A Study and Analysis on Image Indexing and Retrieval in Texture Base using Haar Wavelets

Mrs. Sheema. A. S[1]
1.Assistant Professor,
.Department of CSE & IT,
YCET, Kollam

Mr. Allwin Stephen[2]
2. Centre of Information Technology and Engineering ,
M.S.University, thirunelveli,kerala.

Abstract:- **Due to the enormous increase in image database sizes, as well as its vast deployment in various applications the need for CBIR development arose. Primitive features of an image are; texture, colour and shape. The extracted features are used as the basis for a similarity check .**
The images are retrieved based on the content. Primitive features of an image are texture, colour and shape. This system based on texture feature extraction . It is based on efficient extraction of texture features in image database based on multi resolution techniques (Gabor Wavelets). A clustering method ROCK is used to cluster the group of images based on feature vectors of images of database by considering the minimum Euclidean distance.

KeyWords: CBIR,,wavelets,indexing,image retrieval.

INTRODUCTION

INTRODUCTION TO THE CONTENT BASED IMAGE RETRIEVAL SYSTEM (CBIR)

The Content Based Image Retrieval System (CBIR) is a system, which retrieves the images from an image collection where the retrieval is based on a query, which is specified by content and not by index or address. The user is interested and wants to find similar images from the image collection. The CBIR system retrieves selected images from an image collection based on automatic derived features. The derived features include primitive features like texture, colour, and shape.

The present method implemented by three steps.
First, for each image in the image collection, a feature vector of size ten, characterizing texture of the image is computed based on the Wavelet transformation method. The Wavelet transformations are used because they capture the local level texture
features quite efficiently, where feature vectors are stored in a feature database.
Second, using clustering algorithm to construct indexed image database based on the texture feature vectors obtained from wavelet transformation, and finally, given a query image, its feature vector is computed and compared to the feature vectors in the feature database, and relevant images to the query image from the image database returned to the user.

The purpose of this project is to make solution to the problem of designing a Content Based Image Retrieval-CBIR system. Due to the increase in image database sizes, as well as its vast deployment in various applications the need for CBIR development arose. Primitive features of an image are; texture, colour and shape. These features are extracted and used for a similarity check between images.

OBJECTIVE OF THE PROJECT

Content-based image retrieval (CBIR) technique uses image content to search and retrieve digital images. Content-based image retrieval system was introduced the problems associated with text-based image.

This proposal will arose due to enormous increase in image database size as well as vast deployment in various applications. The images are retrieved based on the content. Primitive features of an image are texture, colour and shape. This system based on texture feature extraction. The main work identified for the Project are that efficient extraction of texture features in image database based on multi resolution techniques (Haar Wavelets). A clustering method ROCK is used to cluster the group of images based on feature vectors of images of database by considering the minimum Euclidean distanc.

PRINCIPLE OF CBIR

A typical CBIR system automatically extract visual attributes (colour, shape, texture and spatial information) of each image in the database based on its pixel values and stores in a different database within the system called feature database**.** The feature data for each of the image is very much smaller in size compared to the image data. Thus the feature database contains an abstraction (compact form) of the images in the image database; each image is represented by a compact representation of its contents (colour, texture, shape and spatial information) in the form of a fixed length real-valued multi-component feature vectors or signature.

The users formulate query image and present to the system. The system automatically extract the visual attributes of the query image in the same mode as of each database image, and then identifies images in the database whose feature vectors match those of the query image, and sorts the best similar objects according to their similarity .

During operation the system processes less compact feature vectors rather than the large size image data thus giving CBIR its cheap, fast and efficient advantage over text-based retrieval.

CBIR system can be used in one of two ways. First, exact image matching, that is matching two images, one an example image and the other, image in image database. Second is approximate image matching, which is finding most closely match images to a query image

In most current CBIR systems each of the feature extraction techniques used for retrieval are treated with equal emphasis. If the feature most suitable for retrieval of a particular image is used in flexible weighted combination with other features it is expected that a higher level of precision and recall can be achieved. Flexible weighted combination of the image features can provide the basis for elevating current CBIR system to recognise objects and scenes. Since objects concepts are usually related to visual characteristics high-level scene properties may be inferred from weighted combination of low-level image features – texture, shape and spatial information .

