



Comparing the impact of Clustering with Content Based Image Retrieval Approaches for Plant Identification

Komal Asrani¹, Renu Jain²

¹ B.B.D.N.I.T.M., Lucknow.

komalasrani@rediffmail.com

² U,I,E,T., Kanpur

jainrenu@gmail.com

Abstract

Contour Based retrieval of images is an active and challenging field of research. Among various parameters available for contour based image retrieval, shape is considered an important aspect because it is closest to the human perception. Most of the shape based image retrieval methods require large processing time for generating accurate results due to huge database. To reduce the search time, we have divided the database into clusters on the basis of eccentricity of leaf using K-Means approach. After making the clusters, different contour based approaches are applied for leaf/plant identification and results are compared. The leaf image is processed to generate feature vectors which are stored in database. We have used Swedish leaf image database (SLID) consisting of 15 species with 75 leaves per class and total of 1125 leaf images. In this paper, we compare results of contour based retrieval approaches with and without clustering. From these results, it is found that by incorporating clustering, performance of contour based retrieval approaches remains same but retrieval time is reduced.

Index Terms

Contour Based Image Retrieval, Shape descriptors, Plant Identification, Clustering, Content Based Image Retrieval, Leaf recognition, Image Retrieval.

ACADEMIC DISCIPLINE AND SUB-DISCIPLINES

Computer Science (Image Processing)

SUBJECT CLASSIFICATION

Image Processing and Retrieval

TYPE (METHOD/APPROACH)

Experimental

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 9, No 1

editor@cirworld.com

www.cirworld.com, member.cirworld.com



1. INTRODUCTION

With the fast growth of the image capturing devices, we are left with huge collection of images. But these collections of images could be useful only if we are able to retrieve exact images when required. Content Based Image Retrieval (CBIR) is an important field which assists in image retrieval on the basis of salient features of the image. The image is processed to extract the information defining the image to be expressed in terms of feature vectors. For CBIR to be effective, it is necessary that the feature vectors extracted should be such that it is invariant to translation, rotation and affine transformation. Shape is considered as an important parameter used in CBIR for identifying objects [1]. This is because shape contains more intrinsic information than other features like color, color layout and texture. Moreover, considering human perception, it is considered as the basic criteria for defining and representing object. However, the most challenging issue for defining shape is the accurate extraction and representation of the shape. Among the performance metrics available for representation, qualitative measures focus on measuring the retrieval performances of the feature vectors extracted from the shape whereas quantitative measures lay stress on the amount of data used for representing the image in the form of feature vectors to ensure improved retrieval [2]. Here we are focusing on plant identification. For this, we have used the shape of leaf image. We have done clustering of the SLID database and compared the results against contour based image retrieval approaches.

We have projected comparative results in terms of recall and precision. Here we are focusing on plant identification. For this, we have used the shape of leaf image. We have done clustering of the SLID database and compared the results against contour based image retrieval approaches. We have projected comparative results in terms of recall and precision. The paper has been organized into seven sections. First section is Introduction where we present our motivation and objectives. Related work and techniques are presented in second section. Clustering is discussed in third section and contour based image retrieval approaches are discussed in the fourth section. Section fifth deals effectiveness measures and sixth illustrate implementation details and present results. Finally, the last section discusses results and concludes with future work.

