Plant leaf recognition based on SVM

CS220 Final Report

Wenting wu

(151518)

DECEMBER 7, 2016

KAUST

1. Introduction

Plant leaves are very common in our daily life and it plays a very crucial role in human society sharing a close relationship of human beings, for example, some plant leaves can be used as medicinal materials. A very urgent situation is that some plants are at the risk of extinction (Wu et al., 2007). However, since there are hundreds of thousands of plants exist on earth, it is difficult for ordinary people to recognize. So providing a quick recognition and classification method is necessary. Plant leaves classification has become a hot area and there are many methods presented in literature. Some researches still involve human interference in feature extraction which means that those methods require prior domain knowledge of non-expert users. Obviously, it is not the best solution to classify plant leaves. For this reason it's of great importance to develop an automatic image processing technique to extract leaf feature without manually operations. In this work, we train a SVM (support vector machine) using easy-to-extract features. There are numbers of features and parameters having been proposed for plant leaf classification. Sakai extracted some geometrical parameters to classify leaf types such as length, width, area, perimeter (Sakai et al., 1996). Wu combined 5 basic geometric features and 8 digital morphological features using probabilistic Neural Network to classify different leaf species (Wu et al., 2007). Apart from geometrical features, Munisami employed a distance map, convex hull area and perimeter, color histogram with k-nearest neighbour classifiers (Munisami et al., 2015). In this paper, the main improvements are on good feature extraction and a good multi-class SVM classifier in order to ensure a higher accuracy as well as the speed of classification.

The organization of the rest of this paper is as follows. Section 2 shows the image pre-processing. Section 3 describes how 19 features are extracted. The classification steps are discussed in section 4. Experimental results are given in section 5 and conclusions in section 6.

2. Image pre-processing

Plant leaf image pre-processing is crucial for feature extraction. We use Flavia dataset which is shared by Wu et al. (Wu et al., 2007). In this dataset almost all the images present a white and uniform background which is not difficult task to deal with. The whole process of proposed scheme including image pre-processing, feature extraction and classification is shown as Fig.1.

2.1 Converting RGB image to binary image

Firstly we have converted the 2 dimensional image of plant leaf to gray scale image using the function Gray = 0.2999 * R + 0.578 *G+ 0.114 * B where R, G, B correspond to the color of the pixel, respectively. The image then transformed into a gray scale image. To make the pre-processing easier and faster, we use a predetermined threshold 0.95 calculated according to the RGB histogram which is the value for converting gray scale image into binary image.

