Image and annotation retrieval via image contents and tags

Prof D.D. Pukale, Miss. Laxmi P. Dabade, Miss. Varsha B. Patil, Miss. Chandani P.Lodha, Miss. Nikita S. Ghode

Department of Computer engineering, Bharati Vidyapeeth's college of Engineering for Women, Pune, Maharashtra, INDIA

Abstract- At present, tags are extensively used to describe the images, so that utilizing these tags for image retrieval system is today's need. So, our system incorporates image similarity graph with image-tag bipartite graph using visual features of image like color, shape & texture .For color feature extraction HSV model, for shape feature extraction Sobel with mean and median filter, for texture feature extraction Framelet methods are used. Initially we extract features by these three methods and all features are combined for matching between images in the database and query image. Combination of the features of the three methods along with CBIR and TBIR are used to balance the influence between image content tags. The system is able to retrieve the images related to the query as well as annotating the query image.

Index Terms- Content-based image retrieval, image annotation, text-based image retrieval.

I. INTRODUCTION

In this paper we develop a user friendly application in which a user can easily and quickly retrieve the image that he wants to retrieve. In this paper following two level data fusions are used to retrieve the image and annotate the query image based on image contents and tags:

1) A unified graph is built to fuse the visual feature-based image similarity graph with the image tag bipartite graph.

2) A cbir along with tbir is used to utilize a fusion parameter to balance the influence between the image contents and tags.

Also automated image annotation technique is used to make huge unlabeled digital photos indexable by existing text based indexing and search solutions. An image annotation task consists to assign a set of semantic tags or labels to a novel image based on some models Learned from certain training data.

To take advantages of both the visual information and user-contributed tags for image retrieval, in this system we are going to incorporate both image content and tag information in to image Retrieval and annotation tasks. This system is useful to all users who search for respective image i.e. in biomedical, police department.

Another approach to the semantic gap issue is to take advantage of the advance in computer vision domain, which is closely related to object recognition and image analysis. Duygulu *et al.*[12] present a machine translation model which maps the keyword annotation onto the discrete vocabulary of clustered image segmentations. Moreover, Blei and Jordan extend this approach through employing a mixture of latent factors to generate keywords and blob features. Jeon *et al.* [13] reformulate

the problem as cross-lingual information retrieval, and propose a cross-media relevance model to the image annotation task. In most recent, bag-of-words representation [7] of the local feature descriptors demonstrated promising performance in calculating the image similarity. To deal with the high-dimensionality of the vector feature space, the efficient hashing index methods have been investigated in [7] and [12]. These approaches did not take consideration of the tag information which is very important for the image retrieval task. Most recently, Jing and Baluja [14] present an intuitive graph model-based method for product image search. They directly view images as documents and their similarities as probabilistic 464 IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 12, NO. 5, AUGUST 2010 visual link. Moreover, the likelihood of images is estimated by a similarity matching function on the image similarity graph.

However, the image-tag [8] and video-view graphs [4] based approaches did not take consideration of the contents of images or videos, which lose the opportunity to retrieve more accurate results. In [13], a re-ranking scheme is developed using similarity matching over the video story graph. Multiple-instance learning can also take advantage of the graph-based representation [13] in the image annotation task. Apart from its connection with research work in content based image retrieval, our work is also related to the broad research topic in graph-based methods. Graph-based methods are intensively studied with the aim of reducing the gap of the visual features and semantic concept. In [10], the images are represented by the attributed relational graphs, in which each node in the graph represents an image region and each edge represents a relation between two regions. An image is represented as a sequence of feature-vectors characterizing low-level visual features, and is modeled as if it was stochastically generated by a hidden Markov model, whose states represent concepts.
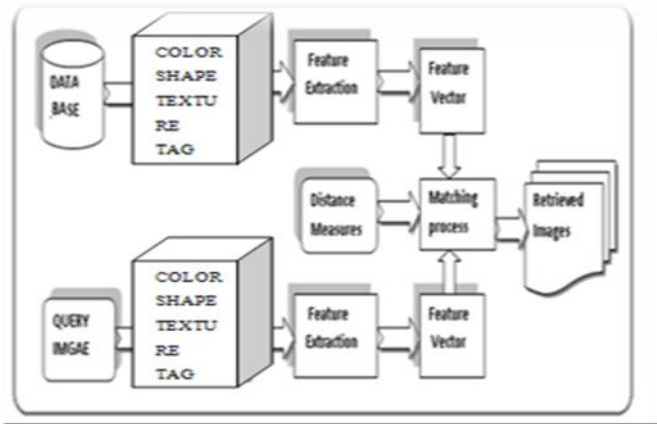
Fig1.System architecture design

Fig 2. HSV model

II. PROPOSED SYSTEM

In this paper following two level data fusions are used to bridging the gap between the image contents and tags:

• A unified graph is built to fuse the visual feature-based image similarity graph with the image tag bipartite graph.