2. RELATED WORK

Several techniques have been explored for shape representation. Casanova et. al.[7] explored different aspects of the leaf and used different classifier for each aspect. The aspects considered were GPS coordinates, Gabor filters, volumetric fractal dimensions, local binary pattern and geometric features. Finally, the results generated from these different classifiers were subjected to a final classifier. The approach provided good results but at the cost of high computational complexity. Mouine et. al[8] identified leaves combining local and shape based feature. In this paper, an extension of shape context was done in terms of voting sets which defined the coarse arrangement of the shape and computing sets which contained the shape context points. But the notion of the saliency depended on the application. Lee et. al[6] used the main vein of the leaf and frequency domain data. The feature vectors were extracted using Fast Fourier Transform, geometrical digital features, morphological features and convex hull. T. Bernier[3] focuses on contour points of an object with their distances and angles relative to the center point of the object to form a shape representation, which is invariant to translation, rotation and scaling but yields poor result when edge detection is incomplete as the computation is based on main vein extraction. Costa et al. [4] proposed the use of shape saliencies for object representation which was implemented using Image Foresting Transform. The saliencies of a shape are defined as the maximum influence areas of its higher curvature points. But this approach requires detailed calculations. Petrakis [5] defines shapes, which are represented as a set of segments between two consecutive inflection points. Segments are considered at different levels of shape resolution and matching is achieved by minimizing a cost function. The success of any method depends strongly on the choice of the features considered and their extraction. The method is good only for presenting an overall representation but not for defining fine details. Ling and Jacob [15] introduced the Inner Distance, which they combine with the shape contexts. They showed that the combined descriptor, called IDSC, outperforms many other approaches on leaf classification and identification. IDSC samples points along the boundary of a shape and builds a 2D histogram descriptor at each point. This histogram represents the distance and angle from each point to all other points, along a path restricted to lie entirely inside the leaf shape. Belhumeur et al. [14] proposed a faster version of the same method that avoids finding correspondences by aggregating the similarities between each point and its nearest point in the feature space on the other shape. Hu et al. [16] proposed a fast algorithm for plant leaf recognition that uses the matrix of pairwise distances between points sampled on the leaf boundary. The approach used Euclidean distances as well as inner distances for measuring similarity between images.

Most of the shape based approaches discussed above deal with complex and elaborate computation and handle huge database. In this paper, we make an effort to reduce the search space by forming clusters and ensure retrieval of images on the basis of simplified and easy computation. In our earlier work [17], we had performed clustering of the leaf images. This paper is an extension of our earlier work where we are measuring the impact of performing clustering.

3. CLUSTERING

Clustering has found its wide application in the field of pattern recognition and image classification [11]. The idea of considering clustering in image classification is that clustering helps to group the objects belonging to the same class. Hence when we are dealing with huge database, it is necessary that we need some mechanism using which we are able to group similar objects effectively. Clustering helps in grouping objects into subsets such that similar objects are grouped

together whereas distinct objects belong to different groups. In this paper, we have implemented clustering of images using K-Means and taking eccentricity as clustering parameter. The choice of K-Means among various clustering techniques is because it is not dependent upon data ordering, it allows straightforward parallelization and can be extended easily and effectively for scaling data. The choice of eccentricity parameter is done because it is independent of scale, translation and rotation invariant and it is considered to be a critical requirement for image recognition.

The K-means can be understood as: Let x_1, x_2, \dots, x_n are data points and each data point will be assigned to one and only one cluster. Let $C(i)$ denotes cluster number for i th data point. It uses Euclidean distance for dissimilarity measure [12], [13]. The k-means minimizes within cluster point scatter by considering

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 \quad (1)$$

K-Means Clustering Results:

For SLID of 1125 leaves, we have generated 24 clusters. The number of data in each cluster is variable. The results of some of the clusters obtained by implementing K-Means clustering using eccentricity are as follows:

Cluster 1 and associated leafs':



Cluster 2 and associated leafs':



For a query image given by the user, the relevant cluster to which it is matched with least difference is identified. Then the contour based image retrieval approaches are applied to the specific cluster identified and the results are further refined. Thus, this approach helps to reduce the computation time of comparing with the huge database and ensures better results.

