ISSN:0975-9646

Minimization of Image Retrieval Time using Clustering

Mohammad Naved – Ul – Haque
M.E. Scholar, PRMIT&R,
Badnera, India.

Prof. Dr. S. S. Sikchi
Asst. Prof. PRMIT&R,
Badnera, India.

Abstract—Image Mining is a apecial technique used to study different aspects in detail of an image. In case of image mining there are certain algorithms with which different details of an image can be studied. Some of them are k-means, c-means, fuzzy approaches. Out of which we have adopted the technique of c-means clustering to achieve our target. In this algorithm we generally make different clusters on the basis of similarity and differences of images and form an image set. This approach depends on the membership function which clarifies the similarity and dissimilarity between the images. It depends on the mathematical value that specifies the membership function. Values of membership function is either near zero or one. Zero means similarity between images is very less and one means images are highly similar.

Index Terms— *Keywords:* String Patterns, image mining, color similarity, color space.

1. INTRODUCTION

Image mining represents some special characters because of richness of data in an image. Effective evaluation of the results of image mining techniques by its content requires different aspects that the user point of view is used on the parameter of performance. Comparison is to be done between different mining by similar systems, possess various challenges owing to the great variety of methods implemented to represent similarities and the dependence that the results represent of the used image set. Other major obstacle is the lagging of parameters for comparing experimental performance. In this paper we propose an evaluation framework for comparing the influence of the distance function on image mining by color and also a way to mine an image with its name. Experiments with color similarity mining by quantization on color space and measures of likeness between a sample and the image results have been carried out to illustrate the proposed scheme. Important aspects of this type of mining are also described. A more efficient approach is gathered when image example is given by the user on input to the mining process. Automatically generate matching is required then for an efficient image mining. The basic idea is to extract characteristic features similar to that of object recognition schemes. After matching, images are ordered with respect to the query image according to their similarity measures and displayed for viewing. in this work, we present a framework for considering the influence of this distance function on color mining. The framework accesses a system's quality from the point view of users; it provides a basic set of attributed to characterize the ultimate utility of systems. Then we analyze example of mining by color and present some conclusions.

2. LITERATURE REVIEW

Earlier techniques which were used were not generally based on visual features but on the textual annotation of images. It means images were first annotated with text and then search using text based approach from traditional database management system. Image retrieval system based on text uses traditional database techniques to manage images. Images can be organized through text description, by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. However since automatically generating descriptive text for wide spectrum of images is not fisible, most of the text based image retrieval system requires manual annotation of an image. Obviously, annotating images manually increases complexity and an expensive task for large image databases, and is often subjective, context sensitive and incomplete. As a result it becomesextremely difficult for traditional text based method to support variety of task dependant queries.

3. MINING METHODS

3.1 COLOR BASED IMAGE MINING:

Color based image mining in visual database is extremely different from standard alphanumeric image mining. Current approaches, features vectors per image are computed for evaluation distance function on the feature space. Then this function is used to retrieve images from a given image set. Images with distance less than a predefined threshold or within a predefined number are retrieved. These feature vectors facilitate and help color based mining, based on texture, geometric properties, shape, volume, spatial constraints, etc.

In this system we have proposed, that we are having two approaches for images retrieval that is contents color based mining and contents text based mining. Images are mostly RGB, Gray Scale and Binary types which are easily possible to segments as per their pixels values.RGB images are segmented as per the color content, Gray Scale images are segmented as per their average RGB pixels value and Binary Images as per their ON/OFF pixels status. Image mining by color contents include search for images based on their segmented color contents. Let see Dataflow Diagram for proposed image mining technique based on their color contents is shown in figure 1.

Segmentation of images based on their color contents is a major area of research. In a proposed method we are using C-means clustering technique to clusterize an image based on their RGB colors as per equation.

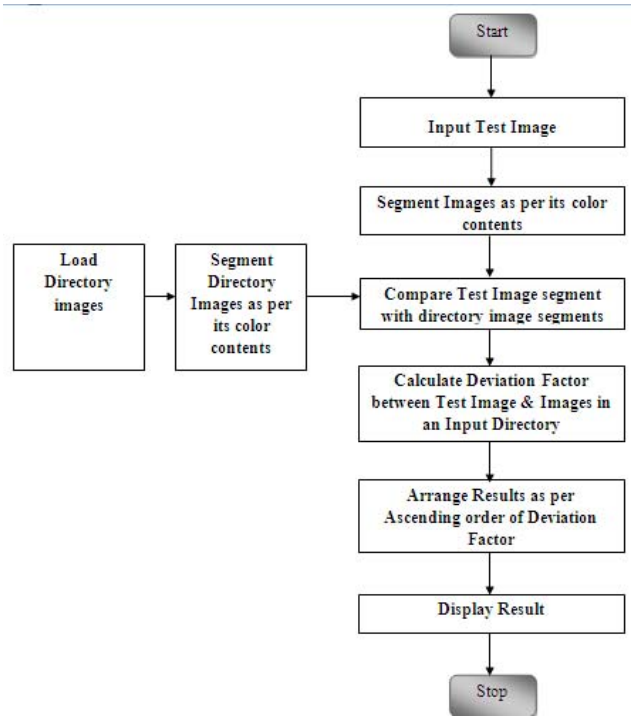

Diagram 3.1 (DFD for Mining by Color)

$$\text{If } |P_i - P_{i+1}| = 0 \quad \Big\{ \quad P_{i+1} = P_i$$

$$\text{Else}$$

$$P_i - P_{i+1} \neq 0 \quad \Big\} \quad P_i = 0 \ldots\ldots \text{ Eq}^{n.}\ 3.1.1$$

Where,
Cn= C means cluster.
n= No of Clusters.
Pi= Backward pixel
Pi+1= Forward Pixel
M= No of Pixels in an image.

