ISSN:0975-9646

A.Ramesh Kumar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 242 - 245

Content Based Image Retrieval Using Color Histogram

A.Ramesh Kumar, D.Saravanan

Sathyabama University,
Chennai ,Tamil Nadu, India

Abstract - Content-based image retrieval (CBIR) scheme searches the most-similar images of a query image that involves in comparing the feature vectors of all the images in the database with that of the query image using some pre-selected similarity measure, and then sorting of the results. On querying an image, a reduced set of candidate images which have the same Grid Code as that of the query image is obtained. The color histogram for an image is constructed by quantizing the colors within the image and counting the number of pixels of each color. The feature vector of an image can be derived from the histograms of its color components and finally can set the number of bins in the color histogram to obtain the feature vector of desired size. Thus the grid code of an image is obtained through the quantization of the feature vector derived from the histogram of the desired color component of the image. In order to have similar features of the images the grid code must be same for all Images in the grid. Experimental result show confirms that the proposed method is most effective.

Key word-: CBIR, Color Histogram, Data Mining, Query Image, Grid Code.

I INTRODUCTION

The main goal of this project is to retrieve the most similar image from the data base images by grid-based indexing. The main scope of the project is retrieving the image with good accuracy. In content-based image retrieval (CBIR) searching for *k* most-similar images to a query image involves comparing the feature vectors of all the images in the database with that of the query image Using some pre-selected similarity measure, and then sorting the results. However, this requires Linear time with respect to the size of the database and quickly becomes impractical for large databases. Similar to text retrieval, indexing can be performed in vector space to improve retrieval speed. Different algorithms have been proposed such as k-d tree, SS-tree and R*-tree. However, all of these algorithms have running times which increase exponentially with dimension *d* of the vector. In this paper, we propose a two-stage CBIR approach to reduce the computation effort and make the retrieval fast. In our approach, the image is first converted into the desired color space. Then the feature vector derived from the color histograms of the image is used to generate a grid code via vector quantization, so as to group the database images into grids, each of which contains the same grid code and visually similar patterns. It is called grid-based indexing (GBI) in the paper. In the coarse classification stage, a reduced set of candidates with the grid codes adjacent to that of the query image can be obtained through GBI. In the fine matching stage, only the remaining candidates with similar attributes are examined for similarity comparison. Both the search space and the dimensionality of the feature space are thus reduced. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging.

Closely related to image processing are computer graphics and computer vision. In computer graphics, images are manually made from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from natural scenes, as in most animated movies. Computer vision, on the other hand, is often considered high-level image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans).

In modern sciences and technologies, images also gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data). Examples include microarray data in genetic research, or real-time multi-asset portfolio trading in finance.

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems.

II EXISTING SYSTEM

The image query is given by keywords. The results will be of query based since the result retrieved will be of more irrelevant. Finding of most relevant information makes the system harder. The images matched to that are information is collected. Matching with pixel by pixel is a slow process. Accuracy is low and searching time will be higher.

Many CBIR system have been developed: QBIC [5], Photo book [6], MARS [7], NeTra [8] PicHunter [9], Blovworld [10], and others In CBIR system, low-level visual image features (e.g., color, texture and Shape) in a typical CBIR system, low-level visual image features (e.g., color, texture, and shape) are automatically extracted for image descriptions and indexing purposes. To search for desirable images, a user presents an image as an example of similarity, and the system returns a set of similar images based on the extracted features.

Content-Based Image Retrieval (CBIR), also known as Query by Image Content (QBIC) and Content-Based Visual Information Retrieval (CBVIR) is the application of image retrieval problem, searching for digital images in large database. “Content-based" means that the search will analyze the actual contents of the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. Retrieving images based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values that humans express as colors.

III PROPOSED SYSTEM

The proposed system decides which images in the image database is the most similar one with the query image. To speed up the retrieval, use the proposed grid-based indexing to obtain the nearest neighbors of the query image, exact images are retrieved. The images related to that information is also collected. Indexing can be performed in vector space to improve retrieval speed.