4. CONTENT BASED IMAGE RETRIEVAL

CBIR approach is based on capturing the outline contours defining the shape, without any concern to the interior details. There are two approaches for CBIR: contour based and region based approach. Contour based image retrieval concentrates only on the boundary details whereas region based approach details the finer details of the complete image. Some of the contour based image retrieval approaches are as follows:

4.1 Centroid Distance Approach

In this approach, the distances from centroid are calculated. The values of centroid are calculated as defined in equation (1). For a given set of coordinates representing the image, the centroid distances can be calculated as follows:



$$\begin{cases} M_{Ex} = \frac{1}{m} \sum_{i=1}^m x_i \\ M_{Ey} = \frac{1}{m} \sum_{i=1}^m y_i \end{cases} \quad (2)$$

where M_{Ex} defines centroid for x-point and M_{Ey} defines centroid for y point. Hence, for a discrete pixel boundary image, it is simply the average of the boundary coordinates. Once the contours of the leaf image are identified, the coordinates representing the shape are generated. These coordinate values are summed and center of gravity is computed using (2). Using these values, centroid distances are computed as follows:

$$\text{Centroid distance} = ((x_n - x_c)^2 + (y_n - y_c)^2)^{1/2} \quad (3)$$

Here each leaf contour is processed and feature vector representing each leaf is an array of distances as computed using equation (3). The size of the array representing the shape contour depends on the number of pixels representing the boundary of the shape.

Centroid Distance Method Results:

The following data shows centroid distances for leaf_id 100 computed using equation (3):

16.873055,16.967018,17.118979,17.279657,16.904623,16.581495,16.31336,16.102964,15.952591,15.86395,15.838075,17.842974,17.307554,16.814611,18.25751,17.621475,18.928854,18.206627,19.830946,19.03802,20.933975,20.085096,21.316097,20.43418,22.70112,21.783495,23.269842,22.330814,24.868664,23.908789,25.585056,24.612494,25.375568,26.192825,26.06508,25.975235,25.92368,25.910645,25.936188,26.000196,26.102386,25.255468,25.4394,24.461046,24.691439,23.724394,24.003567,23.381088,23.748676,22.829353,23.248735,22.353601,22.825596,21.958769,21.103727,21.649286,20.829578,20.02626,20.64886,19.888569,19.150324,19.850925,19.16348,18.504559,19.913898,19.28064,18.679476,18.1136,17.388866,17.871983,17.867935,18.76961,19.030333,19.390951,20.7616,21.328371,22.933262,23.65988,25.432

4.2 Tangent Angle Approach

For a given set of coordinate values, the tangential angle helps in capturing the curvature details. The tangential angle at any point can be calculated as:

$$\theta_n = \arctan \frac{(y_{(n+1)} - y_{(n)})}{(x_{(n+1)} - x_{(n)})} \quad (4)$$

where $(x_{(n)}, y_{(n)})$ and $(x_{(n+1)}, y_{(n+1)})$ represent the successive coordinate locations representing the shape. The feature vector generated using this method is an array of the angles, hence defining the angular dispersions with respect to x-axis. The size of the array representing the shape contour depends on the number of pixels representing the boundary of the shape.

Tangent Angle Method Results:

The following shows tangential angles (partial results) for leaf id 100 computed using equation (4):

1.3617734,1.3677511,1.3329604,1.2882414,1.2966288,1.3045443,1.2036225,1.2160907,1.2277724,1.2387369,1.2490457,1.2587543,1.141034,1.1562895,1.1705557,1.0899091,1.1071488,1.0164888,1.0370882,0.9600704,0.98279375,0.87605804,0.9025069,0.81569195,0.844154,0.7551044,0.7853982,0.69473827,0.7266424,0.6000502,0.63502675,0.5028432,0.5404195,0.57637525,0.32175055,0.36397895,0.4048918,0.4444192,0.4825133,1.3258177,1.3284632,1.3179022,1.3205957,0.2536741,0.28805545,0.32175055,0.35470566,0.37433362,0.4048918,0.4214192,0.4497596,0.4636476,0.48995733,0.47646743,0.5016041,0.5260664,0.5123895,0.53581125,0.5585993,0.54486394,0.5667292,0.5532943,0.5743048,0.5947593,0.5813802,0.60107374,0.6202495,0.638914,0.62548506,0.6435011,0.66104317,0.6781193,0.6647745,0.6812982,0.6973908,0.71306086,0.7283174,0.7431698,0.7443254,0.75837773,0.77206564,0.75908846,0.7724119,0.7853982,1.112357,1.1423101,1.1615504,1.1760052,1.19029,1.204402,1.2183387,1.2285007,1.2422432,1.2524587,1.2659934,1.2853335,1.3258177,1.3360738,1.336541,1.3448371,1.3557814,1.3644724,1.3756379,1.3847382,1.3920033,1.4016951,1.4097044,1.418147,1.4270576,1.4364749,1.4429264,1.4444882,1.442473,1.440393,1.4382448,1.4546547,1.4527043,1.4506876.

