Proceedings of The National Conference
On Undergraduate Research (NCUR) 2008
Salisbury University
Salisbury, Maryland
April 10–12, 2008

Classification of Textures in Microphotographies of Leaves Epidermis

Joyce M. Fernandez
Department of Mathematics
University of Puerto Rico at Humacao
Humacao, Puerto Rico

Faculty Advisors: Elio Ramos, Denny S. Fernández

Abstract

Microscopic images of leaves, collected from Mona Island dry forest (which is located between Puerto Rico and the Dominican Republic), were analyzed. For each leaf side an image was obtained at two magnifications (200x and 400x). This resulted in four samples of images showing a wide variety of textures and stomata patterns. For each group of images we used the gray-level co-occurrence method to characterize the observed gray level patterns. From the GLCM matrix several texture features were calculated among others: the angular second moment (ASM), the contrast, correlation, inverse difference moment (ISM), and entropy. Visual inspection indicates the formation of three groups of images at 200x magnifications based on the observed patterns. The results of the GLCM analysis indicate consistency between the texture features and the isotropic and anisotropic patterns observed in the leaves
Keywords: microscopic images, texture features, patterns

1. Introduction

The classification and characterization of textures in digital images is of great interest in areas like artificial vision and pattern recognition. The natural world provides a wide variety of examples of textures and patterns that can be observed at different spatial scales. This characterization is very important in many areas of the biological sciences. In taxonomy, for instance, the traditional approach for the discrimination between species is based in the observation of the characteristics. Furthermore, these characteristics may be related to more fundamental issues like the relation between the observed structures and the functions. Through the extensive availability of digital technology and image processing methods, many tasks, like classification, that used to be handled manually, can be performed in an automatic or semi-automatic way using many statistical and artificial intelligence methodologies. This possibility may be attractive but the actual implementation is difficult, mainly due to the correct selection of discriminating features and in many cases the heterogeneous quality of the images.

This paper presents the results of an image analysis for a sample of microphographies of leaves epidermis. The samples show an extensive variety of textures, spatial patterns, cell structures, and stomata configurations. The main objective of this work is to combine several image processing and statistical techniques so that the original group can be divided in sub-samples with similar features. The following section describes the image data sets, and then the data analysis section explains the Gray Level Co-occurrence Matrix (GLCM) method that was used to obtain a texture features matrix. Several multivariate statistical methods were applied to the texture features matrix including principal component and cluster analysis which are described in subsequent sections. Finally, the results and conclusions are presented.

2. Data Set

The data set consisted of four groups of images of leaves epidermis of 1600x1200 pixels at two magnifications (200x and 400x) and sides. The first group (20x_E) consisted of 69 images, the second group (20x_H) consisted of 39 images, the third group (40x_E) consisted of 70 images, and the last group (40x_H) consisted of 60 images. In this paper the results for the 20x_E and 20x_H samples are presented. The leaves were collected from the Mona Island dry forest, which is located between Puerto Rico and the Dominican Republic. The epidermis is the outermost cellular layer that covers the whole plant structure and typically can be observed as a set of closely packed cells without intercellular spaces[4]. Besides the epidermal cells a prominent structure known as the stomata can be observed. The stomata are basically a pore surrounded by two bean shaped cells known as the guard cells (Figure 1). The epidermis has many functions being the most important to allow the sunlight to pass through the chloroplasts which is crucial for the photosynthesis process and to avoid an excessive loss of water from the inner tissues. The stomata allow the gas exchange between the plant and the environment which again is necessary for photosynthesis and respiration. For different plant species a wide variety of patterns of cell epidermis and stomata configurations can be observed. In this sense the observed structures in the images can be used as a discriminator between species or group of species. Traditionally this type of task is performed manually by visual inspection of the images and the corresponding classification. In this paper an automatic procedure is presented that is able to measure some features from the images, followed by a method that allows the construction of groups based on the features.

(a) Cissus trifolata (b) Ipomea triloba (c)Tricholaena rosea

Figure 1: Some examples of images prototypes that were found by visual inspection.

3. Data Analysis

Starting with the raw images the data analysis procedure consists of several steps which are described in the following sections.

