Use of Content Based Image Retrieval System for Similarity Analysis of Images

[1]Dipalee J. Fendarkar
[1]Research Student, Government College of Engineering, Aurangabad

Abstract **-** This paper is based on the development of content based image retrieval system (CBIR) and sketch based image retrieval system (SBIR). The aim of CBIR is to extract visual features and display the relevant images. This paper introduces the presents problems and challenges concerned with the design and the creation of CBIR systems, which is based on a free hand sketch (i.e. SBIR). The use of the existing methods, describe a possible results, how to search a number of images and implement a task specific descriptor, which can handle the informational gap between a sketch a colored image to make an opportunity for the efficient search. The CBIR system computes the similarity between the query and the images stored in the database. This paper introduces and presents the result of primitive features of images like textures, colors and shapes. The description of feature extraction, feature based matching and indexing which represent the base of retrieving images and it allows the comparison of database images with queries containing various levels of detail and similarity analysis of image, representation of the database images and also introduced the creation of the mosaic of the images and compared the methods of matching sketches descriptor. This result paper is focus on retrieval of images based on the visual content of the input (query) object, which demands on the quite wide methodology variety on the area of the Image Processing.

Index Terms **- Content Analysis and Indexing, Indexing methods, Digital Images, Image Descriptors, K-means algorithm, Image Database.**

I. INTRODUCTION

In computer technology, Database and Internet are the most basic part of the searching information. These are mostly based upon the text search. Although there are many advanced machines to handle queries and storing the vast amount of information. The efficiency of search cannot be same. The methods of searching using text based content are not efficient because most of the visual Information cannot be expressed textually. This situation can be avoided by using CBIR technology [1].

The CBIR technology is necessary for the image; the image search is made efficient and dependable. CBIR is allows to extract information which itself represent the image. Content-based image retrieval (CBIR) is also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR). Application of computer vision to the image retrieval problem is the problem to search for digital images in the large databases [1, 2]. The term 'content-based' means that the search can be analyzing the actual content of the image. Content must be relay on metadata to examine the image content, searches without the facility like captions or keywords. This type of metadata must be generated by a human and stored alongside each image in the database. The rise of interest in these techniques for retrieving images on the basis of automatically-derived features such as color, texture and shape. This technology generally referred to as Content-Based Image Retrieval (CBIR) [1, 2].

The aim of this paper is to introduce the current state of the art in content-based image retrieval. The word 'content' in CBIR refers to feature of images like color, shape, texture, edges etc. which can be extracted from that image itself. Thus a system which uses features of image will give better indexing and give accurate results. The work is done on inventing an easier and smarter system called Sketch-Based Image Retrieval (SBIR). It is a system that basically operates to show the user's query by line-based hand drawing. The query picture can be well expressed by SBIR through shape only while it needs more than that to be expressed by CBIR. The SBIR system adds new interesting variety of hand drawing style and even more flexibility to the query users. Below figure shows the flow diagram of content based image retrieval system [1, 2].

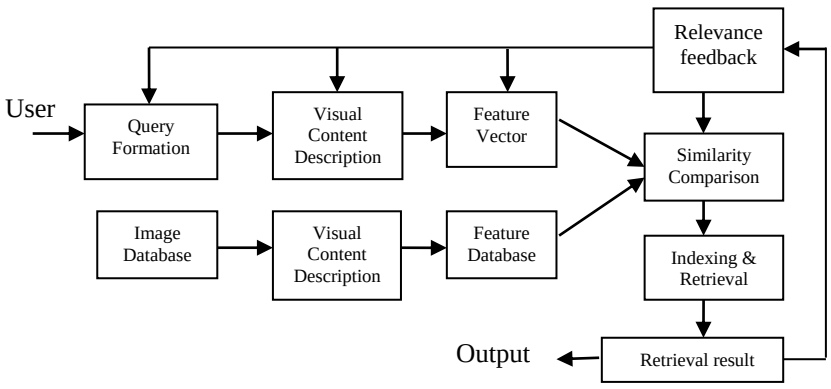

Fig. 1 Diagram for content-based image retrieval system

II. RELATED WORK

Jing Huang ed al discussed new feature called color correlogram for image indexing and comparison. This new feature computed efficiently and shows that performance very well. Sim, D. G., H. K. Kim and R. H. Park "the image retrieval scheme

for JPEG formatted image is presented. Content based image retrieval for JPEG images has attracted many people's attention and a series of algorithms directly based on the discrete cosine transform domain. And to take full advantage of DCT coefficients and consider the color and texture information for the retrieval of JPEG formatted images. Here decompressing the images and then performing in the spatial domain. The feature vectors are computed from several DCT coefficients. And this operation is performed in the partial decoded domain. It can greatly decrease the retrieval complexity [3, 4, 5]**.**

