Oriental Journal of Computer Science & Technology Vol. **3**(1), 31-36 (2010)

Leaf recognition for plant classification using GLCM and PCA methods

ABDOLVAHAB EHSANIRAD and SHARATH KUMAR Y. H.

Department of Computer Science, University of Mysore, Mysore - 570 006, India.

(Received: April 12, 2010; Accepted: June 04, 2010)

ABSTRACT

In this paper, the image processing techniques has been used in order to classify the plants by applying on the leaves images. To extract the leaves texture features, the Gray-Level Co-occurrence matrix (GLCM) and Principal Component Analysis (PCA) algorithms have been considered. The Algorithms are trained by 390 leaves to classify 13 kinds of plants with 65 new or deformed leaves images. The result indicates that the accuracy for the GLCM method is 78% while the accuracy for the PCA method is 98%.

Keywords: Classification, GLCM, PCA, feature extraction.

INTRODUCTION

Leaf recognition is a pattern recognition task performed specifically on leaves. It can be described as classifying a leaf either "known" or "unknown", after comparing it with stored known leaves. It is also desirable to have a system that has the ability of learning to recognize unknown leaves.

Computational models of leaf recognition must address several difficult problems. This difficulty arises from the fact that leaves must be represented in a way that best utilizes the available leaf information to distinguish a particular leaf from all other leaves.

Compared with other methods, such as cell and molecule biology methods, classification based on leaf image is the first choice for plant classification. Sampling leaves and photogeny them are low-cost and convenient. One can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques. Some systems employ descriptions used by botanists. But it is not easy to extract and transfer those features to a computer automatically.

It is difficult job to tell the just one algorithm alone is the best and successful at recognizing any and all variation of the same object. And it is more difficult to tell the same algorithm to be able to differentiate between different objects. Many research has done for the leaf classification with some texture feature extraction methods[3,9,10,7].

Leaf classification process method

The conventional method of leaf classification involves two main steps. The first step is obtaining a priori knowledge of each class to be recognized. Normally this knowledge encompasses some sets of texture feature of one or all of the classes. Once the knowledge is available and texture feature of the observed image are extracted, then classification techniques, for example nearest neighbors and decision trees, can be used to make the decision[5], that is the second step. Such a procedure is illustrated in Fig. 1, the tasks that texture classification has been applied to include the classification of plant leaves images[2].

Currently there are a huge number of texture feature extraction methods available and most of the methods are associated with tunable parameters. It is difficult to find the most suitable feature extraction methods and their optimal parameters for a particular task. In addition, performance of classification methods also depends upon the problems, which makes selecting an optimal "feature extraction + classification" combination a difficult assignment.

Fig. 1. Conventional Plant Classification Process

The conventional method of plant classification involves two main steps. The first step is obtaining a priori knowledge of each class to be recognized. Normally this knowledge encompasses some sets of texture features of one or all the classes. Once the knowledge is available and texture feature of the observed image are extracted, then classification techniques, for example nearest neighbors can be used to make the decision. That is the second step. Such procedure is illustrated in Figure 1.

Feature Extraction

Different features are chosen to describe different properties of the leaves. Some leaves are with very distinctive shape, some have very distinctive texture patterns, and some are characterized by a combination of these properties.

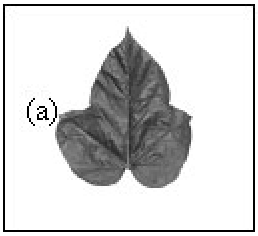
Gray-Level Co-occurrence Matrix

This method was first proposed by Haralick in 1973 and still is one of the most popular means of texture analysis [8]. The key concept of this method is generating features based on gray level co-occurrence matrices (GLCM). The matrices are designed to measure the spatial relationships between pixels. The method is based on the belief that texture information is contained in such relationships.

Co-occurrence features are obtained from a gray-level co-occurrence matrix. We used 22 features that extracted from GLCM matrix in this study. (Table 2) [8,4,1].

Textural Features Extracted from Gray-level Co-occurrence matrices

Our initial assumption in characterizing image texture is that all the texture information is contained in the gray-level Co-occurrence matrices. Hence all the textural features here are extracted from these gray-level Co-occurrence matrices. The equations which define a set of 22 measures of textural features are given in this paper. Some GLCM Extracted textural features are illustrated in Table 1 for two different leaf images.



