Image Mining Using Image Feature

Varsha Kundlikar[1], Meghana Nagori[2].

Abstract— In this paper, features of image used for mining images from database. Basic features of images such as colour of pixel in image, texture of image, shapes in images and edges of shapes present in image are extracted from image and stored in database to match with query for mining images. This is implemented by using association rule of data mining with specific class.

***Keywords*—image mining, Pre-processing, feature extraction, texture and association rule.**

I.INTRODUCTION

There are many applications where images play an important role for discovering information from it. Image contains information such as RGB colour, texture of image for background and different edges. By extracting these important features, image mining is possible for required purpose from large image database. Many of existing databases allow formulating queries by submitting appropriate input. Then system identifies those stored images whose feature values math and displayed output images. Colour feature are represented as histogram intensity of the colour pixels. Low level feature called edge detection is also considered to extract knowledge from images. By detecting edges, different shapes may form. So geometric shapes may form that can be stored as extracted pattern for mining images. Texture features are extracted by using Gabor filter.

Image mining is a process of mining knowledge, pattern, information and required data from large databases of images. Image mining uses some operations of image processing such as image acquisition, image pre-processing, feature extraction and stored feature matching. Result of this system is images are mined according to colour, edge/shape and textures as well as by combining all feature.

II. IMAGE CHARACTERISTICS/ FEATURES USED FOR MINING

There are main three features of image that are used to extract and store into database for matching with query. These are color, shape and texture.

1. Color Feature:

A computer image is a matrix of pixels. The value of each pixel is proportional to the brightness of corresponding point in the scene. Color images are represented by three intensity component. These components generally correspond to red, green and blue (RGB).Image can associate an integer value with each pixel that can be used as an index to a table that stores the intensity of each color component. The histogram plots the number of pixels with a particular brightness level against the brightness level. For 8-bit pixels, the brightness ranges from zero (black) to 255 (white). The operations on color characteristics on are histogram normalization, histogram equalization and thre-sholding.

ISSN: 2278 – 1323

2. Shape or edge feature.

Edge is simply a large change in frequency. Many approaches to image interpretation are based on edges, since analysis based on edge detection is insensitive to change in the overall illumination level. Edge detection highlights image *contrast*. Detecting contrast, which is difference in intensity, can emphasize the boundaries of features within an image.

3. Texture feature.

Texture is defined as a neighborhood feature as a region or a block. The variation of each pixel with respect to its neighboring pixels defines texture. The Quasi-Gabor filter is explored to present the image texture features.

IMAGE MINING MECHANISM

In this paper image mining is used for mining images that related to given input values from stored images in database. Image mining is an application of image processing and data mining. Feature extraction is step of image processing. Data mining automates the detection of relevant patterns in a database, using defined approaches and algorithms to look into current and historical data that can then be analyzed to predict future trends.

Following general procedure is used for image mining. The four major image mining steps are as follows:

1. Feature extraction. Segment images into regions identifiable by region descriptors (blobs). Ideally one blob represents one object. This step is also called segmentation.
2. Object identification and record creation. Compare objects in one image to objects in every other image. Label each object with an id. We call this step the per-proccessing algorithm.
3. Create auxiliary images. Generate images with identified objects to interpret the association rules obtained from the following step (html page creation).
4. Apply data mining algorithm to produce object association rules.

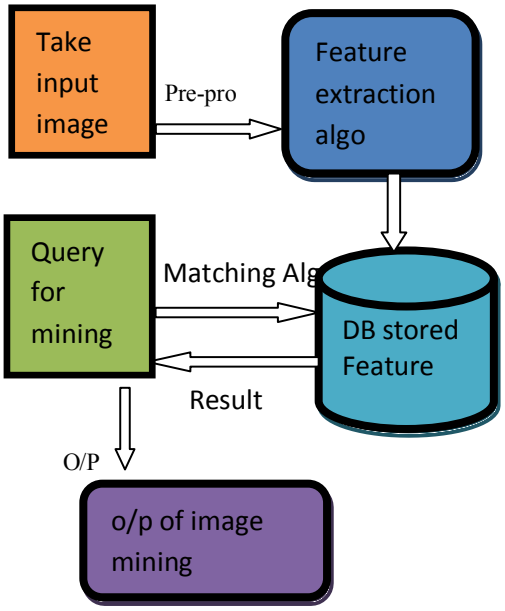

Fig 1. block diagram

General block diagram for image mining is shown above in fig1.

