BIOSYSTEMS ENGINEERING 139 (2015) 66–75

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Automatic classification of plants based on their leaves

Aimen Aakif [a,*], *Muhammad Faisal Khan* [b,1]

[a] *Department of Electrical Engineering, College of Telecommunication Engineering, National University of Sciences and Technology, Pakistan*

[b] *Department of Computer Software Engineering, College of Telecommunication Engineering, National University of Sciences and Technology, Pakistan*

ARTICLE INFO

Article history:
Received 5 November 2014
Received in revised form
3 July 2015
Accepted 7 August 2015
Published online 29 August 2015

Keywords:
Classification of plants
Leaf segmentation
Artificial neural networks
Feature extraction
Shape defining feature

The proposed algorithm identifies a plant in three distinct stages i) pre-processing ii) feature extraction iii) classification. Different leaf features, such as morphological features, Fourier descriptors and a newly proposed shape-defining feature, are extracted. These features become the input vector of the artificial neural network (ANN). The algorithm is trained with 817 samples of leaves from 14 different fruit trees and gives more than 96% accuracy. To verify the effectiveness of the algorithm, it has also been tested on Flavia and ICL datasets and it gives 96% accuracy on both the datasets.

1. Introduction

Automatic classification of plants is an important step for solving general problems like yield prediction, growth estimation and health prediction. Traditionally, botanists classify plants based on their floral parts (Mishra, Maurya, Singh, & Misra, 2012), fruits and leaves. Flowers and fruits may not be the best choice for automatic plant identification as they appear during a limited period. However, leaves are numerous in number and are present for most of the year, which make them suitable for computerised plant classification.

Leaf features that have been used in the literature include different statistical features like eccentricity, aspect ratio, circularity, roundness perimeter etc., and leaves have also been identified using Fourier descriptors (Neto, Meyer, Jones, & Samal, 2006; Singh, Gupta, & Gupta, 2010), wavelets (Gu, Du, & Wang, 2005; Im, Nishida, & Kunii, 1998); vein structure, saw-toothed pattern, texture of a leaf (Rashad, El-Desouky, & Khawasik, 2011); centroid contour distance

* *Corresponding author*. College of Telecommunication Engineering (NUST), Humayun Road, Rawalpindi, Pakistan. Tel.: +92 3335386564.
E-mail addresses: aimen@mcs.edu.pk (A. Aakif), mfaisalkhan@mcs.edu.pk (M.F. Khan).

[1] Dr. Muhammad Faisal Khan is currently working as Assistant Professor in the Department of CSE, College of Telecommunication Engineering (NUST), Humayun Road, Rawalpindi, Pakistan. Tel.: +92 3215250737.

http://dx.doi.org/10.1016/j.biosystemseng.2015.08.003

(Bong, Sulong, & Rahim, 2013; Wang, Chi, & Feng, 2003); and leaf colour (Kadir, Nugroho, Susanto, & Santosa, 2011).

Du, Wang, and Zhang (2007) extracted different morphological features and Hu moments but did not use the contour information. Some of the features like aspect ratio and rectangularity were based upon the physiological length and width of the leaf, which require a human operator to mark the extreme points on the leaf image. Wu et al. (2007) used morphological features and the vein features to classify plants. Although the selected features gave accuracy as high as 90%, the features still required physiological length and width.

Neto et al. (2006) used Elliptic Fourier technique to analyse the shape of a leaf to classify different plant species at different stages of growth. Although the system was fully automatic, the accuracy it could achieve was almost 88.4%.

Beghin, Cope, Remagnino, and Barman (2010) combined the shape and texture features to make a feature set. The accuracy they could achieve was almost 81%. Different classification techniques like support vector machines (SVM) (Zhang, Yanne, & Liang, 2011), probabilistic neural network (PNN) (Kadir et al., 2011; Wu et al., 2007), moving media centres hyperspheres (MMC) (Du et al., 2007; Zhang et al., 2004) and ANN with back-propagation (Heymans, Onema, & Kuti, 1991; Satti, Satya, & Sharma, 2013) have been used for plant classification.

Here we propose an algorithm which automatically classifies a plant based on its leaf. For this research, we have created a database in the laboratory consisting of leaves of 14 different fruit trees. We identify different morphological features, Fourier descriptors and a newly developed feature, called shape-defining feature (SDF), for classification. SDF's basic purpose is to retrieve the shape of a leaf along with its fine serrations. The proposed algorithm uses ANN with back propagation as a classifier as it is more general than PNN and less sensitive to noise than SVM. To verify its effectiveness the algorithm has also been tested on Flavia (Wu et al., 2007) and ICL (Intelligent Computing Laboratory, Chinese Academy of Sciences) datasets.

2. Materials

We have used three different leaf datasets to test our algorithm. These datasets include Flavia, ICL and our own dataset. Our dataset consists of leaves of 14 different seasonal fruit trees that grow in particular regions of Asia. Some fruit trees like apple and apricot grow at higher places and in cold regions like Azad Kashmir (Kotli – a small district of Azad Kashmir, Pakistan) and Murree (Pakistan). We collected leaves from the 3–4 different trees during the month of September (year 2012 and year 2013). Fig grows in warm and plain regions like Punjab (Pakistan). We collected leaves from three different fig trees located in Rawalpindi (Pakistan) during the month of July (2012 and 2013). Leaves of oranges were collected from four different trees located in Rawalpindi (Pakistan) during winter season. These trees belong to different areas and produce fruit in different seasons. We collected these leaves over two years (year 2012 and year 2013). These leaves were collected from different parts of the tree, from different heights and also of different sizes. Criteria for selecting a leaf were that it should be green in colour and it should not be ruptured. Table 1 shows the names of the trees and the number of their sample images. We intend to publish this dataset for future references.