• Along with query image we have provided tag as input and that is built to utilize a fusion parameter to balance the influence between the image contents and tags. In this paper automated image annotation technique is used to make huge unlabeled digital. Photos index able by existing text based indexing and search solutions. An image annotation task consists to assign a set of semantic tags or labels to an image. Our database consists of multiple folders of various types of images. Each folder consists of a particular set of similar images, for e.g. roses, horses, airplanes etc. We are assigning common tags to only one image in a set. Then with the help of similarity matching function, the other images in the same set will be annotated automatically. The database is dynamic, i.e., could be updated at runtime.

Due to text based indexing mechanism, whenever a query is fired, image as well tag is given as input, due to which it will directly search in only that specific folder itself. And thus, the results are prioritized accordingly.

In case of unmatched image-tag combination, it will search the complete database till a match is found for image.

III. FEATURE EXTRACTION

A)COLOR FEATURE EXTRACTION METHODS:

1)Grid Color Image : It extract the color features of the image, in the traditional way they uses the RGB, as compared to RGB, HSV has better performance, so we use HSV for the Grid Color moment feature extraction.[5]
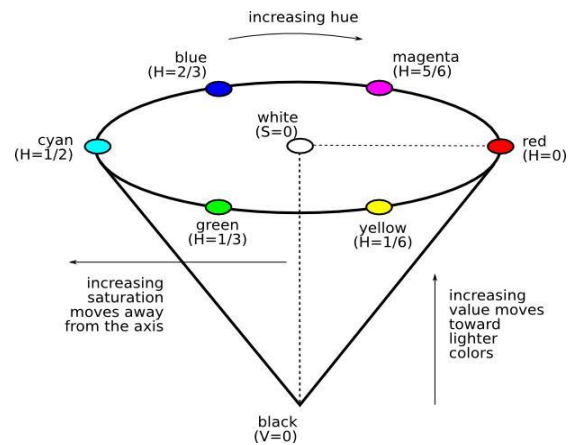
We evaluate the content based image retrieval HSV color space of the images in the database. HSV stands for the Hue, Saturation and Value, provides the perception representation according with human visual feature. The HSV model, defines a color space in terms of three constituent components: Hue, the color type Range from 0 to 360. Saturation, the "vibrancy" of the color: ranges from 0 to 100%, and occasionally is called the "purity". Value, the brightness of the color: Ranges from 0 to 100%. HSV is cylindrical geometries, with hue, their angular dimension, starting at the red primary at 0°, passing through the green primary at 120° and the blue primary at 240°, and then back to red at 360° [10, 7]. The HSV planes are shown as Figure 1. The different planes of HSV color space, the quantization of the number of colors into several bins is done in order to decrease the number of colors used in image retrieval, and J.R. Smith [8] designs the scheme to quantize the color space into 166 colors. Li design the non-uniform scheme to quantize into 72 colors. We propose the scheme to produce 15 non-uniform colors. The formula that transfers from RGB to HSV is defined as below:

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{((R-G)2) + (R-B)(G-B)}} \right\}$$

$$S = 1 - (3/R + G + B)[\min(R, G, B)]$$

$$V = 1/3(R + G + B)$$

The R, G, B represent red, green and blue components respectively with value between 0-255. In order to obtain the value of H from 0 to 360, the value of S and V from 0 to 1, we do execute the following formula:

$$H = ((H/255 * 360)) \bmod 360$$

$$V = V/255$$

$$S = S/255$$

The various steps to retrieve images are given below:

Step 1: Take an image from the available database.
Step 2: Resize the image for m [256. 256].
Step 3: Convert the RGB color space image to HSV by using above given formula.
Step 4: Generate the histogram of hue, saturation and value
Step 5: Quantize the values generated into number of bins.

Step6: Store the quantized values of database images into a file.
Step 7: Load the Query image given by the user.
Step 8: Apply the procedure 2-6 to find quantized HSV values of Query image.
Step 9: Sort the distance values to perform indexing.
Step 10: Display the retrieved results on the user interface.

B) SHAPE FEATURE EXTRACTION
SOBEL WITH MEAN AND MEDIAN FILTER

Sobel has two main advantages compared to other edge detection operators: Sobel has some smoothing effect to the random noise of the image since the introduction of average factor. Because it is the differential of two rows or two columns, the elements of the edge on both sides has been enhanced, so that the edge seems thick and bright. Edge detection is usually carried out by use of local operator in airspace. What is usually used are orthogonal gradient operator, directional differential operator and some other operators relevant to second-order differential operator. Sobel operator is a kind of orthogonal gradient operator. Gradient corresponds to first derivative, and gradient operator is a derivative operator. For a continuous function f(x, y), in the position (x, y), its gradient can be expressed as a vector (the two components are two first derivatives which are along the X and Y direction respectively):

Method	Advantages	Disadvantages
Sobel, Prewitt	Detection of edges and their Orientations	Inaccurate
Laplacian of Guassian (LoG)	Finding the Correct places of the Edges	Corners and curves where the gray Level intensity function varies
Canny	Using probability for finding error rate	Complex Computations and False Zero Crossing

To overcome this disadvantage of sobel, we use mean or median filter with sobel method to de-noising the noise effect. So we can tackle this problem of noising with the help of mean and median result, to get better results.

