Image Mining: A New Approach For Data Mining Based On Texture

Miss. Helode Priyanka S. [#1], Prof. Bharat Choudhary [*2]

[#] *Department of Computer Engineering,*
Dr. V. B. Kolte College of Engineering, Malkapur, Amravati University, Maharashtra State, India
[1]priyankahelode52@gmail.com

[*] *Department of Computer Engineering,*
Dr. V. B. Kolte College of Engineering, Malkapur, Amravati University, Maharashtra State, India
[2]cedeptvbkcoe@gmail.com

Abstract— **Image data mining can be done manually by slicing and dicing the data until a pattern becomes obvious. Or, it can be done with programs that analyse the data automatically. Colour, texture and shape of an image have been primitive image descriptors in Content Based Image Retrieval (CBIR) system. Primitive features of an image used to identify and retrieve closely matched images from an image database. It is very difficult to extract images manually from image database because they are very large.**

This paper presents a novel framework for texture information of an image and achieves higher retrieval efficiency than the shape features of an image. There is a trade-off between accuracy and computational cost. The trade-off decreases as more efficient algorithm is used to solve the problem and increases the computational power and will decreases the cost of the whole system as well.

Keywords— visiual content, features extraction, Histrogram, similarity measures, TBIR(Texture Based Image Retrieval), energy, entropy, contrast , image database

I. INTRODUCTION

In the early 1990s, as a result of advances in the Internet and new digital image sensor technologies, the volume of digital images produced by scientific, educational, entertainment, medical, industrial, and other applications available to users increased dramatically. Storage of such image data is relatively straightforward, but an accessing and searching image database is intrinsically harder than their textual counterparts. A major approach directed towards achieving this goal is to use low-level visual features of the image data to segment, index and retrieve relevant images from the image database[1, 2]. The problems of image retrieval are becoming widely recognized, and the search for solutions an increasingly active area for research and development [3]. Digital images are convenient media for describing and storing spatial, temporal and physical component of information contained in a variety of domains [4]. CBIR also known as query by image content (QBIC) and content based visual information retrieval (CBVIR) is the application of computer vision to the image retrieval problem of searching for digital images in the large databases [5]. Content Based Image Retrieval describes the process of retrieving desired images from a large image database on the basis of features (such as color, texture and shape) that can be extracted from the images themselves[1]. It is a combination of different areas of knowledge, such as pattern recognition, object matching, machine learning, wavelet filtering and so on. CBIR is devoted to understanding visual characteristics of images without any text descriptions. It involves two steps [5]:

Feature Extraction: The first step in the process is to extract image features to a distinguishable extent.

Matching: The second step involves matching these features to yield a result that is visually similar.

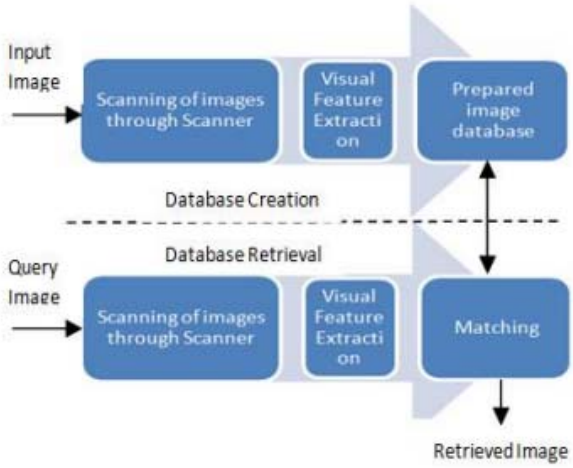


Fig 1: Image Retrieval Model [4]

In typical content based image retrieval systems (Fig. 1), the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. To retrieve images, users provide the retrieval system with example images or sketched figures. The system then changes these examples into its internal representation of feature vectors. The similarities /distances between the feature vectors of the query

example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. The indexing scheme provides an efficient way to search for the image database. Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results.