4.3 Quadrant Based Sectors Approach

The given shape is decomposed into four quadrants, with the origin coinciding with the centroid of the shape. The number of pixels in each quadrant are identified and defined as feature vectors. The size of the feature vectors representing the leaf image is always four as the image is segmented into four sections. The accuracy of the shape can be improved if higher sections are defined. This will help to retain more details of the shape.



Quadrant Sector Approach Results:

Following table details the results generated for various leaves and count of pixels in each of the four quadrants.

Table 1: Results generated for subset of leaves using Quadrant Based Sector Approach

Leaf_ID	Quad_1	Quad_2	Quad_3	Quad_4
100	29	47	60	66
101	183	169	40	100
102	192	49	52	91
107	200	81	60	122
108	140	100	60	104
125	18	135	67	13
161	49	34	50	35

4.4 Circle Based Sectors Approach

The given shape is decomposed by using concentric circles. The concentric circles are centered at the centroid of the given shape. The count of the number of pixels in each of the concentric circles is stored in the form of feature vectors. The radius of the largest circle is calculated from the diameter of the shape. The radius is distributed evenly into 3 circles and count of the number of pixels in each circle is identified. The array size of the feature vectors representing is always 4. The effectiveness of this method can be improved by increasing the number of circles.

Circle Sector Approach Results:

Following table details the results generated for various leaves and the count of pixels in each of the four concentric circles.

Table 2: Results generated for subset of leaves using Circle Based Sector Approach

Leaf_ID	Count_1	Count_2	Count_3	Count_4
100	102	30	60	0
101	10	400	39	43
102	301	23	61	7
107	278	90	73	12
108	116	271	12	5
125	132	16	54	31
161	81	23	52	12

5. EFFECTIVENESS MEASURE

To measure the effectiveness of the various image retrieval approaches, recall and precision can be used. Recall is defined as the ratio of number of relevant retrieved images to the number of the total relevant images. Mathematically, recall is expressed as:

$$\text{Recall} = \frac{\text{Number of relevant retrieved images}}{\text{Number of all relevant images}} \quad (5)$$

Precision is defined as the ratio of number of relevant retrieved images to total number of retrieved images. Mathematically, precision is expressed as:

$$\text{Precision} = \frac{\text{Number of relevant retrieved images}}{\text{Total number of retrieved images}} \quad (6)$$

6. EXPERIMENTS

A complete application has been generated using Java NetBeans. A total of one thousand leaves had been collected from various botanical gardens and scanned against a plain background. The database has been generated using Oracle 10g. A user interface is available which enables the user to load the data, where it is processed and then feature vectors are generated. Once the feature vectors are generated, information is stored in database. For measuring effectiveness, recall and precision are considered. The following table shows results generated by various approaches for the given leaf ids. The recall and precision values are computed and projected for five contour based approaches-Global descriptors, centroid distance, tangent angle, quadrant sector and circle sector approach.



Recall values of contour based approaches:

Table 3: Recall values generated for subset of leaves on contour based image retrieval approaches.