Sobel filtering is a three step process. Two 3× 3 filters (often called kernels) are applied separately and independently.

The weights these kernels apply to pixels in the 3 × 3 region are depicted below:

-1	0	1
-2	0	2
-1	0	1

1	2	1
0	0	0

-1	-2	-1

Again, notice that in both cases, the sum of the weights is 0. The idea behind these two filters is to approximate the derivatives in x and y, respectively. Call the results of these two filters Dx (x, y) and Dy (x, y). Both Dx and Dy can have positive or negative values, so you need to add 0.5 so that a value of 0 corresponds to a middle gray in order to avoid clamping (to [0..1]) of these intermediate results.

The final step in the Sobel filter +approximates the gradient magnitude based on the partial derivatives (Dx (x, y)) and Dy (x, y)) from the previous steps. The gradient magnitude, which is the result of the Sobel Filter S(x, y), is simply:

$$S(x, y) = \sqrt{((Dx\ (x,\ y))2 + (Dy\ (x,\ y))2)}$$

So, in summary, the three steps are:
1) Compute the image storing partial derivatives in x (Dx (x, y)) by applying the left 3 × 3 kernel to the original input image.
2) Compute the image storing partial derivatives in y (Dy (x, y)) by applying the left 3 × 3 kernel to the original input image
3) Compute the gradient magnitude S(x, y) based on Dx and Dy.

Two further things to notice about Sobel filters: (a) both the derivative kernels depicted above are separable, so they could be split into disjoint x and y passes, and (b) the entire filter can actually be implemented in a single-pass GLSL filter in a relatively straightforward manner.

C) TEXTURE FEATURE EXTRACTION
THE PROPOSED ALGORITHM USING FRAMELET TRANSFORM

The basic steps involved in the proposed CBIR system as follows[11].
1) Feature vector (f) Decompose each image in Framelet Transform Domain.
2) Calculate the Energy, mean and standard deviation of the Framelet transform Decomposed image.

$$Energy = 1/M{\times}N \sum Nj{=}1\sum Mi{=}1(\ |Wk\ i,j\ |)$$

Standard Deviation $(\sigma k) = \sqrt{1/(M{\times}N)} \sum Nj{=}1\sum Mi{=}1\ (Wk\ (i,j){-}\mu k)^2 \mu k$

- Mean value of the kth Framelet transform sub band co-efficient of kth Framelet transform sub band. $M{\times}N$ is the size of the decomposed sub band.
3) The resulting $f = [\sigma1, \sigma2, …., \sigma n, E1, E2……En]$ is used to create the feature database.
4) Apply the query image and calculate the feature vector as given in step (2) & (3).
5) Calculate the similarity measure.
6) Retrieve all relevant images to query image

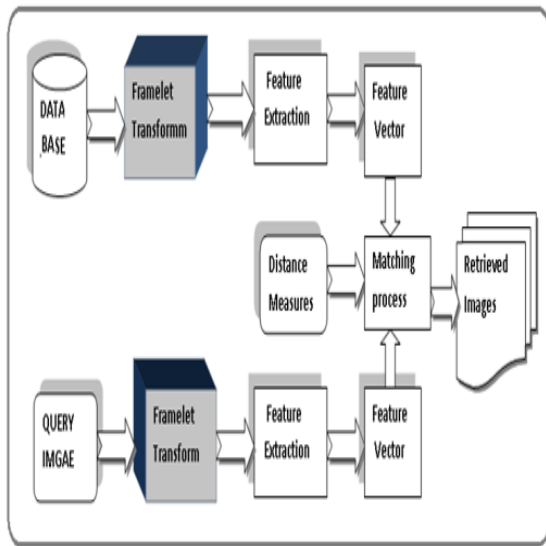## FLOW OF ALGORITHM USING FRAMELET TRANSFORM

Fig 3.Framelet Transform Flow diagram

D) IMAGE ANNOTATION

Automated image annotation has been an active and challenging research topic in computer vision and pattern recognition for years. Automated image annotation is essential to make huge unlabeled digital photos indexable by existing text based indexing and search solutions. In general, an image annotation task consists to assign a set of semantic tags or labels to a novel image based on some models learned from certain training data. Conventional image annotation approaches often attempt to detect semantic concepts with a collection of human-labeled training images. Due to the long-standing challenge of object recognition, such approaches, though working reasonably well for small-sized test beds, often perform poorly on large dataset in the real world. Besides, it is often expensive and time-consuming to collect the training data. In addition to the success in image retrieval, our framework also provides a natural, effective, and efficient solution for automated image annotation tasks. For every new image, we first extract a feature vector through the method described above in Section III-A,B,C. Then, we find the top- similar images using eq(1) , and link the new image to the top- images in the hybrid graph.Finally built image node, and return the top- tags as the annotations to this image.

