Aug. 31
International Journal of advanced studies in Computer Science and Engineering
IJASCSE, Volume 3, Issue 8, 2014

SbLRS: Shape based Leaf Retrieval System

Komal Asrani
Department of Information Technology
B.B.D.E.C.,
Lucknow, India

Renu Jain
Deptt. of C.S.E
University Institute of Engineering and Technology,
Kanpur, India

Abstract--**In this paper, we present an effective image based retrieval system SbLRS (Shape based Leaf Retrieval System) for identification of plants on the basis of their leaves. In this system, a user gives query in the form of a digital leaf image (scanned against plain background) and the retrieval system matches it with its database and displays the names of those plants which have similar type of leaves. Several boundary based approaches has been implemented for matching. However, it is observed that our iterative retrieval approach SbLRS gives best matching results and the size of its database is also diminutive in comparison to other boundary based approaches.**

Keywords—Plant recognition; leaf shape; image retrieval; Contour based image retrieval; shape recognition.

I. INTRODUCTION

There are millions of plants/trees existing in the ecosystem and we always feel the need of a system which can help us in identifying the plants. We, human beings, often try to identify a plant on the basis of its plant size, shape of its leaf, overall structure of leaves, its flower, etc. But, it can be observed that, most of the time, it is possible to identify a plant only by examining its leaf. Besides, leaf of a plant is easily available throughout the year and can be studied as a two-dimensional object. SbLRS is a system that does plant recognition based on shape of leaf.

There are various parameters available for CBIR (Content based Image Retrieval) approaches, like- shape, color and texture. It is well known that the color of most of the leaves is green and capturing the texture of leaves and differentiating among the textures of leaves is very difficult even for human beings. But, the shape and size of leaf varies from plant to plant and also leaf can be easily captured two-dimensional object. Shape representation approaches exploit the boundary details as in boundary based approaches and interior details as in region based retrieval approaches [1,2,3]. Some of the boundary based approaches include boundary chain code[5], fourier transform[8] and region based approaches include moment invariants [4], Zernike moments[11] and curvature scale space[12,13,14].In this paper, we have designed and developed a plant recognition system taking leaf image as input. We have implemented different boundary based approaches and compared the results. We observed that using combination of different approaches at different stages yield best results. The main contribution of this paper is that we propose an effective hierarchical model to ensure fast and effective retrieval. Its effectiveness is compared in terms of recall and precision with the available boundary based approaches. On the basis of experiments conducted, it is observed that our method requires comparatively less time to provide accurate results.

II. BOUNDARY BASED IMAGE RETRIEVAL APPROACHES

In this section, we briefly discuss various shape based image retrieval methods. These approaches were applied once the images were preprocessed as depicted in Figure 2.We have implemented and tested all the methods discussed below in our system for matching the leaves.

A. Centroid Distance Method: The distances of the boundary points, representing the image is computed with respect to the centroid of the image. The centroid is computed using Equation (1).

$$\begin{cases} M_{Ex} = \frac{1}{m}\sum_{i=1}^{m} x_i \\ M_{Ey} = \frac{1}{m}\sum_{i=1}^{m} y_i \end{cases} \quad (1)$$

where M_{Ex} defines centroid for x-point and M_{Ey} defines centroid for y point and (x_i , y_i) are boundary coordinates representing the image and ($\bar{x}$,$\bar{y}$) are the centroid coordinates of the shape , computed using equation (1). The distance between two points is computed using Equation (2).

$$CD = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 + (y_i - \bar{y})^2} \quad (2)$$

The dimension of the image representing the boundary points defines the dimension of the feature vector.

B. Tangential Angle Method: This method is used to measure the turning angle of the consecutive pixels[9]. The turning angle is computed using Equation (3).

$$\tan \varnothing = m = \frac{y_{n+1} - y_n}{x_{n+1} - x_n} \quad (3)$$

This approach is quite significant as it captures the detailed curvature of the image irrespective of the orientation.

C. Angular Partitioning Method: This approach segments the image into four quadrants and counts the number of pixels in each quadrant. This method seems inappropriate because just the count of pixels per quadrant loses boundary information and it cannot be used for as the similarity criteria of shape.

D. Circular Partitioning Method: This approach works on the concept of concentric circles. The image is enclosed within the circle of largest diameter of the image. Thereby, four concentric circles are generated at equidistant spaces. The boundary information of the image is extracted and the number of boundary points lying in each of the concentric circles is identified.