Leaf ID	Global Descri.	Centroid Distance	Tangent Angle	Quadrant Sectors	Circle Sectors
101	0.34	0.47	0.45	0.14	0.09
102	0.41	0.33	0.34	0.18	0.16
103	0.29	0.59	0.28	0.13	0.13
104	0.41	0.5	0.31	0.1	0.06
105	0.35	0.5	0.33	0.16	0.07
106	0.39	0.39	0.43	0.05	0.1
107	0.42	0.51	0.41	0.01	0.14
108	0.28	0.25	0.29	0.16	0.04
109	0.21	0.48	0.3	0.11	0.03
110	0.45	0.37	0.39	0.14	0.05
111	0.41	0.19	0.42	0.09	0.12
112	0.43	0.33	0.37	0.19	0.17
113	0.38	0.38	0.27	0.12	0.14
114	0.37	0.4	0.31	0.15	0.06
115	0.42	0.27	0.25	0.17	0.18
116	0.43	0.26	0.29	0.08	0.1
117	0.37	0.4	0.34	0.18	0.18

Precision values of contour based approaches:

Table 4: Precision values generated for subset of leaves on contour based image retrieval approaches.

Leaf ID	Global Descri.	Centroid Distance	Tangent Angle	Quadrant Sectors	Circle Sectors
101	0.42	0.41	0.44	0.24	0.16
102	0.375	0.45	0.43	0.19	0.13
103	0.43	0.42	0.47	0.28	0.26
104	0.35	0.5	0.51	0.19	0.18
105	0.375	0.39	0.50	0.26	0.28
106	0.25	0.45	0.45	0.23	0.32
107	0.35	0.43	0.49	0.12	0.38
108	0.41	0.39	0.51	0.26	0.25
109	0.37	0.50	0.58	0.27	0.38
110	0.425	0.55	0.33	0.32	0.26
111	0.55	0.52	0.39	0.19	0.16
112	0.48	0.46	0.49	0.29	0.28
113	0.475	0.45	0.47	0.29	0.23
114	0.55	0.42	0.49	0.31	0.21
115	0.45	0.48	0.58	0.32	0.13
116	0.475	0.48	0.56	0.19	0.18
117	0.575	0.45	0.45	0.31	0.39



Recall values of clustered contour based retrieval approaches:

Table 5: Recall values generated for subset of leaves on clustered contour based retrieval approaches

Leaf ID	Centroid Distance	Tangent Angle	Quadrant Sectors	Circle Sectors
101	0.61	0.83	0.56	0.26
102	0.72	0.87	0.59	0.29
103	0.73	0.89	0.55	0.25
104	0.74	0.81	0.61	0.32
105	0.68	0.83	0.48	0.16
106	0.71	0.90	0.3	0.28
107	0.69	0.82	0.52	0.28
108	0.73	0.89	0.49	0.12
109	0.76	0.90	0.47	0.10
110	0.67	0.86	0.39	0.12
111	0.62	0.82	0.42	0.22
112	0.66	0.89	0.61	0.31
113	0.75	0.85	0.51	0.21
114	0.68	0.86	0.34	0.14
115	0.67	0.82	0.27	0.27
116	0.75	0.87	0.36	0.16
117	0.64	0.85	0.46	0.26

Precision values of clustered contour based retrieval approaches:

Table 6: Precision values generated for subset of leaves on clustered contour based retrieval approaches

Leaf ID	Centroid Distance	Tangent Angle	Quadrant Sectors	Circle Sectors
101	0.57	0.89	0.39	0.33
102	0.52	0.88	0.38	0.32
103	0.48	0.83	0.23	0.31
104	0.67	0.85	0.19	0.29
105	0.57	0.87	0.26	0.31
106	0.61	0.85	0.12	0.39
107	0.57	0.84	0.14	0.48
108	0.48	0.86	0.26	0.37
109	0.65	0.82	0.32	0.40
110	0.72	0.90	0.23	0.37
111	0.62	0.95	0.12	0.34
112	0.56	0.92	0.28	0.36
113	0.73	0.85	0.34	0.34
114	0.59	0.88	0.28	0.29
115	0.74	0.85	0.34	0.29
116	0.71	0.79	0.21	0.21
117	0.72	0.79	0.39	0.40