FEATURE DEFINED

Feature is anything that is localized, meaningful and detectable. In an image noticeable features include corners, lines, objects, colour, shape, spatial location, motion and texture. Features extracted from images define and describe the image content . No particular visual feature is most suitable for retrieval of all types of images. Colour visual feature is most suitable for describing and representing colour images. Texture is most suitable for describing and representing visual patterns, surface properties and scene depth. CBIR system using texture is particularly useful in satellite images, medical images and natural scenes like clouds, while shape is suitable for representing and describing boundaries of real world objects and edges. In reality no one particular feature can completely describe an image.

VISUAL FEATURES EXPLAINED

Visual features of an image is the output of the human visual system as represented by human visual perception of images, and these include color, texture, shape, spatial information and motion.

TEXTURE DEFINED

According to the American Heritage dictionary texture is

- A structure of interwoven fibres or other elements such as repetitive patterns.
- The distinctive physical composition or structure of something, especially with respect to the size, shape, and arrangement of its parts.
- The appearance and feel of a surface.
- Distinctive or identifying quality or character.

Texture as repeated patterns of pixels over a spatial domain, of which the addition of noise to the patterns and their repetition frequencies result in textures that can appear to be random and unstructured. Texture properties are the visual patterns in an image that have properties of homogeneity that do not result from the presence of only a single colour or intensity.

The different texture properties as perceived by the human eye are regularity, directionality, smoothness and coarseness.

TEXTURE FEATURE EXTRACTION

The commonly used methods for texture feature description are statistical and transform-based methods .

STATISTICAL METHOD

Statistical methods analyse the spatial distribution of grey values by computing local features at each point in the image, and deriving a set of statistics from the distribution of the local features. They include co-occurrence matrix representation, statistical moments, grey level differences, autocorrelation function and grey level run lengths.

TRANSFORMED-BASED METHOD

Transform methods analyse the frequency content of the image to determine texture features. Examples include the use of Fourier transform to describe the global frequency content of the image and multi-resolution analysis (wavelet transform and Gabor wavelets) that uses a window function whose width changes as the frequency changes. Multi

ISSN: 2278-0181
NCICN-2015 Conference Proceedings

resolution analysis, the representation or analysis of signals at different scales subjects the image to a linear transform followed by energy measurement.

PROBLEM DESCRIPTION

In the twentieth century, introduction of computer and advances in science and technology gave birth to low cost and efficient digital storage devices and the worldwide web, which in turn became the catalyst for increasing acquisition of digital information in the form of images. In this computer age virtually all spheres of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design and historical research are in need of, and use of images for efficient services. A large collection of images is referred to as image database . Image database is a system where image data are integratedly stored [**2**]. Image data include the raw images and information extracted from images by automated or computer assisted image analysis.

The fundamental issues in the design of CBIR including its image database[**6**], which are

- Efficient selection of images which satisfy a user's query
- Data modeling
- Feature extraction
- Selecting appropriate features for content representation
- Query languages, and
- Indexing techniques.

The advantage by applying indexing method in CBIR system is to identify appropriate image features within the image that best represent the image and represent or combine them , during such that retrieval process meaningful result can be obtained. Identifying suitable image feature for describing a particular type or class of image reduces storage size of indexing features used in programming, leading to efficient and fast CBIR system. Features are used to represent an image instead of using the original pixel values because of the significant simplification of image representation and the improved correlation with image semantics. The beauty of CBIR is seen from the fact that though computers cannot understand images, comparison of visual features as enabled by the feature vectors [1]enables comparison of the real-world visual scenes. .

Texture is most suitable for describing and representing visual patterns, surface properties and scene depth. CBIR system using texture is particularly useful in satellite images, medical images and natural scenes like clouds, while shape is suitable for representing and describing boundaries of real world objects and edges.