Two different digital cameras, DSLR A100 and DSC TX7, were used to capture the images of the leaves. DSLR A100 uses a 10.2 MP CCD sensor and DSC TX7 uses 10MP CCD sensor.

A leaf without stalk was placed on a white sheet and imaged with the camera. All images were in RGB format. Out of 2525 sample images, 817 images were selected randomly as the training data and of the remaining 1708 were used as test images. Figure 1 shows sample images of 14 types of leaves.

There are some publicly available leaf image datasets such as the Flavia dataset (Wu et al., 2007), the Smithsonian Leaf dataset (Belhumeur et al., 2008), Swedish Leaf dataset (Soderkvist, 2001), CLEF dataset (Goëau et al., 2013) provided by the CLEF and ICL dataset. Among these, we tested our algorithm on Flavia and ICL datasets to verify its effectiveness. Flavia dataset consists of 32 different classes and ICL dataset consists of 20 different types of plants.

3. Methods

The algorithm has three distinct stages i) pre-processing ii) feature extraction and iii) classification, as shown in Fig. 2. In the pre-processing stage, the leaf is segmented from the background using colour segmentation and then it is normalised. Feature extraction follows: the extracted features include different types of morphological features, Fourier descriptors (FD) and a newly proposed feature called shape-defining feature (SDF). These features become the input vector of the artificial neural network (ANN) in the classification stage. ANN with back propagation classifies the leaf based on the extracted features.

3.1. *Pre-processing*

Pre-processing involves two subtasks i.e. leaf segmentation and image normalisation.

3.1.1. *Leaf segmentation*

The input RGB image (I_c (X, Y)) is converted into greyscale image (I_g (X, Y)) using the Equation (1) (Gitelson, Kaufman, Stark, & Rundquist, 2002; Lamm, Slaughter, & Giles, 2002).

$$I_g = \frac{2*G - B - R}{G + B + R} \tag{1}$$

where R, G and B represent the red, green and blue colours respectively of pixel intensities of the input image.

The greyscale image (I_g (X, Y)) is converted into binary image (I_b (X, Y)) using the global thresholding technique (Gonzalez & Woods, 2008) in which we select the valley in between the two peaks and use it as a threshold.

Table 1 – Details of leaf datasets collected over two years (2012 & 2013).

Scientific name	Common name	Training samples		Testing samples	Total
		Year I	Year II		
Ficus carica	Fig	55	37	65	157
Malus domestica	Apple	50	76	60	186
Prunus armeniaca	Apricot	60	52	62	174
Grewia asiatica	Phalsa	55	47	87	189
Syzygium cumini	Java Plum	57	47	97	201
Vitis vinifera	Grapes	68	68	98	234
Psidium	Guava	80	60	80	220
Mangifera indica	Mango	60	25	75	160
Citrus aurantium	Orange	50	45	48	143
Litchi chinensis	Lychee	70	40	58	168
Citrus limon	Lemon	60	50	114	224
Eriobotrya japonica	Loquat	60	57	60	177
Pyrus	Pear	50	50	60	160
Punica granatum	Pomegranate	42	40	50	132
Total		**2525**			

3.1.2. *Image normalisation*

Normalisation includes i) rotation of a leaf in such a way that its tip should be at the top and the angle between the major axis of the leaf and the major axis of the frame should be zero, ii) the centroid of the leaf and the centroid of the frame should be same and iii) the size of the frame should be the same for all the sample images regardless of the resolution of image capturing devices and the size of the leaf. We have fixed the size of the frame of the image to 4500 × 3000 pixels. In order to normalise an image we rotate the leaf around its centroid and translate it to the centroid of the frame using the following equation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{2}$$

where X, Y represents the original coordinates of the image and Θ is the angle between the leaf and the frame, t_X and t_Y shows the displacements along x-axis and y-axis respectively. Figure 3(a) shows the original image with size 2592 × 3872. Here, the leaf is making an angle of 76.67° with the y-axis. Figure 3(b) shows the normalised image.

3.2. *Feature extraction*

In the proposed algorithm three different types of features, namely morphological features, Fourier descriptors and a newly proposed feature called shape-defining feature are computed. Morphological features include aspect ratio, eccentricity, roundness and convex hull.

3.2.1. *Morphological features*

The morphological features are extracted on the basis of the central moments (Teague, 1980).

Aspect Ratio (AR): It is defined as the ratio of length of the major axis to that of the minor axis of the leaf. These lengths are calculated using the image moments (Teague, 1980).

Eccentricity (E): It is a measure that shows how much an object deviates from a circle. Eccentricity of a complete round object is 0 and that of a line is 1.

Roundness (R): Roundness of a leaf of an irregular shape is closer to zero.

Convex Hull (CH): In the field of mathematics area of convex hull (CH) of a set of points, is the area of the minimum

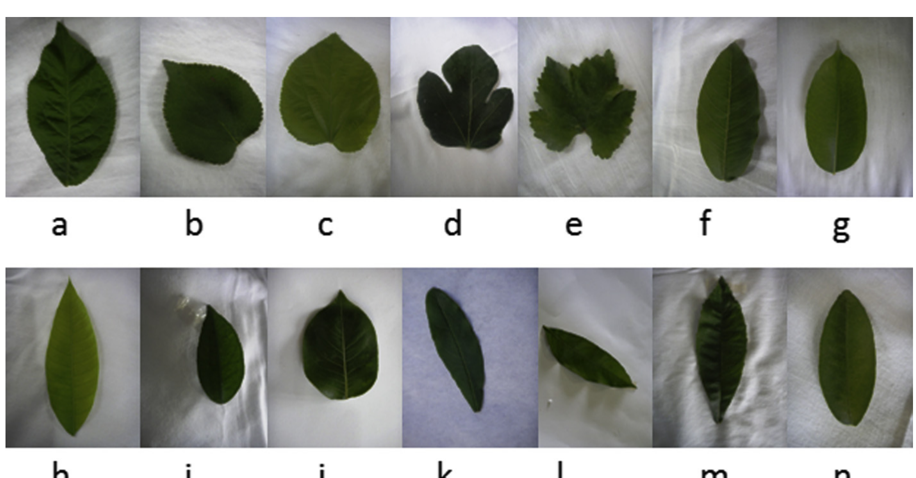

Fig. 1 – Sample images of leaves of each type (Table 1): From a to n: Apple, Apricot, Phalsa, Fig, Grapes, Guava, Java Plum, Mango, Orange, Pear, Pomegranate, Lychee, Loquat, and Lemon respectively.

