Volume 5, Issue 5, May 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Image Object Retrieval Using Semantic Feature Discovery and Tags

Gaurav Patel, S. T. Patil
Computer Department, V.I.T., Pune University, Pune,
Maharashtra, India

Abstract— The exponential growth in the field of Web 2.0 applications and services, the tags which are used widely to describe the contents of the image spread over the Web. As there is a noisy and general nature found in human made tags, how to use these tags for retrieving tasks of an image is a trending research field. As the visual features that are low-level can provide valuable information, so they are used to make better the image retrieval process. But, it is a challenge to bridge the difference between image contents and tags semantics. The tremendous growth in images and videos worldwide with the service of efficient devices and the social networking services such as Facebook and Flickr. To retrieve particular images, user issues a general query to any search engine by either using an image or keywords. Therefore, the final result for retrieving images depends on the content of the image (e.g., low-level features) or the texts surrounding the image (e.g., tags or the descriptions) only. Such solution mostly suffers from the low recall rates because even if there is a slight differ in lights or the viewpoints or some vibrations the overall performance degrades marginally. Due to such degradation of performance at some level the better way to retrieve image can be acquired by using the active contours and level set method. In this paper, the active contours are usedfor retrieving images, which is basically the concept of medical imaging which requires very precise and accurate retrieval.The basic aim is to extract different image features which will enhance the searching. Basically, contours are extracted from the images and are stored in a database.The proposed works generates better results with less processing time compared to existing technique. The framework can be applied to different application, like image object retrieval, image search based on keywords and tag refinement. The precision and control procedure proves the proposed method shows better results compare to previous existing method.

Keywords—Image retrieval, image query, active contours, level set method, image annotation, tag refinement, affinity propagation.

I. INTRODUCTION

The continuous interest developing in the field of CBIR can be seen in all these years with the hope that the various problems might be finally solved. The major issue inCBIR is to find a group of certain image features for better indexing and to carry out similarity evaluation.The typical features that are usually used for image indexing are color, texture, shape and relationships between segmented objects. More users are every now and then helps to make a contribution to related tags by commenting on the photos for better photo management and social interaction [1]. Such user-provided valuable information allows for a better research opportunities for finding and understanding the different images spread over social media. Over vast image collections the retrieval techniques either content-based or keyword based is the key technique for better management of growing media. Many applications are developed like annotation by search [3] that helps in better accuracy and providing efficiency in content-based image retrieval. These days, the current image search engine's uses not only the texts around the images but also the content of the image to retrieve different images (e.g., Bing Google, Yahoo).

For content-based image retrieval systems, bag-of-words (BoW) model is used widely and is considered most effective [5]. Quantization of high dimensional features (local) into discrete visual words (VW's) is represented by BoW model. However, the issues which are related to the noisily quantized visual features which have low lightning conditions, occlusions surrounding the image is a major problem in large image databases. Thus, such methods many times suffer from low recall rate. Due to varying capturing conditions and with the large visual word vocabulary (e.g., large dataset with around 1 million word vocabulary), the features for the images that are target might have all different visual words., it is also hard to find these VWs by expansion of query or other different varying quantization methods because of the large differences in appearance (visual) between the query and the final objects.

The tags are more semantically relevant/same than visual features for keyword-based image retrieval in social sites. But it is difficult to retrieve each and every target images only by keywords because varied users might write non-specific keywords such as "destination" [3]. And we know the text description for images over the internet are randomly entered without any relevance, without any type or categorization. Hence, texts are not accurate and are wrong or ambiguous.

Considering the above challenges for both content based and keyword based image retrieval on social networking sites,we propose a framework which will consider visual information in an unsupervised manner. In particular we augment each image and find the contour of the particular image which will increase the performance of the existing framework. The earlier method tried to augment auxillary visual word, by propagating and selecting auxiliary visual words from every image which increases the complexity of the retrieval process thus we are using the active contour model with level set method to extract the contours of every image.

The main aim of content based image retrieval is to provide efficiency in indexing of image and image retrieval from a database without any human made assumption on various domains like textures, shapes. The proposed framework is applied to two tasks, image object retrieval and the later task is tag refinement. The semantic feature discovery framework is directly applied to extract auxillary visual words. The later stage is tag refinement where semantically related tags are augmented.

