

Introducing New Meta Search Model through Content based Image Retrieval Algorithm

Kanak Giri

M.Tech. Scholar

Computer Science &
Engineering Department,

Rajasthan Technical
University,

Swami Keshvanand Institute of
Technology, Management &
Gramothan,
Ramnagar, Jaipur

Kapil Sharma

M.Tech. Scholar

Computer Science &
Engineering Department,

Rajasthan Technical
University,

Swami Keshvanand Institute of
Technology, Management &
Gramothan,
Ramnagar, Jaipur

Pankaj Dadheech

Sr. Lecturer

Computer Science &
Engineering Department,

Rajasthan Technical
University,

Swami Keshvanand Institute of
Technology, Management &
Gramothan,
Ramnagar, Jaipur

ABSTRACT

This paper presents an approach to model content based meta-search engine which search all the content related images present in the dataset. Searching through the keywords in an image database require a lot of Meta data (keywords about the image) to be stored for each image in a separate database. This does not lead to an effective search mechanism. This mechanism basically works for the relevant output for the input query image. The process based on the image extraction by the implemented functions like “edge histogram”, “image feature extraction i.e. sharpness, smoothness, color etc.

Keywords

Content based image retrieval, IMAGINGMSE, Color Histogram, Edge Histogram, Correlation Matching heuristic, Image Feature Extraction

1. INTRODUCTION

Image based retrieval based on pictorial queries is always a challenging but interesting problem. During the last few years Content based image retrieval (CBIR) has emerge as a powerful tool for efficient retrieval of image visually similar to a query image. Due to availability of good storage space, large number of image have been produced and stored in database or server, and with this large database, everyone wants to search it and make us of image in it. Finally, there is competition between the retrieval of image and researchers are trying to find out accurate and efficient ways of searching images.

The similarity in the context of the content is measured by as many features given below:

- a) **Color** is an important cue for image retrieval. Color not only adds beauty to images but also gives more information, hence used as a powerful tool in content based image retrieval. Color searches do not depend on image size or orientation. It usually involves comparing color histograms for color based search [7].
- b) **Texture** measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image [7].

- c) **Shape** does not refer to the actual shape of an image but to the shape of a particular region that is supposed to search out. Shapes will often be determined first applying segmentation or edge detection to an image.
- d) **Dominant Color** is the term given to those colors in the image which are most frequent and shown as high peaks in the color histogram.
- e) **Correlation Matching heuristic** is the criteria to filter the search results after the combined feature matching score has been obtained.
- f) **Many new** features are being proposed day by day, many of which (like entropy, smoothness, skewness, etc.) we will be incorporating to complete the content based matching task.

2. PROPOSED FEATURES

2.2 Algorithm:

Here we have proposed a wavelet-based approach for content-based image retrieval (see figure: 2). Also histograms are used for color feature extraction for CBIR, and these color and edge histogram features are combined to improve the retrieval efficiency.

2.1 Color Feature

Color not only adds beauty to objects but also more information [1], which is used as powerful tool in content-based image retrieval. In color indexing, given a query image, the goal is to retrieve all the images whose color and texture compositions are similar to those of query image. In color image retrieval there are various methods, but here I will discuss some prominent methods. Typical characterization of color composition is done by color histograms. Color histograms are way to represent the distribution of colors in images where each histogram bin represents a color in a suitable color space. A distance between query image histogram and a data image histogram can be used to define similarity match between the two distributions. They used a coarse comparison of the color histograms of the query and model images in the Distance method they proposed. Most color histograms are very sparse and thus sensitive to noise.

2.3 Color Extraction

Various color spaces such as RGB, HSV, YCbCr, CIE LAB, CIE LUV, etc. can be used in CBIR systems. However, no color space is dominant in all applications. The RGB color space is used as input image color space. It has three components R (Red), G (Green) and B (Blue). The reason to select this color space is its real time availability and natural images occur in this color space only.