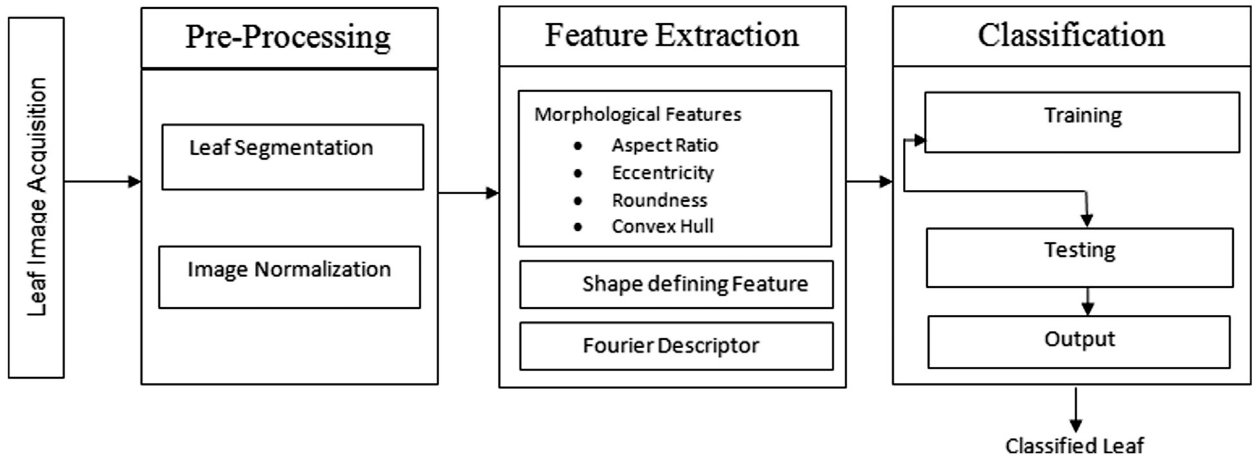

Fig. 2 – Block diagram of the proposed algorithm showing the three stages.

convex set which enclose all these points Quick hull algorithm is used to extract the minimum convex set (Barber, Dobkin, & Huhdanpaa, 1996).

3.2.2. *Shape-defining feature* (SDF)

This is a new idea to reconstruct a shape of an object. In this method a set of horizontal lines and vertical lines is drawn over the body of a leaf. The endpoints of these lines are extracted. Once all the boundary points have been extracted then the slope between the consecutive points is calculated. This slope is used to retrieve the shape of a leaf.

As the leaves are of different sizes, we must determine the distance between two consecutive lines. For this purpose, we use the following formula

$$s_v = \frac{l_{min}}{N_v} \quad , \quad s_h = \frac{l_{maj}}{N_h} \tag{3}$$

where s_v, s_h represent step size between the vertical lines and horizontal lines respectively, l_{min} and l_{maj} are the lengths of minor axis and major axis respectively, N_v and N_h are the number of lines to be drawn vertically and horizontally, respectively.

Initially a line is drawn between the two endpoints of the major axis of the leaf, passing through the centroid, say C_{leaf} (x,y). Compute c_{2x} by adding step size s_v in the x-coordinate of C_{leaf} and keeping its y-coordinate constant. Now draw another straight vertical line passing through c_{2x}.

$$c_{2x}(x', y) = C_{leaf}(x + s_v, y) \tag{4}$$

The line should extend in both directions until and unless it finds a black pixel (as we are processing binary images). These are actually the boundary pixels. Repeat this procedure by reducing the x-coordinate by step size s_v.

Similarly, a straight horizontal line, passing through the centroid, $C_{leaf,}$ is drawn between the endpoints of minor axis. Compute c_{2y} by adding s_h in the y-coordinate of C_{leaf} while keeping the x-coordinate constant.

$$c_{2y}(x, y') = C_{leaf}(x, y + s_h) \tag{5}$$

Now draw a straight horizontal line passing through the c_{2y}. Repeat the procedure by reducing the y-coordinate by s_h.

Once we have drawn the lines horizontally and vertically, we have store their endpoints for future reference. The next

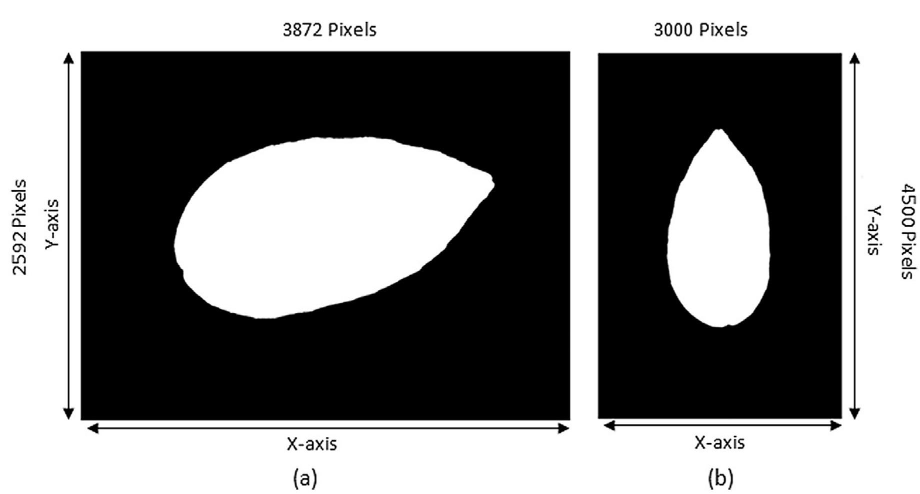

Fig. 3 – (a) Original position of the leaf in binary image and (b) Normalised image.

