Content Based Image Retrieval System Using Color and Texture Features

Prachi R. Kore[1], Prof. Mrs. Y. A. Sadawarte [2]

[1]PG Scholar, Department of Electronics Engineering, BDCE, Sevagram, Wardha –442102, INDIA
[2]Assistant Professor, Department of Electronics Engineering, BDCE, Sevagram, Wardha –442102, INDIA

Abstract- *Content based image retrieval (CBIR) for general purpose image databases is a highly challenging problem because of the large size of the database, the difficulty of understanding images, both by people and computers, the difficulty of formulating a query, and the issue of evaluating results properly. The common method for CBIR systems is to extract a signature for every image based on its pixel values and to define a rule for comparing images. The signature serves as an image representation in the view of a CBIR system. The components of the signature are called features. After extracting signatures, the next step is to determine a comparison rule, including a querying scheme and the definition of a similarity measure between images. For most image retrieval systems, a query is specified by an image to be matched. Color histogram as a global color feature and histogram intersection as color similarity metric combined with texture have been proved to give approximately good retrieval results.*

Keywords: **color moment, color histogram, Content Based Image Retrieval (CBIR), query image.**

1. INTRODUCTION

In this computer age, virtually all spheres of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design, and historical research use images for efficient services. A large collection of images is referred to as image database. An image database is a system where image data are integrated and stored. Image data include the raw images and information extracted from images by automated or computer assisted image analysis. The police maintain image database of criminals, crime scenes, and stolen items. In the medical profession, X-rays and scanned image database are kept for diagnosis, monitoring, and research purposes. In architectural and engineering design, image database exists for design projects, finished projects, and machine parts. In publishing and advertising, journalists create image databases for various events and activities such as sports, buildings, personalities, national and international events, and product advertisements. In historical research, image databases are created for archives in areas Efficient Content Based Image Retrieval that include arts, sociology, and medicine. In a small collection of images, simple browsing can identify an image. This is not the case for large and varied collection of images, where the user encounters the image retrieval problem. An image retrieval problem is the problem encountered when searching and retrieving images that are relevant to a user's request from a database. To solve this problem, text-based and content-based are the two techniques adopted for search and retrieval in an image database. In text-based retrieval, images are indexed using keywords, subject headings, or classification codes, which in turn are used as retrieval keys during search and retrieval. Text-based retrieval is non-standardized because different users employ different keywords for annotation. Text descriptions are sometimes subjective and incomplete because they cannot depict complicated image features very well. Examples are texture images that cannot be described by text. Textual information about images can be easily searched using existing technology, but requires humans to personally describe every image in the database. This is impractical for very large databases, or for images that are generated automatically. The Content Based Image Retrieval (CBIR) technique uses image content to search and retrieve digital images. Content-based image retrieval systems were introduced to address the problems associated with text-based image retrieval. Content based image retrieval is a set of techniques for retrieving semantically-relevant images from an image database based on automatically-derived image features. The main goal of CBIR is efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process. The computer must be able to retrieve images from a database without any human assumption on specific domain (such as texture vs. non-texture, or indoor vs. outdoor). One of the main tasks for CBIR systems is similarity comparison; extracting feature signatures of every image based on its pixel values and defining rules for comparing images. These features become the image representation for measuring similarity with other images in the database. An image is compared to other images by calculating the difference between their corresponding features. Content-based retrieval uses the contents of images to represent and access the images. A typical content-based retrieval system is divided into off-line feature extraction and online image retrieval. A conceptual framework for content-based image retrieval is illustrated in Figure 1. In off-line stage, the system automatically extracts visual attributes (color, shape, texture, and spatial information) of each image in the database based on its pixel values and stores them in a different database within the system called a feature database. The feature data (also known as image signature) for each of the visual attributes of each image is very much smaller in size compared to the image data, thus the feature database contains an abstraction (compact form) of the images in the image database. One advantage of a signature over the original pixel values is the significant compression of image representation. However, a more important reason for using

the signature is to gain an improved correlation between image representation and visual semantics. In on-line image retrieval, the user can submit a query example to the retrieval system in search of desired images. The system represents this example with a feature vector. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature database are then computed and ranked. Finally, the system ranks the search results and then returns the results that are most similar to the query examples. If the user is not satisfied with the search results, he can provide relevance feedback to the retrieval system, which contains a mechanism to learn the user's information needs.