The fundamental issues in the design of CBIR including its image database, which are

1. Efficient selection of images which satisfy a user's query .
2. Texture Features extraction based on statistical method by considering local features at each point in the image, and deriving a set of statistics from the distribution of the local features. It does not consider frequency content.
3. Selecting appropriate features for content representation. Texture is most suitable for describing and representing visual patterns, surface properties and scene depth. CBIR system using texture is particularly useful in satellite images, medical images and natural scenes like clouds, while shape is suitable for representing and describing boundaries of real world objects and edges.
4. By applying indexing techniques fast retrieval is possible..

THE PROPOSED SYSTEM

This proposal will arose due to enormous increase in image database size as well as vast deployment in various applications. The images are retrieved based on the content. Primitive features of an image are texture, colour and shape. This system based on texture feature extraction . It is based on efficient extraction of texture features in image database based on multi resolution techniques (Gabor Wavelets). A clustering method ROCK is used to cluster the group of images based on feature vectors of images of database by considering the minimum Euclidean distance.

IMAGE INDEXING AND RETRIEVAL

The advantage by applying indexing method in CBIR system is to identify appropriate image feature(s) within the image that best represent the image and represent or combine them , during such that retrieval process meaningful result can be obtained.

Statistical methods analyse the spatial distribution of grey values by computing local features at each point in the image, and deriving a set of statistics from the distribution of the local features. By applying the co occurrence matrix the processing speed is very slow , so the multi resolution analysis is selected .

The proposed approach for solving the above problem will be on the following directions:

The multi-resolution analysis (wavelet transform and Gabor wavelets) that uses a window function whose width changes as the frequency changes. Multi resolution analysis, the representation or analysis of signals at different scales subjects the image to a linear transform followed by energy measurement.

The present method implemented by three steps. First, for each image in the image collection, a feature vector of size ten, characterizing texture of the image is computed based on the Wavelet transformation method. The Wavelet transformations are used because they capture the local level texture features quite efficiently, where feature vectors are stored in a feature database. Second, using clustering algorithm to construct indexed image database based on the texture feature vectors obtained from wavelet transformation, and finally, given a query image, its feature vector is computed and compared to the feature vectors in the feature database, and relevant images to the query image from the image database returned to the user.

The steps involved in the methodology are listed below:

- Haar Wavelet transformation is used for feature extraction.
- Pre computing the texture feature vectors for all the images in the database using Haar wavelet Transformation.
- Clustering the images based on feature vectors using modified ROCK clustering algorithm.
- Computing the feature vector of the query image as and when presented.
- Comparing query image with indexed database, identifying the closest cluster for the query image and retrieves those images.

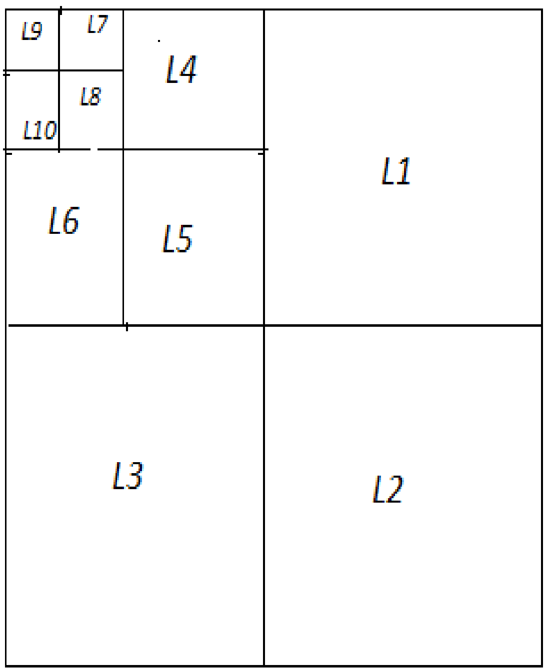
EXTRACTION OF FEATURE VECTOR

Texture is another important property of images. Various texture representations have been investigated in pattern recognition and computer vision. Texture representation methods can be classified into two categories: structural and statistical. Structural methods, including morphological operator and adjacency graph, describe texture by identifying structural primitives and their placement rules.