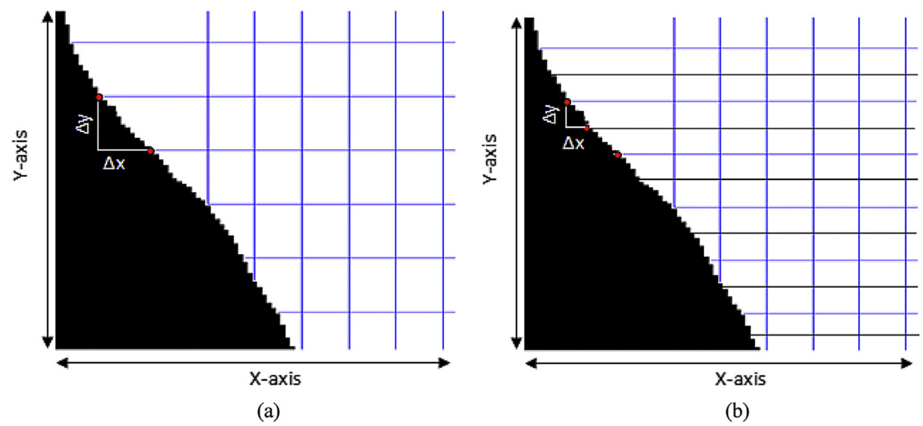

Fig. 4 – A portion of a sample leaf.

step is to calculate the length of the lines using the distance formula:

$$dist = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{6}$$

As we have taken leaves of different sizes from the same tree, in order to normalise their size we take the ratio of these lines with their major axis.

$$R_{maj} = \frac{l_{Maj1}}{L_{Maj}} \tag{7}$$

where l_{Maj1} represents length of one of the lines drawn parallel to the major axis and L_{Maj} is the length of major axis, R_{maj} is the ratio of these two lines. Similarly, the ratio for all the lines drawn parallel to the major axis is calculated. The same procedure is repeated for the lines drawn horizontally.

In the proposed algorithm, 200 lines are drawn horizontally and 200 lines are drawn vertically, so in total 400 lines are drawn over the body of a leaf. All the endpoints of these lines are extracted.

The next step is to determine the slope between the two consecutive endpoints using the following formula:

$$m = \frac{\Delta y}{\Delta x} \tag{8}$$

where Δy is the difference between the two consecutive points along y-axis and Δx is the difference between two consecutive points along x-axis.

We normalised the values of the slope between 0 and 100: 0–49 is the range for negative slopes and 51–99 is the range for positive slopes, where 0 is the value for the maximum negative value and 99 is the value for the maximum positive value.

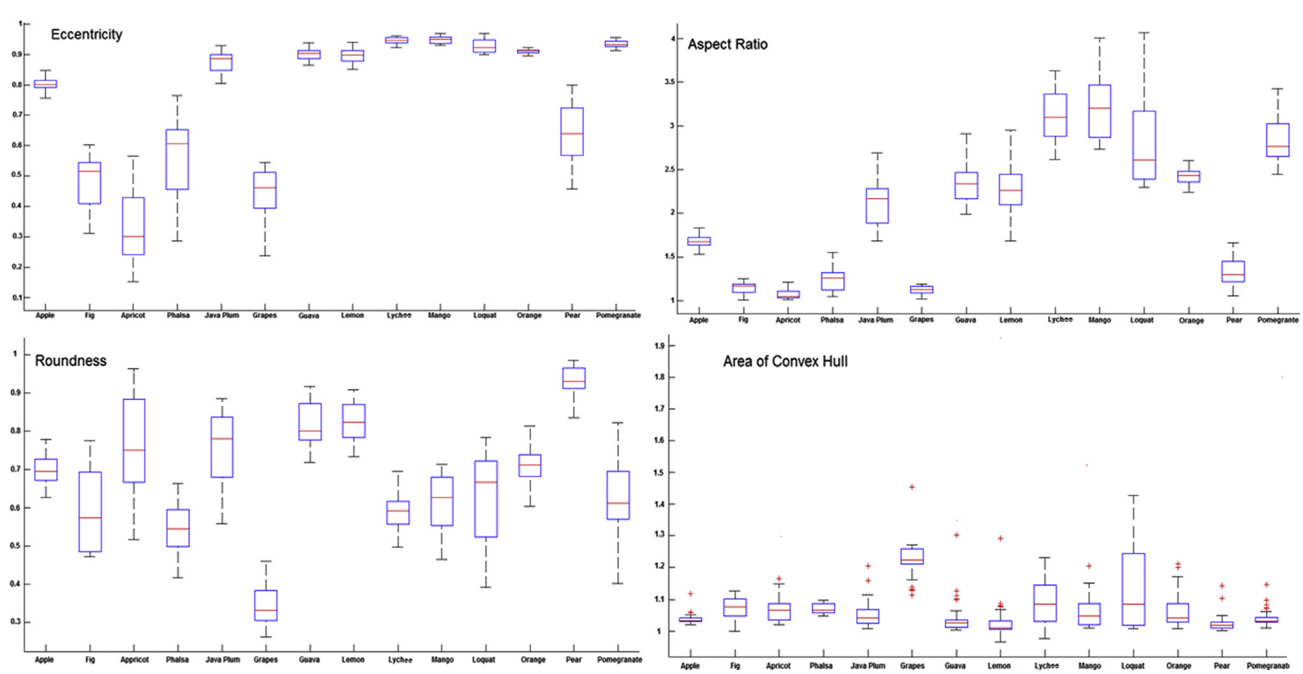

Fig. 5 – Box plot of eccentricity, aspect ratio, roundness and area of convex hull of samples of leaves as listed in Table 1.