In c-means clustering technique the pixels are scanned as per Raster Scan method from left to right and top to bottom way. Whenever difference occurs with backward and forward pixel we set it as OFF else we proceeds our scanning as per the equation

Deviation Factor Df is a difference between number of segments of directory image and input test image. Deviation factor is a measure of similarity and difference between two images which can be represented with an equation :

$$Df = |Ctr - Cts| \ldots\ldots \text{Eq.}^{n.}\ 3.1.2$$

Where,
Ctr = Directory Image Segments.
Cts = Input Test Image Segments.

An overall searching time for color based image mining is depended on number of color segments in an image depended on number of color segments in an image.

$$O(tc) \alpha\ n \ldots\ldots\ldots \text{Eq}^{n.}\ 3.1.3$$

Where O (tc) = searching time for color based image mining

We use best fit searching technique to find out an image segment express with

$$O(C)\ \alpha\ O(n) \ldots\ldots\ldots \text{Eq}^{n.}\ 3.1.4$$

Where O(C) = no of segments comparisons.

3.2 MINING BY TEXT :

From the literature survey, it is observed that, image mining by text is a complicated process and fails in a domain where single image contains multiple images content (like single image with nature, man, aero plane etc.).To have accuracy in image mining process through text is quite impossible. We proposed a novel method that uses any one of the input name image as a reference image for search and rest of the searching mechanism as per section 3.1.Let see a data flow diagram for image mining based on the text as follows

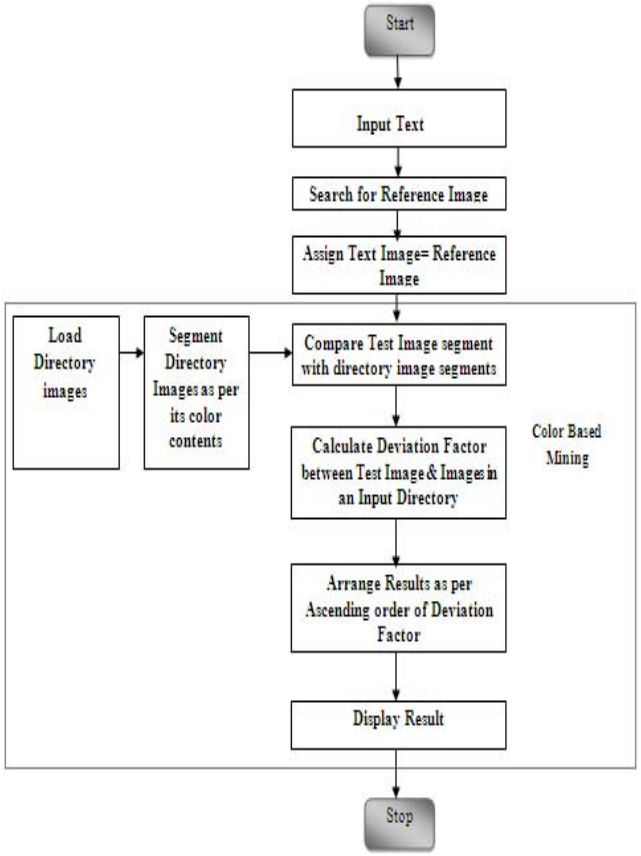

Diagram 3.2 (DFD for Mining by Text)

REFERENCES :

[1] M. J. Swain, D. H. Ballard, "Color Indexing", Int. J.Comp.Vision, Vol. 7 , No. 1, 11-32, 1991.

[2] P. K. Kaiser, R. M. Boyton, Human Color Vision, Second Ed., Washington, D.C.: Optical Society of America, 1996.

[3] J. R. Bach, C. Fuller, A Gupta, A Hampapur, B. Horowits, R. Humphrey, R. C. Jain, C. Shu, "Virage image search engine: an open framework for image management", Symposium on Electronic Imaging: science and technology-storage & retrieval for image and video databases IV, IS&T/SPIE, 76-87, 1996. - http://www.virage.com

[4] A. Pentland, R. W. Picard, S. Sclaroff, "Photobook: content-based manipulation of databases", Int. J. Computer Vision, Vol. 18 , No. 3, 233-254, 1996. http://www-white.media.mit.edu/~tpmink/photobook

[5] V. N. Gudivada, V. V. Raghavan, "Content-Based Image Retrieval Systems", IEEE Computer,September, 18-22, 1995

[6] M. Flickner, H. Sawhner, W. Niblack, J. Ashley, Q.Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petrovic, D. Steele, P. Yanker, "Query by image video content: the QBIC system", IEEE Computer, September, 23-32, 1995.http://wwwqbic.almaden.ibm.com/~qbic/

[7] S. A. Stricker, "Bounds for discrimination power of color indexing techniques", Proc. SPIE, pp. 15-24, 1994.

[8] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner and W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions", IEEE Trans. Pattern Analysis Machine Intell., Vol. 17, No.7, pp. 729-736, 1995.

[9] B. Hill, Th. Roger, F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula",ACM Transaction on Graphics, Vol. 16, No. 2, April, 109-154, 1997.

[10] H. Zhang, Y. Gong, C. Y. Low, S. W. Smoliar, "Image retrieval based on color features: an evaluation study: Proc. of SPIE, 2606, pp. 176-187, 1997.

[11] C. Z. Ren, R. W. Means, "Context Vector Approach To Image Retrieval", Proc. IEEE ICIP, Vol. 1, No 407, 1997.