II. TEXTURE FEATURES

Texture is that instinctive property of all surfaces that defines visual patterns, each having characteristics of homogeneity. It contains essential information about the structural collection of the surface, such as; bricks, fabric, flowers, floor etc. It also defines the relationship of the surface to the surroundings [6]. It is a feature that describes the unique physical composition of a surface. Texture properties include: Coarseness, Contrast, Directionality, Line-likeness, Regularity, Roughness, Entropy, Energy, Contrast Texture is one of the most important defining features of an image. It is characterized by the spatial distribution of gray levels in a neighbourhood [7]. In order to capture the spatial dependence of gray-level values, which contribute to the perception of texture, a two-dimensional dependence texture analysis matrix is taken into consideration. A variety of techniques has been used for measuring texture similarity; the best-established rely on comparing values of what are known as second-order statistics calculated from query and stored images. Essentially, these calculate the relative brightness of selected pairs of pixels from each image. From these it is possible to calculate measures of image texture such as the degree of contrast, coarseness, directionality and regularity or periodicity, directionality and randomness. Alternative methods of texture analysis for retrieval include the use of Gabor filters and fractals. A recent extension of the technique is the texture thesaurus developed by Ma and Manju, which retrieves textured regions in images on the basis of similarity to automatically-derived code-words representing important classes of texture within the collection. [8]

Methods of Representation

There are three principal approaches used to describe texture; statistical, structural and spectral…

Statistical techniques characterize textures using the statistical properties of the grey levels of the points/pixels comprising a surface image..

Structural techniques characterize textures as being composed of simple primitive structures called “texels” (or texture elements). These are arranged regularly on a surface according to some surface arrangement rules.

Spectral techniques are based on properties of the Fourier spectrum and describe global periodicity of the grey levels of a surface by identifying high-energy peaks in the Fourier spectrum [8].

For optimum classification purposes, what concern us are the statistical techniques of characterization. This is because it is these techniques that result in computing texture properties.

The most popular statistical representations of texture are:

• Co-occurrence Matrix

• Tamura Texture

• Wavelet Transform

Haralick proposed the following texture features [7]: Angular Second Moment ,Contrast ,Correlation, Variance Inverse Second Differential, Moment, Sum Average, Sum Variance, Sum Entropy, Entropy.

By observing psychological studies in the human visual perception, Tamura explored the texture representation using computational approximations to the three main texture features of: coarseness, contrast, and directionality [2, 12]. Each of these texture features are approximately computed using algorithms…

• Coarseness is the measure of granularity of an image [8], or average size of regions that have the same intensity [9].

• Contrast is the measure of vividness of the texture pattern. Therefore, the bigger the blocks that make up the image, the higher the contrast. It is affected by the use of varying black and white intensities [8].

• Directionality is the measure of directions of the grey values within the image [8].

Textures can be modeled as quasi-periodic patterns with spatial/frequency representation. The wavelet transform transforms the image into a multi-scale representation with both spatial and frequency characteristics. This allows for effective multi-scale image analysis with lower computational cost [2]. Unlike the usage of sine functions to represent signals in Fourier transforms, in wavelet transform, we use functions known as wavelets. Wavelets are finite in time, yet the average value of a wavelet is zero [2]. In a sense, a wavelet is a waveform that is bounded in both frequency and duration. The wavelet transforms convert a signal into a series of wavelets, which can be stored more efficiently due to finite time, and can be constructed with rough edges, thereby better approximating real-world signals [10].

III. METHODOLOGY USED

The method initially extracts the primitive features of a query image and compares them to those of database images. The image features under consideration are texture based features. Thus, using matching and comparison algorithms, the texture features of one image are compared and matched to the corresponding features of another image. This comparison is performed using characteristics such as energy, entropy and contrast features of an image. In the end, these characteristics are extracted one after another, so as to retrieve database images that are similar to the query. The similarity between the characteristics of the features are calculated using algorithms one for each specific feature for both extraction
and matching.