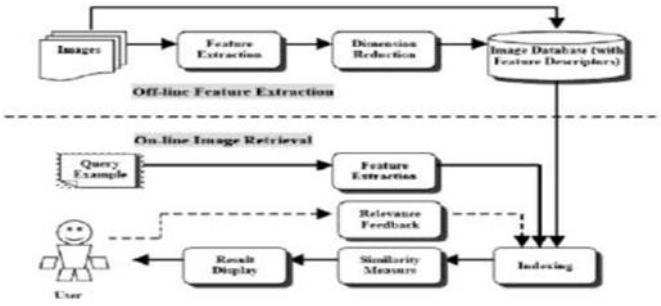

Fig-1: Conceptual framework for content-based image retrieval

The word transform refers to a mathematical representation of an image. There are several texture classifications using transform domain features in the past, such as discrete Fourier transform, discrete wavelet transforms, and Gabor wavelets. Wavelet transform is a good multi-resolution approach that represents the texture of an image in an effective way using multiple orientations and scales. This approach has a spatial property that is similar to mammalian perceptual vision, thereby providing researchers a good opportunity to use it in image processing. On the contrary, there is Content Based Image Retrieval(CBIR). This is different from previous one. This system will take input as a image rather than text. We can call it is as a query image. Now system will calculate texture and features of all the image like database images also. Now we can use many other transform like Gabor , Har Wavelet transform to extract texture features from all the images. We can also use color histogram to extract color features from all the images. A CBIR system is composed of a query interface for the acquisition of the query image, databases for storing indexing data and metrics, similarity and retrieval system. Now let's talk about applications of Content Based image retrieval(CBIR). Content-Based Image Retrieval has been used in several applications, such as medicine, fingerprint identification, biodiversity information systems, digital libraries, crime prevention, historical research, among others.[8]

The number of medical images produced by digital devices has increased more and more. For instance, a medium-sized hospital usually performs procedures that generate medical images that require hundreds or even thousands of gigabytes within a small space of time. The task of taking care of such huge amount of data is hard and time-consuming. That's one of the reasons that has motivated research in the field of Content-Based Image Retrieval. In fact, the medical domain is frequently mentioned as one of the main areas where Content Based Image Retrieval finds its application. Biologists gather many kinds of data for biodiversity studies, including spatial data, and images of living beings. Ideally, Biodiversity Information Systems (BIS) should help researchers to enhance or complete their knowledge and understanding about species and their habitats by combining textual, image content-based, and geographical queries.[8]There are several digital libraries that support services based on image content. One example is the digital museum of butterflies, aimed at building a digital collection of Taiwanese butterflies. This digital library includes a module responsible for content-based image retrieval based on color, texture, and patterns.[9]

2. LITERATURE REVIEW

In this paper, a method combining both color and texture features of image is proposed to improve the retrieval performance. Given a query, images in the database are firstly ranked using color features. Then the top ranked images are re-ranked according to their texture features. Results show the second process improves retrieval performance significantly. However, image retrieval using color features often gives disappointing results because in many cases, images with similar colors do not have similar content.[1]

An experimental comparison of a number of different texture features for content-based image retrieval is presented in this paper. The primary goal is to determine which texture feature or combination of texture features is most efficient in representing the spatial distribution of images. In this paper, authors analyze and evaluate both Statistical and Structural texture features. For the experiments, publicly available image databases are used. Analysis and comparison of individual texture features and combined texture features are presented. The First-order statistics, second-order statistics, Gabor transform and 2D Wavelet transforms were considered for retrieval. The retrieval efficiency of the texture features was investigated by means of relevance. According to the results obtained it is difficult to claim that any individual feature is superior to others. The performance depends on the spatial distribution of images.[2]

In this paper, authors presented a CBIR system that uses Ranklet Transform and the color feature as a visual feature to represent the images. Ranklet Transform is proposed as a preprocessing step to make the image invariant to rotation and any image enhancement operations. To speed up the retrieval time, images are clustered according to their features using k-means clustering algorithm. To evaluate the proposed system, authors used each image in our database to be a query image and submit it to the system and calculated the precisions for each query in all classes. Then for each class average of all precisions was calculated.[3]

2. EXISTING SYSTEM

TBIR(Text Based Image retrieval) and CBIR. Now firstly there is text based image retrieving , in this system there is a input of text rather than image. But text can be the inappropriate input of the system. That is suppose text can be the name of an image, we can call it as a metadata, annotation others.

2.1 Drawbacks of existing system:

- Unexpressed feelings and emotions when end user enters any image name.
- Human or can say end user can enter misspelling of image name.

3. PROPOSED SYSTEM

Content-Based Image Retrieval has been used in several applications, such as medicine, fingerprint identification, biodiversity information systems, digital libraries, crime prevention, and historical research, among others. It will accept input as an image, can say query image.