E) SIMILARITY MATCHING

For every set of similar images in the database, only single images are be assigned tags. The remaining images of the set are be annotated automatically using similarity matching method. Thus, we don't have to manually assign tags to every image.

Image with Tag: is build to utilize a fusion parameter to balance the influence between the image contents and tags and finally retrieve the images rank wise as well as retrieve the annotations for the query image.

Hybrid graph construction

$$Sim(dp, dq) \ = \ \frac{dp \cdot dq}{||dp|| \, ||dq||} \qquad (1)$$

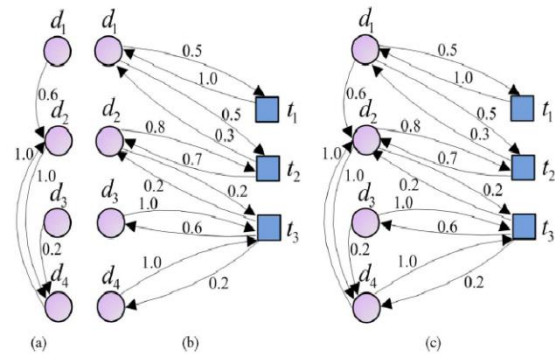Where, dp and dq represent image feature vector of corresponding image dp, and dq

Fig 5.Hybrid graph

Hybrid Graph : On the basis of the feature extracted from the images we are match the similarity between the images by using hybrid graph, for that first we build the image to image graph from extracted features & image to tag graph from the database and by combining both we are generate the bipartite graph to match the similarity.

We can then apply our framework to several application areas, including the following.

1) Image-to-image retrieval. Given an image, find relevant images based on visual information and tags. The relevant documents should be ranked highly regardless of whether they are adjacent to the original image in the hybrid graph.

2) Image-to-tag suggestion. This is also called image annotation. Given an image, find related tags that have semantic relations to the contents of this image.

3) Tag-to-image retrieval. Given a tag, find a ranked list of images related to this tag. This is more like the text-based image retrieval.

4) Tag-to-tag suggestion. Given a tag, suggest some other relevant tags to this tag. This is also known as tag recommendation problem.

PRECISION = Number of Relevant images retrieved /Total Number of images retrieved.