II. RELATED WORK

The earlier content-based image retrieval systems we can classify into two different categories as per the input query: pictorial query or textual query. Systems based on text query, the keywords and captions were used to characterize images. The text features provides powerful description as a query, if more specific textual descriptions are provided for images in a database. However, this manual annotation of description for every image is a time consuming task.

The visual query can be generated by many visual queries. The task of a query method should be to extract meaningful results by capturing the enough and valuable information from the user. The systems based on pictorial queries, generally uses a basic example of the desired image. Image features such as colour or texture are used which can be extracted easily to retrieve similar images based on the example.

The basic CBIR system has to perform two important tasks. The first task to achieve is a feature extraction, in which a bunch of features also called feature vector is extracted to correctly represent the actual content of each image in a certain database. A feature vector generated is of much smaller size than the actual original image. The second task to achieve is a similarity measurement (SM), the top "closest" images can be retrieved by computing the distance between each image in a database to the query image using their features.

The block diagram of general CBIR system is shown.

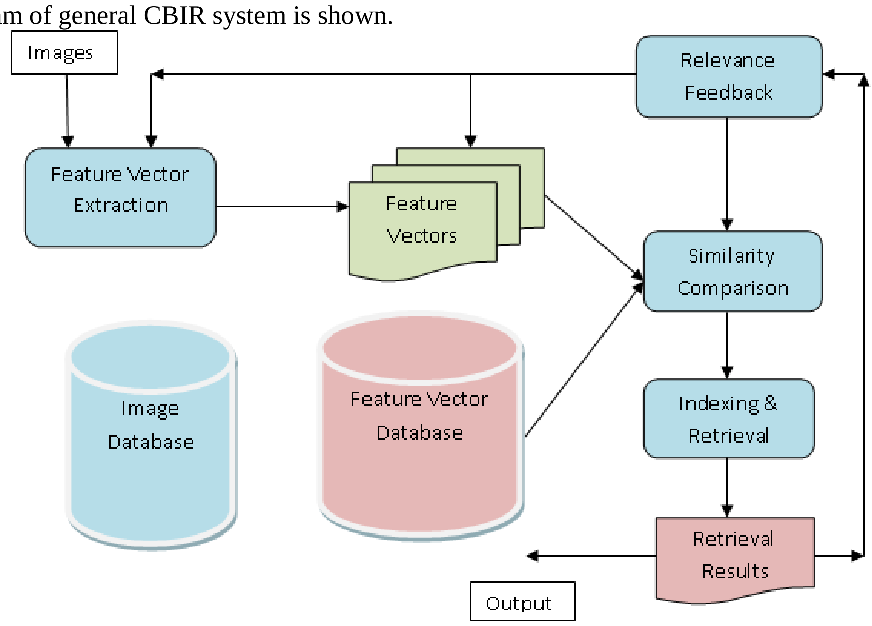

Fig.1A typical Content Based Image Retrieval System

Image retrieval based on content for image database is a challenging problem because the database is usually very large, the complexity of understanding the images by computers as well as humans, the complexity of query formation, and problem of proper evaluation of results. Various image search engines have been implemented and developed commercially, QBIC was the post earlier system. Recently, VIR[9] system have been developed in a commercial domain. MIT Photobook is the earlier systems that were developed in the academic domain. Moat well known systems are Webseek[8] and Visualseek. The basic for the content based image retrieval is extraction of various features from an image based on its pixel value and deciding/defining a way to compare images. Feature components supports better compression for image representation as compared to storing original pixel values.

Most of the existing CBIR systems come under three categories based on the method to extract feature components: color layout, histogram and region-based search technique. Many systems are also designed to combine the retrieval results by individual performing algorithms by a method called as weighted sum matching metric[12].

Semantic information of image was calculated where the information of an image was considered as a blob set in a Cross Media Relevance Model[4]. In CMRM the blob set were erected depending on discrete region clustering which extracts loss of features(vision) so the annotated results were accurate. The solution to this problem was proposed in[4] the Continuous Relevance Model (CRM).

Most of the retrieval systems adopts the feature transform descriptor SIFT (scale-invariant feature transform)[16] for capturing the local information and adobt the traditional BoW model[5] for object matching. The BoW model uses K-means clustering for the generation of visual words such as textual information[19], visual constraints[18]. But, this method needs manual information for the process of learning which may be a problem for larger image database.