E. Shape Matrix: This method captures the detailed pixel count of the image. The image is superimposed by a matrix of 4x6. The count of the number of pixels is maintained in each cell, corresponding to the grid [6, 7]. As compared to Circular Partitioning and Angular Partitioning, this method captures finer details of the image. The resultant matrix is affected by noise.

F. Quad Centroid Distance Variation Method: This method divides the image into four quadrants and computes the distances from centroid for each of the four quadrants [10]. Then, coefficient of variation is applied on the distance values and four values are generated. The coefficient of variation is defined using Equation (4),(5) and (6).

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{where: } n = \text{Number of samples (total)}, \; x_i = \text{Value of each sample} \quad (4)$$

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n}{n-1}} \quad (5)$$

$$\text{Coefficient of Variation (\%)} = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100 \quad (6)$$

*G. Mean-Max-Min (MMM) approach:*This method is an extension of centroid distance approach. Here image is divided into four quadrants using centroid of the image as the origin. For each quadrant, distances from centroid are computed and from these values, maximum, minimum and average values are calculated. Hence, for complete image, only 12 (3 values per quadrants) values are stored as feature vectors.

III. SbLRS System

SbLRS software is a stand-alone application implemented in Java on Windows platform. It recognizes leaf images on the basis of shape parameters. The application was developed using NetBeans 7.3.1 and Oracle 10g. A database summing to about 500 leaves was created by considering approximately 15-20 leaves of each plant. These images were preprocessed and for each method, feature vectors were created. Fig.1 illustrates overall architecture of SbLRS. Fig. 2 depicts preprocessing stages for a given image before the feature vectors are computed.