Table 2 – Mean square error between the reconstructed image using 512 coefficients and the reconstructed image using 52 (first 26 and the last 26) coefficients.

Specie name	Mean square distance
Apple	0.00144
Fig	0.00157
Apricot	0.001092
Phalsa	0.00139
Grapes	0.00193
Java Plum	0.00108
Guava	0.0009914
Lychee	0.00089
Lemon	0.00118
Loquat	0.0013
Mango	0.00106
Orange	0.00125
Pear	0.00121
Pomegranate	0.000743

Value 50 refers to slope zero. In the case where the denominator is zero, we take the slope to be 255, maximum value of a slope, and incorporate it as a special feature. Figure 4 shows a portion of leaf with the lines drawn on it, with dots showing the ends of two consecutive lines.

Here, some serrations are also visible in Fig. 4a. But if we increase the number of lines then the two points becomes closer and reducing the values of Δx and Δy as shown in Fig. 4b. In Fig. 4b more lines (black in colour) have been introduced. The main purpose of SDF is to retrieve the shape of the leaf along with the serrations shown in Fig. 4, which is only possible if the slope between the boundary points is known and the boundary points should be closer. If we increase the number of lines, the distance between the consecutive lines decreases, thus decreasing the Δx and Δy.

3.2.3. *Fourier descriptor (FD)*

Fourier descriptors are obtained by applying a Fourier transform on a shape signature (Yang, Zhao, & Lan, 2010).

Let $r(n)$ is the distance of the boundary coordinates from the centroid

$$r(n) = \sqrt{(x(n) - \bar{x})^2 + (y(n) - \bar{y})^2} \tag{9}$$

Applying 1-D Fourier transform to the sequence r (n):

$$a_k = \frac{1}{B}\sum_{n=0}^{B-1} r(n)e^{i\theta(n)k} \quad k = 0,\ 1, \ldots,\ B-1 \tag{10}$$

where $\theta(n) = 2\pi\frac{n}{B}$, for $n = 0, 1, \ldots, B-1$ and a_k are the coefficients of the FDs. In this research, 512 coefficients are computed.

3.3. *Classification*

In the proposed algorithm, a multilayer neural network with back propagation is employed for classification.

3.3.1. *Architecture of artificial neural network*

The size of input layer is equal to that of the input feature set. Each input is weighted with the appropriate weight. Sigmoid function is used as a transfer function as it gives output in the range of 0 and 1 (Grossberg, 1982; Williams, 1983).

$$sigmoid = \frac{1}{1 + f^{-\sum(v_i w_{ij}) + b}} \tag{11}$$

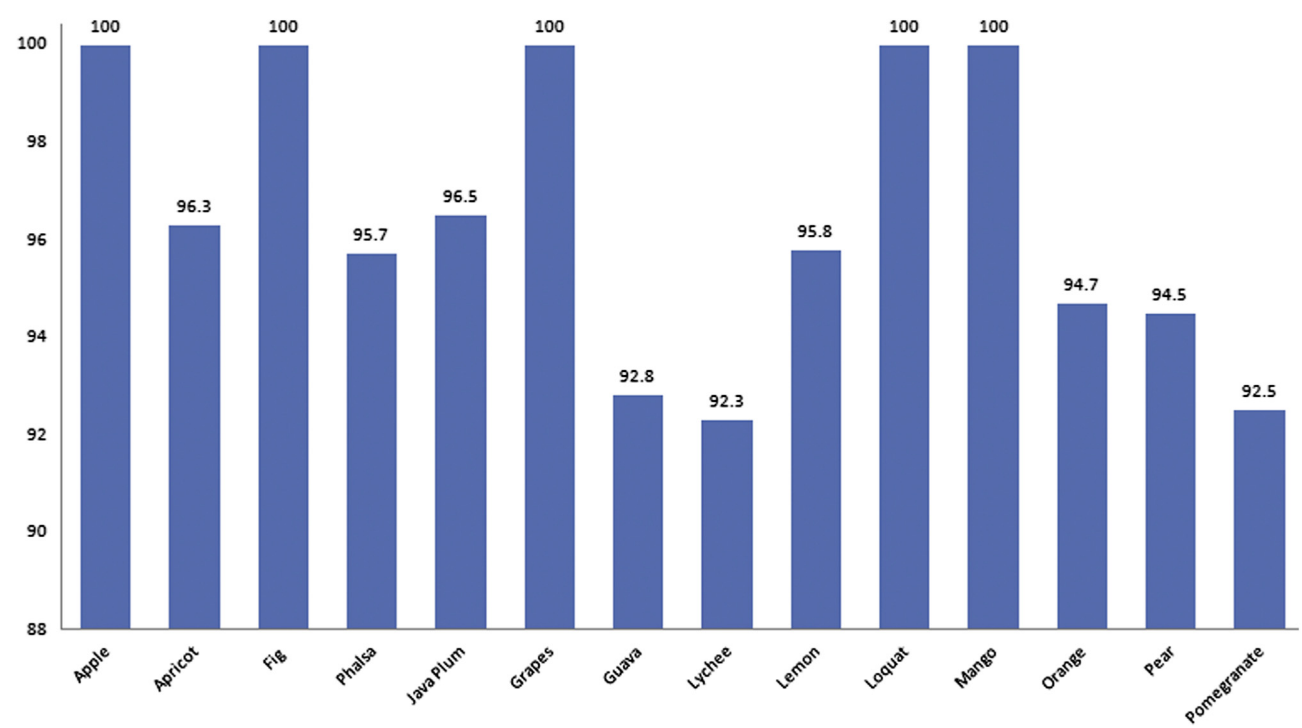

Fig. 6 – Accuracy on own dataset.