Some researchers used original visual word vocabulary instead of newly generated visual word vocabulary[20]. The method selects useful features by comparing the neighbour images for better feature description.

Researchers in [11] adopted unsupervised learning method by considering both visual and textual information just like we did, but they considered just the global features and the method called random walk like was used for after processing in the process of image retrieval. As compared to earlier work, we used local features as well as contours and expanded it to visual as well as textual graphs for image object retrieval.

In image retrieval and several content-based image retrieval tasks BoW representation is widely used these days. Most of the existing systems apply the BoW model neglecting the sparseness of the visual word space. Visual words are the visual appearance and are not semantic descriptions thus lacks the retrieval of more varied results. The proposed method will target these issues.

III. PROPOSED METHOD

Based on the observations we discovered above, discovery of semantic feature for every image is necessary. Previous works constructed the features only in one particular single domain. We are proposing a general framework based on different modalities like contents of the images and tags. We proposed a standalone application for unsupervised semantic feature discovery. We are obtaining the auxillary visual words for each image and combining various visual features. Auxillary visual word can improve the efficiency in time and improve the recall rate. Proposed work comprises of different modules, each modules forwards towards getting the final output. The open source database is considered for the image retrieval which will retrieve desire images based on the image query. The system architecture for the unsupervised semantic feature discovery framework is shown below;

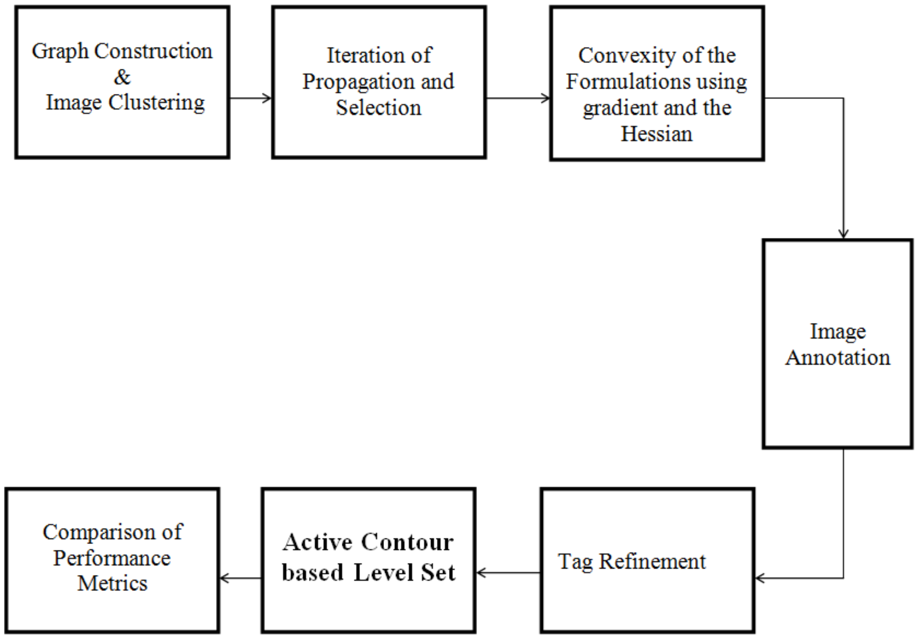

Fig. 2 System Architecture

Our work focuses on the extended feature extraction of contours which will further increase the retrieval accuracy and speed.

A. Graph Construction and Image Clustering

The visual words[1] which are nothing but any type of interesting information in any feature space are extracted by using the octave three level generation methodology. The methodology extracts the key points from an image by inter-blocking splitting of the image. After extraction of all the keypoints, maximum and minimum points are considered as our key point extracted values. Keypoints for objects in an image are extracted and are stored in a database. These maximum and minimum keypoints are visual words. Combination of visual features in the form of vectors in a matrixes and constructing graphs based on vectors.

*1) Auxillary Visual Word Propagation:*The BoW model has the limitation of considering only visual word thus KUO et al propagated auxillary visual word were each image is augmented with additional visual words considered from various visual as well as texturefeatures, by considering both visual and texture feature descriptions of every image, the retrieval is further benefitted.

