International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 1, Issue 2, July – August 2012 ISSN 2278-6856

Analysis of Web Information Gathering Based on Sketch Image Retrieval System

Sushma Annam[1], M. Kavitha[2], K.C. Ravi Kumar[3]

[1]Pursing M.Tech.(CSE), Sri Devi Women's Engineering College, India .

[2]Asst. Prof., Sri Devi Women's Engineering College, India .

[3]HOD, Sri Devi Women's Engineering College, India .

Abstract - ***Image processing is the popular research area is based on the content image retrieval system, image search tools such as Webopedia, Google Images and Yahoo, image search based on textual annotation of images. Images are manually annotated with keywords and then retrieved using text-based search methods. Our analysis introduces the content based on a free hand sketch with the help of the existing methods which describes sketch colored image search is efficient, sequence of preprocessing steps that the transformed full color image and the sketch. Sketch based image retrieval system can be used as digital libraries, crime prevention etc, we compare this with previous technology and also analyzed the algorithm such a system has great data in suspects and identifying victims in forensics sketch to shot images which demands wide spectrum on the image processing.***

Keywords – Image Processing, K-Means, Sketch, Content based image Retrieval

1. INTRODUCTION

The act of selecting a subset of an image database corresponding to a description given by the user query Image retrieval has been extremely active research area over the past but first to review access methods in image databases appeared already in the recent 80years the state of the art of the corresponding years and contain references to a large number of systems descriptions of the technologies implemented. The extensive description of image archives various indexing methods and common searching tasks using text based searches on annotated images. A user produces a query representing the images wants to retrieve from the database and submits it to the content based image retrieval system, the system computes the similarity between the query and the images stored in the database is done according to the internal description of query and databases image and returns a list of images sorted according to their similarity to the query, the user modifies the query or uses part of the result to form a new query.

Before spreading of information technology a huge number of data had to be processed and stored was also textual and visual information, simultaneously the appearance and quick evolution of computers an increasing measure of data had to be managed. The increasing growth of data storages and revolution of internet had changed the world efficiency of searching in information set is a very important point of view. If we want to search efficiently some data have to be recalled the human is able to recall visual information more easily using example the shape of an object or arrangement of colors and objects is visual type we look for images using other images and follows the approach also at the categorizing. We search using some features of images and these features are the keywords unfortunately at the moment there are not frequently used retrieval information of a sample image reason may be that the text is a human abstraction of the image. The purpose of this is to analyze content based image retrieval system which can retrieve using sketches in frequently used databases. User has a drawing area where can draw those sketches which are the retrieval method.

In the sketch based image retrieval system the user draws color sketches and blobs on the drawing area, the image were divided into grids and the color, texture features were determined. The grids were also used in other algorithms example like in the edge histogram descriptor method defect these methods is that they are not invariant opposite rotation, scaling and translation and other application of fuzzy logic or neural networks invest to determine suitable image features.

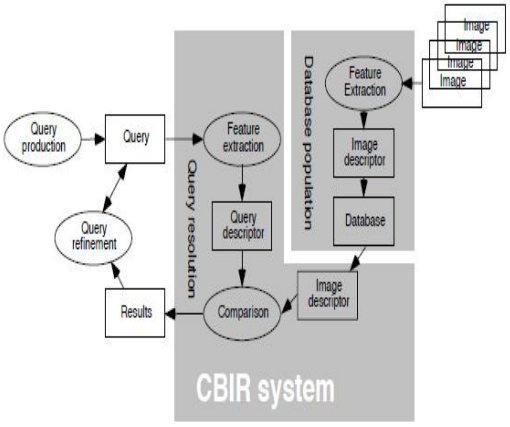

Figure 1: Shows the Content Based Image Retrieval System.

The interaction between the user and content based image retrieval system can help in achieving better retrieval results and interaction ranges from simply allowing the user to submit a new query based on a existing one to giving the user the possibility to select part of the result image as relevant and non-relevant to allow the user visually arrange a small set of the database images into clusters of similar images and rearrange the whole database according the actions.

2. RELATED WORK

Finding images in a database that satisfy a particular specification that were stamp collector is composed of several thousand stamps and that want to be able to make lists of our stamps according to the issuing nation the date of issue the nominal value current market value the shape the pictorial content the condition the series they belong to or similarity to particular stamp. Being given an image database what we would like from content system is the ability to select a subset of the database according to a query submitted to the system, the size of the subset can range from the empty set to the whole database. Since image databases are based similarity rather than on matching as pointed. This subset will in general be sorted according to similarity to the query.

Content based Image Retrieval applications are listed below.

Galleries and museum management
Architectural and engineering design
Interior design
Remote sensing and management of earth resources
Geographic information systems

Images contained in databases can be disparate kinds ranging from a 16*16 two bit pattern to a 1200 dpi 32-bit color scan of an A4 size page. The databases containing a more uniform kind of images will in general be easier to handle and will allow more precise searches than heterogeneous databases specialized algorithms or domain experts will be extract the available wanted data.