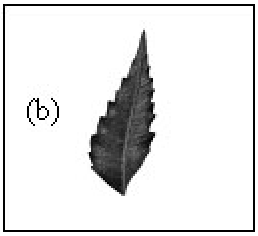


Table 1. GLCM Extracted textural features for two different leaf images.

some texture Features extracted from Leaf image (a)					
Angle	Autocorrelation	Entropy	Contrast	Correlation	Homogeneity
0°	455748	1.4811	0.3184	0.9638	0.6144
45°	452799	1.4828	0.4458	0.9486	0.6060
90°	456190	1.3886	0.2301	0.9738	0.6166
135°	452932	1.4716	0.4192	0.9526	0.6074

some texture Features extracted from Leaf image (b)					
Angle	Autocorrelation	Entropy	Contrast	Correlation	Homogeneity
0°	548540	0.8972	0.4361	0.9401	0.8214
45°	547371	0.9132	0.4488	0.9396	0.8199
90°	549797	0.8405	0.1845	0.9747	0.8267
135°	546610	0.9810	0.5961	0.9189	0.8172

Leaves Classification Using Eigenspace

In this study, we have followed the method which was proposed by M. Turk and A. Pentland [6] inorder to develop a leaves classification system based on the eigenspace approach. If a multitude of leaf images can be reconstructed by weighted sum of a small collection of characteristic features or eigenpictures, perhaps an efficient way to learn and recognize leaves would be to build up the characteristic features by experience over time and recognize particular leaf by comparing the feature weights needed to approximately reconstruct them with the weights associated with known leaves. Therefore, each leaf is characterized by a small set of feature or eigenpicture weights needed to describe and reconstruct them. This is an extremely compact representation when compared with the images themselves.

Table 2. All of 22 feature extraction equations of the GLCM

Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\right\}, \lvert i-j\rvert = n$	Information Measure of Correlation1	$\frac{HXY - HXY1}{\max\{HX, HY\}}$
Correlation	$\frac{\sum_i\sum_j (i,j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$ Where μ_x, μ_y, σ_x, and σ_y are the means and std. deviations of p_x and p_y, the partial probability density functions.	Information Measure of Correlation2	$(1 - \exp[-2(HXY2 - HXY)])^{1/2}$ Where $HXY = -\sum_i\sum_j p(i,j)\log(p(i,j))$, HX, HY are the entropies of p_x and p_y, $HXY1 = -\sum_i\sum_j p(i,j)\log\{p_x(i)p_y(j)\}$, $HXY2 = -\sum_i\sum_j p_x(i)p_y(j)\log\{p_x(i)p_y(j)\}$
Entropy	$-\sum_i\sum_j p(i,j)\log(p(i,j))$	Max Correlation Coefficient	$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$
Homogeneity	$\sum_i\sum_j \frac{1}{1+(i-j)^2}p(i,j)$	Energy	$\sum_i\sum_j p(i,j)^2$
Angular Second Moment	$\sum_i\sum_j p(i,j)^2$	Autocorrelation	$\sum_i\sum_j (ij)p(i,j)$
Sum of Squares : Variance	$\sum_i\sum_j (i-\mu)^2 p(i,j)$	Dissimilarity	$\sum_i\sum_j \lvert i-j\rvert \cdot p(i,j)$
Sum Average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$	Cluster Shade	$\sum_i\sum_j (i+j-\mu_x-\mu_y)^3 p(i,j)$
Sum Variance	$\sum_{i=2}^{2N_g} (i-SE)^2 p_{x+y}(i)$	Cluster Prominence	$\sum_i\sum_j (i+j-\mu_x-\mu_y)^4 p(i,j)$
Sum Entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i)\log\{p_{x+y}(i)\}$	Maximum Probability	$\max_{i,j} p(i,j)$
Difference Variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$	Inverse difference normalized	$\sum_{i,j=1}^{G} \frac{C_{ij}}{1+\frac{\lvert i-j\rvert^2}{G^2}}$
Difference Entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i)\log\{p_{x-y}(i)\}$	Inverse difference moment normalized	$\sum_{i,j=1}^{G} \frac{C_{ij}}{1+\frac{(i-j)^2}{G^2}}$

Experimental results and discussions

The experiment is designed to illustrate the performance of two feature extraction methods, Gray Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA) algorithms for plant leaves classification purpose.

The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. The classification steps are illustrated in Fig. 2.

In the first experiment after changing the color image to gray-level image with using of the GLCM texture feature extraction we extracted the 22 features[8,4,1] of each leaf images.

Figure 2. Classification Steps in GLCM method.