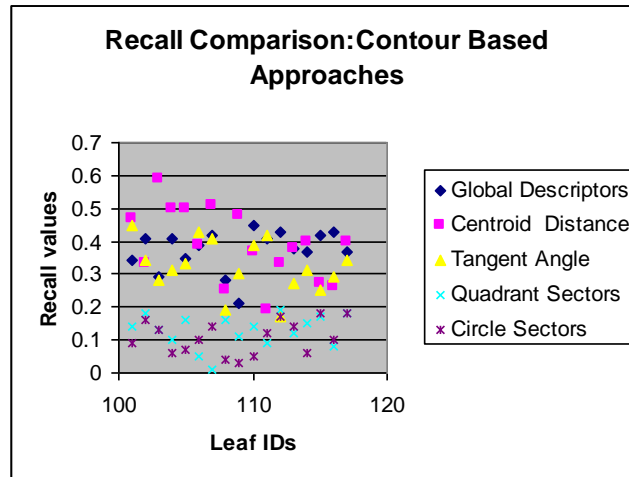


Fig. 1 Recall comparison for Contour Based Retrieval Approaches

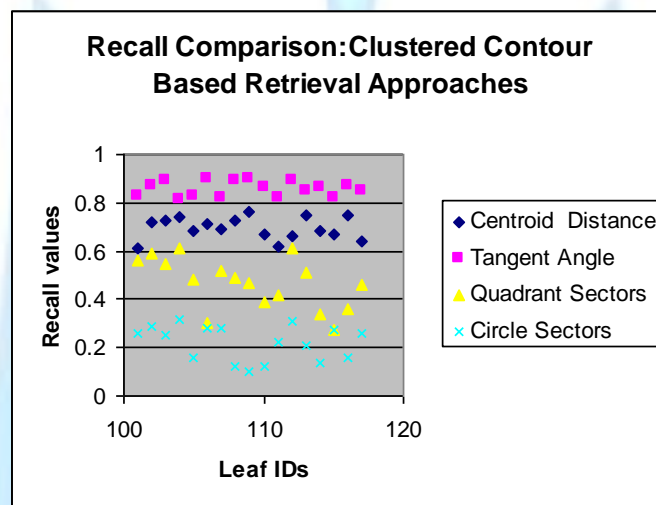


Fig. 2 Recall comparison for *Clustered Contour Based Retrieval Approaches

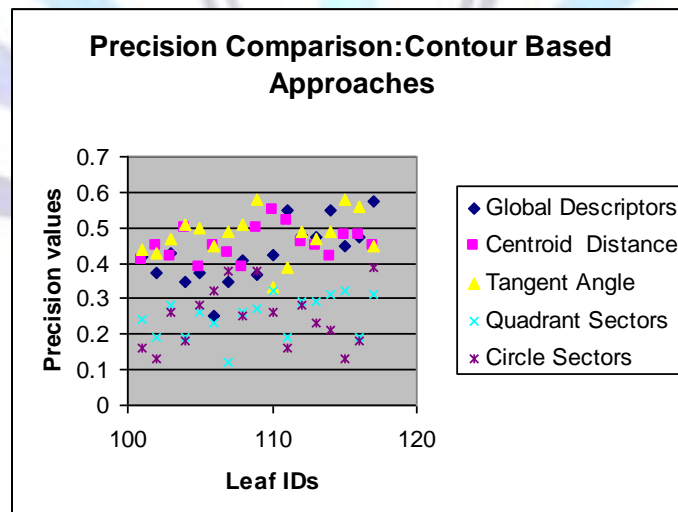


Fig. 3 Precision comparison for Contour Based Retrieval Approaches

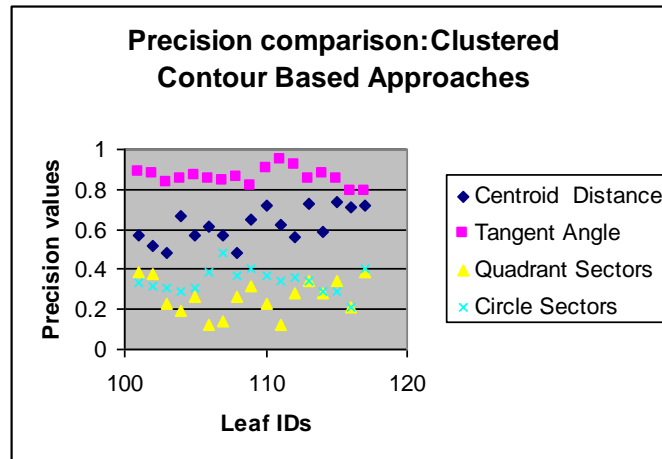


Fig. 4 Precision comparison for *Clustered Contour Based Retrieval Approaches

* referred as Clustered Contour Based retrieval approaches because first clustering is performed on the complete database and then contour based image retrieval approaches are implemented.

Retrieval Time for Contour Based Retrieval Approaches:

Table 7: Average retrieval time for contour based retrieval approaches

Approaches	Time(sec)
Centroid Distance	18
Tangential Approach	17
Quadrant Sector	11
Circle Sector	10

Retrieval Time for Clustered Contour Based Retrieval Approaches:

Table 8: Average retrieval time for clustered contour based retrieval approaches

Approaches	Time(sec)
Clustered Centroid Distance	11
Clustered Tangential Approach	12
Clustered Quadrant Sector	5
Clustered Circle Sector	6

It has been observed from Table 5 and 6 that recall and precision is comparatively better than results projected in Table 3 and 4. Due to clustering, it is inferred that recall and precision results are improved by 12-15% average for all the methods. Table 7 and 8 project the retrieval time. It is observed that the retrieval time for tangential approach and centroid distance approach is comparatively large as compared to quadrant and circle sector approach. This is because in both centroid distance and tangent angle approach, the array representing depends on the number of pixels representing the boundary shape (approximately 200 points). Whereas quadrant and circle sector approach requires an array of size four and this marks the difference in retrieval time.