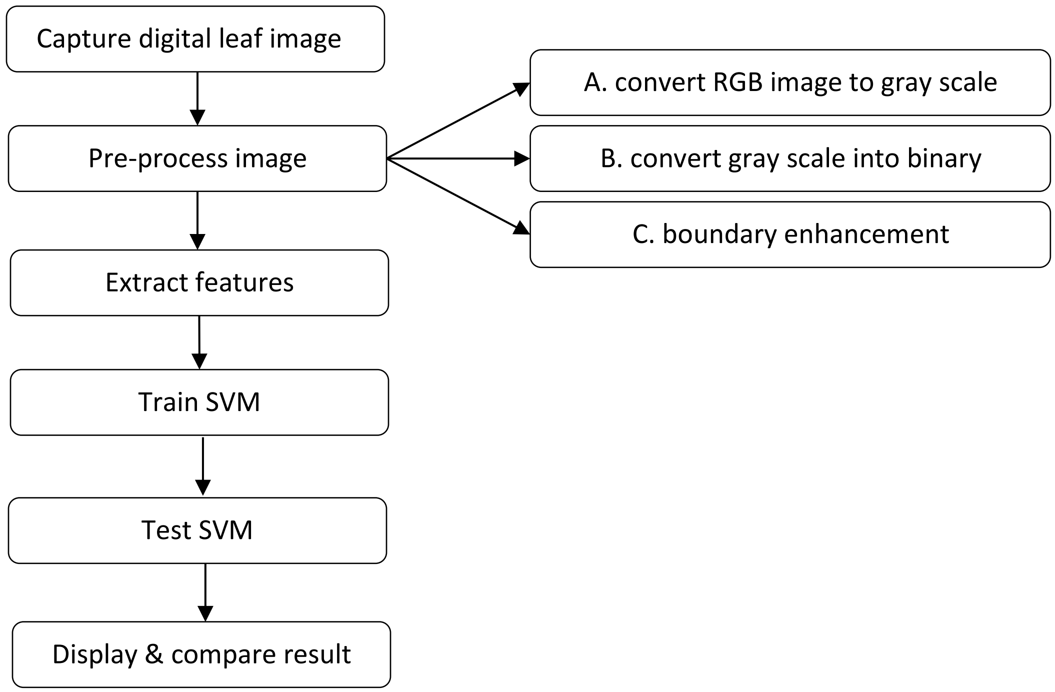

Fig 1. Flow diagram of proposed scheme

2.2 Boundary enhancement

In this stage, we use a rectangular averaging filter of size 3*3 to filter noises. However, after averaging filter, the image still not very clean. There are some dark areas in background and white areas inside the leaf. Therefore an opening operation and a hole filling operation are needed. We use the command bwareaopen() and imfill() in Matlab to fill all the holes in leaf itself. The last operation before feature extraction step is the leaf boundary extraction. In order to get the margin of leaf, we use a Laplacian filter of following 3*3 spatial mask:

$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$

Then we can get the boundary of leaf image. A flow of image pre-processing is illustrated in Fig.2. as follows.

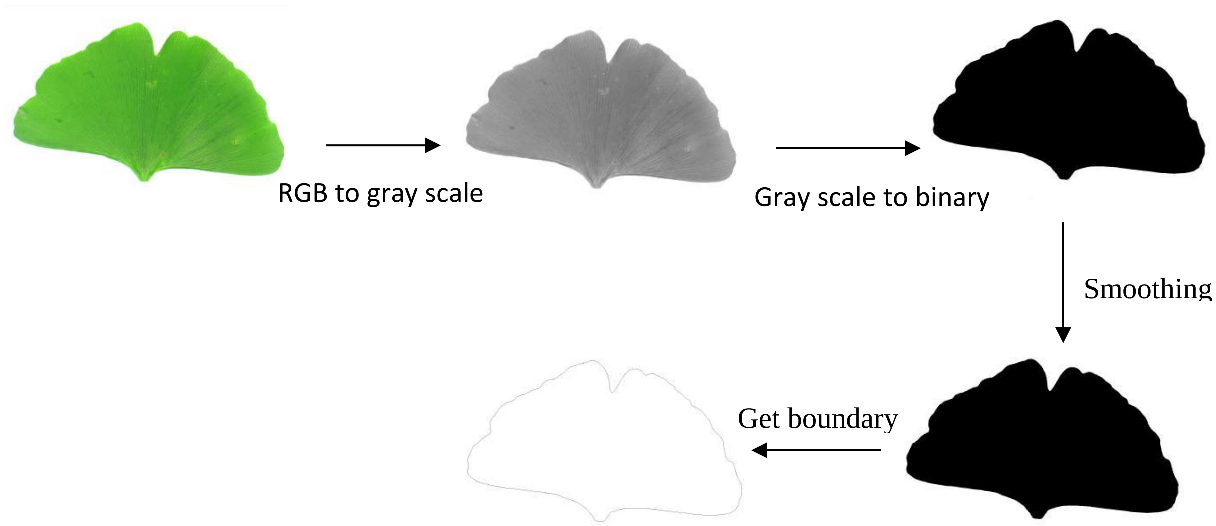

Figure 2. A pre-processing example

3. Feature extraction

3.1 Morphological information

The length and width of the leaf is calculated by minimum bounding rectangle. The perimeter of the leaf is calculated using the boundary points while the area is calculated by binary image. Here, we obtain 4 basic features. F_i denotes the number of features.