We have tried the GLCM method with Distance 1 (d=1) and degree 0°, Distance 1 (d=1) and degree 45°, distance 1 (d=1) and degree 90° and distance 1(d=1) and degree 135°. The performance accuracy of each one is shown in Table 4.

Table 3. Some features extracted from some chosen leaf image of each leaves classes in (d=1) and degree 0°

Features	Sample leaf from leaves Classes												
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12	Class 13
Autocorrelation	5527373	49.59994	60.29225	5431234	45.25222	50.65949	41.42991	61.86218	46.57263	51.48354	54.85598	52.6443	50.78359
Contrast	0.312278	0.368838	0.322488	0.308646	0.310127	0.333228	0.385127	0.334494	0.315823	0.46519	0.438078	0.363924	0.28481
Correlation	0.965284	0.981478	0.87083	0.966296	0.961898	0.959497	0.971515	0.786661	0.970818	0.951302	0.940084	0.947037	0.962174
Dissimilarity	0.126899	0.174367	0.123101	0.138608	0.172468	0.151266	0.172152	0.107911	0.148519	0.181646	0.159962	0.182595	0.15443
Energy	0.699192	0.528028	0.828287	0.696948	0.387419	0.546983	0.377983	0.893974	0.498494	0.574198	0.688862	0.580049	0.516229
Entropy	0.857156	1.288197	0.574384	1.034478	1.52748	1.168368	1.514552	0.379659	1.155161	1.211928	0.892223	1.208832	1.352482

GLCM method in leaf recognition for the degrees 0° and 90° gave the same accuracy and same result. Here the poor result is in the 45° degree. Because any changes in the neighboring distance or the neighboring degree it will change the value of extracted texture feature.

The GLCM method is very sensitive for the any changes in the images such rotation, scale and etc. in Tables 3 you can see the difference in extracted features in different neighborhood degrees. The computation time for GLCM method is less and recognition of this method is very fast.

Table 4. The performance of GLCM method in different degrees with neighborhood distance 1 and performance of PCA method.

Average recognition rate (%)	degree
78.46	0°
49.23	45°
78.46	90°
70.76	135°
98.46	PCA

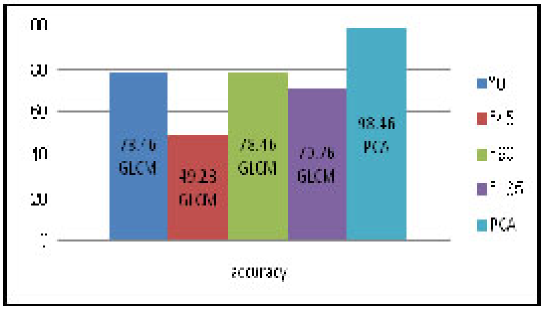


Fig. 3. GLCM and PCA accuracy chart in different degrees

PCA method mostly using for the face recognition purpose but we tried as leaf recognition. In PCA also image should be change to gray level that can reduce the image dimension. In our experience the PCA method gave the efficient performance and very good result. It was the just one wrong recognition out of 65 test image in our test. But the test speed is not much good and computation time is high for recognizing one test image. Compare with GLCM it's very slow but the performance of PCA method is efficient (Figure 3).

Database

The database used in our experiment is collected by our self. We pluck the leaf from the plant in the fields near our campus and around University of Mysore, which consists of intact and fresh leaf images in different rotation for 13 plant species class and constructed by our self. We taken 390 images as training set and each plant class contains the 30 leaf images in different degree of rotation and different leaf images. The test set contains the 65 of deformed and new leaf images and for each class has 5 leaf images for test. The sample dataset of leaf images and related classes are illustrated in Figure 4.