Table 3 – Accuracy comparisons of features.

Feature name	Accuracy
Aspect ratio, Roundness, Eccentricity, Convex hull	68.3%
Fourier descriptors	77.8%
SDF	83.6%
All combined	96.5%

where v_i refers to ith input, w_{ij} is the weight factor for the jth node, b is the bias and f is actually the activation function.

3.3.2. Network training

Training and learning functions are mathematical procedures used to automatically adjust the network's weights and biases to optimise the network performance. The network performance is measured by the average of the squared error (mse) between the network outputs and the targets.

$$mse = \frac{1}{L}\sum_{i=1}^{L}(C_i - O_i)^2 \tag{12}$$

where C_i represents the target class and O_i represents the actual output. The error is propagated back towards the hidden layer. The next step is to update the weights and biases using the error derivatives. In this experiment, 817 training samples have been used to train the network for 14 classes. Initially the weights were set to some random values. The training set consists of N data pairs (v_i, C) as where v_i refers the input and C is the corresponding target class. The weights are updated after each training cycle in order to minimise the overall network error. The training of the network continues until the overall error has crossed the threshold (10^{-5}) level or fixed number of iterations (e.g 1000) has been completed.

4. Results and discussion

We randomly selected 1708 samples from the dataset to evaluate the proposed algorithm. The algorithm uses different morphological features and Fourier descriptors to recognise the type of the leaf. Box plots of eccentricity, aspect ratio, roundness and area of convex hull of 14 different types of leaves, are shown in Fig. 5. As evident from the plot of eccentricity, the samples of apple, lychee, mango, orange, guava, lemon and pomegranate shows less dispersion than the others. In the case of java plum, the median is closer to the 3rd quartile and that of loquat is closer to the 1st quartile. Fig, apricot, phalsa and pear show the maximum variation and spread. This is due to the huge variation that exists among the leaves of the same class even if they belong to the same tree. As is evident from the box plot of aspect ratio, apple, fig, apricot, grapes and orange show the minimum variance. Loquat has the maximum spread and its upper whisker has the maximum value. However, its median is closer to its 1st quartile. Here lychee, mango and loquat overlap each other. As is visible from the box plots of roundness, grape shows minimum roundness because of its irregular shape. Guava and lemon again overlap each other and their upper whisker is close to 0.9 as the younger leaves of guava and lemon are loosely round in shape. Apricot and loquat show the maximum spread and variance. Orange, apple and pear show the minimum variance. Mango, lychee, loquat and

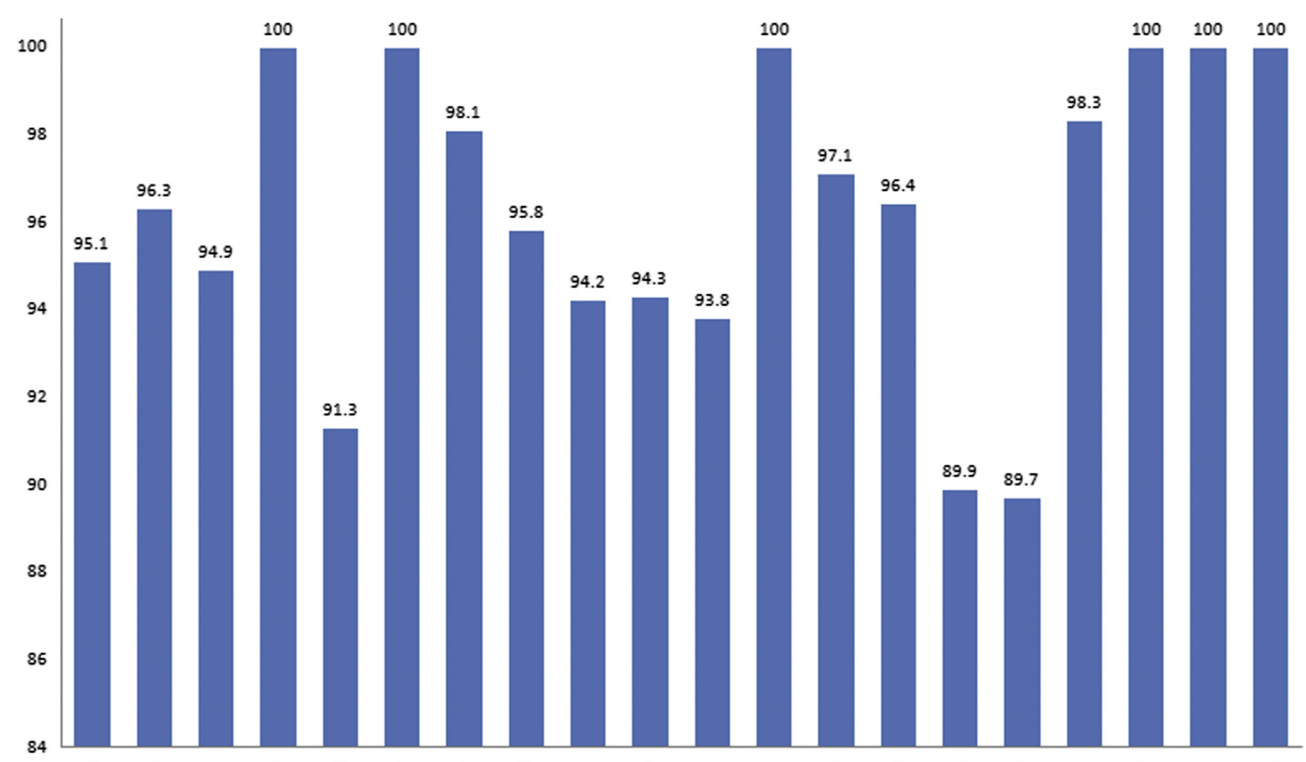

Fig. 7 – Accuracy on ICL dataset.