F1. *Area:* The area of leaf represents number of pixels in the leaf region.

F2. *Perimeter:* Perimeter is the distance around the margin of leaf region.

F3. *Length:* The length of minimum bounding rectangle.

F4. *Width:* The width of minimum bounding rectangle.

3.2 Convex hull information

A convex hull is formed using the boundary points of the leaf which connected with numbers of vertices (see Fig 3). Convex hull represents the smallest convex polygon that encapsulates the leaf region. Here, we obtain 3 features as follows:

F5. *HullArea:* The area of convex hull.

F6. *HullPerimeter:* The perimeter of convex hull.

F7. *SubAreaConvex:* SubAreaConvex denotes the value of HullArea minus Area.

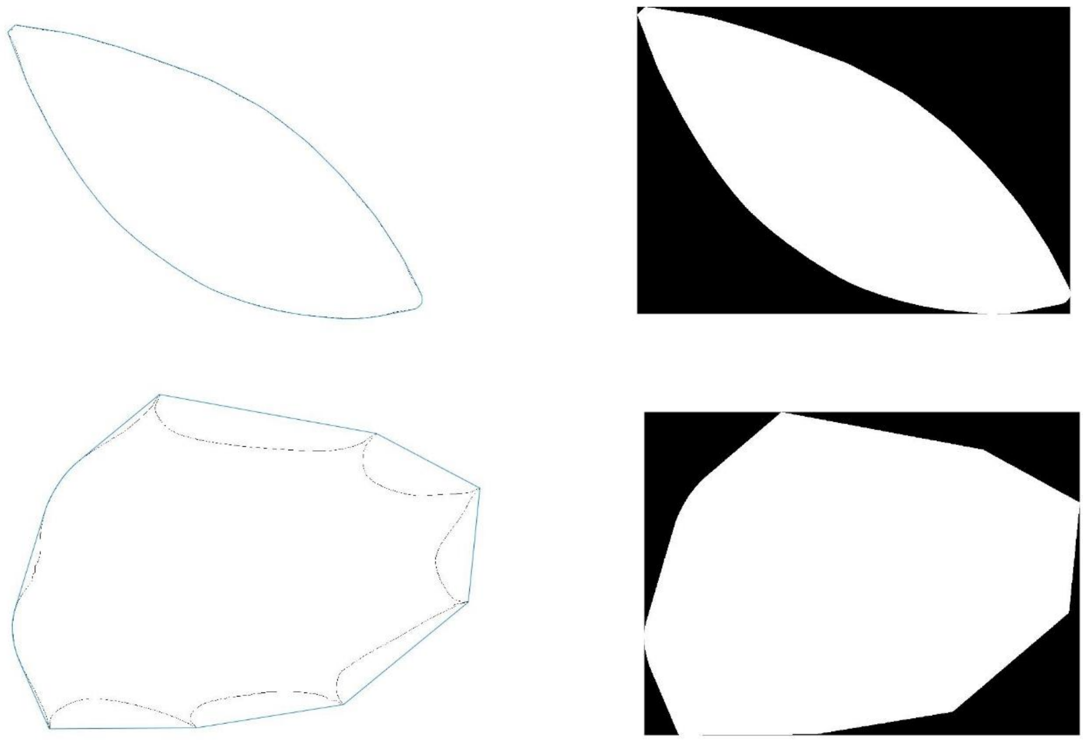

Fig 3. Different convex hull areas in different leaf species

3.3 Other feature information

Using the function regionprops() in Matlab, we can obtain more properties about leaf region. Thus, we get 8 features as follows. The definitions of those features come from the documentation of Matlab R1016b. Math Works Inc. Here is the link https://cn.mathworks.com/help/images/ref/regionprops.html.