After getting the auxillary visual words, graph is constructed and then, the images are clustered for applying the BoW model, k-means clustering is used to create the cluster of vectors based on the similarities basis of visual words.

Image clustering is carried out by using the k-means clustering algorithm in Matlab, for BoW model.Bow method is applied to these vectors by clustering images, by dividing these vector matrixes. We used k-means clustering for the division of vector matrixes. Images are clustered based on division of frontend and backend objects and division based on type of object. We applied the k-means by considering the 3-color based cluster and applying on the given image then reshaping the output clusters we obtained the pixel labels. This pixel labels are nothing but the color features of our image.

B. Iteration of Propagation and Selection

This module calculates the respective affine matrix of the query image. Affine function is applied to calculate the affine which generates the graph which is a straight line. An affine function shows an affine transformation that is same as compared to a linear transformation followed by a translation. The affine is calculated and the graph is generated. Affinity function can be directly applied to rows and columns for calculating the affine. The graph based on affinity propagation is constructed.

We calculated the affine to get the further affine output image. Affine matrix further minimizes the complexity of matrix created after clustering. Propagation and selection is calculated by converting the matrix into affine matrix.

For affinity we considered the square root of rows and columns with specific distances.We created the final minimised vector matrix with less complicated structure.

These operations mentioned above are performed iteratively to obtain relevant visual words which are semantic in nature which improves recall rate and by proper selection operation irrelevant visual words are removed which further improves efficiency and memory usage.

C. Convexity of the Formulations Using the Gradient and the Hessian

We converted the final minimized vector obtained after applying affine function into convex form. Convex form is nothing but the equation form for feature extracted vectors. We are converting it into the equation form for easy identification which makes the retrieval process simpler.

Gradient and Hessian is a general equation which is used in various applications like gender identification, face recognition, speaker identification for considerably limiting the execution process.

Affine function calculated is applied into Gradient and Hessian evaluation for error prediction. The noisy part of an image is identified and is optimized using Hessian evaluation [H,g,z]=hessiancsd(f,x); . The gradient and hessian evaluation removes the noisy part and converts the affine function into convex form.

Gradient method was used to find the local minima and local maxima by using the first derivative of image. The sudden sharp intensity also called as noise is handled by gradient based method. We used canny operator for edge detection and converting the pixel labels into binary form.

D. Tag Propagation and Tag Selection

We proposed of propagating the semantically relevant tags through its visual identical and similar images to obtain more semantic features. Traditional content based image retrieval systems many times retrieves photos with noisy descriptions, hence it is necessary to apply image annotation and tag refinement. These methods helps in improving the tag quality.

We proposed a semantic feature discovery framework, where we considered the image graphs to carry out the propagation and selection of auxillary visual words.

Image annotation is used to define the extracted values in a form of arrays called tags. These tags are different to each and every image but nearly equal to matched. Tag propagation defines the extracted values in a form of array called tags. Tags values are the final values used for matching and retrieving. Tags are generated by using image annotations. Image annotation calculates the edge of the objects in the image.

Image annotation is applied to the pixel labels generated by gradient and hessian method. Image annotation will generate the tags based on red, green, blue components of the image.

For tag selection the tags are ranked on the similarity basis by image annotation and thus important tags are sustained and noisy tags are discarded.The images are ranked based on the similarity basis and the images are retrieve based on these values.

For tag refinement, we are going to consider the annotation image with 10 iterations as inputs. Tag refinement is applied to the above generated tags for graph based feature extraction. We considered the respective column of the image and then calculated the respective nodes or scores. We converted the image into double precision matrix and gave input as alpha =0.2, iterations=10, threshold = 0.0001. After defining the initial scores we are finding and calculating the relevant scores. Graph is constructed based on those feature vectors.

E. Active Contour Based Level Set

Image segmentation includes partitioning of certain image into homogeneous regions which shares some of the common properties. Image segmentation consist of two main approaches: edge and region-based. Discontinuities in the intensity of a certain image is anedge-based segmentation. Uniformity within the sub-region which is based on properties like intensity, texture and color is a task of region-based segmentation.

Our approach is based on the active contour model which can help to obtain robust segmentation model by integrating adaptive region information. This active contour method can be applied to improve better image feature extraction by segmenting images and this method stores lesser features for a particular image thus it will help in better and fast processing and lower computation time.