3.1. pre-processing

Due to the poor contrast in some of the images a normalization procedure (contrast stretching) was carried out for each set of images. In this sense we were able to obtain consistency in the ranges of the pixel values for all the images. Furthermore, the original images were converted to 8 bit gray scale images and the size of the images was reduced in 50% in order to improve the efficiency of the image processing methods. A preliminary visual inspection of the images revealed the formation of three images prototypes which are shown in Figure 1.

3.2 the gray scale co-correspondence matrix

In order to characterize statistically the texture patterns observed in the images the Gray Scale Co-ocurrence Matrix (GLCM) was calculated[6]. The GLCM is a tabulation of how often different combinations of gray levels occurs in a matrix. For an image g we can construct a N x N gray level co-occurrence matrix $M_{d,\theta}$. The elements of $M_{d,\theta}$ represent the probability of the co-occurrence of gray value i,j at points p_1, p_2, separated by distance d and angle θ.

$$M_{d,\theta}(i,j) = Prob\{g(p_1) = i, g(p_2) = j\} \text{ given that } d(p_1, p_2) = \theta \quad (1)$$

where $p_i = (x_i, y_i)$,and i=1,2 and

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$

$$\angle(p_1, p_2) = \arctan((y_2 - y_1)/(x_2 - x_1)) \quad (3)$$

Where Equation (1) is the co-ocurrence matrix, Equation (2) is the distance between pixels and Equation (3) is the angle between the selected pixels. From $M_{d,\theta}$ different Haralick texture features can be calculated from the following equations (Table 1). The textural feature were calculated by averaging the co-ocurrence matrix at four angles (0^o, 45^o, 90^o, and 135^o) and at fixed distance of d=1 pixel.

Texture feature	Formula	Texture feature	Formula
Absolute value	$\sum_{i,j=0}^{N-1} \lvert i-j \rvert M_{d,\theta}(i,j)$	Entropy	$\sum_{i,j=0}^{N-1} M_{d,\theta}(i,j) \log_2 [M_{d,\theta}(i,j)]$
Inverse difference	$\sum_{i,j=0}^{N-1} \frac{M_{d,\theta}(i,j)}{1+(i-j)^2}$	Correlation	$\frac{\sum_{i,j=0}^{N-1} ijM_{d,\theta}(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$
Homogeneity	$\sum_{i,j=0}^{N-1} \frac{M_{d,\theta}(i,j)}{1+\lvert i-j \rvert^2}$	Energy	$\sum_{i,j=0}^{N-1} M_{d,\theta}(i,j)^2$
Contrast	$\sum_{i,j=0}^{N-1} (i-j)^2 M_{d,\theta}(i,j)$		

Table 1: Texture features utilized in the GLCM analysis of the images with the corresponding formula[2]

The GLCM and the corresponding features from Table 1 were calculated for each of the 4 groups of images. The GLCM features were complemented with four first order statistics namely the average gray intensity, the standard deviations and the corresponding skewness and kurtosis. The GLCM method was implemented as a plugin with the ImageJ[1] public domain Java image processing software. The plugin, `GLCM.java` was developed in Java and based on the texture analysis plugin developed by Julio E. Cabrera from NIH[3]. The correspondence between the image number and the plant species (from visual identification) are given in the next two figures.