The three features of texture which have been used in the accomplishment of the proposed solution from the problem statement of the work undertaken in this paper are:

1. Energy
2. Entropy
3. Contrast

It will store the extracted values into an excel sheet in the form of matrix of order N by 3 where N is the number of

images in the image database. It will then compares the extracted values of the query image with the images stored in excel sheet.

IV. ALGORITHM FORMULATED

The method to extract the similar images from the database based on texture features were discussed earlier i.e. on the basis of "energy, entropy and contrast" is implemented using algorithm i.e.

1. Insert the query image by selecting from the database.
2. Convert the image from RGB to grey to use it by using the method rgb2grey ('imagename')
3. Calculate the total pixel of image by Tp=xx*yy
4. Extract all the feature of image like

```
gray_img = rgb2gray (img_data);
texture_img = rangefilt(gray_img);
texture_energy = sum (texture_img(:).^2)/tp;
texture_entropy = entropy (texture_img);
cmap_img = contrast (texture_img);
[x1 y1] = size (cmap_img);
cont_img = sum(cmap_img(:))/(x1*y1);
```

5. Calculate the feature matrices' and store it into a excel sheet

```
features = [texture_energy, texture_entropy, cont_img];
data_val = xlsread('DataValues');
```

6. Now compare these feature matrices with already exist feature of the database images

```
If(feature are matched)
    Show all the related images on the GUI window
Else
    Report an error that match not found
```

7. Go to step 2 for every new process.
8. End of process.

This algorithm works on the basis of the flow of control illustrated below.

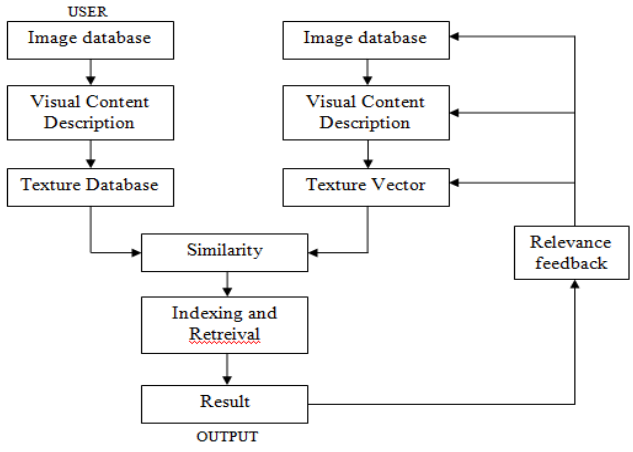


Fig 2: Algorithm to extract similar images from Database [9]

V. PERFORMANCE ANALYSIS USING MATLAB

The performance analysis of all the three modules is being done on the basis of two criteria's:

1. Precision
2. Recall

Precision = number of relevant images retrieved /total number of images retrieved.

Recall= number of relevant images retrieved /total number of relevant images.

This performance is being shown in both the tabular as well graphical form below:

