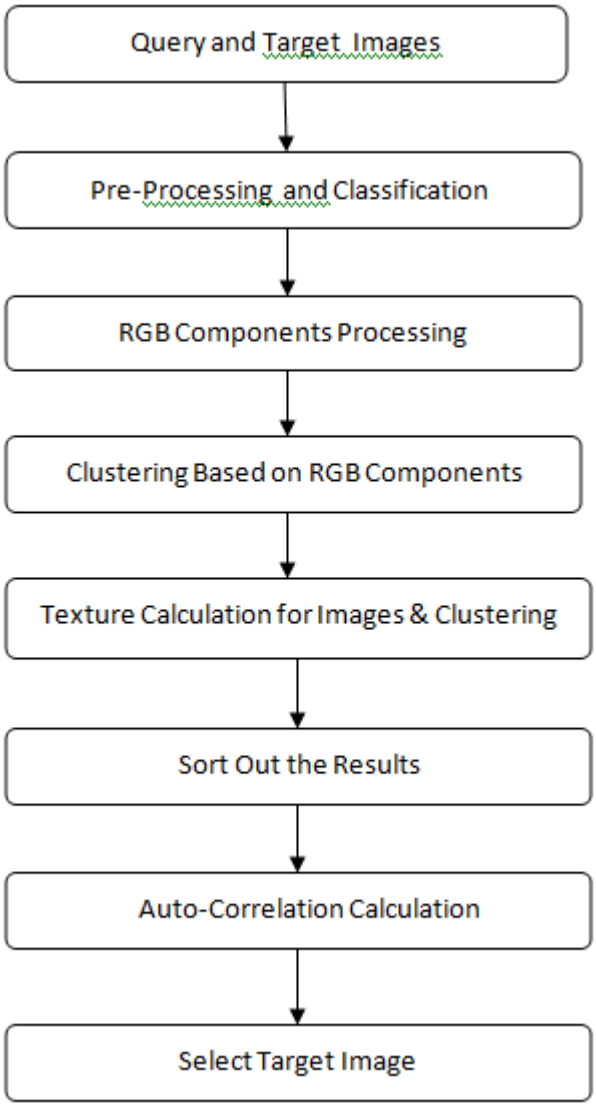

Figure 2: Image retrieval system

4.1 Pre-processing

Pre-processing is the name used for actions on images at the lowest level of abstraction. The main goal of the pre-processing is an improvement of the image that suppresses unwilling distortions or enhances some image features, which are important for future processing of the images. This step focus on image feature processing.

Filtering is a technique used in pre- processing for modifying or enhancing an image. The image is filtered to highlight certain features or remove other features. The noise in the image is filtered using linear and non-linear filtering techniques. Lucy - Richardson filtering is used here to reduce the noise.