M. Flickner et.al proposed Color histograms are computationally efficient, and generally insensitive to small changes in camera position. However, a color histogram provides only a very coarse characterization of an image, An images with similar histograms can have dramatically different appearances. Here , to describe a method which imposes additional constraints on histogram based matching [6].

J. Zhang et.al suggests the image retrieval based on the textural information of an image, such as orientation, directionality, and regularity. Here, utilize texture orientation to construct the rotated Gabor transform for extraction of the rotation-invariant texture feature. The rotation invariant texture feature, directionality, and regularity are the main features used in the proposed approach for similarity assessment. Using these features, we finally propose an efficient mechanism for CBIR and examine it through some applications. the system can now compare features of the query with features of images in the collection based on some matching criterions. Because three features are used in this work, three matching scores need to be computed. A weighted average of the matching scores is then calculated to get a final score for each image. Finally, rank images based on these final scores and top ranked images are displayed to the user as the result of retrieval [7].

III. CONTENT BASED IMAGE RETRIEVAL

In Content Based Image Retrieval, visual content of image is used to match similar images. Visual content refer to image's characteristics such as its color, shape, texture, object or any other information which can be derived from image itself. Content Based Image retrieval (CBIR) is desirable because most web-based image search engines purely reliable on metadata and this produces a lot of garbage in the results. Humans manually entering keywords for the images in large databases. It can be inefficient, expensive and may not capture every keyword that describes the image. Hence a system that can filter images based on their content would provide better indexing and return more accurate results [8].

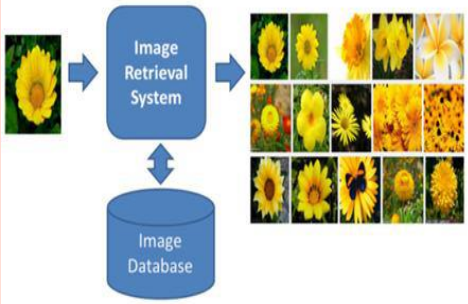

Fig. 2 CBIR.

Content-based image retrieval is also known as query by image content. Will study the actual information content based image search means. A database of images from earlier experiments Auto Recover feature was the color and size. Primitive features used for restoration or could mean, but the extraction process must be predominantly automatic [9]. Semantic extraction can be done automatically and correctly, but the image retrieval system cannot expect to find the exact images. Users can be select the desired images to the most similar images should choose. Set the number of images to retrieve perceptual similarity measure proposed in this paper by applying the similarity measure can be reduced. A typical CBIR system consists of three major components used variations of them depending on the features [9].

i. Extraction facility - the facility to check the raw image data to extract specific information.
ii. Feature storage - also help to improve the pace of discovery, offering efficient storage of the extracted information.
iii. Similarity measure - to determine the significance between the images to measure the difference between the images [9].

IV. EXPERIMENTAL DETAILS

4.1 The K-Means Algorithm

The k-means algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intra cluster similarity is high but the inter cluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's center of gravity. "How does the k-means algorithm work" [10].

The k-means algorithm proceeds as follows. First, it randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean. It then computes the new mean for each cluster [10].

Algorithm: k-means. The k-means algorithm for partitioning based on the mean value of the objects in the cluster [11].
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 [11].

4.2 Indexing

Indexing the whole set of images using K-means Clustering algorithm. Indexing is done using an implementation of the Document Builder Interface. A simple approach is to use the Document Builder Factory, which creates Document Builder instances for all available features as well as popular combinations of features (e.g. all JPEG features or all avail-able features) [12].

In content based image retrieval system, target images are sorted by feature similarities with respect to the query (CBIR).In this indexing use to K-means clustering for the classification of feature set obtained from the histogram. Histogram provides a set of features for proposed for Content Based Image Retrieval (CBIR). Hence histogram method further refines the histogram by splitting the pixels in a given bucket into several classes. Here is computed the similarity for 8 bins and similarity for 16 bins. Standard histograms, because of their efficiency and insensitivity to small changes, are widely used for content based image retrieval. But the main disadvantage of histograms is that many images of different appearances can have similar histograms because histograms provide coarse characterization of an image [12].