1 Achasp1011(20x)E	19 Cistri962(20x)E	37 Jatgos957(20x)E	55 Schfru998(20x)E
2 Agasis981(20x)E	20 Cocdiv1014(20x)E	38 Jatmul1003(20x)E	56 Sespor989(20x)E
3 Alover978(20x)E	21 Cocmic993(20x)E	39 Krufer1000(20x)E	57 Sidobo953(20x)E
4 Amyele1027(20x)E	22 Cocuvi949(20x)E	40 Leuleu1024(20x)E	58 Stajam958(20x)E
5 Antacu965(20x)E	23 Comdod954(20x)E	41 Morcit1006(20x)E	59 Stastr1030(20x)E
6 Barasi990(20x)E	24 Comele1029(20x)E	42 Nepmul979(20x)E	60 Stiema963(20x)E
7 Bousuc1008(20x)E	25 Corglo956(20x)E	43 Pasto(1)984(20x)E	61 Surmar946(20x)E
8 Bursim968(20x)E	26 Cyphum1018(20x)E	44 Phyama980(20x)E	62 Swimah1013(20x)E
9 Caebon985(20x)E	27 Eupcor973(20x)E	45 Pilmar1019(20x)E	63 Tercat952(20x)E
10 Caemon1012(20x)E	28 Euppet1001(20x)E	46 Pisalb1026(20x)E	64 Thepop947(20x)E
11 Caklan1031(20x)E	29 Ficcit977(20x)E	47 Pluobt997(20x)E	65 Tilutr1023(20x)E
12 Canros988(20x)E	30 Goshir955(20x)E	48 Porole1016(20x)E	66 Toumic961(20x)E
13 Canwin1015(20x)E	31 Guadis1017(20x)E	49 Porrub974(20x)E	67 Triros982(20x)E
14 Capbif950(20x)E	32 Guaoff1004(20x)E	50 Preagg975(20x)E	68 Uromax1032(20x)E
15 Capfle1002(20x)E	33 Guasan1005(20x)E	51 Psymon994(20x)E	69 Vercin986(20x)E
16 Cenvir969(20x)E	34 Guekru1020(20x)E	52 Ranacu1028(20x)E	
17 Chanic1009(20x)E	35 Hipman1007(20x)E	53 Raunit945(20x)E	
18 Chialb1022(20x)E	36 Ipotri983(20x)E	54 Reynuc960(20x)E	

Table 2: List of the 69 images at 200x magnification by E side and its corresponding numbers. The name of the files correspond to the abbreviated scientific name of the identified species.

1 Achasp1011(20x)H	11 Cocdiv1014(20x)H	21 Nepmul979(20x)H	31 Stajam958(20x)H
2 Agasis981(20x)H	12 Cocmic993(20x)H	22 Pasto(1)984(20x)H	32 Stastr1030(20x)H
3 Alover978(20x)H	13 Cocuvi949(20x)H	23 Phyama980(20x)H	33 Styham971(20x)H
4 Boeere987(20x)H	14 Comele1029(20x)H	24 Pilmar1019(20x)H	34 Surmar946(20x)H
5 Caklan1031(20x)H	15 Cyphum1018(20x)H	25 Pisalb1026(20x)H	35 Thepop947(20x)H
6 Canros988(20x)H	16 Eupcor973(20x)H	26 Porrub974(20x)H	36 Tilutr1023(20x)H
7 Capbif950(20x)H	17 Goshir955(20x)H	27 Raunit945(20x)H	37Tricis951(20x)H
8 Cenvir969(20x)H	18 Ipotri983(20x)H	28 Schfru998(20x)H	38 Triros982(20x)H
9 Chanic1009(20x)H	19 Jatgos957(20x)H	29 Sespor989(20x)H	39 Uromax1032(20x)H
10 Cistri962(20x)H	20 Leuleu1024(20x)H	30 Sidobo953(20x)H	

Table 3: List of the 39 images at 200x magnification by H side and its corresponding numbers. The name of the files correspond to the abbreviated scientific name of the identified species.

3.3 Principal Component Analysis

The Principal Component Analysis (PCA) is one of the main tools of exploratory multivariate data analysis. A very common situation in multivariate data analysis, like in the features table obtained from the GLCM analysis, is the availability of several variables (features) for a single observation. In this sense each observation is a point in a multidimensional space. However, in general multidimensional data is very difficult to visualize and consequently hard to identify patterns.

The PCA is a method that reduces data dimensionality by performing a covariance analysis between factors. This method involves a mathematical technique to solve for the eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigenvalue has the same direction as the first principal component. The eigenvector associated with the second largest eigenvalue determines the direction of the second principal component. The sum of the eigenvalues equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows or columns of this matrix.