IV HISTOGRAM GENERATION

Histogram search algorithms , characterize an image by its color distribution or histogram. A histogram is nothing but a graph that represents all the colors and the level of their occurrence in an image irrespective of the type of the image. Few basic properties about an image can be obtained from using a Histogram. It can be used to set a threshold for screening the images. The shape and the concentration of the colors in the histogram will be the same for similar objects even though they are of different colors. Identifying objects in a grey scale image is the easiest one as the histogram is almost similar as the objects have the same colors for same objects. In order for identifying the objects in the images or generating the histogram the system has to obtain the array values.

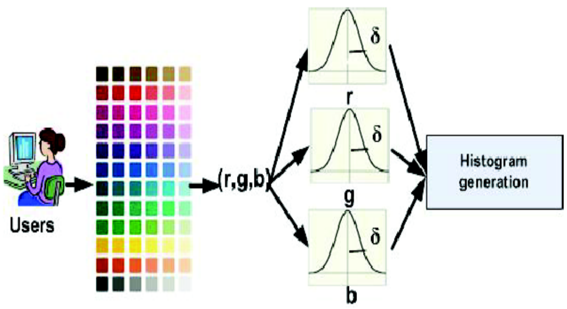
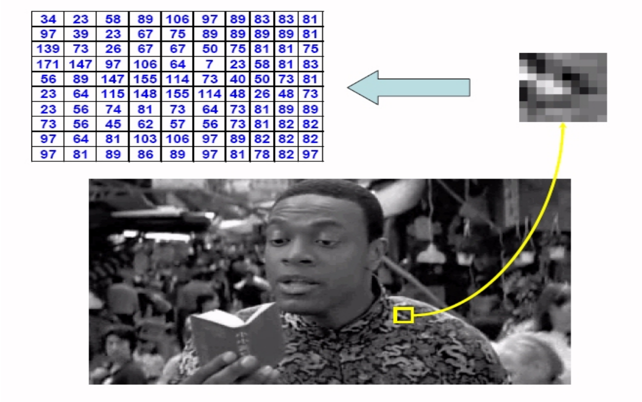


Fig 1. Color Histograms

In general any image contains useful and unwanted information. The system has to differentiate between the both. Consider the below image where the person reading a book is the useful information and the background, people and the market is the unwanted data. The system has to group together the repeated pattern to identify the objects in the image. For example below is given the array for the part of the shirt and this pattern is repeated again



34	23	58	89	106	97	89	83	83	81
97	39	23	67	75	89	89	89	89	81
139	73	26	67	67	50	75	81	81	75
171	147	97	106	64	7	23	58	81	83
56	89	147	155	114	73	40	50	73	81
23	64	115	148	155	114	48	26	48	73
23	56	74	81	73	64	73	81	89	89
73	56	45	62	57	56	73	81	82	82
97	64	81	103	106	97	89	82	82	82
97	81	89	86	89	97	81	78	82	97

Fig 2. Histograms Generation

Consider the above image where the small part of the person’s) shirt is enlarged and the respective representation in the form of the array is given.

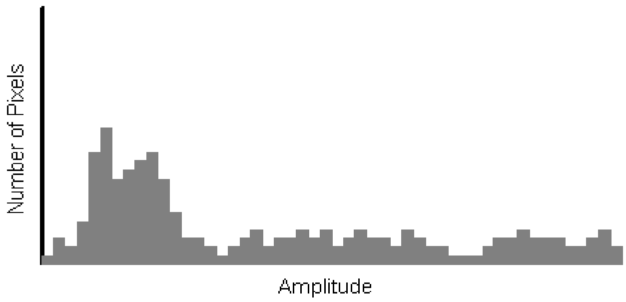


Fig 3 .An example of image histogram

V EXPERIMENTALS ALGORITHM

The Algorithm for CBIR using Principal Component Analysis the steps for CBIR using Principal Component Analysis are as follows:

A .Representation of Each Image

Obtain cancer images A_1, A_2…………A_M and represent every image as a one dimensional vector. Every image in the must be of the same size (NxN).

B. Computation of the Mean Cancer Image

The cancer images must be mean centered. Mean value of the pixels intensities in each image is calculated and subtracted from the corresponding image. The process is continued for all images in the database.

C. Normalization Process of Each Image in the Data Base.

Normalization is done by subtracting the mean image from the training cancers.