Characteristics of an image are the size of the image aspect ratio, color depth in bits conditions under which it was illumination distance between object and camera, the number of objects portrayed Knowledge about what kind of object can be in the image and its origin, the file format, whether the objects are in front of known background The variance of these image characteristics within database allows us to perform a classification of the databases.

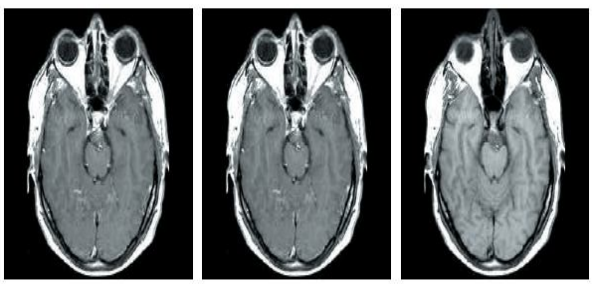

Figure 2: Shows the database image retrieval system.

3. PROBLEM DEFINITION:

To retrieve the image from the database, need to preprocess all the characteristics in the structured subsystem and the functionality of subsystems. To identify the process of feature vector and retrieval subsystem from stock index database management system is aim to analyze the functionality.

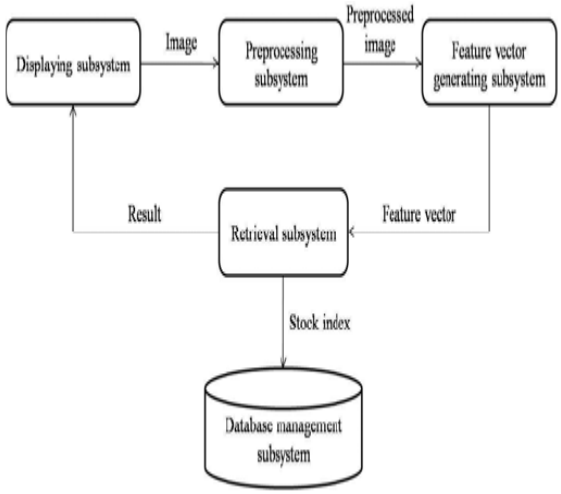

Figure 3 shows the problem definition

Even though in the sketch based image retrieval increases there is no widely used sketch based image retrieval system the user has a drawing area where can draw all shapes and moments which are expected to occur in the given location and with a given size.

3.1. Clustering in Image Retrieval System: Cluster is a number of similar objects grouped together. It can also be defined as the organization of dataset into homogeneous and/or well separated groups with respect to distance or equivalently similarity measure. Cluster is an aggregation of points in test space such that the distance between any two points in cluster is less than the distance between any two points in the cluster and any point not in it. There are two types of attributes associated with clustering, numerical and categorical attributes. Numerical attributes are associated with ordered values such as height of a person and speed of a train. Categorical attributes are those with unordered values such as kind of a drink and brand of car.

Clustering is available in flavours of i)Hierarchical and ii) Partition (non Hierarchical).

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partition stakes place, which may run from a single cluster containing all objects to n clusters each containing a single object [9].Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groupings.

For the partitional can be of K-means and K-mediod.

The purpose solution is based on K-means (Unsupervised) clustering combine with Id3 Decision Tree type of

Classification (Supervised) under mentioned section describes in details of K-means & Decision Tree. K-means [10] [11] is a centroid based technique. Each cluster is represented by the
center of gravity of the cluster so that the intra cluster similarity is high and inter cluster similarity is low. This technique is scalable and efficient in processing large data sets because the computational complexity is O(nkt) where n-total number of objects, k is number of clusters, t is number of iterations and k<<n and t<<n

3.2. Algorithm to be applied:

The k-means algorithm:

Algorithm: k-means. The k-means algorithm for partitioning based on the mean value of the objects in the cluster.

Input: The number of clusters k and a database containing n objects.

Output: A set of k clusters that minimizes the squared-error criterion.

Method:

(1) arbitrarily choose k objects as the initial cluster centers:

(2) repeat

(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;

(4) Update the cluster means, i.e., calculate the mean value of the objects for each cluster;

(5) Until no change.

In earlier days, image retrieving from large image database can be done by following ways. We will discuss briefly about the image retrieving of various steps Automatic Image Annotation and Retrieval using Cross Media Relevance Models Concept Based Query Expansion Query System Bridging The Semantic Gap For Large Image Databases Ontology-Based Query Expansion Widget for information Retrieval Detecting image purpose in World-Wide Web documents *Benefits* Relevance feedback is an interactive process that starts with normal CBIR. The user input a query, and then the system extracts the image feature and measure the distance with images in the database. An initial retrieval list is then generated.

User can choose the relevant image to further refine the query, and this process can be iterated many times until the user find the desired images.