III. ASSOCIATION RULE

An *association rule* is an implication of the form X⇨Y, where X ,Y ∈ I, and I is item set. For a given image database, construct a database with records containing the following structure:

ISSN: 2278 – 1323

(imageID, C_1, C_2, …, C_n, T_1, T_2, …, T_m, S_1, S_2, …, S_k, F_1, F_2, …, F_l), where imageID is a unique identification of the image; C_1, C_2, …, C_n, are the values of the color characteristics; T_1, T_2, …, T_m, are the values of texture characteristics; S_1, S_2, …, S_k, are the values of shape characteristics; F_1, F_2, …, F_l are the high level semantic features, given by an expert in the field. The mining process is divided into two steps. First we find the frequent multidimensional value combinations and find the corresponding frequent features in the database. The combination of attribute values that occurs two or more times are called multidimensional pattern. The second step includes mining the frequent features for each multidimensional pattern. They constitute the obtain rule base set for the high level semantic features.

Multidimensional association-rules mining

For a given image database we construct a database with records containing the following structure: (imageID, C_1, C_2, …, C_n, T_1, T_2, …, T_m, S_1, S_2, …, S_k, F_1, F_2, …, F_l), where imageID is a unique identification of the image; C_1, C_2, …, C_n, are the values of the color characteristics; T_1, T_2, …, T_m, are the values of texture characteristics; S_1, S_2, …, S_k, are the values of shape characteristics; F_1, F_2, …, F_l are the high level semantic features, given by an expert in the field. First we find the frequent multidimensional value combinations and find the corresponding frequent features in the database. The combination of attribute values that occurs two or more times are called multidimensional pattern. For mining such pattern a modified BUC algorithm is used. The second step includes mining the frequent features for each multidimensional pattern. They constitute the obtain rule base set for the high level semantic features.

IV. IMPLEMENTATION PROCESS

Color Feature Extraction:

Some of the techniques tried were – Average color in Gray scale, Average color in RGB format and Average color in YCBCR (Y is the luminance and CB, CR are the chrominance components). We evaluated the various methods using Precision and Recall (introduced in the next section which compares the Precision and Recall values of the methods), and found that YCBCR performs better than the other two. Hence we used it as the basis of color extraction.

Average color $=\Sigma$(intensity of all pixels in the current block)/(total pixels in the block)

The output of this procedure would be a region matrix, of *30X30* (for 10X10 block or 37X37 for 8X8) size, with '1' in the areas corresponding to the presence of color match and '0' in the areas without color match.

Texture Features Extraction:

In this use of histogram without bins, with bins , Normalized Histogram with bins and Discrete

ISSN: 2278 – 1323

Cosine Transform is included. In the Histogram methods, the template is compared with each image block in terms of its histograms. The difference between the individual peaks is taken and the mean squared difference is determined. Each block with relatively smaller difference matches the template and hence can be extracted as part of that texture. If the difference between the template and the current block is smaller than a particular threshold, then that feature is marked YES for that particular block. As in the case of color, we also evaluated the performance of these methods and found that for the training image the Histogram with bins method was the most accurate.