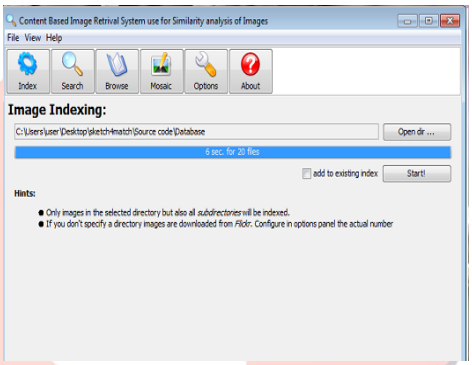

Fig.3 Process of Image Indexing

4.3 Searching

This module performs intensive searching of images from the database. User gives his query image and based on various algorithms a set of images are generated. The query is taken up and according to the algorithm given a set of related images is generated. This module searches for the images in the database according to the specified algorithm. To generate more relevant images, the number of search images can be decreased. Now if the user is not satisfied by the images generated, he/she can perform the search test again and again, until it generates. After indexing database, if we want to search images in the database. We have to provide external input image. With this module we provide the input image for comparing it in the database. For obtaining desired images, system apply the various algorithm on the input image and it compare with the images in the database [12].

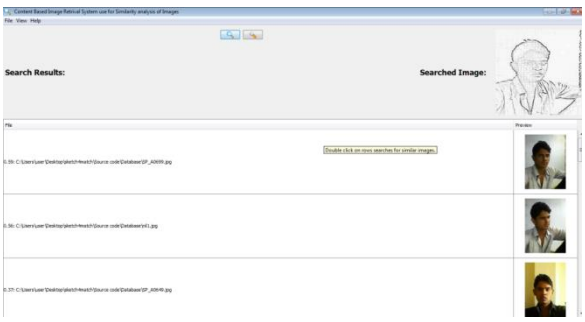

Fig 4 Searching process of Images

4.4 Browse panel

In the browse module, browse the number of objects, among which one is used to show the total number of indexed image files. The other thing is object is used to view the indexed image files one after other in the jTextField object. The image obtained in the jTextField object can be searched directly instead of using search panel [12].

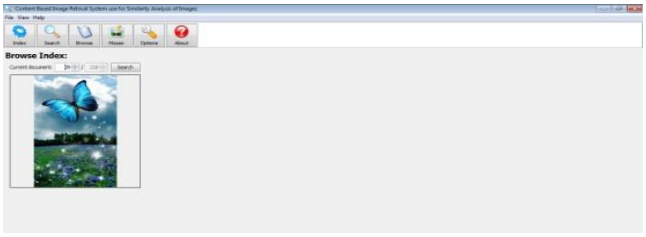

Fig. 5 Browse Image i.e. no. of images it has considered

4.5 Option

This module provide the various ways to modifying the indexing options, while indexing there are various parameters provided for modifying indexing database.

For indexing or importing database we have to choose the type of index searcher, type of index searcher it is one of the various parameter for indexing [8].

Type of Index Searcher:

It provides the various algorithms by which we can import the database images. Like algorithms are Edge Histogram, CEDD, FCTH, Auto color Correlogram and All MPEG-7 Descriptor.

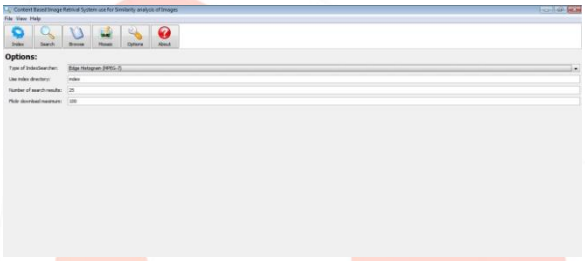

Fig. 6 Option Module

I) EHD Algorithm

In EHD method, find the edges in image and these edges are used to store the feature value. To restrict the edge distribution to a certain area of the image, this image space is divided into 4 x 4 sub – images. An edge histogram generated for each sub-image to characterize sub-image edge distribution. [9].The sub-image is more divided into small square blocks called image-blocks. For classify different edge types [9].