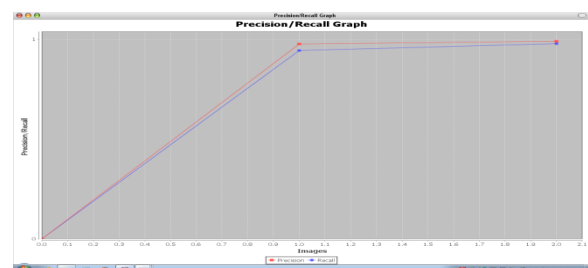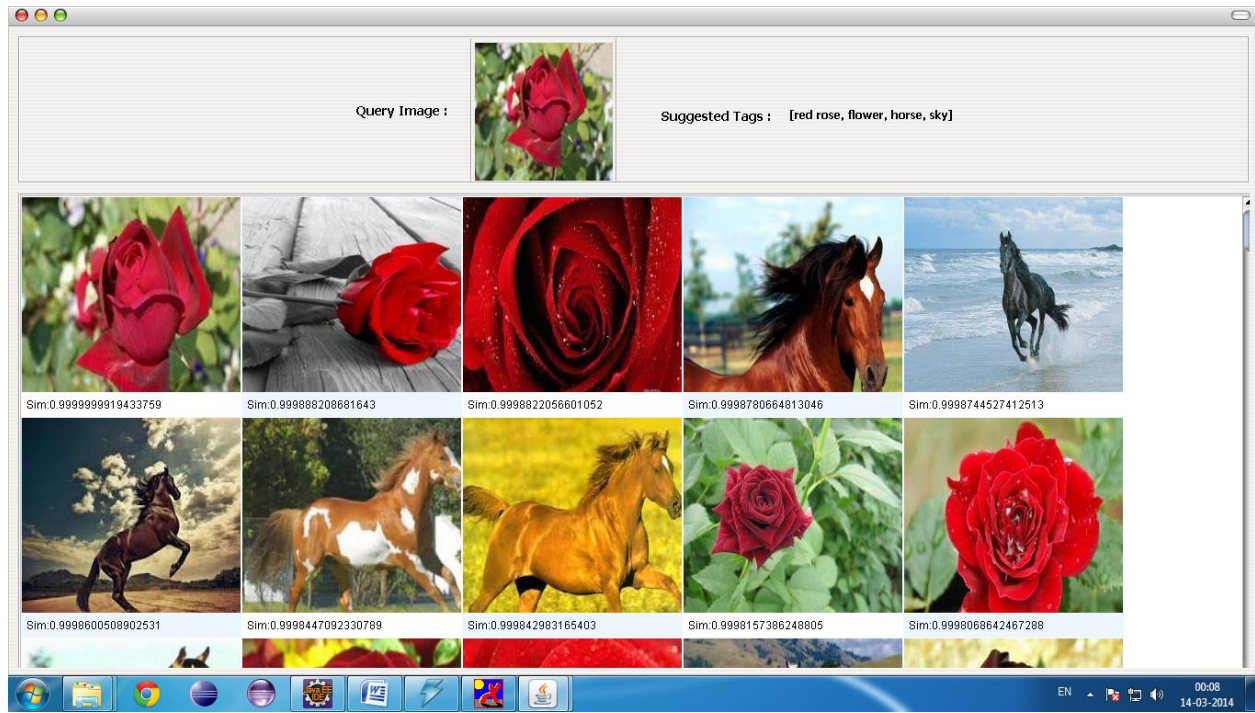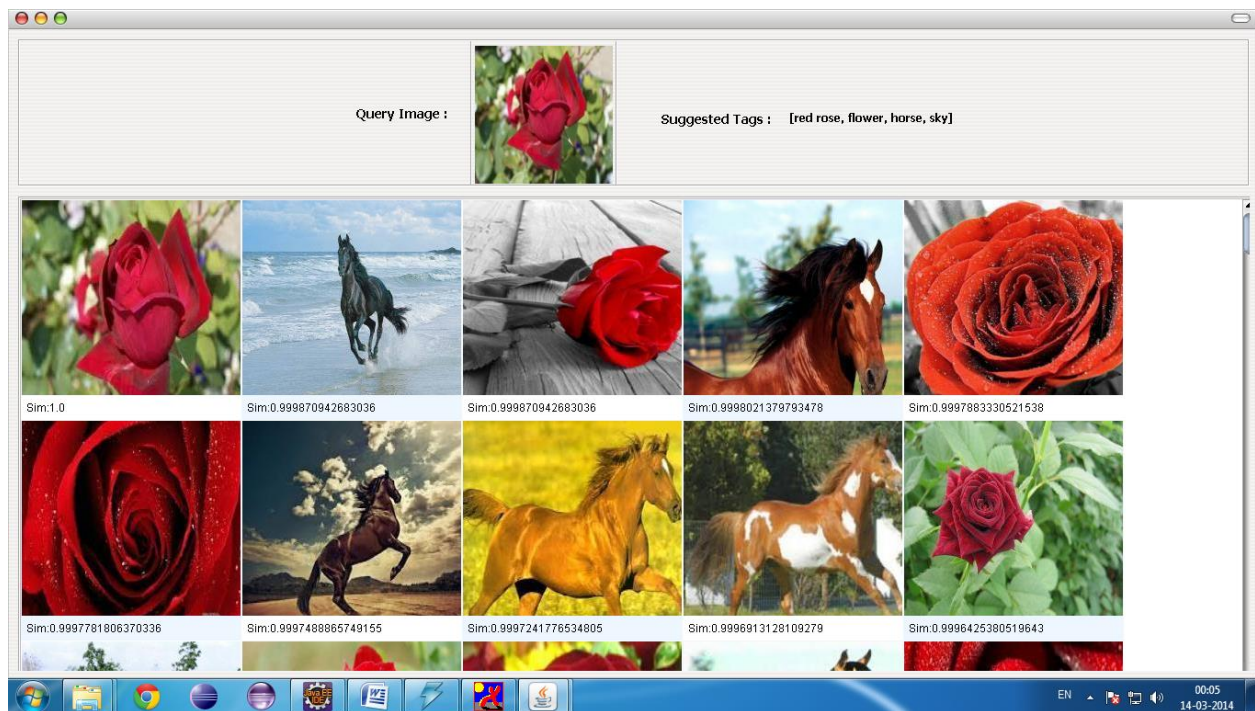**RECALL**= Number of Relevant images Retrieved/Number of relevant images in the database.