And according to the similarity between query image and database images this system will retrieve most precised images.

3.1 Advantages of proposed system:

- It's reliable and power efficient.
- System is portable and required less processing time.

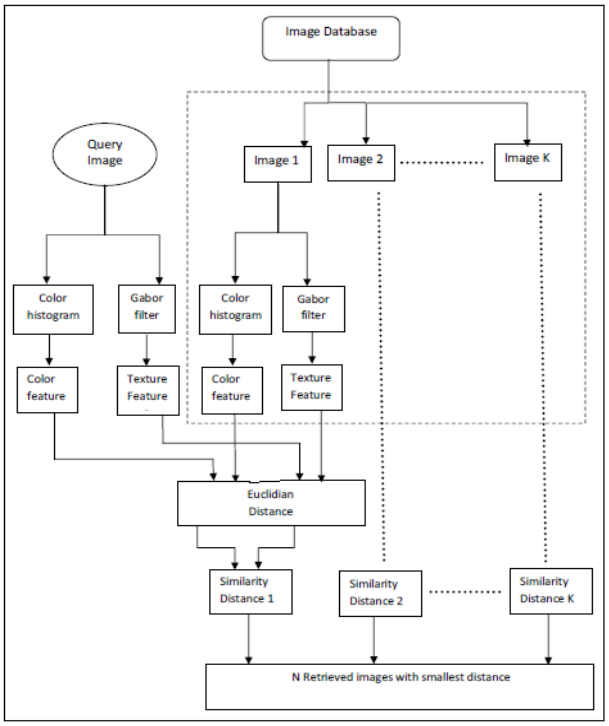

Fig 2: Implementation of CBIR system

4. SOFTWARE USED

4.1 MATLAB

Millions of engineers and scientists worldwide use MATLAB [11] to analyze and design the systems and products transforming our world. The MATLAB platform is optimized for solving engineering and scientific problems. The matrix-based MATLAB language is the world's most natural way to express computational mathematics. Built-in graphics make it easy to visualize and gain insights from data. A vast library of prebuilt toolboxes lets you get started right away with algorithms essential to your domain.

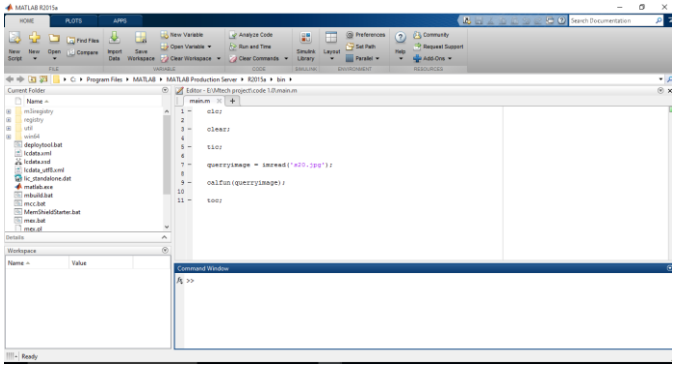

Fig 3: MATLAB software

Key Features of MATLAB are

- High-level language for scientific and engineering computing
- Desktop environment tuned for iterative exploration, design, and problem-solving
- Graphics for visualizing data and tools for creating custom plots
- Apps for curve fitting, data classification, signal analysis, and many other domain-specific tasks
- Add-on toolboxes for a wide range of engineering and scientific applications
- Tools for building applications with custom user interfaces
- Interfaces to C/C++, Java®, .NET, Python®, SQL, Hadoop®, and Microsoft® Excel®
- Royalty-free deployment options for sharing MATLAB programs with end users

Signal Processing Toolbox provides functions and apps to generate, measure, transform, filter, and visualize signals. The toolbox includes algorithms for resampling, smoothing, and synchronizing signals, designing and analyzing filters, estimating power spectra, and measuring peaks, bandwidth, and distortion. The toolbox also includes parametric and linear predictive modeling algorithms. You can use Signal Processing Toolbox to analyze and compare signals in time, frequency, and time-frequency domains, identify patterns and trends, extract features, and develop and validate custom algorithms to gain insight into your data.

5. RESULT AND DISCUSSION

In this Content Based Image Retrieval System end user has to give input or query image with proper extension.