CONCENTRATION OF NORMALIZED IMAGES:

From the matrix A, which is the concatenated form of all normalized images and is of the order N^2 x M [A_1 A_2 ……..A_m]

D. Calculation of Covariance

The covariance matrix which is of the order N^2 x N^2 is calculated as given by $C = AA^T$

E. Calculation of Eigen Values and Eigen Vectors

Now, find the Eigen values of the covariance matrix C by solving the equation ($C\lambda$ - I) = 0 To calculate the eigen values $\lambda_1,\lambda_2,\ldots,\lambda_{N^2}$ For specific Eigen values λ_i solve the system of N^2 equations $(C\lambda_i - I) = 0$ To find the eigenvector X repeat the procedure. Where, X_i indicates corresponding Eigen values.

F. Sorting the Eigen Values and Eigen Vectors.

The Eigen vectors are sorted according to the corresponding Eigen values in descending order. The eigenvector associated with the largest Eigen value is one that reflects the largest variance in the image.

G. Choosing the First 'K' Eigen vectors and Eigen Values.

Selecting only 'k' Eigen vectors corresponding to the 'k' largest eigen value. Then, each cancer in the training set can be represented as a linear combination of the first 'k' Eigen vector.

H. Projection of Training Images.

A facial image can be projected onto k dimensions by computing $Y_i = X^T A_i$

I. Forming Feature Space

Feature space is obtained as $y = [y_1, y_2 \ldots, y_m]$, where, M is the no. of images in the database.

J. Procedure for Query Image.

Given an unknown cancer image, represents the image as a vector Q. This image must be of the same size like the training cancers and it must be mean centered. The normalization procedure is subtracting the mean image from the query cancer. $\Phi = Q - m$.

K. Finding the Euclidean Distance.

Euclidean Distance, a nearest neighbor classifier is used for classification. Find the maximum distances between the feature database and projected query image as given by

$\mathbf{d} (\mathbf{y_i}, \mathbf{y_t}) = \sum \| \mathbf{y_j}^{I} - \mathbf{y_j}^{t} \|^2$ Where $\| y_j^{I} - y_j^{t} \|^2$ denotes the Euclidean distance between the training feature matrices and query features matrix.

L. Sorting the Distance and Retrieval of Required Images.

Now the distances are sorted and the images corresponding to the top ten retrieved images are retrieved as the images closer to the query image given.

VI Experimental Setup

A. Image pre-processing Module
B. Feature extraction Module
C. Training of images
D. Retrieval of Image Module

A Image Pre-Processing Module

The image is digitized using three level processes: 1. Low level process 2.Mid-level process, 3. High level process.

B. Feature Set Module

The image is set to database and it is compared with another image using Texture, shape and color format.

C. Training of Images.

After extracting the image features like texture and matrix conversion the pixel values are trained in the database by labeling the features of the images. The matrix conversion us done by giving intensity at each point x, y and RGB values are found. A matrix will be formed having M rows and N columns.

D. Image Retrieval

The two images are matched using the features and reorganizing the objects.

VII Experimental Result

Fig 4Login Page

Fig 5.Training the image

Fig 6. Matrix Values of the image

Fig 7. Input and Retrieve Image

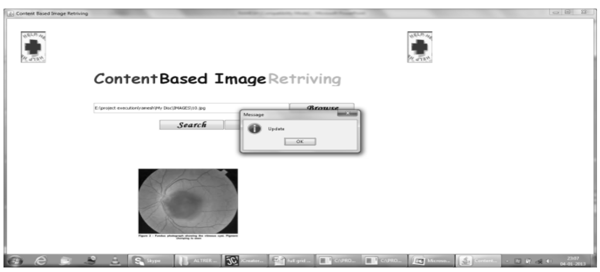


Fig 8. Image Matching

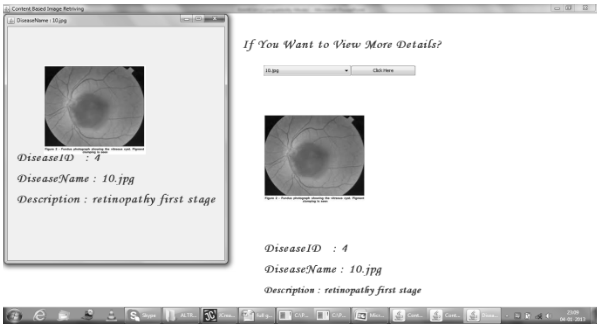


Fig 9. Result of Input Image

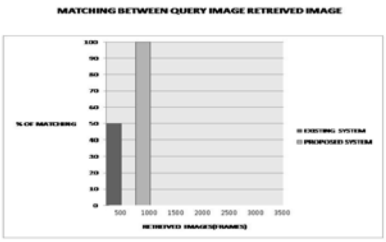


Fig 10. Comparison Grahp Between Query Image and Retreived image Proposed Method.