4. COMPARATIVE ANALYSIS:

Query by Image Content is one of the earliest commercial attempts at an engine providing content-based image indexing. It was built at the IBM Almaden Research Center, and started out based primarily on color-histogram image indexing. The features extracted for images are primarily global features like color histograms, global texture and the average values of the color distribution. Images in QBIC are represented as whole images, but can also have manually outlined objects. Retrieval is performed by measuring the similarity between the user's query and the database images. Specific similarity functions are defined for each feature. Although this is one of the least sophisticated image querying systems technologically, it nevertheless has sufficient to support as a commercial system. It should be noted here that most of the other systems mentioned here, or covered in the literature, are not used commercially. Compared to query by image content image retrieval system using the sketch based image retrieval system is having the capability of searching a database of images has becomes as crucial and should become as natural as text search has become for text databases. The medium supporting the search should be the same as the medium of the documents to be retrieved. Text-based searching of images is therefore unit. Searching for three people standing in front of a truck" would require that the database maintainer has pre-indexed every image for each of its elements, including actions such as standing" and spatial relations such as in front of". Alternatively, if the indexing process is to be automated, the system should either be able to recognize any type of object in a database to be indexed and know a name for it, or create mental representations of words in the query. Neither of these questions has been solved yet. Based on this discussion, content-based retrieval cannot and should not be done based on a text interface, unless computer vision research is able to reliably construct image models from either text or images. Therefore, the user input should be an image itself. Alternatively, it might be simpler to let the user choose from a set of images already in the database the ones that more closely match the image to be retrieved. This would allow pre-computation on these images and feature extraction online, thus speeding up the query time, but unfortunately limiting the user's versatility. If full freedom is to be left to the user, the only practicable medium for a query input is a sketch.

5. CONCLUSION

The objectives of system performs a test sketch-based image retrieval system, main aspects were the retrieval process has to be unconventional and highly interactive. The robustness of the method is essential in some degree of noise, which might also be in case of simple images. The drawn image without modification not be compared with color image, or its edge representation, alternatively will go for transform. The simple smoothing and edge detection based method was improved, which had a similar importance compare to as the previous system.

REFERENCES

[1] S. Santini, R. Jain, ÒIntegrated Browsing and Querying for Image DatabasesÓ, *submitted to IEEEMultimedia*, 1999.

[2] G. Cha, C. Chung, ÒObject-Oriented Retrieval Mechanism for Semistructured Image CollectionsÓ,*Proceedings of the 6th ACM Multimedia Conference, Bristol, UK*, 1998.

[3] Y. Cheng, ÒMean Shift, Mode Seeking, and ClusteringÒ, IEEE Transactions on Pattern Analyisisand Machine Intelligence, Vol. 17, pp. 790-799, 1995.

[4] L. Cinque, F. Lecca, S. Levialdi, S. Tanimoto, ÒRetrieval of Images Using Rich Image DescriptionsÓ, *Proceedings of the ICPR*, 1998.

[5] G. Ciocca, I. Gagliardi, R. Schettini, ÒRetrieving Color Images by ContentÓ, *Proceedings of the Image and Video Content-Based Retrieval Workshop*, 1998.

[6] G. Ciocca, R. Schettini, ÒUsing a Relevance Feedback Mechanism to Improve Content-Based Image Retrieval Ó, *Proceedings of the third Conference on Visual Information Systems (VISUAL99), Amsterdam, The Netherlands*, 1999, pp. 107-114.

[7] D. Comaniciu, P. Meer, ÒRobust Analysis of Feature Spaces: Color Image SegmentationÓ, *Proceedingsof IEEE Conference on Computer Vision and Pattern Recognition CVPRÕ97, PuertoRico*, 1997, pp. 750-755.

[8]Commission Internationale de lÕEclairage, ÒRecommendations on Uniform Color Space, Color- Difference Equations, and Psychometric Color TermsÓ, *Supplement No. 2 to Publication CIE*, No.15, 1978.

[9] Text Book of Data mining Techniques by Arun K PujariUniversities Press (India) Private Limited.

[10]. Intoduction to hierarchical clustering, A tutorial on clustering,A Tutorial on Clustering Algorithms ... Hierarchical Clustering- Interactive demo.

[11]. K-Means+ID3: A Novel Method for Supervised AnomalyDetection by Cascading K-Means Clustering and ID3 DecisionTree Learning Methods, Shekhar R. Gaddam, Vir V. Phoha,and Kiran S. Balagani, IEEE Transactions on Knowledge andData Engineering, VOL. 19, NO. 3, March 2007.

[12]. Karl-Heinrich Anders, A Hierarchical Graph-ClusteringApproach to find Groups of Objects, IEEE Transactions.

Sushma Annam pursuing M.Tech Computer Science Engineering from SriDevi Women's Engineering College. Master of Computer Applications from N.N.S Vidya College of P.G Studies Acharya Nagarjuna University. Her interested areas include Network Security, Image Processing, Retrieval Systems.

M.Kavitha M.Tech Software Engineering from JNT University currently she working as Assistant Professor at SriDevi Women's Engineering College and having 6 years of Academic Experience. Her interested areas include Database Management Systems, Operating Systems, Data warehousing and Data Mining.

K.C.Ravi Kumar M.Tech CSE from JNTU Hderabad currently he is the head of department for M.Tech CSE programme in SriDevi Women's Engineering College having 17 years of Academic Experience. He is life member of IEEE & IST areas of research include Data Mining & Data Warehousing Information Retrival Systems Information Security.