Similarity measure =

$$\sqrt{\Sigma[}$$ (Means of bins for template) - (Means of bins for block)$$]$$

Edge Feature extraction:

Edge features are particularly important for some of the darker images. Fortunately, the training image was of normal quality and hence i did not use the edge feature. However, it does use it for some of the darker images in the set for testing. The Canny edge detection method with default threshold (0) was used. Edge feature alone has very little efficiency; hence it needs to combine it with a stronger feature, like color. It is combined with the color feature to describe the boundaries and inner regions of image as shown in fig. 2

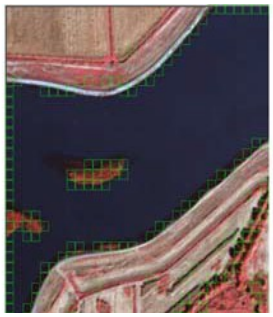

Fig2 Edge feature extraction

V. CONCLUSION

In this paper author tries to mine images using traditional features of image that are used for extraction and matching. Image mining is an emerging research for different applications such as web image mining. Using association rule, image mining is easier as it looks for similarity and difference.

VI. ACKNOWLEGMENT

I would like to express my sincere thanks to my guide Prof. Meghna Nagori for her valuable guidance and cooperation. This

ISSN: 2278 – 1323

work has not been completed without cooperation our head of department and my friends.

VII. REFERENCES

*[GW92] R. Gonzalez and R. Woods. Digital Image Processing, Addison-Wesley publications Co,*March 1992.

[JV96] A. Jain, A. Vailaya. *Image Retrieval using Color and Shape*. Pattern Recognition, 29(8): 1233-1244, August 1996.

[MAT02] MATLAB Image Processing Toolbox User's guide, Version 3, 1993-2001 by The Mathworks Inc., www.mathworks.com

[MDM97] W. Ma, Y. Deng and B. S. Manjunath. *Tools for texture/color based search of images.*

SPIE International conference - Human Vision and Electronic Imaging: 496-507, February 1997.

[NB94] W. Niblack, R. Barber, and et al. *The QBIC project: Querying images by content using color, texture and shape*. In Proc. SPIE Storage and Retrieval for Image and Video Databases, Feb 1994

[RHC99] Y. Rui, T. Huang and S. Chang. *Image retrieval: current techniques, promising directions and open issues.* Journal of Visual Communication and Image Representation, 10(4): 39-62, April 1999.

[RTC02] Romberg Tiburon Center for Environmental Studies. http://romberg.sfsu.edu/~egeria.

[PPS96] A. Pentland, R.W. Picard, and S. Slaroff. Photobook: *Content based manipulation of databases*. Int. J. Comput. Vis., 18(3): 233-254, 1996

[SC96] J. Smith and S. Chang. *VisualSEEK: A fully automated content-based image query system.* ACM Multimedia: 87-98, November 1996.

[WF00] I. Witten and E. Frank. *Data Mining: Practical Machine Learning tools and techniques with Java Implementations*, Morgan Kaufmann publications, 2000.

[ZHL01] Ji Zhang, Wynne Hsu, Mong Li Lee. *An Information-driven Framework for Image Mining*, in Proceedings of 12th International Conference on Database and Expert Systems Applications (DEXA), Munich, Germany, 2001.

[ZHL001] Ji Zhang, Wynne Hsu, Mong Li Lee. *Image Mining: Issues, Frameworks and Techniques*, in Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), San Francisco, CA, USA, 2001. *Peter Stanchev,*Kettering University Flint, Michigan, USA pstanche@kettering.edu www.kettering.edu/~pstanche

ISSN: 2278 – 1323

Ms. Meghana Nagori passed her bachelor degree in Computer Science and Engineering from Government Engineering College of Aurangabad of BAMU University Aurangabad. She is completing her research work on computer science for doctorate degree. Currently she is working as Associate professor in an autonomous Government Engineering College of Aurangabad.

Ms. Varsha Kundlikar passed her B.E in information technology from Government Engineering College of Aurangabad of BAMU University Aurangabad in 2006. She is working as lecture in Government Polytechnic Aurangabad and appeared to complete her Master degree in Computer Science.

ISSN: 2278 – 1323