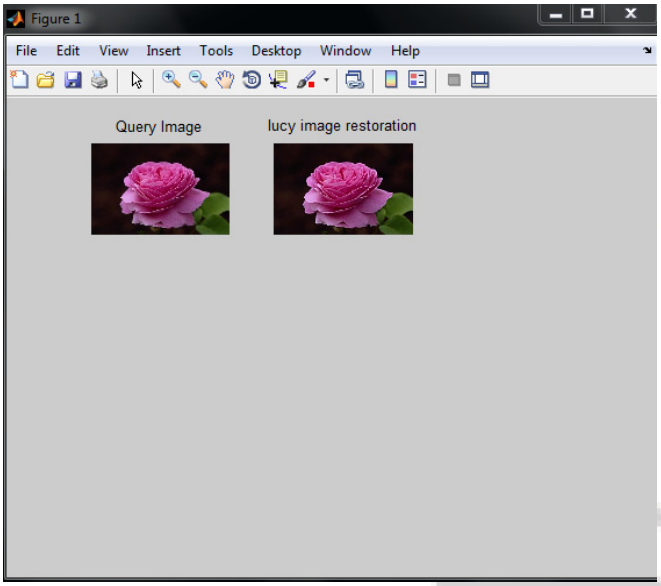

Figure 3: Pre-processing

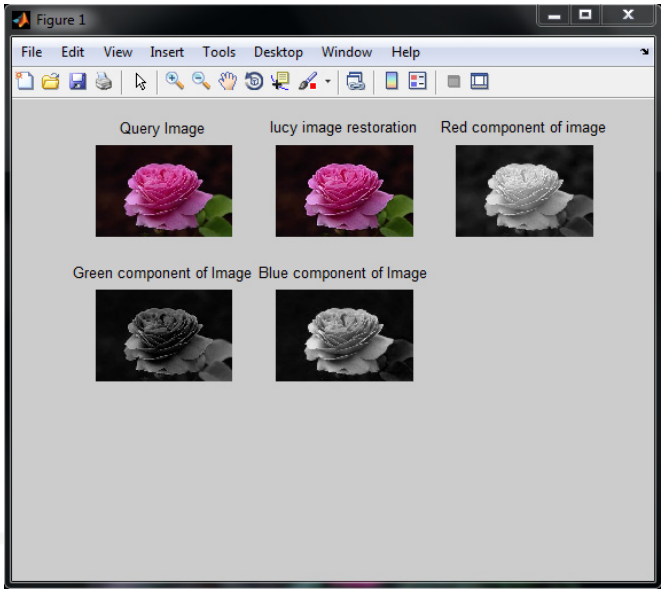

Figure 4: RGB components processing

4.2 RGB components processing

An RGB colour image is an M*N*3 array of colored pixels, where each color pixel is a triplet equivalent to the red, green, and blue components of an image. An RGB image can be seen as the stack of three grayscale images that, when inserted into the red, green, blue inputs of a color monitor, generate the color image on the screen. By custom the three images form an RGB images are known as red, green and blue components.

The mean values for the RGB components are calculated for all images:

$$\text{Red Mean (rm)} = \frac{\text{sum of all the red pixels in the image R(y)}}{\text{No. of pixels in the image P(y)}}$$

$$\text{Green Mean (gm)} = \frac{\text{sum of all green pixels in the image G(y)}}{\text{No. of pixels in the image P(y)}}$$

$$\text{Blue Mean (bm)} = \frac{\text{sum of all the blue pixels in the image B(y)}}{\text{No. of pixels in the image P(y)}}$$

Where R (y) = RED component pixels,

G (y) = GREEN component pixels,
B (y) = BLUE component pixels,
P =No. of pixels in the image.

After calculating the mean value of Red, Blue and Green components, the values are to be compared with each other to find the apogee value of the components. For example, if the value of Red component is Higher than the two, then we can conclude that the image is Red Intensity oriented image and which can be clustered into Red Group of Images.

Whenever the query image is already given, calculate the Red, Green, Blue components average values. Then compare this with the already stored values.

4.3 Auto-correlation

The texture represents the energy content of the image. If an image contains high textures, then the energy will be high compared to the average and low texture images. There are various texture parameters to be considered. However, here, the texture parameter Auto-Correlation is highly focused and which is to be calculated for the query and target images.

Auto-correlation introduces to the correlation of a time series with its own past and future values. Auto-correlation is also sometimes called “lagged” or “serial” correlation, which mention to the correlation between members of a series of numbers arranged in time. Positive auto-correlation might be considered a particular form of “diligence”, a tendency for a system to remain in the same state from one observation to the next observation.

4.4 Image clustering

Image Clustering will be a widely advantage for reducing the searching time of images in the database. Fuzzy c-means (FCM) is a technique of clustering which allows one piece of data to belong to two or more clusters. In fuzzy clustering data elements can belong to more than one cluster, and with each element a set of membership levels is associated. These indicate the strength of the corporation between that data element and a particular cluster. Fuzzy clustering is a technique of assigning these membership levels, and then using levels to assign data elements to one or more clusters.

In this clustering, each point has a degree of belongings to clusters, as in fuzzy logic, rather than belonging completely

too just one cluster. Thus, the points on the edge of a cluster may be in the cluster to a lesser degree than points in the centre of the cluster. FCM sort data in specific number of clusters.

5. Related Work

In this section we will look into the review of Content based image retrieval for image databases. It describes the previous work which had been done on a CBIR system using texture feature extraction and other techniques.

By Kannan in 2010[1] In this paper content based image retrieval method was proposed. It uses the feature of the image for its retrieval. The entropy texture feature is used here.

By Manesh B. Kokare, M.S.Shirdhonkar in 2010[2] In this paper document based image retrieval method was proposed. It includes the current state of the art of the research in document image retrieval based on images such as signature, machine-print, different fonts etc.

By Kun-Che in 2009[3] In this paper Pixel-wised image characteristics were extracted and changed into a database like table which permits a variety of data mining algorithms to make explorations on it.