It tend to be most effective when applied to textures that are very regular. Statistical methods such as co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and Haar wavelet transform, characterize texture by the statistical distribution of the image intensity. The Extraction of feature vector is the most crucial step in the whole CBIR system. This is because these feature vectors are used in all the subsequent modules of the system. It is to be realized that the image itself plays no part in the following modules. It is the feature vectors that are dealt with. The quality of the output drastically improves as the feature vectors that are used are made more effective in representing the image. The fact that the quality of the output is a true reflection of the quality of the feature vector is very much evident in our experiments. The Feature vector generation [9] has been tried in two different ways. One way was to use wavelets [2, 7, 9] to compute energies whose values were classified.

HAAR WAVELETS

The Wavelets are useful for hierarchically decomposing functions in ways that are both efficient and theoretically sound. Broadly speaking, a wavelet representation of a function consists of a coarse overall approximation together with detail coefficients that influence the function at various scaled. The wavelet transform has excellent energy compaction and de-correlation properties, which can be used to effectively generate compact representations that exploit the structure of data. By using wavelet sub band decomposition, and storing only the most important sub bands (that is, the top coefficients), we can compute fixed-size low-dimensional feature vectors independent of resolution, image size and dithering effects. In addition, wavelets are robust with respect to colour intensity shifts and can capture both texture and shape information efficiently. Furthermore, wavelet transforms can be computed in linear time, thus allowing for very fast algorithms. We compute feature vectors using Haar wavelets because they are the fastest to compute and have been found to perform well in practice [9, 10]. Haar wavelets enable us to speed up the wavelet computation phase for thousands of sliding windows of varying sizes in an image. They also facilitate the development of efficient incremental algorithms for computing wavelet transforms for larger windows in terms of the ones for smaller windows. One disadvantage of Haar wavelets is that it tends to produce large number of signatures for all windows in image. We proposed the modified Haar wavelet transformation that reducing signatures only by calculating 10 for the image in our method. In our feature vector computation process, we applied Wavelet Transformations only three times to get 10 sub images of input image in the following way.



In each iteration Li2...4 images are saved and Li1 sub image is again subjected to wavelet Transformation instead of entire image for three iterations, by which 10 sub images of input image are obtained. Sub image L11 is further divided into sub images L 21 ... L24 in the second iteration. The sub image L 21 is further divided into L31 L32 L33 L34 in the third iteration. All sub images are normalized to maintain the uniform size.

ALGORITHM FOR CALCULATING WAVELET SIGNATURES

1. Let L be the image of size w×w
2. Divide the image L into four bands L1,L2,L3,L4 based on Haar wavelet of size

   w/2×w/2
3. Compute Signatures fr for L2,L3,L4
4. Now take the image L1 and divide it into 4 bands namely L11,L12,L13,L14 of size w/4×w/4
5. Compute signatures fr for L12,L13,L14
6. Again take the L11 and divide it into 4 bands namely L111,L112,L113,L114 of size w/8×w/8.
7. Then we obtain 10 signatures then stop the process.

The Wavelet signature (texture feature representation) is computed from sub image as follows,

$$f\,r = \sqrt{\sum C_{ij}^{2} / i \times j}$$

Where fr is the computed Wavelet signature (texture feature representation) of the sub image, Cij is the representation of the intensity value of all elements of sub image and i × j is the size of the sub image.