Total number of images =100

Number of relevant images=10

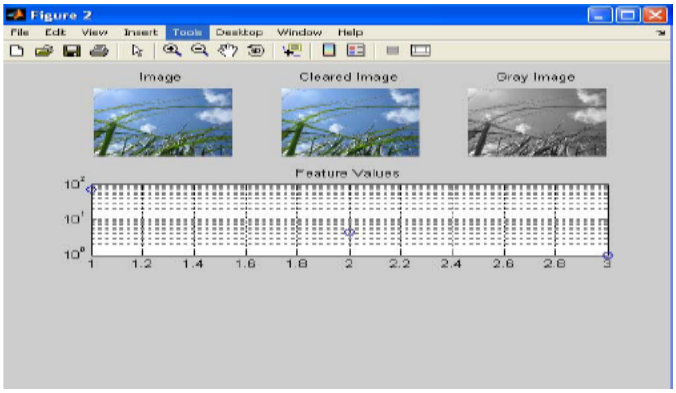
TABLE I
TABULAR FORM

TECHNIQUE	PRECISION	RECALL
Texture based image retrieval	0.6	0.06

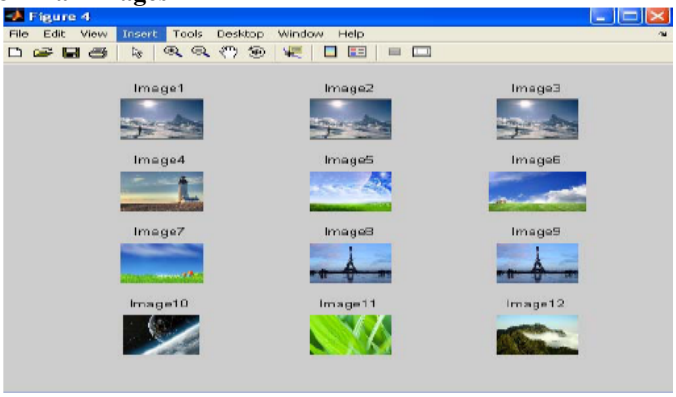
VI. RESULTS AND DISCUSSION

A. *Graphical User Interface:*

A Graphic based User Interface was designed and developed using MatLab GUIDE or Graphical User Interface Design Environment. Using the layout tools, screen shot of graphical user interface for TBIR application is shown below:



Similar Images



Precision=0.6
Recall=0.06

VII. CONCLUSION

Image data mining at present is still very much a research topic. The technology is existing but not complete to retrieve efficient and reliable results, so can be further expand with respect to demand, time and research in this area. A method for image data mining processing is based on texture content of the image i.e. "Energy, Entropy and Contrast" which is differ from the existing methods for the query image processing and get the results. The results from this received from this work and performance analysis of the images for texture based image retrieval technique given an average precision of 0.6. The results are good but can be further improved with the use of other features of images like shape and color.

REFERENCES

[1] V.N.Gudivada and V.V.Raghavan.: Special issue on content-based image retrieval systems - guest eds. IEEE Computer. 28(9) (1995) 18-22.

[2] M.De Marsicoi, L.Cinque, and S.Levialdi.: Indexing pictorial documents by their content: a survey of current techniques. Image and Vision Computing. 15(2) (1997) 119-141.

[3] Image retrieval using color and shape by Anil K. Jain and Aditya Vilaya , May 1995.

[4] Sharmin Siddique, "A Wavelet Based Technique for Analysis and Classification of Texture Images," Carleton University, Ottawa, Canada, Proj. Rep. 70.593, April 2002.

[5] http://cbir.fastsite.ru/about/3/

[6] Thomas Seidl Hans-Peter Kriegel "Efficient User-Adaptable Similarity Search in Large Multimedia Databases"

[7] http://www.vldb.org/conf/1997/P506.PDF

[8] http://km.doc.ic.ac.uk/pr-p.techasith-2002/Docs/OSE.doc

[9] "Texture," class notes for Computerized Image Analysis MN2, Centre for Image Analysis, Uppsala, Sweden, Winter 2002, at: http://www.cb.uu.se/~ingela/Teaching/ImageAnalysis/Texture2002.pdf.

[10] Gudivada and Raghavan, 1995a

[11] Content-based Image Retrieval by John Eakins Margaret Graham University of Northumbria at Newcastle www.jisc.ac.uk/uploaded_documents/jtap-039.doc

[12] FOLDOC, Free On-Line Dictionary Of Computing, "cooccurrence matrix," May 1995, [Online Document], http://foldoc.doc.ic.ac.uk/foldoc/foldoc.cgi?cooccurrence+matrix

[13] Pravi Techasith, "Image Search Engine," Imperial College, London, UK, Proj. Rep., July 2002, Found at: http://km.doc.ic.ac.uk/pr-p.techasith-2002/Docs/OSE.doc.

[14] Bjorn Johansson, "QBIC (Query By Image Content)", November 2002,[OnlineDocument],Availableat:http://www.isy.liu.se/cvl/ Projects/VISIT-bjojo/survey/surveyonCBIR/node26.html