Fig 6. Precision Recall graph
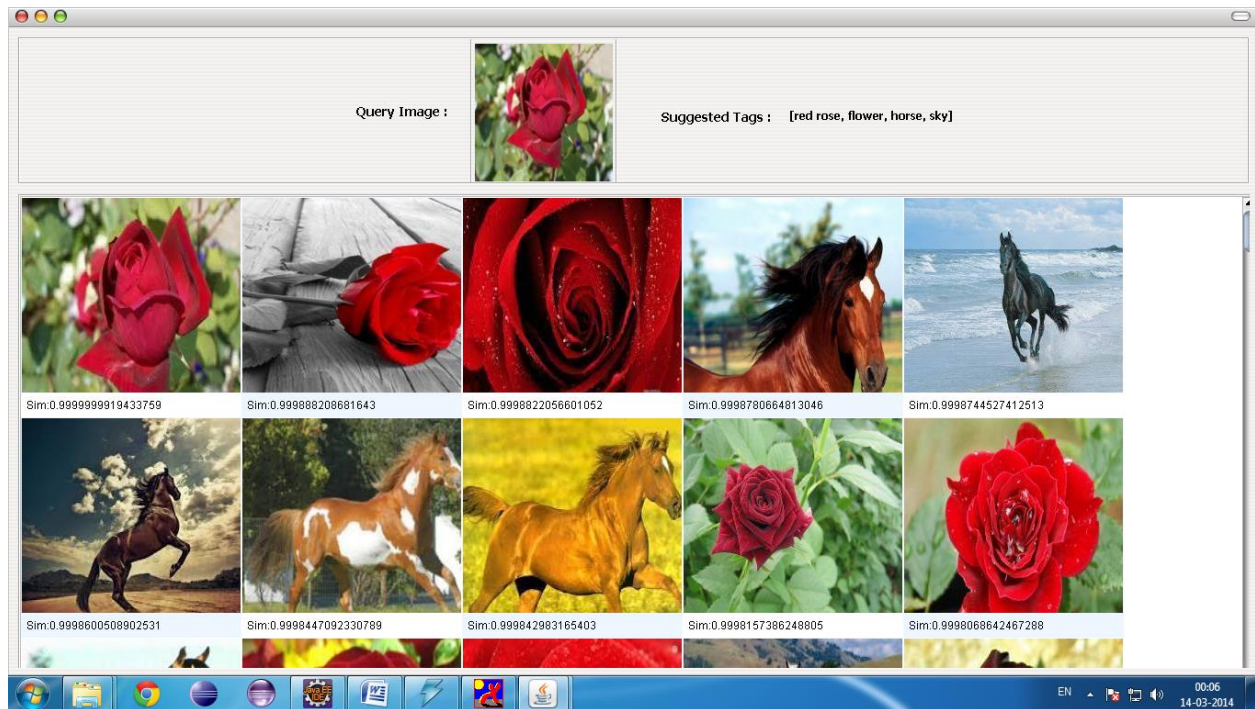
IV.   IMPLEMENTATION RESULT
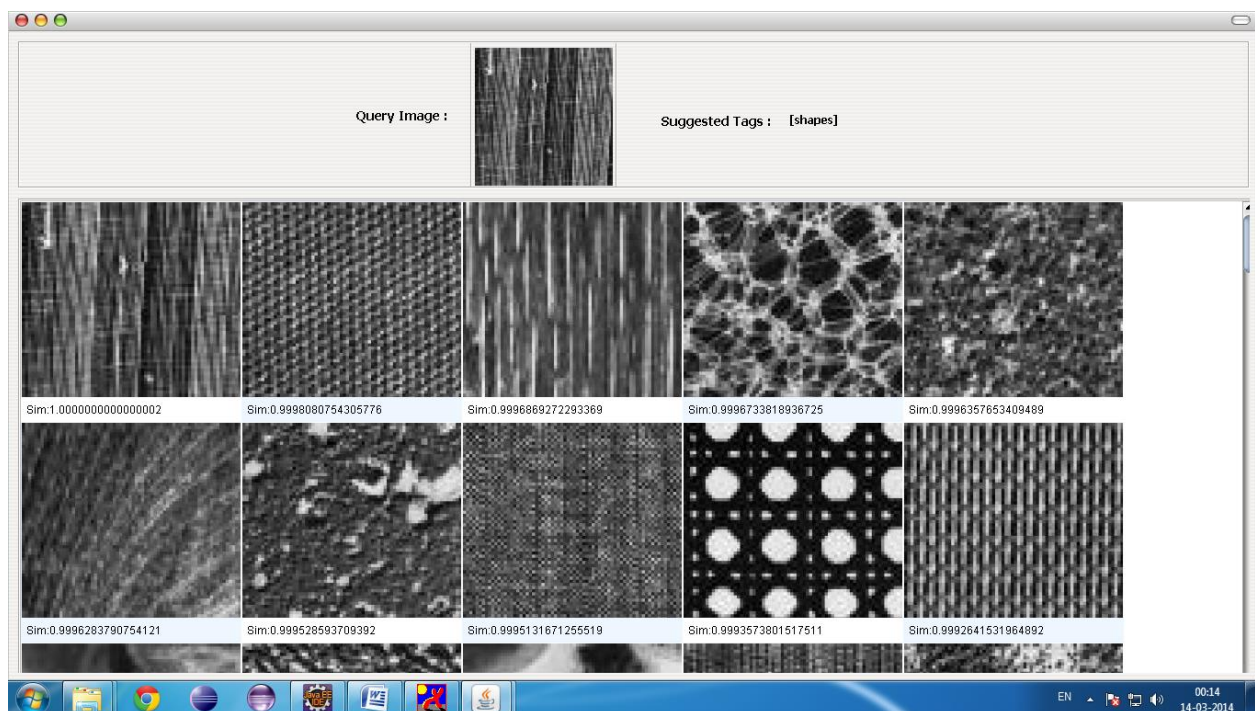
Search By HSV Method for color Feature