2.4 Color Spaces

The different color spaces used in MPEG-7 include the familiar monochrome, RGB, HSV, YCrCb, and the new HMMD. The monochrome (intensity only) space is also supported. This corresponds to the component in the YCrCb space. It is possible to define RGB with reference chromaticity primaries, if available from the capture process. The conversion from normalized RGB (where the values of each of the spectral components range from 0 to 1) to the other color spaces are shown in figure 1, distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

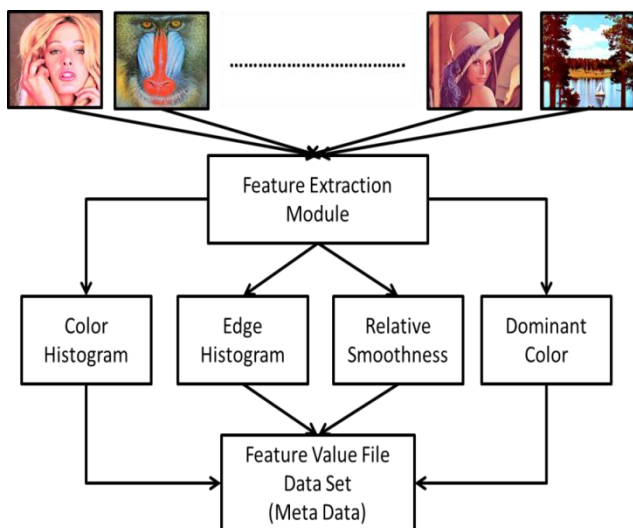


Figure 1: Algorithm flowchart for feature extraction

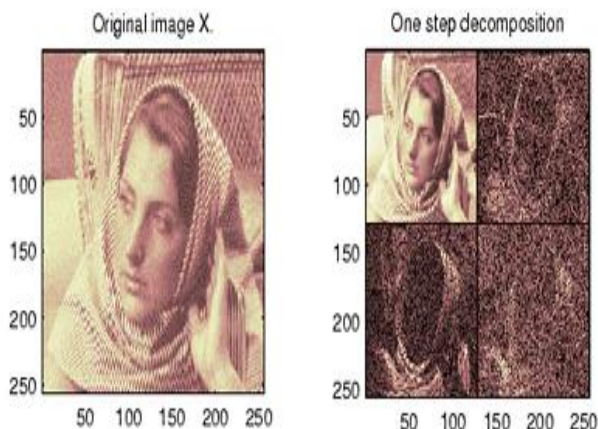


Figure 2: First level decomposition by DWT

2.5 Edge Histogram Descriptor

The edge histogram descriptor captures the spatial distribution of edges. The distribution of edges is a good texture signature that is useful for image to image matching even when the underlying texture is not homogeneous. The computation of this descriptor is fairly straightforward. See figure 3.

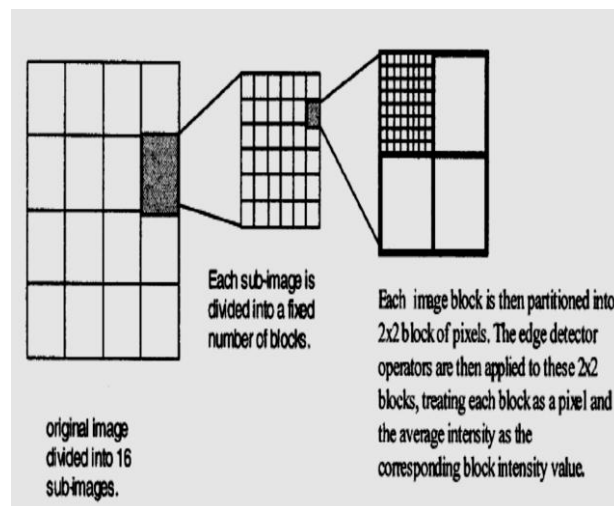


Figure 3: Computing the Edge Histogram descriptor

3. PERFORMANCE ANALYSIS

3.1 Simulation Tool

The algorithms discussed in previous chapters have been implemented on MATLAB v7.4 using image processing toolbox. It is a powerful tool for image processing since it gives the exact picture of the results we suppose to get. Monitoring of the variables and debugging of the programs is very easy. The programs are modular in nature so writing complex simulations can be done very easily. All the features are calculated using MATLAB function (*.m) files. Various results in this thesis have been shown using the Mat-lab figures generated as a result of algorithm application. The feature dataset has been created which is stored in the MATLAB data files. These files store the variables in the form of data structures. We have used various inbuilt functions of the Mat-lab to read the image find the features and storing them in the Mat-lab data files. Inbuilt functions to calculate DWT, histogram have been used.