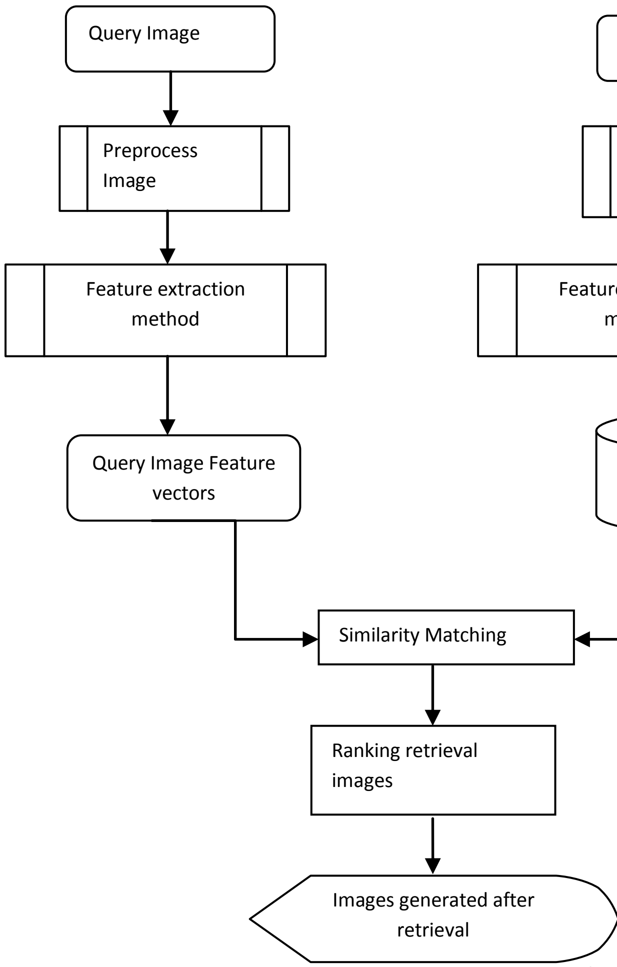

Figure 1.System architecture of SbLRS

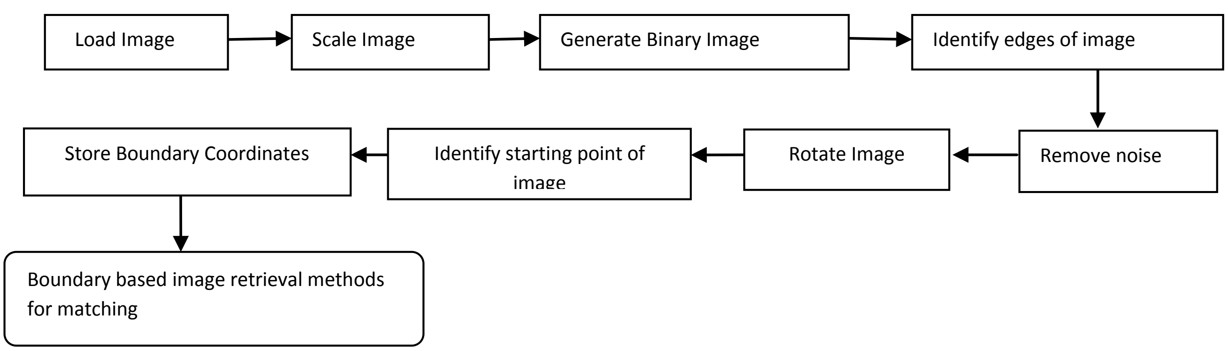

Figure 2. Pre-processing stages for colored image

Functionalities managed by administrator is depicted in Figure 3 and feature vectors for boundary based approaches can be generated by using interface as depicted in Fig. 4.

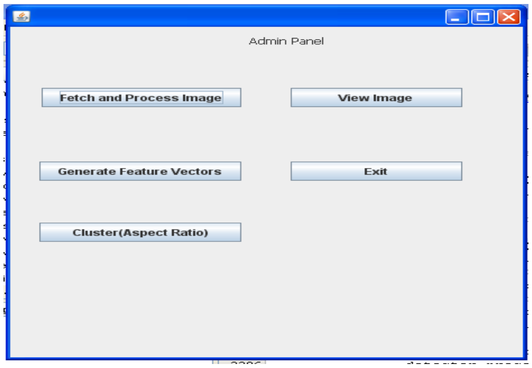

Figure 3.Interface for Administrator

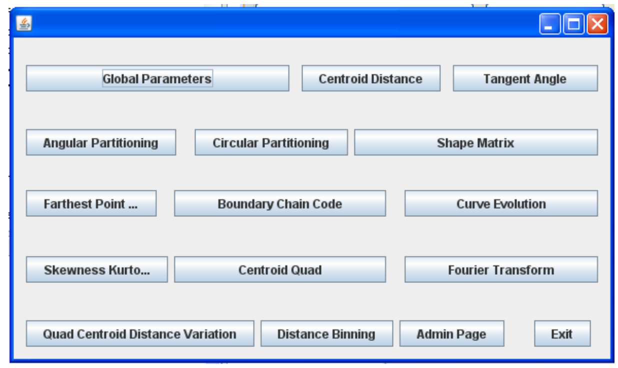

Figure 4. Feature vectors methods for Administrator interface

User can query from SbLRS using various boundary based approaches as shown in fig. 5.

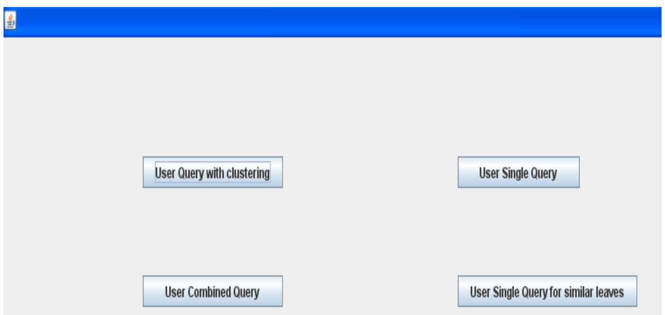

Figure 5. Interface for User

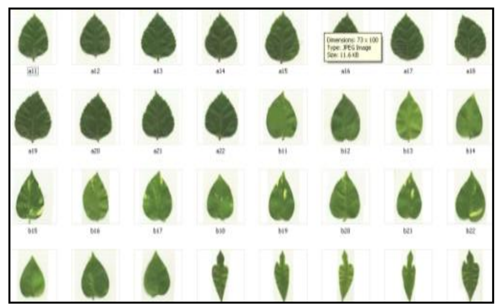

Figure 7.Database Snapshot

The iterative approach used in SbLRS is explained in Figure 6.

The database has been created by scanning the leaves against plain background. It contains multiple leaves of the same plant , approximately 15-20 leaves of each plant. In all, there are approximately 500 leaves. The leaves were scanned using flatbed scanner VistaScan Astra 3600. A snapshot of the leaves is depicted in figure 7.

IV. Experimental Results

The queries were run on “Leaves Database” and results generated were compiled.