1) *Active Contour Model:*Segmentation by using active contours model also called as snakes was introduced by Kass et al [6]. The concept of active contours is a very simple idea. The user initially has to specifya guess for the contour, this initial contour is moved towards the desired objects by image driven forces. This active contour model considers two types of forces- the internal forces which keep the model smooth when the deformation process is undergoing and the external forces are the forces which moves the model towards the object boundary or the features which are desirable within the image. The implementation of active contours is made flexible and convenience by using the level set method.
2) *Level Set Method:*The level set method was for the first time brought by Osher and Sethian [6]. The level set method is a theoretical and numerical tool for the propagation of interfaces. The basic idea is by starting with a closed curve in 2-Dimension (2D) or can be started in surface in 3D and then by allowing this curve at a prescribed speed to move perpendicular to itself. Level set method is most commonly used as a segmentation tool for the propagation of a certain contour by using the original properties of an image. The initial application based on this method was the detection of edges in an image [6], but with the advancement of the technology the applications can now detect shapes, colors, textures etc. The basic idea of how the active contour model and level set method can help in detecting the contours and thus segmenting the image in a smaller dimension can be seen in following images,we considered the 10 iterations for applying the active contour method:

Algorithm pseudo code steps in extracting the contours:
Input: Query image
Output: Final Contour with retrieved similar images.

1. *Resize the image into [255 255] square matrix.*
2. *Apply median filter to denoise.*
3. *Set the initial level set function 'u'*
   Truncated_len=50; //initial length for contour
   c0=2; //divisional equivalence condition
4. *Initialize the level set coefficients lambda1=1;lambda2=1;timestep=.1;v=1;epsilon=1;*
5. *Show the intial 0-level-set contour*
   imshow(Img, []); //hold on;axis off,axis equal
   title('Initial contour');
6. *Initialise the contour color as red and pause the figure until the contour coverage is not over*
   [c,ih] = contour(u,[0 0],'r');
   pause(0.1);
7. *Main loop for contour level set , for n=1:iterNum*
8. *Update the contour on the image for different time steps based on iterations*
   u = biurgc(u0, Img, timestep, mu, v, lambda1, lambda2, pc, epsilon, numIter)
9. *Apply if mod(n,10)==0, for plotting in time of iterations*
10. *Display the final contour after the completion of the 10 iterations*
    totalIterNum=[num2str(n), ' iterations'];
    title(['Final Contour, ', totalIterNum]);
11. *Evaluate all the features and formation of the final comarision variable*
    *D1=sum(sum(Tagr))+floor(sum(sum(H))*sum(sum(g))*1e-8)*ceil(sum(sum(Imgout))*1e-7);D2=sum(sum(sum(ain)));*
12. *Compare the database on range based search for retrieving the similar images.*
13. *Visually and semantically similar images are retrieved.*

We combined all the features extracted by the framework with the contours extension and we found that the semantic feature discovery framework works well when we add the contour part. Next we see the comparison of performance metrics.

IV. RESULT AND ANALYSIS

A. *Comparisons of Performance Metrics*

Our work aims at retrieving the images with high precision and recall rate from the dataset. Various performance evaluation parameters were applied to our work which validates effectiveness of our method compared to the existing method.

1) *Requirements:* In our proposed work all the utility measures and algorithms were worked on 2.40 GHz Inter Core i3 machine, with 4 GB dedicated RAM with Windows 8 professional. We used MATLAB 2013a simulator for implementing our work.
2) *Dataset:*For the moment we collected different types of images and created the small-scale artificial dataset which belongs to different category of images.

3) *Aspects of System:* We extracted different features for our framework and we included our extension of active contour model. We further considered different parameters to prove that the framework works better with our contour inclusion in the framework.