Quadrant and circle sector based approaches are based on pixel densities. These approaches are useful in determining the proportionality of the objects in terms of pixel count, but are not at all the true representatives of the image shape. Further, simple global descriptors are useful only for capturing overall details of the shape in terms of perimeter, eccentricity, area and circularity. These values do not help in determining the image shape details but are helpful in differentiating shapes with larger differences. Tangent angle approach helps in determining the angular details but this approach does not define any other details. Hence angular dispersion is not sufficient for true representation as the location details are undefined. Further, centroid distance approach is comparatively better enough as compared to the rest of the above discussed approaches but it involves lengthy decimal computation, which adds to the storage requirement and adds to computation time. Thus, these approaches cannot be considered as standalone descriptors; however they can be effectively used to couple with other techniques to improve its effectiveness. So, here we have coupled clustering with contour based approaches. Clustering results in improving the methods, especially in case of tangent angle and centroid distance. Also the retrieval time involved is marking the difference. Thus, clustered contour based retrieval approaches are fast in execution, efficient in recognition and easy in implementation.



7. CONCLUSIONS

In this paper, we presented impact of clustering on contour based approaches. We tried to explore the advantages and shortcomings of the contour based approaches. Emphasis was made on speeding up the retrieval process. For this, we had incorporated clustered contour based approaches. Results show that clustered contour based retrieval approaches helps in reducing the retrieval time with simplified computation and provides effective results. We look forward for exploring methods by which the retrieval of plants using leaves can be improved and the feature vector size can be confined. Also, the finer details defining the shape of the leaves need to be focused.