F8. *MinBoundRectArea:* The area of the minimum bound rectangular.

F9. *MinBoundRectPerimeter:* The perimeter of the minimum bound rectangular.

F10. *MajorAxisLength:* A scalar that specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.

F11. *MinorAxisLength:* A scalar that specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.

F12. *Eccentricity:* Eccentricity is a scalar that specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. Its value varies between 0 and 1.

F13. *EquivDiameter:* A scalar that specifies the diameter of a circle with the same area as the region. Computed as sqrt(4*Area/pi).

F14. *Solidity:* A scalar specifying the proportion of the pixels in the convex hull that are also in the region.

F15. *Extent:* Scalar that specifies the ratio of pixels in the region to pixels in the total bounding box. Computed as the area divided by the area of the bounding box.

3.4 Feature radios

The features below shows the ratios that have been calculated from the previous measurements.

F16. *Aspect ratio:* The aspect ratio is defined as the ratio of the length of leaf to the width of leaf, thus *Length/Width*.

F17. *Hullarea-Area ratio:* This feature is defined as the ratio of the area of leaf to the area of convex hull, thus *Area/HullArea.*

F18. *Perimeter-Area ratio:* Ratio of perimeter to area, representing the ratio of leaf perimeter and leaf area, is calculated by *Perimeter/Area.*

F19. *Perimeter to hull:* The perimeter to hull ratio is defined as the ratio of the perimeter of convex hull to the perimeter of leaf, thus *HullPerimeter/Perimeter.*

After illustrating the definition of 19 features, now, we have finished the step of feature acquisition and go on classification section.

4. Classification

Many classifiers have been chosen for plant leaf recognition. Among all approaches, Artificial Neural Network (ANN) as the fastest speed method, has been used in many researches. Based on Flavia dataset, Satti et al. proposed a plant recognition system using the information of shape and color to get 93.3% accuracy with the method of ANN and getting 85.9% accuracy with k-Neighbour classifier (kNN) (Satti et al., 2013). More recently, Wang et al. tested on the Swedish leaf dataset with kNN and achieved 96.6% accuracy (Wang et al., 2014). Another neural network method named Probabilistic Neural Network (PNN) which also has been used by Wu (Wu et al., 2006) and Huang (Lin and Peng, 2008). In 2012, based on the bisection of leaves, Uluturk and Ugur achieved an accuracy of 92.5% with PNN (Uluturk and Ugur, 2012). Hossain and Amin obtained 91.4% accuracy by using a ten-fold cross validation technique to train PNN (Hossain and Amin, 2010). Kadir et al. extracted vein, shape, colour and texture features of the leaf and obtained around 95% accuracy with PNN (Kadir et al., 2012). Zhang and Lei used a modified locally linear discriminant embedding algorithm (MLLDE) on the LCL plant leaf database shared by Ren et al. (Ren et al., 2012) and achieved an accuracy of 93.5% (Zhang and Lei, 2011).

All of those methods above can obtain a satisfy result. However, it still has possible to improve the accuracy of plant leaf recognition. So the purpose of this paper is to get a higher accuracy than other methods using a different classifier which is SVM. In our work, we trained SVM model using the entire feature space by cross validation technique. Since there are more than 2 classes in our dataset, we should use multi-class SVM by one against one approach rather than using a binary SVM classifier.

5. Experimental result

In Flavia dataset, it contains a total of 1907 leaf images with regard to 32 plant species which is showed in Fig.4 and the number below each leaf denotes as its label. Each specie's common name and the number of species samples are listed in Table 1.