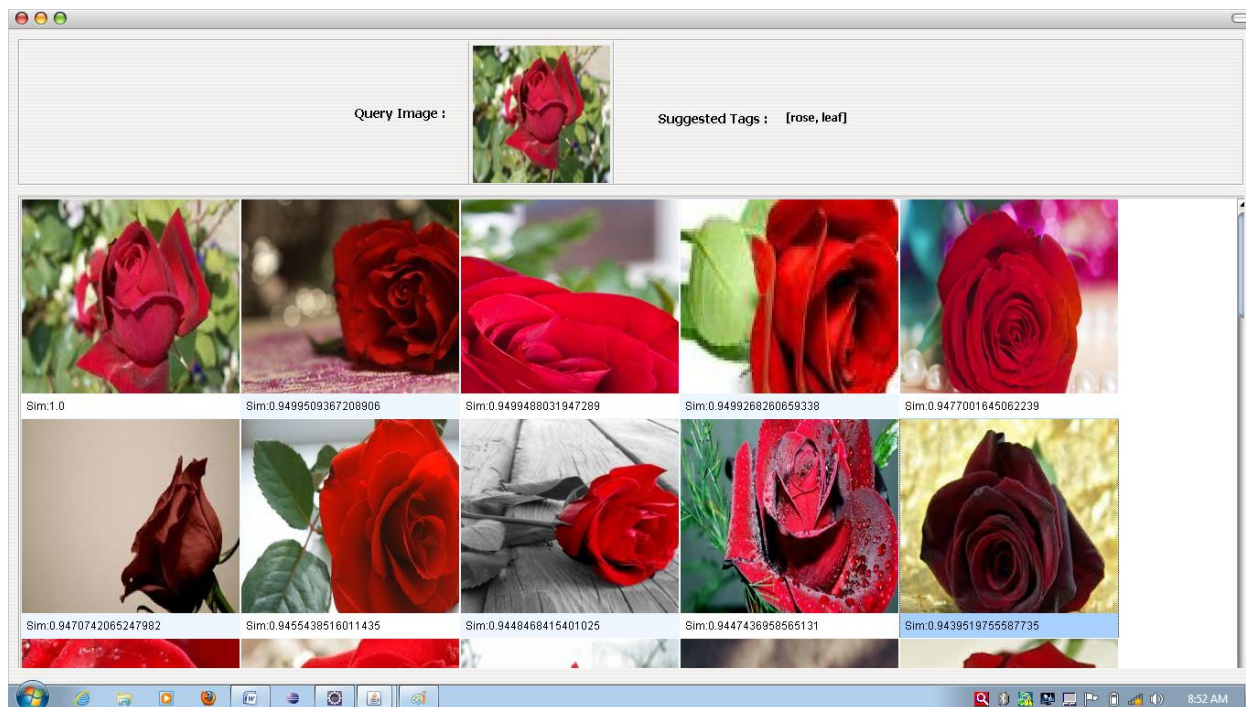### Search by Modified sobel for shape Feature

Search by modified sobel and HSV for shape and color Feature

Search By Framelet for Texture Feature

Search By All Methods Feature

Analytical Table: Pr=Precesion Re=Recall

Feature Extraction Method	No. of images	Color (HSV)%		Shape (Sobel with mean and median filter)%		Texture (Framelet)%		Color+Shape (HSV+Modified sobel)%		Color+Shape+Texture(HSV+Modified sobel +Framelet)%	
Images		Pr	Re	Pr	Re	Pr	Re	Pr	Re	Pr	Re
Horse	90	94.12	93.79	98.506	98.32	87.90	86.11	97.85	96.87	99.28	98.50
Rose	100	96.35	95.71	99.54	98.26	89.340	88.82	99.95	99.94	99.19	98.00
Aeroplane	100	93.70	92.17	94.23	93.69	88.60	87.45	96.13	95.45	98.33	97.21
Bortratz	108(108*8)	100	95.481	88.27	87.32	100	95.58	91.12	90.23	95.23	94.09
Fish	70	93.80	92.36	85.56	84.29	86.56	85.44	95.54	94.68	97.86	96.92
Apple	100	91.60	90.491	87.24	86.55	87.65	85.96	91.25	91.05	96.98	94.90
TOTAL	1324	94.928	93.33	92.22	91.405	90.00	88.22	95.306	94.703	97.81	96.60

V.   CONCLUSION

In this paper, we present a novel framework for one thousand image retrieval tasks. The proposed frameworks retrieve images with their annotation. Our method can be easily adapted to very large datasets. For every set of similar images in the database, only single images are be assigned tags. The remaining images of the set are be annotated automatically using similarity matching method. Thus, we don't have to manually assign tags to every image. This helps to achieve performance and retrieval efficiency and thus decreases time complexity.