For measuring similarity in image retrieval, recall, precision and f-score were used. Recall is the ratio of the number of relevant retrieved images to number of all relevant images.

$$Recall = \frac{Number\ of\ relevant\ retrieved\ images}{Number\ of\ all\ relevant\ images} \qquad (7)$$

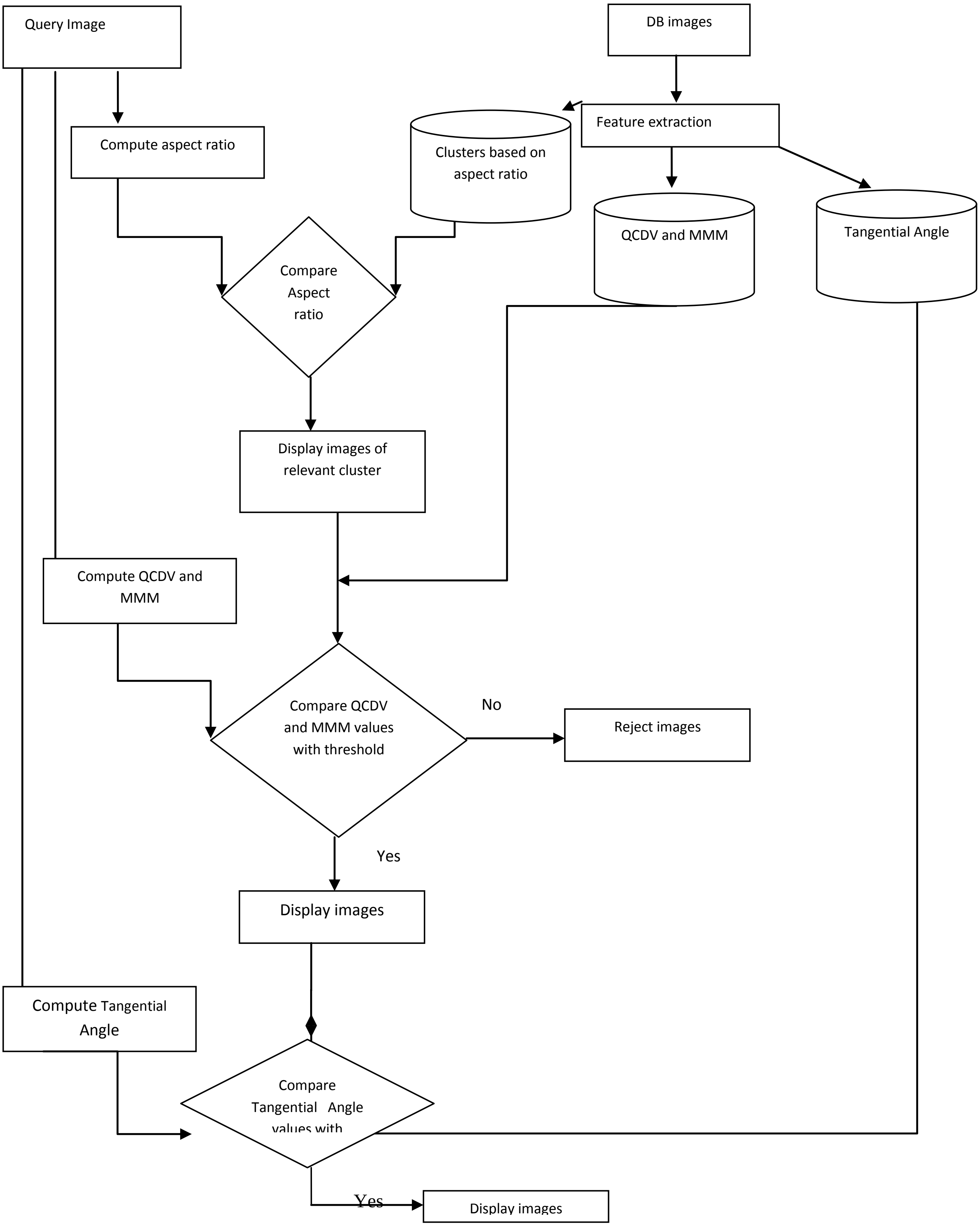

Figure 6. Detailed Iterative Image Retrieval process in SbLRS

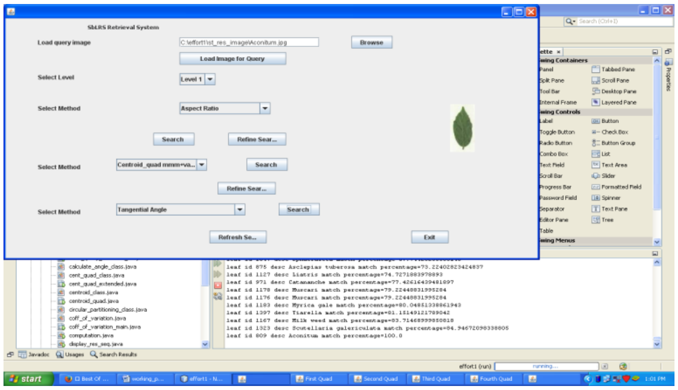

Figure 8. User interface for generating query

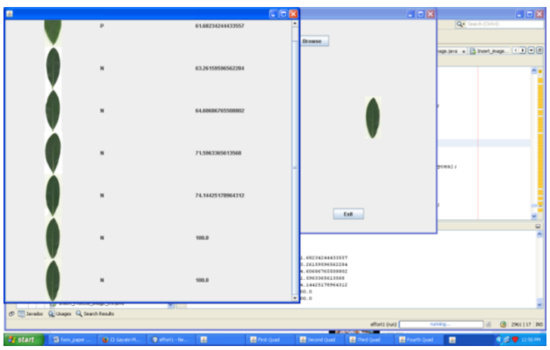

Figure 9. Results generated by SbLRS

Precision is ratio of number of relevant retrieved images to number of retrieved images.

$$Precision = \frac{Number\ of\ relevant\ retrieved\ images}{Number\ of\ retrieved\ images} \qquad (8)$$

F-score is harmonic mean of recall and precision. It is referred as F_1 measure because it gives equal weightage to recall and precision.

$$F\ score\ = 2\frac{Precision * Recall}{Precision + Recall} \qquad (9)$$

Table 1 shows the average F1 score of various boundary based approaches for plant identification different threshold percentage:

TABLE 1. AVERAGE F SCORE RATE OF BOUNDARY BASED APPROACHES

Threshold %age	Centroid Method	Tangential Angle	Angular Partitioning	Circular Partitioning	Shape Matrix	SbLRS
90%	3.00	2.28	1.31	1.32	3.00	3.00
85%	2.63	1.96	1.40	1.08	2.85	2.81
80%	2.19	1.68	1.16	0.97	2.81	2.81
75%	1.72	1.53	0.93	0.83	2.33	2.81

Table 2 shows the level of filtering of images at each subsequent level of iterative retrieval for our approach.

TABLE 2. LEVEL OF FILTERING IN ITERATIVE RETRIEVAL

Level	Stages	Reduction in DB size
1	Clustering using Aspect Ratio	90%
2	QCDV and MMM approach	40%
3	Tangential Angle	68%

Figure 10 shows plot of precision versus recall at different threshold percentage: 90%, 85%, 80% and 75%

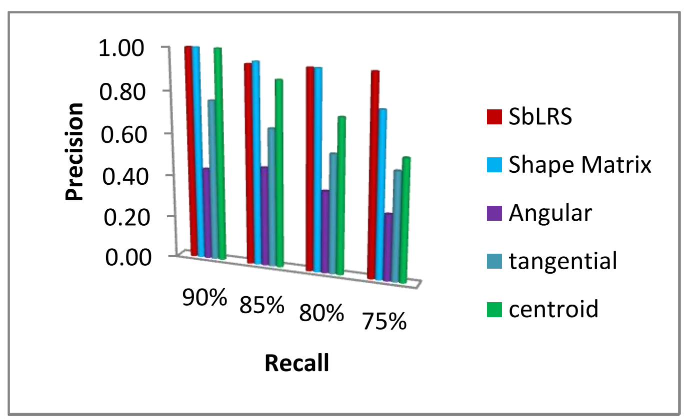

Fig. 10 Precision versus recall plot at 90%, 85% , 80% and 75% for various boundary based approaches

Figure 11 shows precision versus recall comparison for the first five results generated when query was executed for various boundary based approaches.

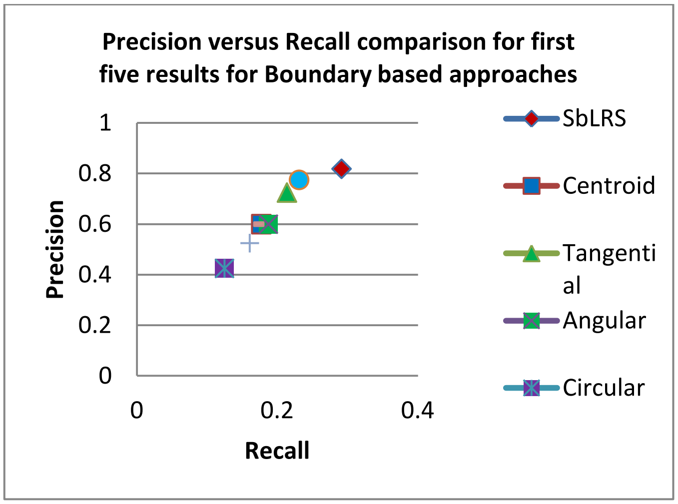

Fig. 11 Precision versus recall plot for first five result for boundary based approaches