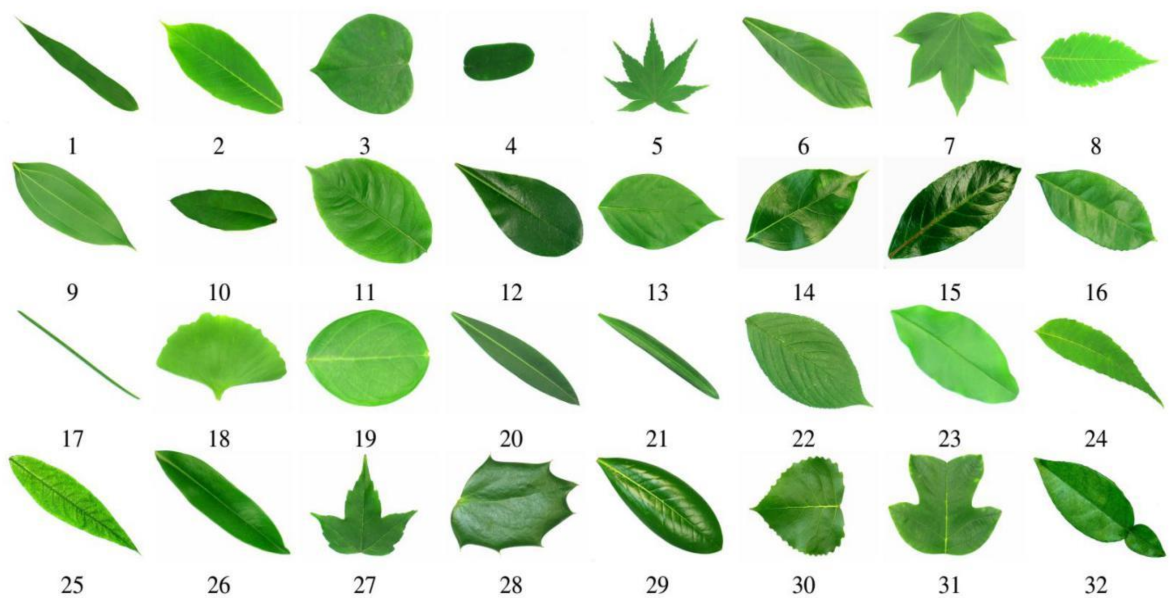

Fig 4. The 32 plant species in Flavia dataset

As we can see from Table 1, on average, the number of samples for each species is approximately 60. So there is no huge gaps between the samples of each species.

Table 1. Details about leaf name, label and number

Common name	label	Species samples
Pubescent Bamboo	1	59
Chinese Horse Chestnut	2	63
Chinese Redbud	3	72
True Indigo	4	73

Japanese Maple	5	56
Nanmu	6	62
Castor Aralia	7	52
Goldenrain Tree	8	59
Chinese Cinnamon	9	55
Anhui Barberry	10	65
Big-fruited Holly	11	50
Japanese Cheesewood	12	63
Wintersweet	13	52
Camphortree	14	65
Japanese Viburnum	15	60
Sweet Osmanthus	16	56
Deodar	17	77
Ginkgo Maidenhair Tree	18	62
Crape Myrtle	19	61
Oleander	20	66
Yew Plum Pine	21	60
Japanese Flowering Cherry	22	55
Glossy Privet	23	55
Chinese Toon	24	65
Peach	25	54
Ford Woodlotus	26	52
Trident Maple	27	53
Beale's Barberry	28	55
Southern Magnolia	29	57
Canadian Poplar	30	64
Chinese Tulip Tree	31	53
Tangerine	32	56

Since different kernel functions and different parameters will have a great effect the performance of SVM model. We used the same training and testing dataset to compare different results between different kernel functions which is RBF, Linear, Quadratic and Polynomial kernel function, respectively (see as Table 2). Through a comparative analysis, we finally chose RBF kernel function with the best value of parameter sigma which is 6 and the best value of parameter C which is 2^8. Based on this, we achieved an accuracy of 96.07% with 1729 training samples and 178 testing samples which were randomly chose.

Table 2. Accuracy between different kernel functions

(Based on 1729 training samples & 178 testing samples)

Kernel	RBF	Linear	Quadratic	Polynomial
Accuracy	96.07%	93.82%	92.7%	91.01%

By using RBF kernel with the parameter as above mentioned, we chose different training samples to test our model and got the follow accuracy curve, shown as Fig 5.