II) Auto Color Correlogram (ACC)

It deals with color distribution of pixels and spatial correlation of pairs of colors. Let I be an image that comprises of pixels f(i, j). Each pixel has certain color or gray level. Let [G] be a set of G levels g1, g2, .. .g. G that can occur in the image. For a pixel f let I(f) denote its level g, and let Ig correspond to a pixel f, for which I(f) = g. Histogram for level gx is defined as [4, 5]:

$$hgx (I) \equiv Pf \in I \, |f \in Igi|$$

Auto correlogram captures the spatial correlation of identical levels [13, 14].

III) All MPEG-7 Descriptor

MPEG-7 is also called as Multimedia Content Description Interface. It is a new standard for describing the content of multimedia data [ISO00, MPE99, MPE00, and MPE01]. MPEG-7 is a means that the attach metadata to multimedia content. It specifies a standard set of depiction tools, which can be used to describe various types of multimedia information. These tools shall be linked with the content itself to allow efficient and effective searching for multimedia material of users interests. MPEG-7 is a basic standard with broader application area [13, 14].

IV) Color Layout Descriptor

The CLD is a very compact and resolution invariant representation of color for high speed image retrieval [7]. It captures the spatial layout of the representative colors on a grid superimposed on a region or image based on the Discrete Cosine Transform (DCT). It is expressed in the YCbCr color space. The size of the array is fixed to 8x8 elements to ensure scale invariance. It is then transformed using DCT followed by zig-zag re-ordering. This is shown in the figure given below. It can be used for fast searching of databases as well as filtering in broadcasting applications. Another application is description of video clips, where CLD is combined with the Time Series structure [13, 14].

V) Scalable color Descriptor

SCD is derive from a color histogram define in the HSV color with fixed color space quantization. It uses a Hear transform coefficient encoding allowing scalable representation of description as well as complexity scalability of feature extraction & matching procedure. It is scalable because precision of histogram bin value can vary 16 to1000 bits per bin for different requirements.

4.6 Mosaic module

In the Mosaic module, arrange the number of images or photograph in the database to form a composite picture. This module shows the number of images at a time in one page. This is used to large database. In mosaic creation to use the titles for how many titles of images show in one row at a time.

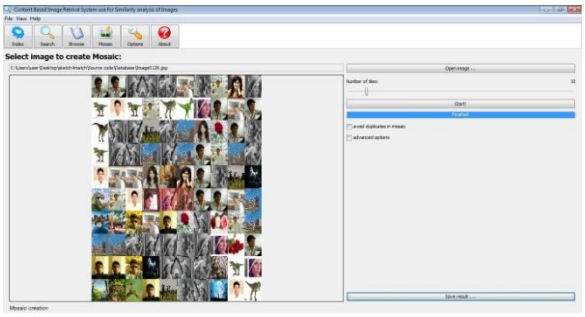

Fig. 7 Mosaic Creation

4.7 Image Database

There are many image databases are available for image processing purpose such as flicker, Google, Wang Database etc. These databases are specially used for image processing fields, psychological studies etc. These databases contain simple images with decent background so as to focus on main object. All the images in these databases are of various objects which have basic simple shape [15]. Some of the database images are shown in the Fig. 8.

Fig. 8 some of the database images

V. IMPLEMENTATION AND EXPECTED RESULT

The Image Retrieval system is the main technique for Precision and Recall which represents the main way to evaluate and measure the performance [16].

To evaluate the effectiveness of the system forming method and comparing different applied methods. This compression can be done easily through Metrics. To evaluate effectiveness & accuracy of the system, Precession & recall rate to be calculated.

Precision (A) = No. of images displayed with similar Shape (X) / No. of images Displayed (Y)

Recall (B) = No. of images displayed with similar shape (X) / no of images with similar shape in whole database (Z).

Calculate or check the accuracy using the below formula. The formula is given by,

Accuracy= (A+B)/ 2

where the precision gives information about the relative effectiveness of the system.
where the recall gives information about the absolute accuracy of the system[16].