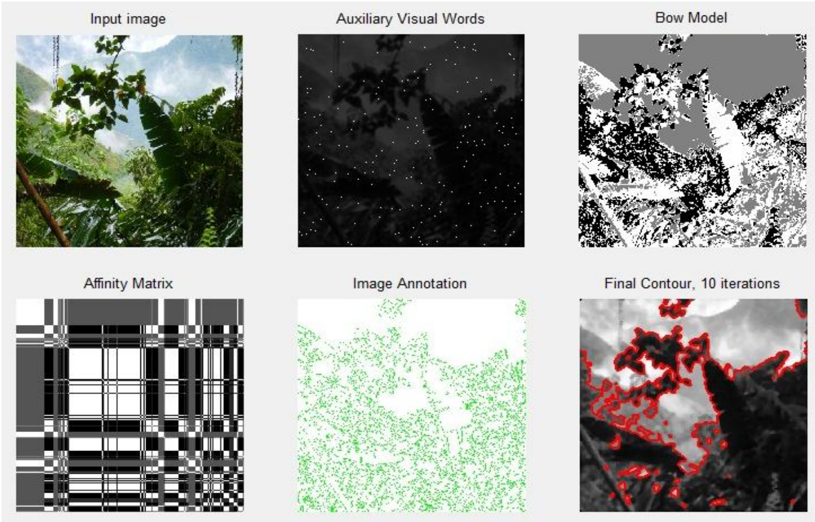

Fig. 3 Different steps in proposed method

Fig. 4 Retrieved results

4) *Evaluation Parameter:* We plotted receiver operating characteristic (ROC) curve to visualize the quality of our ranking on the basis of true-positive ratio and false-positive ratio. The curve proved that our contour addition to the framework outputs better result than without contours addition.

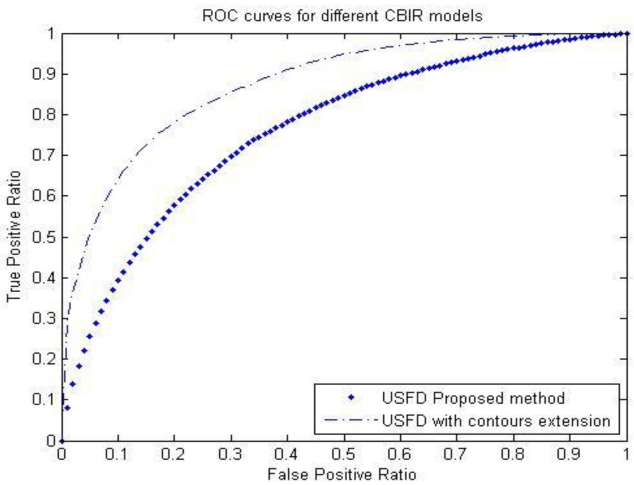

Fig. 5 ROC curve

We constructed Precision-Recall and F-measure graphs for measuring the accuracy of our retrieval system. We adopted procedure of precision and control method for obtaining the precision which states that, the ratio of number of the relevant images that are retrieved to the number of irrelevant and relevant images retrieved. We calculated the similarity of the query image with the other images in the database.

We measured how similar our query image is with the database images. We have similarity measures (ranks) for each of the database images in relation to our query image and then we sort these similarities in descending order.

For a good retrieval system, you would want the images that are the most relevant (i.e. what you are searching for) to all appear at the beginning, while the irrelevant images would appear after.

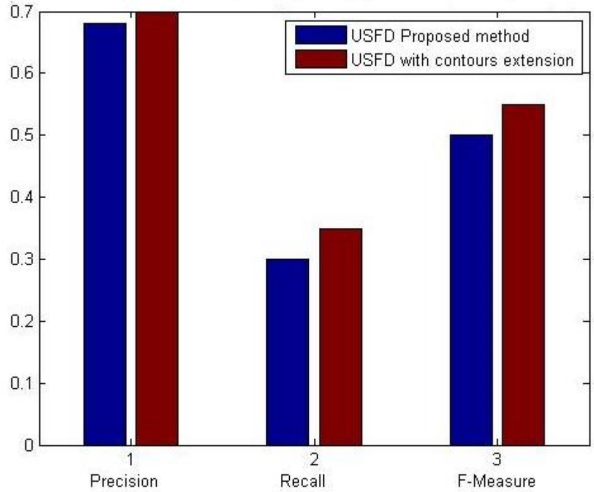

Fig. 6 Overall performance evaluation for local, global and final saliency maps of the system.