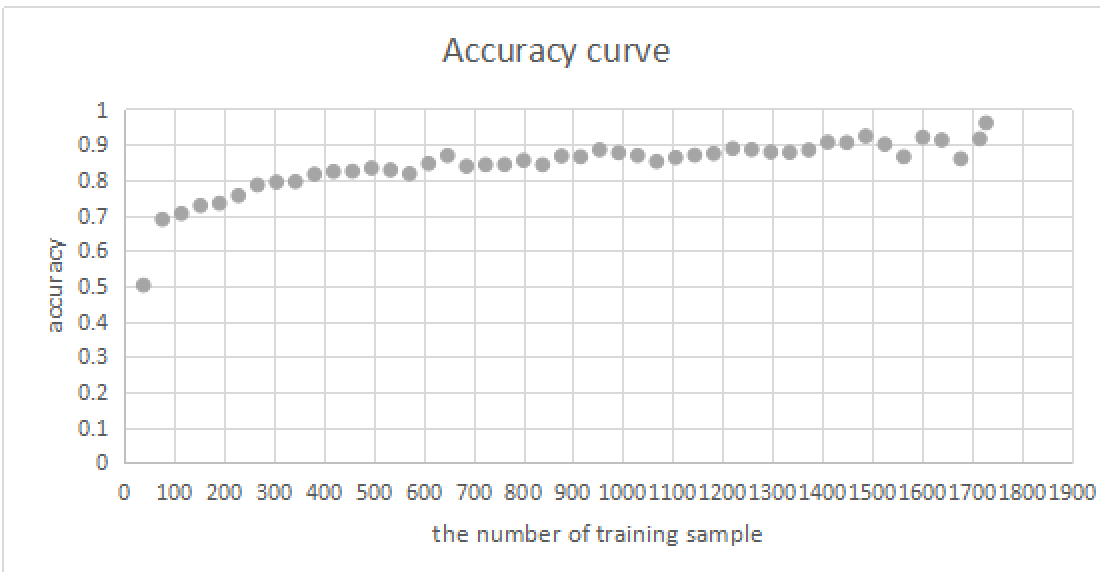

Fig 5. The accuracy curve with different training sample

As we can see from Fig 5, even decreasing training samples from 1729 to only 100, our method can still achieve accuracy more than 70% which can be considered as a relatively good result. We also compared our method with other researchers in classifier, accuracy, whole dataset number, training, testing, species and samples number, shown as Table 3.

Table 3. Comparison between our method and other methods present in literature

Authours	Classifiers	Accuracy	Dataset	Training	Testing	Species	Samples
Wenting wu	**SVM**	**96.07%**	**1907**	**1729**	**178**	**32**	**50-77**

Satti et al. (2013)	ANN	93.3%	1907	1742	165	33	40-60
Chaki et al. (2015)	ANN	87.1%	930	620	310	31	30
Arun et al. (2012)	SGD, kNN, SVM, DT, RF	94.7%	250	175	75	5	50
Uluturk and Ugur. (2012)	PNN	92.5%	1280	1120	160	32	40
Hossain and Amin. (2010)	PNN	91.4%	1330	1200	130	30	40
Kadir et al. (2012)	PNN	95.8%	6900	5700	1200	60	95
		95.0%	1600	1280	320	32	50
Zhang and Lei. (2011)	MLLDE	93.5%	750	500	250	50	15
Amin and Khan. (2013)	kNN	71.5%	1600	1120	480	100	16
Wang et al. (2014)	kNN	96.6%	1125	750	375	15	75
Larese et al. (2014)	PDA	88.4%	150	10-fold cross validation			3
		84.1%	866				

From Table 3, we can see that our method has a higher classification accuracy than almost all the other methods by only using 19 basic features which is easy to be extracted and has fast speed on training and simple structure.