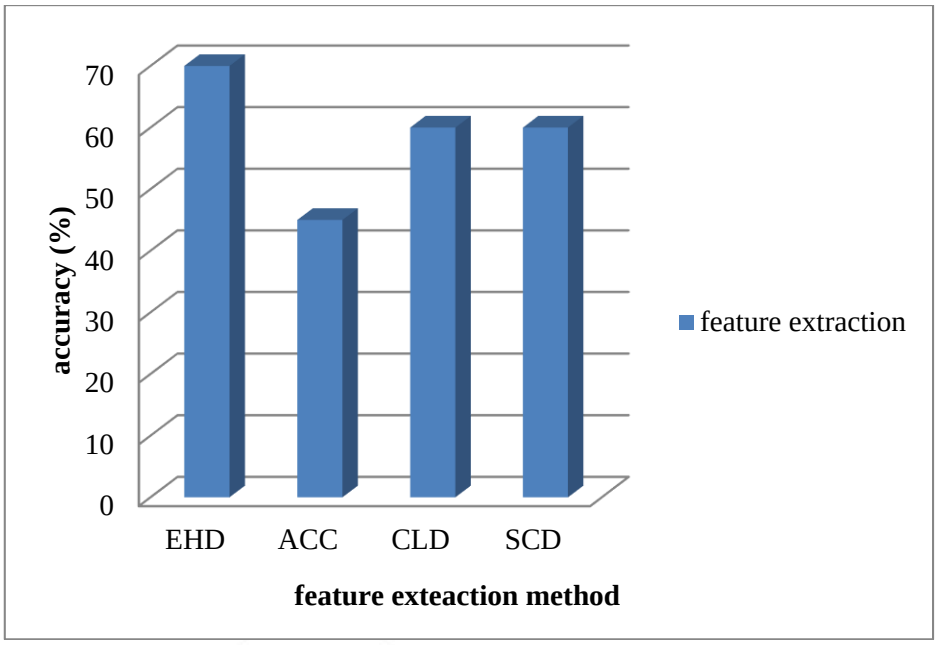

Fig. 9 comparison of different methods used for image retrieval

Comparison between Proposed method and Existing method

Graph 4.5 gives the persons visualizes results of Base paper (HOG+CLD) and proposed method. It is clear that for a feature extraction method is proposed method shows better results than Base paper (HOG+CLD). Base paper method shows accuracy of 54 % and proposed method shows visualized precision and recall accuracy of 60% from the EHD and SCD method. This graph shows the Comparison between proposed method and existing method. In proposed method, EHD and Scalable descriptor have more accuracy than the earlier methods. The color and shape descriptor gives the better efficiency than the texture descriptor. In proposed method user can search color jpeg images as well as sketches with the help of CBIR method and also shape and color feature extraction methods.

The comparison between proposed method and existing method is shown in graph 4.5.

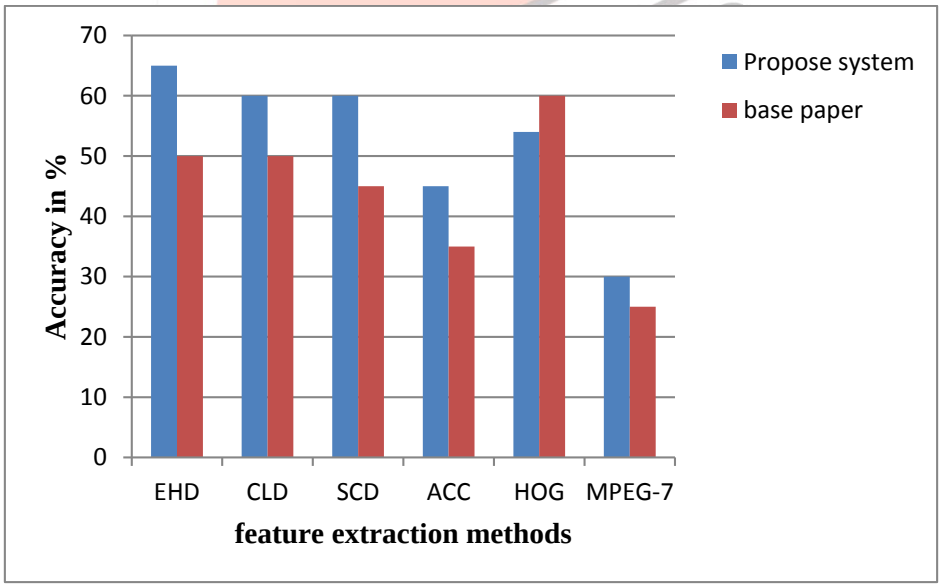

Fig.10 comparision between Proposed system and Base paper

VI. Conclusion

Among the objectives of this paper present to design, implement and test a sketch-based image retrieval system. In all cases of different descriptor algorithm was studied. These algorithms check the similarity of images in the database. Based on the test result with many databases EHD is better than other cases. EHD algorithm gives accuracy in Similarity of Images. The retrieval process has to be unconventional and highly interactive. The robustness of the technique is important in some scale of noise, which might also be in case of simple images. The drawn image without modifications cannot be compare with color