REFERENCES

- [1] Persoon, E., Fu, K.: "Shape discrimination using Fourier descriptors", IEEE Trans. Syst. Man Cybern., 1997, 7, (3), pp. 170–179
- [2] Belongie, S., Malik, J., Puzicha, J.: "Shape matching and object recognition using shape contexts", IEEE Trans. Pattern Anal. Mach.Intell., 2002, 24, (4), pp. 509–522
- [3] T. Bernier, J.A. Landry, "A new method for representing and matching shapes of natural objects", Pattern Recognition 36(2003) 1711-1723.
- [4] R. da S. Torres, A.X. Falcao, L. da F. Costa, Shape description by image foresting transform, in: 14th International Conference on Digital Signal Processing, vol. 2, Santorini, Greece, 2002, pp. 1089–1092.
- [5] E. Petrakis, E. Milios, "Shape retrieval based on dynamic programming", IEEE Transactions on Image Processing 9 (1) (2000) 141–147.
- [6] Kue- Bum Lee, Kwang-Woo Chung, Kwang- Seok Hong "An implementation of Leaf Recognition System" ISA 2013, ASTL Vol. 21, pp. 152- 155,2013.
- [7] Casanova,Joao Batista Florindo, Wesley Nunes Goncalves ,Odemir Martinez Bruno "IFSC/USP at ImageCLEF 2012: Plant identification task"
- [8] S. Mouine, I. Yahiaoui, and A. Verroust-Blondet, "Advanced shape context for plant species identification using leaf image retrieval," in Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, ser. ICMR '12. New York, NY, USA: ACM, 2012, pp. 49:1– 49:8. [Online]. Available: <http://doi.acm.org/10.1145/2324796.2324853>
- [9] Mouine Sofiene, Yahiaoui Itheri, Blondet Anne Verroust "A shape based approach for leaf classification using multiscale triangular representation" Third ACM International Conference on Multimedia Retrieval 2013.
- [10] L.Yong, J. Walker, J. Bowie,"An analysis technique for biological shape", Comput. Graphics Image Process. 25 (1974) 357–370
- [11] V. Capovleas, G. Rote, and G. Woeginger, Geometric "Clusterings, J. Algorithms", vol. 12, pp. 341-356, 1991.
- [12] E. Forgey, "Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classification", Biometrics, vol. 21, p. 768, 1965.
- [13] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations", Proc. Fifth Berkeley Symp. Math. Statistics and Probability, vol. 1, pp. 281-296, 1967.
- [14] P. Belhumeur, D. Chen, S. Feiner, D. Jacobs, W. Kress, H. Ling, I. Lopez,R. Ramamoorthi, S. Sheorey, S. White, and L. Zhang, "Searching the world's herbaria: A system for visual identification of plant species," in Computer Vision ECCV 2008. pp. 116–129.
- [15] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance" IEEE Trans. Pattern Anal. Mach. Intell, vol. 29, no. 2, pp. 286–299, Feb. 2007. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2007.41>
- [16] R. Hu, W. Jia, H. Ling, and D. Huang, "Multiscale distance matrix for fast plant leaf recognition," IEEE Transactions on Image Processing, 2012.
- [17] Asrani, Komal ;Jain, Renu , "Designing a clustered database for identification of leaves", IEEE 3rd International Advance Computing Conference (IACC)Digital Object Identifier: 10.1109/IACC.2013.6514227 , 2013 , Page(s): 237 – 242.



Komal Asrani, female, is serving as Asstt.Professor in the Babu Banarasi Das National Institute of Engineering and Technology, Lucknow. She has completed her M.S in Software Systems from B.I.T.S, Pilani in 2005.Presently, she is pursuing her Ph.D. Her current research focuses on image retrieval, pattern recognition.



Dr. Renu Jain, female is presently H.O.D, Deptt. of C.S.E., U.I.E.T , C.S.J.M., Kanpur. She completed her Ph.D in Computer Science and Engineering from B.I.T.S, Pilani in 1996. Her research areas include Artificial Intelligence, Digital Image Processing and Pattern Recognition.