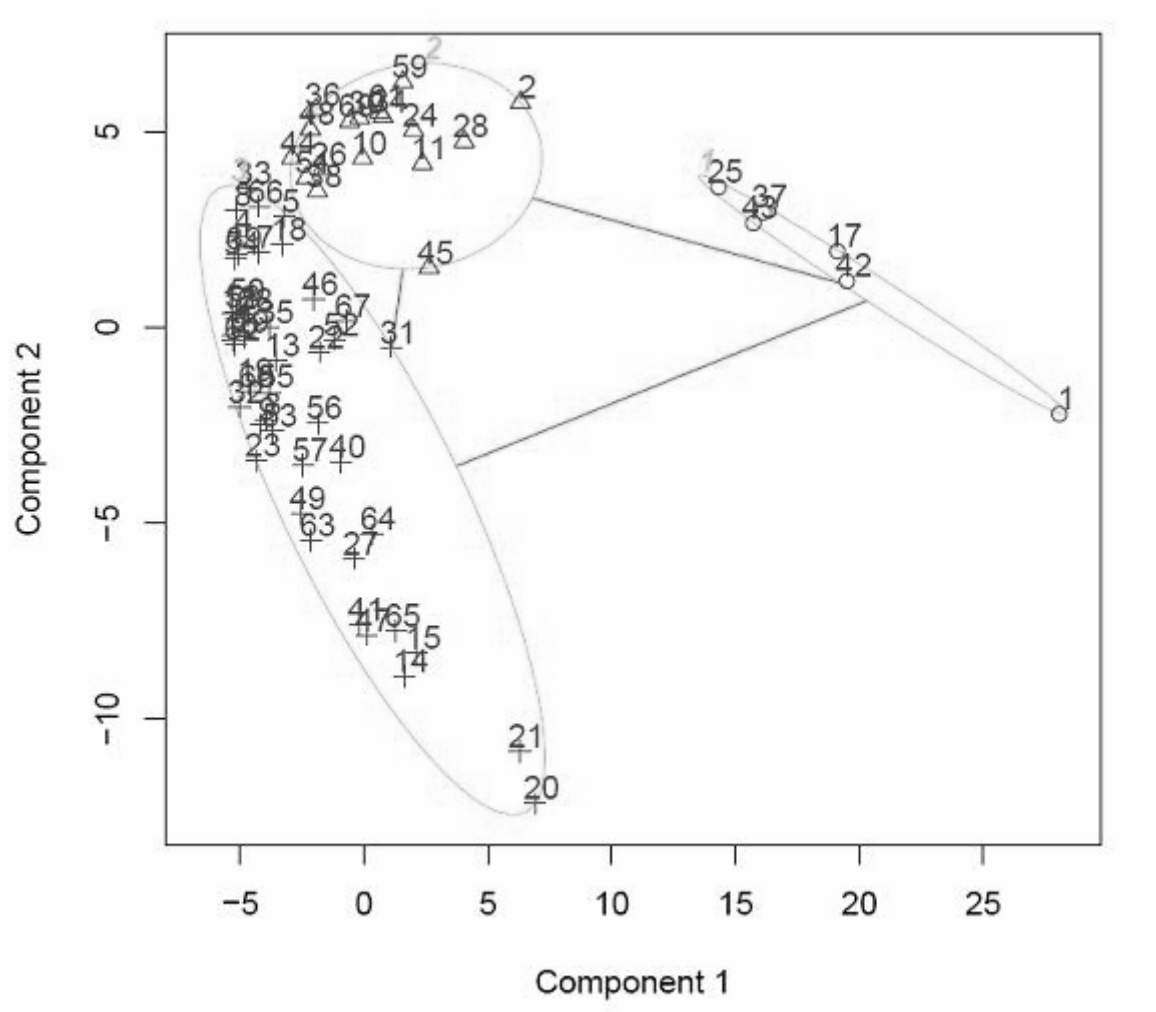

Figure 1: Principal Component Analysis (PCA) Plot at 200x and E side. The two components describe 89% of the variability observed in the features matrix. The groups were obtained from the CLARA optimal partitioning method.