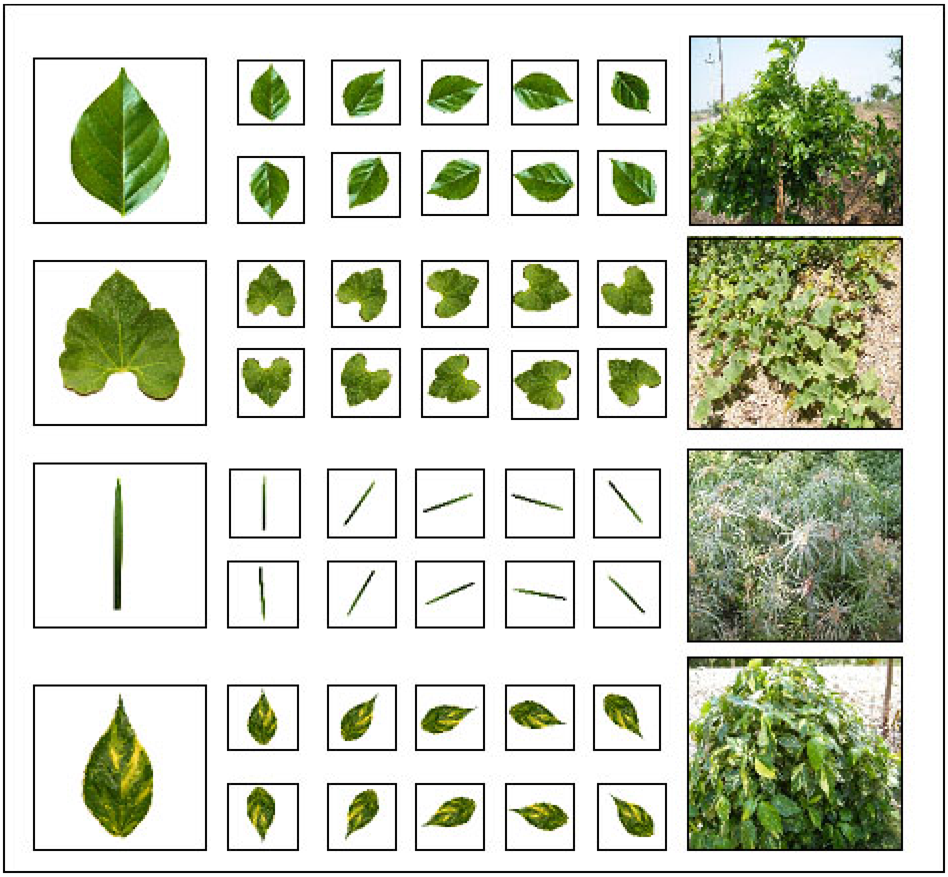


Fig. 4. The sample dataset of leaf images and related classes

CONCLUSION

In this study, the classification based on the recognizing the leaves images with extracted texture features was proposed and performed. The texture features have been extracted with using the Gray-Level Co-occurrence Matrix (GLCM) and the Principal Component Analysis (PCA) algorithms, on the 390 image in dataset and with 65 deformed or new leaf images for test. In addition, different degrees for the GLCM method were used and it was found out to be more efficient in the degree 0° by 78.46 % accuracy. Therefore, it was specified that the GLCM is very sensitive in any changes for images such as deforming or giving the new leaf image as a test. In addition, the PCA method comes out to be more efficient compare to the GLCM method by 98.46 % accuracy.

REFERENCES

1. D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization", *Can. J. Remote Sensing*, **28**(1), 45-62, 2002.
2. F. Dell' Acqua and P. Gamba. Texture-based characterization of urban environments on satellite sar images. *IEEE Transaction on Geoscience and Remote Sensing,* **41**(1): 153-159, January 2003.
3. j. Graham, "Application of the Fourier-Mellin transform to translation-, rotation and scale-invariant plant leaf identification" McGill University, Montreal, July 2000.
4. L. K. Soh and C. Tsatsulis. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. In IEEE Transaction on Geoscience and Remote Sensing, **37**(2), (1999).
5. M. Tuceryan and A.K. Jain. Texture analysis. In C. H Chen, L. F. Pau, and P. S. P. Wang, editors, Handbook of Pattern Recognition and Computer Vision, chapter 2, pages 235-276. World Scientific, Singapore, (1993).
6. M. Turk, A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, **3**(1), 71-86 (1991).
7. P. Tzionas, S. E. Papadakis, D. Manolakis, "Plant leaves classification based on orphological features and a fuzzy surface selection technique ",in Fifth International Conference on Technology and Automation, Thessaloniki, Greece, 365-370 (2005).
8. R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC*-**3**(6): 610-621, (1973).
9. S. G. Wu, F. S. Bao, E. Y. Xu, Y. Wang, Y. Chang and Q. Xiang, "A Leaf ecognition Algorithm for Plant Classification Using Probabilistic Neural Network" arxiv, 0707, 4289v1, [cs.AI], 29 Jul 2007.
10. Z. Wang, Z. Chi, D. Feng and Q. Wang, "Leaf Image Retrieval with Shape Features", R. Laurini (Ed.): VISUAL 2000, LNCS 1929, pp. 477-487, (2000).