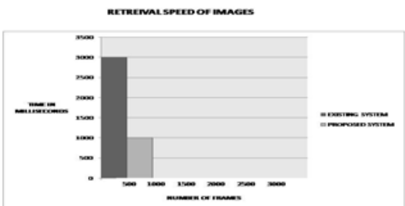


Fig 11. Comparison Graph of Retrieval speed of proposed system

VIII CONCLUSION

In imaging science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. In this paper CBIR using color histograms technique is proposed with help of grid technique to improve the image retrieval performance.

IX FUTURE ENHANCEMENT

In this paper on querying an image, a reduced set of candidate images. The color histogram for an image is constructed by quantizing the colors within the image and counting the number of pixels of each color. The feature vector of an image can be derived from the histograms of its color components and finally can set the number of bins in the color histogram to obtain the feature vector of desired size.

REFERENCES

[1]Bimbo, A. D., *Visual Information Retrieval*, San Francisco: Morgan Kaufmann, 1999.

[2] Castelli, V. and Bergman, L. D., *Image Databases*, New York: John Wiley & Sons, 2002.

[8]D.Saravanan, Dr.S.Srininvasan "Indexing and Accessing Automatic Video Frame by a Statistical Modeling Approach"

[3] Gudivada, V., and Raghavan, V., "Content-Based Image Retrieval Systems", *IEEE Computers*, vol. 28, pp.18-22, 1995.

[4] M. Flickner, H.S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker,"Query by Image and Video Content: The QBIC System," Computer, vol. 28, no. 9, pp. 23-32, Sept. 1995.

[5] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R., "Content-Based Image Retrieval at the End of the Early Years", *IEEE Trans. on Pattern Analysis and Mahine Intelligence*, vol. 22, pp.1349-1380, 2000.

[6] A. Pentland, R.W. Picard, and S. Sclaroff, "Photobook: Content- Based Manipulation for Image Databases," Int'l J. Computer Vision, vol. 18, no. 3, pp. 233-254, 1996.

[7] D.Saravanan, Dr.S.Srinivasan "Video Image Retrieval Using Data Mining Techniques "JCA,Volume V, Issue 1, 2012.

[8] D.Saravanan, Dr.S.Srinivasan,"Data Mining Framework for video Data ",RSTCC 2010,Pages 167-17,Nov 2010.

[9] D.Saravanan, Dr.S.Srinivasan "Video Image Retrieval Using Data Mining Techniques "JCA,Volume V, Issue 1, 2012.

[10] D.Saravanan, Dr.S.Srinivasan,"Data Mining Framework for video Data ",RSTCC 2010,Pages 167-170,Nov 2010.

[11] M. Ortega-Binderberger and S. Mehrotra, "Relevance Feedback Techniques in the MARS Image Retrieval Systems," Multimedia Systems, vol. 9, no. 6, pp. 535-547, 204.

[12] W.Y. Ma and B. Manjunath, "Netra: A Toolbox for Navigating Large Image Databases," Proc. IEEE Int'l Conf. Image Processing (ICIP '97), pp. 568-571, 1997.

[13] I.J. Cox, M.L. Miller, T.P. Minka, T.V. Papathomas, and P.N. Yianilos, "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments," I EEE Trans. Image Processing, vol. 9, no. 1, pp. 20-37, 2000.

[14]C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pp. 1026-1038, Aug. 2002

[15] Wang, J. Z., Content Based Image Search Demo Page. http://bergman.stanford.edu/~zwang/project/imsearch/ WBIIS.html, 1996.

[16] Nadler, M. and Smith, E. P., *Pattern Recognition Engineering*, New York: Wesley Interscience, 1993.

[17] Gray, R. M. and Neuhoff , D. L., "Quantization," *IEEE Trans. on Information Theory*, vol. 44, pp.2325-2383,1998.