6. Conclusion

In this paper, we demonstrated a new method to completely automatic classify 32 kind of species of plant leaf using one against one approach to multi-class SVM classifier. By image pre-processing technology, we extracted 19 features for each species without need for manual intervention by the user. The experimental results indicate that the proposed method has an excellent performance and can achieve a relatively higher accuracy of 96.07%. Future works could focus on feature extraction. In our method, we didn’t use the information of leaf color, texture and vein which could be considered as good features to classify different leaf species. To avoid a narrow set of features, in future work we can add more features like color and texture to further improve the accuracy of leaf recognition. What’s more, in the training process, we only use Flavia dataset. In future, we can test our method on other available datasets to further improve the robustness of our algorithm.

Reference

A.H.M. Amin, A.I. Khan. One-shot classification of 2-D leaf shapes using distributed hierarchical graph neuron scheme with k-NN classifier. In: Elsevier: 17th Asia Pacific Symposium on Intelligent and Evolutionary Systems (2013), pp. 84–96

A. Kadir, L.E. Nugroho, A. Susanto, P.I. Santosa. Performance improvement of leaf identification system using Principal Component Analysis. International Journal of Advanced Science and Technology, 44 (2012), pp. 113–124

C. H. Arun, W. R. Sam Emmanuel, D. Christopher Durairaj. Texture Feature Extraction for Identification of Medicinal Plants and Comparison of Different Classifiers. International Journal of Computer Applications, Vol. 62.12, (2013), pp. 1-9.

C. Uluturk, A. Ugur. Recognition of leaves based on morphological features derived from two half-regions. In: Institute of Electrical and Electronics Engineering(IEEE), International Symposium on Innovations in Intelligent Systems and Applications, (2012) pp. 1-4.

J. Hossain and M. A. Amin, "Leaf shape identification based plant biometrics," in International Conference on Computer and Information Technology (ICCIT 2010), 2010, pp. 458–463.

J. Chaki, R. Parekh, S. Bhattacharya. Plant leaf recognition using texture and shape features with neural classifiers. In: Elsevier: Pattern Recognition Letters, 58 (2015), pp. 61–68

Lin, H. and Peng, H. Machine recognition for broad-leaved trees based on synthetic features of leaves using probabilistic neural network. In Computer Science and Software Engineering, 2008 International Conference on, volume 4(2008), pp. 871–877.

N.Sakai, S.Yonekawa, A.Matsuzaki, Two dimensional image analysis of the shape of rice and its applications to separating varieties, J.Food Eng. 27(1996), pp. 397–407.

S. Zhang, Y.K. Lei. Modified locally linear discriminant embedding for plant leaf recognition. In: Elsevier: Neurocomputing, 74 (14) (2011), pp. 2284–2290

T.Munisami, M.Ramsurn, S.Kishnah, S.Pudaruth, Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers. J.Procedia Computer Science. 58(2015), pp. 740-747.

V. Satti, A. Satya, S. Sharma. An automatic leaf recognition system for plant identification using machine vision technology. International Journal of Engineering Science and Technology (IJEST), Vol. 5.4, (2013), pp. 874-879.

Wu, S., Bao, F., Xu, E., Wang, Y.-X., Chang, Y.-F., and Xiang, Q.-L. A leaf recognition algorithm for plant classification using probabilistic neural network. In IEEE International Symposium on Signal Processing and Information Technology, (2007), pp. 11–16.

X. Wang, B. Feng, X. Bai, W. Liu, L.J. Latecki. Bag of contour fragments for robust shape classification. In: Elsevier: Pattern Recognition, 47 (6) (2014), pp. 2116–2125.

X.M. Ren, X.F. Wang, Y. Zhao. An efficient multi-scale overlapped block LBP approach for leaf image recognition. In: Springer, International Conference on Intelligent Computing (ICIC), 8 (2012), pp. 237–243.