INDEXING OF IMAGES

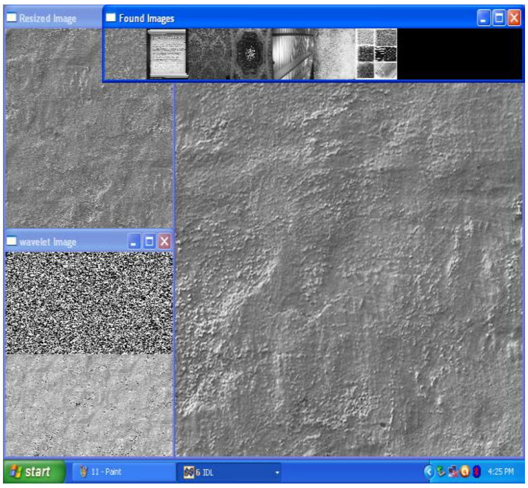
Another important issue in content-based image retrieval is effective indexing[1] and fast searching of images based on visual features. Because the feature vectors of images tend to have high dimensionality and therefore are not well suited to traditional indexing structures, dimension reduction is usually used before setting up an efficient indexing scheme. The basis of the clustering method in indexed image database is that ,the images belonging to the same cluster are similar or relevant to each other when compared to images belonging to different clusters. We clustered the images using ROCK[5].The ROCK allow us to minimize the undesirable results of the ROCK algorithm. The feature vector of each image is a vector of size 10.The Euclidean distance measure is used to measure the similarity between feature vectors of query image and indexed database image. In the present method we calculated representative Feature vector of Cluster (FC) as the minimum Euclidean distance, which resulted in good cluster-matching results. The representative feature vector of cluster(FC) is computed from the following equation.

$$Fci = \min|Fi - \Sigma Fj|$$

Where j =1,2,…..n and j ≠ i, and i=1,2,..n. Fci denotes representative feature vector of cluster i,and Fi,Fj represents feature vector of the given cluster. The system converts the image into an internal representation of features. Images stored in the database with similar features are searched.

RESULT

As a case study the proposed method is applied on different images. The result is given below.



CONCLUSION

CBIR at present is still topic of research industry .Different features are used for images retrieval. By deriving feature vectors from wavelet transformation in three iterations reduces overall time complexity than previous methods. In this paper compute feature vectors using Haar wavelets because they are the fastest to compute and have been found to perform well in practice. The new method proposed in my study for clustering effectively minimizes the undesirable results and gives a good matching pattern, that will be having zero or a minimum set of no relevant images.

FUTURE ENHANCEMENT

The present system operates partially at the primitive feature level. The present system extracts only the Texture feature of an image. This system can be enhanced to extract the other primitive features also.The retrieval efficiency and timing performance can be further increased if the image collection is trained and grouped using supervised learning. This will greatly enhance research time and precision.

ISSN: 2278-0181
NCICN-2015 Conference Proceedings

BIBLIOGRAPHY

[1] Sameer Antani, Rangachar Kasturi, and Ramesh Jain. A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing and Retrieval of Images and Video. Pattern Recognition,35:945–965, 2002.

[2] I Daubechies, Ten Lectures on Wavelets, Capital City Press, 1992.

[3] Faloutsos, R. Barber, M. Flickner, J. Hafner,W. Niblack,D. Petkovic, W.Equitz ``Efficient and effective querying by image content,' Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, vol. 3, no. 3-4, pp. 231-62,July 1994.

[4]Gupta, R. Jain, ``Visual information retrieval,"Comm.Assoc. Comp. Mach., vol. 40, no. 5, pp. 70-79, May1997

[5] Guha S.,Rastogi R., and Shim K.ROCK: A robust clustering algorithm for categorical attributes. In proceeding Conclusions of the IEEE International Conference on data engineering,Sydney,March 1999.

[6] W. Y. Ma, B. Manjunath, ``NaTra: A toolbox for navigating large image databases," Proc. IEEE Int.Conf. Image Processing, pp. 568-71, 1997.

[7] Y. Meyer, Wavelets Algorithms and Applications, SIAM, Philadelphia, 1993.

[8] S. Mukherjea, K. Hirata, Y. Hara, ``AMORE: a World Wide Web image retrieval engine," World Wide Web, vol. 2, no. 3, pp. 115-32, Baltzer,1999.