V. CONCLUSION

Content Based Image Retrieval is a wide and active area and is a very fast advancing area of research. In all these years varied techniques are developed which can make the retrieval process better and the image searching over the internet faster. There's a remarkable development in applications based on content based image retrieval concept. Such a advancement in this particular concept is because of availability of the internet, storage devices, digital sensors. With the advancement in these technical forces the development in the content based image retrieval will keep on getting better and new models, new methods will be continuing in solving the problems related to content based image retrieval.

The framework generates better results as it carry outs operation in an unsupervised manner. The proposed system with the Contour Model and Level Set Method is based on the concept of active contours which is basically used in medical imaging. The framework extracts various features for similarity matching of the images and with the addition of active contours provides the boundary of a region, thus the segmentation of this region can produce more better results and gradually more better the final results of image retrieval. The framework retrieves better, accurate and faster results with the addition of contours.

Content-based Image Retrieval is a trending area of research which still requires much more exploration to the existing method. Future scope can be to extract features much faster with lesser storage and faster retrieval.

REFERENCES

[1] Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement Yin-Hsi Kuo, Wen-Huang Cheng, Member, IEEE, Hsuan-Tien Lin, Member, IEEE, and Winston H. Hsu.

[2] Sridhar, Gowri. 2012. Color and Texture Based Image Retrieval ARPN Journal of System and Software, Vol. 2, No. 1, pp 1?6.

[3] P. Nagarani, R. VenkataRamanaChary and Dr. D. Rajya Lakshmi 2012. Semantic Based Image Annotation Using Descriptive Features and Retagging approach, International Journal of Multimedia Its Applications (IJMA), Vol. 4, No. 1, pp 1526.

[4] Jeon J, Lavrenko V, Manmatha R.," Automatic image annotation and retrieval using cross-media relevance models". Proc. of Int. ACM SIGIR Conf. On Research and Development in Information Retrieval, Toronto, Canada, 119-126,Jul. 2003.

[5] Sowmya Rani, Rajani N., and Swathi Reddyi. 2012. Comparative Study on Content Based Image Retrieval, International Journal of Future Computer and Communication, Vol. 1, No. 4, pp 366368.

[6] M. Airouche, L. Bentabet and M. Zelmat, "Image Segmentation Using Active Contour Modeland Level Set Method Applied to Detect Oil Spills", Proceedings of the World Congress on Engineering 2009 Vol IWCE 2009, July 1 - 3, 2009, London, U.K.

[7] Hao Ma, Jianke Zhu, Michael Rung-Tsong Lyu, and Irwin King. 2010. Bridging the Semantic Gap Between Image Contents and Tags, IEEE Transactions On Multimedia, Vol. 12, No. 5, pp 462473.

[8] M. Kass, A. Witkin, and T. Terzopoulous. Snakes: Active contour models. International Journal of Computer Vision, pages 321-331, 1988.

[9] L. Wu, L. Yang, N. Yu, and X. -S. Hua. 2009. Learning to Tag, In Proceedings of ACM of the 18th International Conference on World Wide Web (WWW 2009), pp 361370.

[10] J. Smith and S. Chang, "Visualseek: A Fully Automated Content-Based Image Query System", Proceedings of the 4th ACM international conference on Multimedia table of contents, Boston, Massachusetts, United States, Nov. 1996, pp. 87-98.

[11] A. Gupta, and R. Jain, "Visual information retrieval," Comm. Assoc. Comp. Mach., vol. 40, no. 5, pp. 70–79, May. 1997.

[12] C. Wang, L. Zhang, and H. -J. Zhang. 2008. Learning to reduce the semantic gap in web image retrieval and annotation, In Proc. SIGIR, pp. 355362.

[13] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos. 2007 Supervised learning of semantic classes for image annotation and retrieval, Transaction on pattern analysis and machine Intelligence (PAMI), pp 394410.

[14] F. Long, H. Zhang, H. Dagan, and D. Feng, "Fundamentals of content based image retrieval," in D. Feng, W. Siu, H. Zhang (Eds.): "Multimedia Information Retrieval and Management. Technological Fundamentals and Applications," Multimedia Signal Processing Book, Chapter 1, Springer-Verlag, Berlin Heidelberg New York, 2003, pp.1-26.

[15] J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2007, pp. 1–8. J. Philbin, M. Isard, J. Sivic, and A. Zisserman, "Descriptor learning for efficient retrieval," in Proc. Eur. Conf. Comput. Vis., 2010, pp. 677–691.