Table 4 – Details of Flavia dataset.

Common name	Wu et al. (2007)	Proposed algorithm
Pubescent Bamboo	100	100
Chinese Toon	90	90
Chinese Horse chestnut	100	100
Trident Maple	90	90
Chinese Redbud	90	100
True Indigo	100	100
Southern Magnolia	100	90
Japanese Maple	100	100
Nanmu	90	100
Castor Aralia	100	100
Oleander	100	90
Goldenrain Tree	100	100
Chinese Cinnamon	80	100
Anhui Barberry	100	100
Big-Fruited Holly	100	100
Japan Viburnum	80	90
Japanese Cheesewood	90	100
Camphor tree	70	90
Winter Sweet	80	90
Beale's Barberry	100	100
Ford Woodlotus	70	90
Crepe Myrtle	100	90
Peach	80	90
Ginkgo	100	100
Glossy Privet	90	100
Sweet Osmanthus	50	80
Yew Plum Pine	100	90
Deodar	100	100
Japanese Flowering Cherry	100	100
Canadian Poplar	70	100
Chinese Tulip Tree	100	100
Tangerine	100	100
Total	91.6%	96%

pomegranate overlap each other as they are elongated in shape. The box plot of the area of convex hull also shows the outliers. The lobed leaves have more surface area of the convex hull than the leaves without lobes. That is why grape shows maximum median of convex hull. Loquat shows the maximum variation in the convex hull. Apple, pear and pomegranate show the minimum variance as compared to others.

The SDF is useful in reconstructing the shape of the leaf including the fine detail of the leaf edge. This feature is helpful in classifying a leaf whose statistical features are being overlapped by other leaves.

In the proposed algorithm we have also used the 512 coefficients of FD to get the refined details of the leaf edges. In order to cut down the size of the feature set we have selected the first 26 and last 26 coefficients using the two fold cross validation. The mean square error between the reconstructed image using 512 coefficients and reconstructed image using 52(first 26 and last 26) coefficients is shown in the the Table 2.

The algorithm is trained on 817 sample images and 1708 images have been selected randomly for testing. Figure 6 shows that apple, fig, mango, loquat and grapes gave 100% accuracy, and on average the algorithm gives 96.5 ± 1% accuracy. There was very little variance between the leaves of apple as evident from the box plot of eccentricity, aspect ratio,

Table 5 – Accuracy comparison on leaf dataset.

Name of algorithm	Type of classifier	Features	Leaf dataset	Accuracy
Wu et al. (2007)	PNN	Geometric features, morphological features, vein structure	Flavia	90%
Du et al. (2007)	MMC hypersphere classifier	Morphological features, Hu moments	20 species of leaves	91%
Zhang et al. (2011)	SVM wavelet transforms	GLCM, geometric features,	Flavia	93.1%
Kadir et al. (2011)	PNN	Morphological features, Texture, Colour	Flavia	93.75%
Arun Priya et al. (2012)	SVM	Morphological features, geometric features, vein structure	Flavia	94.2%
Proposed algorithm	ANN with back propagation	Morphological features, SDF, FDs	Flavia, ICL, own database	96%

roundness and convex hull, leaves of grape and fig are irregular in shape, which make them different from others. Apricot, guava, lemon, phalsa, java plum and Orange show accuracy greater than 90%.

Table 3 gives the comparisons of the features used in this algorithm. If Aspect ratio, Eccentricity, Roundness and Area of Convex Hull are used as a feature set, then they give 68% accuracy. When feature set consists of FDs only, they give the accuracy of 77.8%. If SDF is used alone, it gives 83.6% accuracy. When all of these features are combined we get 96.5% accuracy.

To evaluate the effectiveness of the proposed algorithm, we tested it on two other datasets as well, namely ICL and Flavia datasets. Figure 7 shows the classification accuracy on ICL dataset that consists of 20 different classes. The overall average accuracy that our algorithm shows is 96.3%.

The Flavia dataset consists of 32 different classes and each class has almost 40 leaf samples, out of which 10 samples for each species were selected at random for the testing. On average, the algorithm gives 96% ± 1%accuracy. Table 4 shows the dataset and the percentage of incorrectly classified leaves. Our algorithm shows better performance while classifying camphor tree, sweet osmanthus and Canadian poplar. The accuracy has been improved due to the inclusion of SDF. As the shape of Canadian coplar is very irregular and unique, it can be easily classified. In the case of sweet osmanthus and camphor tree, both have pointed tips, but SDF helps in identifying them, as the width of the Camphor tree leaf is quite different at different levels as compared to sweet osmanthus.

Table 5 gives the comparison of accuracy on different leaf dataset achieved through different techniques and algorithm. Wu et al. (2007) used geometric features, morphological features and the vein structure. Although the algorithm is quite simple and gives good accuracy, the only limitation is the use of physiological length and width, which requires a human to intervene. Du et al. (2007) used morphological features and Hu moments to classify the leaves of 20 species. Zhang et al. (2011) used wavelets and texture of a leaf in addition to the traditional geometric features. The leaf image is decomposed at three levels, thus generating a redundant set of features, which in turn affect the accuracy. Kadir et al. (2011) in addition to morphological features and texture, also used colour moments as a part of the feature set. Colours sometimes may change with the season and depend upon the sunlight and the nutrients a plant receives. Arun Priya, Balasaravanan, and Thanamani (2012), also employed morphological features, geometric features and the vein structure but used SVM as a classifier. Again, the features are dependent upon the physiological length and width. Our algorithm achieves comparable accuracy even without using texture.

Literature review shows that most of the algorithms have used either PNN or MMC as a classifier; in this algorithm, we have used ANN with back propagation and accuracy we achieved is comparable to all these algorithms.