V. Conclusions and Future Work

Two plots for recall versus precision have been presented. Fig. 10. depicts comparison of precision versus recall for various boundary based approaches at different retrieval percentage. Fig 11 depicts comparison of precision versus recall for the first five results generated when different boundary based approaches were executed. It is observed that at 90%, shape matrix and centroid distance method performs equally well. But with reducing threshold percentage, the precision rate remains steady with improving recall rate but drops at 75% in shape matrix. Tangential angle approach shows linear improvement in recall rate and fall in precision rate. As compared to other boundary based approaches, our approach SbLRS maintains comparably satisfactory precision versus recall rate. Precision values are maintained stable values with improving recall rates. From figure 11, it is observed that precision versus recall rate is good for SbLRS for the first five leaves as compared against the boundary based approaches.

The important aspect to be focused of SbLRS is that because it uses iterative search approach, which causes reduction of database size at each consecutive level, hence ensuring faster results and reduces computation time. Though SbLRS provides satisfactory results when compared to other boundary based approaches, but still the results can be further improved by considering aspects of plant leaf. The work can be extended for region based image retrieval for understanding vein structure of the leaves.

References

[1] Arnold. W.M. Smeulders, M. Worring, S.Satini, A.Gupta, R.Jain, “Content Based Image Retrieval at the end of early years” IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, pp. 1349-180, 2000.

[2] X. Fu, Y. Li,R. Harrison, S. Belkasim , “Content based image retrieval using Gabour Zernike features” Proc. Of 18th IEEE International Conference on Pattern Recognition(ICPR’ 06), Vol.2, pp 417-420, 2006.

[3] D. Zhang, G. Lu, “Review of shape representation and description techniques” Pattern recognition, vol. 37, pp. 1-19, 2004.

[4] M. Bober, “MPEG -7 Visual Shape Descriptors”, IEEE transactions on Circuits and System for Video Technology, vol.11, no.6, pp. 716- 719, June 2001.

[5] R. C. Gonzalez, R.E. Woods, Digital Image Processing, Second Edition, Prentice Hall, 2002, pp. 793.

[6] Cui Ming, Peter Wonka, Anshuman Razdan, Jiuxiang Hu, “A New Image Registration Scheme Based on Curvature Scale Space Curve Matching”, Visual Computation ,2007, 23: 607-618, DOI: 10.1007/s00371-007-0164-1.

[7] Cyrus Shahabi, Maytham Safar,” An experimental study of alternative shape-based image retrieval techniques”, Multimedia Tools Application, DOI 10.1007/s11042-006-0070-y.

[8] M. Trimeche, “Shape representation for Image Indexing and Retrieval”, Master of Science Thesis, Tampere University of Technology, May 2000.

[9] D. S. Zhang and G. Lu,”A comparative study on shape retrieval using Fourier descriptors with different shape signatures”, Proc. International Conference on Intelligent Multimedia and Distance Education (ICIMADE01),2001.

[10] Asrani Komal, Jain Renu ,”A novel shape based descriptor for plant identification”, International Journal on Computational Vision and Robotics, International Journal of Computer Applications (0975 –8887) Volume 95–No. 15, June 2014.

[11] M. E. Celebi and Y. A. Aslandogan, “A comparative study of three moment-based shape descriptors”, Proc. of the International Conference of Information Technology: Coding and Computing, pp. 788-793, 2005.

[12] A. C. Jalba, M. H. F. Wilkinson, and J. B. T. M. Roerdink,”Shape representation and recognition through morphological curvature scale spaces”, IEEE Trans. Image Processing, vol. 15(2), pp. 331_341, 2006.

[13] J. Peng, W. Yang, and Z. Cao,”A symbolic representation for shape retrieval in curvature scale space”, Proc. International Conference on Computational Intelligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce, 2006.

[14] D. Zhang and G. Lu,”A comparative study of curvature scale space and Fourier descriptors for shape-based image Retrieval”, Visual Communication and Image Representation, vol. 14(1), 2003.