image, or its edge representation. The simple smooth and edge detection based method was improved. Compare the different methods of feature matching sketch which one is better and check the effectiveness of the methods and also create the mosaics of images.

VII. ACKNOWLEDGEMENT

I express my profound gratitude to Dr. P. S. Adwani, Principal, Government College of Engineering, Aurangabad. I express my gratitude to Prof. A. K. Gulve Associate Professor of the Department, Computer science & Engineering for their constant encouragement.

VIII. REFERENCES

[1] Dipalee J. Fendarkar, A.K. Gulve, "Utilizing effective way of sketches for Content Based Image Retrieval System ", International Journal of Computer Science, Volume 116, number 15.
[2] Surekha Karpe, Sneha Shree, Saurabh Kurkute, Abuzer Shah, Mrs. Archana Chaugule, "Sketch4Match: Content Based Image Recognition using Coloured Sketches", International Journal of Engineering Trends and Technology (IJETT)- Volume 9 Number 10- Mar 2014.
[3] S.R.Surya, G.Sasikala, "Survay on Content Based Image Retrieval" Indian Jounal of Computer Science and Engineering (IJCSE) ISSN: 0976-5166 Vol. 2 No. 5 Oct-Nov 2011
[4] Content-based image retrieval systems. IEEE Computer, 28(9), 1995.
[5] M. Flickner et al. Query by image and video content: The QBIC system. IEEE Computer, 28(9):23–32, 1995.
[6]Huang, Y. L. and R. F. Chang, Texture features for DCT-coded image retrieval and classification, Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.6, pp.3013-3016, 1999.
[7] C.Carson, M. Thomas, S.Belongie, J.Hellerstein, and J. Malik. Blobworld: A system for region based image indexing and retrieval. In Third International Conference on Visual Information Systems. SpringerVerlag, June 1999.
[8] Prajakta Mahendra Ovhah, Prof. B.D. Phulpagar, "A Survey On Various Feature Extraction Techniques Followed By Image Retrieval System", International Journal of Emerging Trends and Technology in Computer Science (IJETTCS) , Volume 3, Issue 5, Sept-oct. 2014 – ISSN 2278-6856.
[9] Swapnil Khobragade, Shobhit Nikose, Mussavir Shaikh , "Content Based Image Retrieval System Use for Similarity Analysis of Images", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1701-1704**.**
[10] M. Karthikeyan, P. Aruna, "An Efficient Model for Content based Image Retrieval", International Journal of Computer Applications", (0975 – 8887) Volume 42– No.2, March 2012 21 .
[11] B. Sz´ant´o, P. Pozsegovics, Z. V´amossy, Sz.
Sergy´an,"Sketch4Match – Content-based Image Retrieval System Using Sketches", SAMI 2011 • 9thIEEE International Symposium on Applied Machine Intelligence and Informatics • January 27-29, 2011 Smolenice, Slovakia.
[12] S.Pavan , Shivani, T.Venkat Narayana Rao , Deva Shekar, " Similarity Analysis Of Images Using Content Based Image Retrieval System", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 1 Jan 2013 Page No. 251-258.
[13] Asmita A. Desai, "Review Paper on Sketch Based Image retrieval System", Proceedings of Sixth IRAJ International Conference, 6th October 2013, Pune, India. ISBN: 978-93-82702-32-0.
[14] Ala'a Al-zoubi, Mohammad Al-zoubi, "Automatic Mapping of MPEG-7 Description to relational Database", The International Arab Journal of Information Technology, volume 9 number 6, November 2012.
[15] V. Venkat Ram1, P. Srinivas, " Content Based Image Retrieval Sketches", International Journal of Emerging Technology and Advanced Engineering ,ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 9, September 2014.
[16] Shilpa Pawar, Sonali Tidke, "Survey on Sketch Based Image Retrieval System", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 8, August 2014.