REFERENCES

[1] The International journal of multimedia and its application Vol 2 "content based image retrieval using exact legendre moments and support vector machine" ,No.2,May 2010

[2] Areiam I. Grosky, "Imtrieval - ExistingTechniques, Content-Based (CBIR) System"s.Department of Computer and Information Science, University of

Michigan-Dearborn, Dearborn,MI,USA,http://encyclopedia.jrank.org/articles/pages/6763/Image Retrieval.html#ixzz0l30drFVs, (referred on 9 March 2010)

[3] Lei Wu Member, IEEE Transactions on Pattern Analysis and machine intelligence "Tag Completion for Image Retrieval". VOL. XX, NO. XX, IEEE, Rong Jin, Anil K. Jain, Fellow, IEEE JANUARY 2011

[4] V.N.Gudivada and V.V.Raghavan.: "Special issue on content-based image retrieval systems - guest eds". IEEE Computer. 28(9) (1995) 18{22ISSN: 2278 – 8875 International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 1, Issue 5, November 2012

[5] Council for Innovative Research International Journal of Computers & Technology " Content Based Image Retrieval using Texture, Color and Shape for Image Analysis" www.ijctonline.com ISSN: 2277-3061 Volume 3, No. 1, AUG, 2012

[6] "A Review on Image Feature Extraction and Representation Techniques". International Journal of Multimedia and Ubiquitous Engineering Vol. 8, No. 4, July, 2013.

[7] O. Chum, M. Perdoch, and J. Matas, "Geometric min-hashing: Finding a (thick) needle in a haystack," in Proc. CVPR'09, 2009, pp. 17–24.

[8] V.N.Gudivada and V.V.Raghavan.: "Special issue on content-based image retrieval systems - guest ed"s. IEEE Computer. 28(9) (1995) 18{22ISSN: 2278 – 8875 International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 1, Issue 5, November 2012

[9] Ch.Srinivasa Rao1 , S.Srinivas Kumar2 and B.Chandra Mohan Department of ECE , Sri Sai Aditya Institute of Science & Technology, Surampale "Texture Based Image Retrieval Using Framelet Transform–Gray Level Co-occurrence Matrix(GLCM)" (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 2, 2013

[10] Hao Ma, Jianke Zhu, Member, IEEE, Michael Rung-Tsong Lyu, Fellow, IEEE, and Irwin King, Senior Member, IEEE "Bridging the Semantic Gap Between Image Contents and Tags" IEEE transactions on multimedia, vol. 12, no. 5, august 2010

[11] J. Tang, H. Li, G.-J. Qi, and T.-S. Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," IEEE Trans. Multimedia, vol. 12, no. 2, pp. 131–141, Feb.2010.

[12] Y.-H. Kuo, K.-T. Chen, C.-H. Chiang, and W. H. Hsu, "Query expansion for hash-based image object retrieval," in Proc. MM'09, Beijing, China, 2009, pp. 65–74.

[13] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," in Proc. SIGIR'03,Toronto, ON, Canada, 2003, pp. 119–126.

[14] Y. Jing and S. Baluja, "PageRank for product image search," in Proc.WWW'08, Beijing, China, 2008, pp. 307–316.

AUTHORS

First Author – Prof D.D. Pukale, Department of Computer engineering ,Bharati Vidyapeeth's college of Engineering for Women, Pune, Maharashtra, INDIA, Email: pukaledd@yahoo.com

Second Author – Miss. Laxmi P. Dabade, Department of Computer engineering ,Bharati Vidyapeeth's college of Engineering for Women, Pune, Maharashtra, INDIA

Third Author – Miss. Varsha B. Patil, Department of Computer engineering ,Bharati Vidyapeeth's college of Engineering for Women, Pune, Maharashtra, INDIA, Email: varshapatil2885@gmail.com

Fourth Author – Miss. Chandani P.Lodha, Department of Computer engineering ,Bharati Vidyapeeth's college of Engineering for Women, Pune, Maharashtra, INDIA

Fifth Author – Miss. Nikita S. Ghode, Department of Computer engineering ,Bharati Vidyapeeth's college of Engineering for Women, Pune, Maharashtra, INDIA, Email: niki_gh@rediffmail.com