By Silakari in 2009[4] In this paper Clustering of images based on color moment and Block Truncation Coding to extract features from an image database. K-means clustering algorithm is conducted to group the dataset in various clusters.

By Pattnaik in 2008[5] in this paper Data mining, clustering method together with Vector Quantization (VQ) is implemented to cluster and compact static color image.

Sheela in 2007[6] Developed a system which uses image mining approaches to categorize the images either as normal or abnormal and then divide the tissues of the anomalous Brain MRI to recognize brain related diseases.

Sanjay in 2007[7] used an Image mining approach using wavelet transform. The wavelet transform is utilized to decompose an image into dissimilar frequency sub bands and a small frequency sub band is used for Principal Component Analysis (PCA).

Abhi Gholap in 2005[8] uses A four-level system to exploit the knowledge of a pathologist with image examination, pattern identification, and artificial intelligence is used.

6. Conclusion

The main objective of the image retrieval is to retrieve the images from database very fast and in an efficient manner. The images are pre-processed with various techniques and the texture calculation is highly focused. Here, the images are clustered based on RGB Components, Texture values and Fuzzy C means Clustering algorithm. Clustering is very efficient and powerful technology to handle large data sets. It assists faster image retrieval and also allows the search for more relevant images in large image databases. Auto-correlation is used to compare the images and to improve the system performance.

References

[1] A.Kannan, Dr.V.Mohan, Dr.N.Anbazhagan, “Image Clustering and Retrieval using Image Mining Techniques”, 2010 IEEE International Conference on Computational Intelligence and Computing Research .
[2] Amanbir Sandhu, Aarti Kochhar, “Content Based Image Retrieval using Texture, Color and Shape for Image Analysis”, International Journal of Computers & Technology ,Volume 3, No. 1, AUG, 2012 .
[3] Saroj Shambharkar, Shubhangi Tirpude, “Fuzzy C-Means Clustering For Content Based Image Retrieval System”, 2011 International Conference on Advancements in Information Technology With workshop of ICBMG 2011, IPCSIT vol.20 (2011) © (2011) IACSIT Press, Singapore.
[4] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, “Image texture feature extraction using GLCM approach”, International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013, ISSN 2250-3153.
[5] J. Eakins, M. Graham, “Content-based image retrieval”, Technical Report, University of Northumbria at Newcastle, vol. 7, 1999.
[6] I.K. Sethi, I.L. Coman, “Mining association rules between low-level image features and highlevel concepts”, Proceedings of the SPIE Data Mining and Knowledge Discovery, vol. III, 2001, pp. 279–290.
[7] Abhi Gholap, Gauri Naik, Aparna Joshi and CVK Rao,“Content-Based Tissue Image Mining”, IEEE Computational Systems Bioinformatics Conference - (CSBW’05), vol.II, 2005, pp.359- 363.
[8] Zhang Ji, Hsu, Mong and Lee, “Image Mining: Issues, Frameworks and Techniques”, Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD’2001), in conjunction with ACM SIGKDD conference, San Francisco, USA, vol.9, 2001, pp.205-209.
[9] Lu Kun-Che and Yang Don-Lin, “Image Processing and Image Mining using Decision Trees”, Journal of information science and engineering, vol. 25, 2009, pp. 989-1003.
[10] Nai-Chung Yang, “A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval”, Journal of Visual Communication Image, vol.19, 2008, pp. 92–105.
[11] S.K. Chang, S.H. Liu, “Picture indexing and abstraction techniques for pictorial databases”, IEEE Trans. Pattern Anal. Mach. Intell, vol.4. 1984, pp. 475–483.
[12] Yanmei liang, “Fuzzy color-image retrieval”, optics communication, vol. 212, 2002, pp. 247- 250.

Author Profile

Er. Mehak Garg received the B.Tech degree in Computer Science Engineering from PTU GZS campus, Bathinda, India in 2012. Now I am doing M.Tech from PTU GZS campus. I have learnt Oracle,

SQL server, C++ . I have done my major project in ASP.net. During my project I have deal with visual studio and sql server also. I have put my project live also.

Er. Jyoti Rani received the B.Tech and M.tech degrees from SLIET Longowal, India and Punjabi University, Ptiala respectively. Her area of specialization is Image Processing. She has a teaching experience of approx 15 years.