5. Conclusion

We have proposed an algorithm that automatically classifies a plant leaf. It incorporates morphological features, FDs and newly developed shape defining-feature, SDF. The experimental results have demonstrated the effectiveness and the robustness of the proposed algorithm. It is applicable to any leaf dataset and we have tested it on three different sets and achieved accuracy greater than 96%.

REFERENCES

Arun Priya, C., Balasaravanan, T., & Thanamani, A. S. (2012). An efficient leaf recognition algorithm for plant classification using support vector machine. In *Proceedings of International Conference on Pattern Recognition, Informatics and Medical Engineering*.

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quick hull algorithm for convex hulls. *ACM Transactions on Mathematical Software, 22*(4), 469–483.

Beghin, T., Cope, J. S., Remagnino, P., & Barman, S. (2010). Shape and texture based plant leaf classification. In *ACIVS, Part II* (pp. 345–353).

Belhumeur, P. N., Chen, D., Feiner, S., Jacobs, D. W., Kress, W. J., Ling, H., et al. (2008). Searching the Worlds Herbaria: a system for visual identification of plant species. In *ECCV*.

Bong, M. F., Sulong, G., & Rahim, M. S. M. (March 2013). Recognition of leaf based on its tip and base using centroid contour gradient. *International Journal of Computer Science Issues, 10*(2).

Du, J. X., Wang, X. F., & Zhang, G. J. (2007). Leaf shape based plant species recognition. *Applied Mathematics and Computation, 185*(2), 883–893. Amsterdam, Elsevier.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithm for remote estimation of vegetation fraction. *Remote Sensing of Environment, 80*, 76–87 (Elsevier).

Goëau, H., Bonnet, P., Joly, A., Bakic, V., Barthelemy, D., Boujemaa, N., et al. (2013). The image CLEF 2013 plant identification task. In *Image CLEF 2013 Working Notes*.

Gonzalez, R. C., & Woods, R. C. (2008). *Digital Image Processing* (3rd ed.). Prentice Hall.

Grossberg, S. (1982). *Studies of mind and brain: Neural principals of learning, perceptions, development, cognition and motor control*. Boston: Reidel Press.

Gu, X., Du, J. X., & Wang, X. F. (2005). Leaf recognition based on the combination of wavelet transform and gaussian interpolation. In *ICIC 2005, Part I, LNCS 3644* (pp. 253–263).

Heymans, B. C., Onema, J. P., & Kuti, J. O. (1991). A neural network for opuntia leaf-form recognition. In *Proceedings of International Joint Conference on Neural Network, 18–21 Nov (1991), Singapore* (Vol. 3, pp. 2116–2121).

Im, C., Nishida, H., & Kunii, T. L. (1998). Recognizing plant species by leaf shapes – a case study of the Acer family. In *Proceedings of Conference on Pattern Recognition* (Vol. 2, pp. 1171–1173).

Intelligent Computing Laboratory, Chinese Academy of Sciences Homepage http://www.intelengine.cn/English/dataset Accessed 02.07.14.

Kadir, A., Nugroho, L. E., Susanto, A., & Santosa, P. I. (2011). Leaf classification using shape, color, and texture features. *International Journal of Computer Trends and Technology, 3*(3), 256–263.

Lamm, R. D., Slaughter, D. C., & Giles, D. K. (2002). Precision weed control for cotton. *Transactions of ASAE, 45*(1), 231–238.

Mishra, P. K., Maurya, S. K., Singh, R. K., & Misra, A. K. (2012). A semi automatic plant identification based on digital leaf and flower images. In *IEEE (ICAESM)*.

Neto, J. C., Meyer, G. E., Jones, D. D., & Samal, A. K. (2006). Plant species identification using Elliptic Fourier leaf shape analysis. *Computers and Electronics in Agriculture*, 121–134.

Rashad, M. Z., El-Desouky, B. S., & Khawasik, M. S. (2011). Plants images classification based on textural features using combined classifier. *International Journal of Computer Science and Information Technology, 3*(4).

Satti, V., Satya, A., & Sharma, S. (2013). An automatic leaf recognition system for plant identification using machine vision technology. *International Journal of Engineering Science and Technology (IJEST)*. ISSN: 0975-5462, *5*(04).

Singh, K., Gupta, I., & Gupta, S. (2010). Svm-bdt pnn and fourier moment technique for classification of leaf shape. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 3*(4), 67–78.

Soderkvist, O. (2001). *Computer vision classification of leaves from Swedish trees*. Master's Thesis. Linkoping University. Available online at http://www2.cvl.isy.liu.se/ScOut/Masters/Papers/Ex3132.pdf.

Teague, M. R. (August, 1980). Image analysis via general theory of moments. *Journal of Optical Society of America, 70*(8).

Wang, Z., Chi, Z., & Feng, D. (2003). Shape based leaf retrieval. *IEEE Proceedings Vision, Image and Signal Processing, 150*(1).

Williams, R. J. (1983). *Unit activation rules for cognitive network models*. ICS Report 8303. Institute of Cognitive Science, University of California at San Diego.

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, IEEE Int. Symp., Cairo, Egypt*.

Yang, J. W., Zhao, Z. X., & Lan, R. S. (July 2010). Generic ring Fourier descriptors for shape description. In *Proceedings of the 2010 International Conference on Wavelet Analysis and Pattern Recognition, Qingdao*.

Zhang, G. J., Wang, X. F., Huang, D. S., Chi, Z., Cheung, Y. M., Du, J. X., et al. (2004). A hypersphere method for plant leaves classification. In *Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP 2004), Hong Kong, China* (pp. 165–168).

Zhang, H., Yanne, P., & Liang, S. (2011). Plant species classification using leaf shape and texture. *American Journal of Engineering and Technology Research, 11*(9).