3.2 Image Query

The search in the text based search engines is done on the basis of text data and description available for the images. So an image without complete description of what it contains cannot appear in the search results. Most text searches are efficient in this way but when it comes to the image search it is not sure that the image which is not fully described will appear in the result or not. To ensure its appearance in the result it is better to search images not on the basis of the description text but the features of the image itself. So we propose to search the image on the basis of the features extracted from the image and compare it with the already stored features of available images as meta-data. We are simulating the whole process. First of all features of all the available images in the dataset are calculated. As soon as new image is added to the dataset its features are calculated and stored as Matlab data file shown in figure 4(a). Now query

image can be browsed or given as text image. This image may or may not be same to the available. Figure 4(b) shows result.

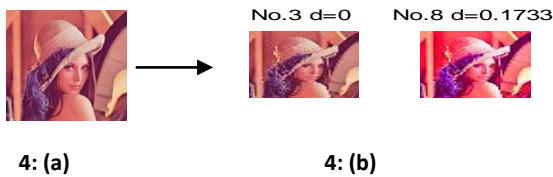


Figure 4: (a) Query image

(b) Result only for Combined Features for Query Image

4. CONCLUSION AND FUTURE WORK

This thesis presents a novel approach to search the images from a data set based on their contents. The strategy includes the application dependent weighted combination of various image content features like color, edges, relative smoothness and dominant colors in the histogram. The strategies are summarized in the flowcharts given in proposed feature chapter. The simulation results depicts that the search results only include the relevant images and all the irrelevant images but very close in terms of features are filtered from the results using the correlation similarity heuristic. Overall the strategy has high ratio of successful search compare to all other strategies since no irrelevant results are in the search results.

The future work for this scheme is to map this scheme for medical image database search. This field is of high importance and medical image search can easily be mapped to our approach easily. So we will look upon it. There are some challenges that we have encountered are the enhanced images give the more accurate results, the meta-data requires some storage and indexing methods for fast search for minimizing storage the feature meta-data can be stored in the image itself but then indexes have to be prepared on the meta-data just once the features have been calculated.

5. REFERENCES

- [1] John Eakins, Margaret Graham, "Content-based Image Retrieval", University of Northumbria at Newcastle, Technical Report, 1999. [66] Pentland A et al (1996) "Photobook: tools for content-based manipulation of image databases" International Journal of Computer Vision 18(3), 233-254.
- [2] Using relevance feedback with short-term memory for content-based spine X-ray image retrieval XiaoqianXu, Dah-JyeLee, SameerK.Antani, L.RodneyLong, JamesK.Archibald 2010. pp.1-4.
- [3] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital Image Processing, Pearson Prentice Hall, 2004.
- [4] LeninaBirgale, ManeshKokare, DharmPalDoye, "Colour and Texture Features for Content Based Image Retrieval," International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06), pp. 1-4, 2006.
- [5] MouradOussalah "Content Based Image Retrieval: Review of State of Art and Future Directions," Image Processing Theory, Tools & Applications, pp. 1-10, 2008.
- [6] Nguyen HuuQuynh, Ngo Quoc Tao, Ngo Truong Giang "An efficient method for content based image retrieval using histogram graph," 10th Intl. Conf. on Control, Automation, Robotics and Vision, pp. 874-878, 2008.
- [7] RitendraDatta, Dhiraj Joshi, Jia Li and James Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, 2007.
- [8] Dr.N.Krishnan, M.SheerinBanu, C.CallinsChristiyana, "Content Based Image Retrieval using Dominant Color Identification Based on Foreground Objects", International Conference on Computational Intelligence and Multimedia Applications, pp. 190-194, 2007.