5.1 case1 :

Let's consider first case , input is some brain image and output is as follows-

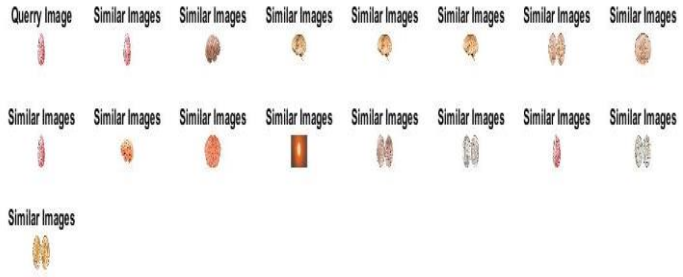

Fig 4: Retrieved images (query image: brain)

5.2 case2 :

Let's consider second case ,input is some sun image and output is as follows-

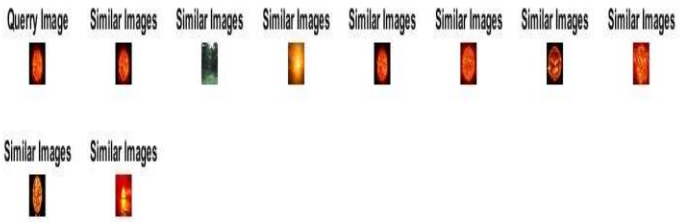

Fig 5: Retrieved images (query image: sun)

5.3 case3 :

Let's consider third case, input is some coin image and output is as follows-

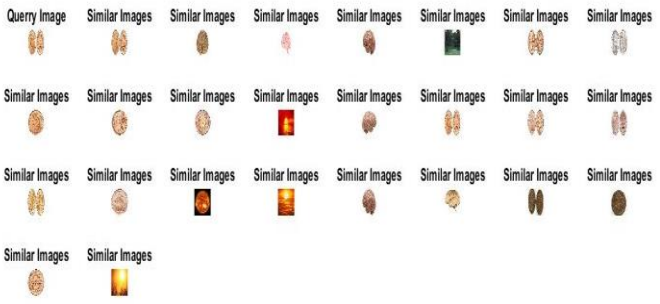

Fig 6: Retrieved images (query image: coin)

6. CONCLUSION

In this work, proposed system will help ongoing research on content based image retrieval. The system would tried to mark the difference between low level and high level approaches to the problem of searching image collections. In this paper, we presented a novel approach for Content Based Image Retrieval by combining the color and texture features called Wavelet-Based Color Histogram Image Retrieval.

ACKNOWLEDGEMENT

Firstly, I would like to thank respected Prof. Mrs. Y. A. Sadawarte mam for giving me such a wonderful opportunity to expand my knowledge and for his huge support. Secondly, I would like to thank my parents who patiently helped me as I went through my work and my friends who helped me to make my work more organized and well-stacked till the end.

REFERENCES

[1] D. Zhang, "Improving Image Retrieval Performance by Using Both Color and Texture Features," In Proc. of IEEE 3rd International Conference on Image and Graphics (ICIG04), Hong Kong, China, Dec.18-20, 2004, pp.172-175.

[2] S.Selvarajah and S.R. Kodituwakku, "Analysis and Comparison of Texture Features for Content Based Image Retrieval", in International Journal of Latest Trends in Computing, Volume 2, Issue 1, March 2011, pp 108 -113.

[3] Ahmed J. Afifi and Wesam M. Ashour, "Image Retrieval Based on Content Using Color Feature", in ISRN Computer Graphics, Volume 2012, pp 1 – 12.

[4] M. Singha, K. Hemachandran A. and Paul, "Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram", in IET Image Processing, vol 6, issue 9, 2012, pp 1221 – 1226.

[5] Ritendra Datta, Jia Li and James Z. Wang, "Content-Based Image Retrieval - Approaches and Trends of the New Age", in Proceedings of the 7th ACM SIGMM International workshop on Multimedia information retrieval, 2005, pp 253 – 262.

[6] S.Mangijao Singh and K. Hemachandran,"Content-Based Image Retrieval using color Moment and Gabor Texture Feature",in International Journal of Computer Science Issues, Vol.9,Issue 5,No 1,September 2012,pp 299-309.

[7]B. Nandhini* , S.Gowri,"A Medical Image Retrieval System using Gabor Barcode", in Journal of Chemical and Pharmaceutical Sciences, ISSN: 0974-2115.

[8]João Augusto da Silva Júnior, Rodiney Elias Marçal Marcos Aurelio Batesta"Image Retrieval:Importance and Applications"

[9] Santos, Ana Paula de Oliveira - Recuperação de Imagens Mamográficas Baseada em Conteúdo. Trabalho de Conclusão de Curso. Fundação de Ensino Eurípides Soares da Rocha, 2006.

[10] H. Kopka and P. W. Daly, A Guide to L ATEX, 3rd ed. Harlow, England: Addison-Wesley, 1999.