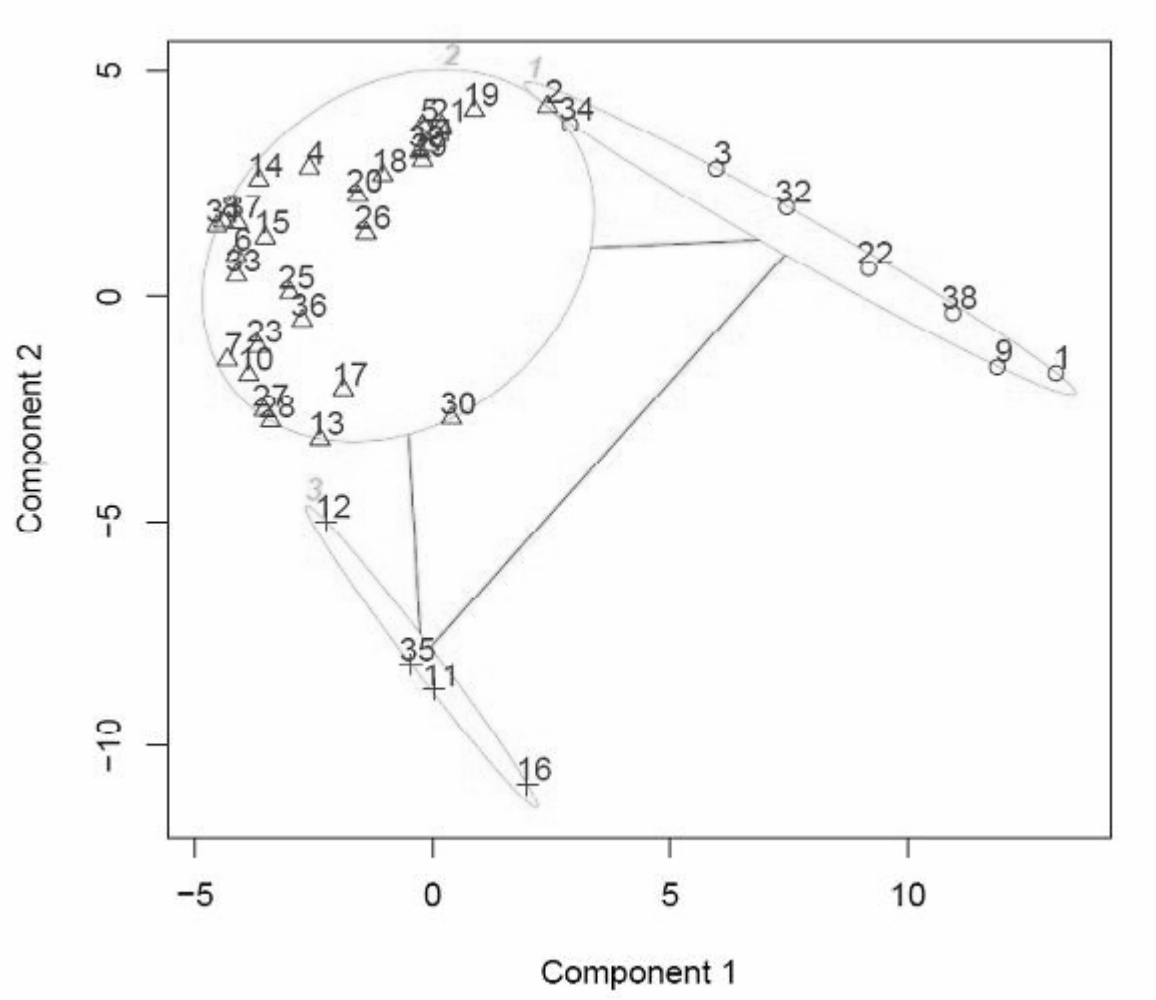

Figure 2: Principal Component Analysis (PCA) Plot at 200x and H side. The two components describe 91% of the variability observed in the features matrix. The groups were obtained from the CLARA optimal partitioning method.

The PCA method was applied to the features table for each of the image samples. A two dimensional plot of the first two components was obtained (see Figure 4 and Figure 5) and subsequently analyzed. Each point in the PCA plot was associated with an image in the sample. The relation between the numbers and the images was obtained from Figure 2 and Figure 3. Once the PCA plots were obtained two optimal partitioning methods[9] (PAM and CLARA) were applied in order to identify groups in the data sets.

4. Results and Conclusions

An image and multivariate data analysis was performed on several samples of images of leaves epidermis in search of patterns that allow an automatic or semi-automatic grouping or characterization of the images. In the first phase of the analysis eleven textural features were estimated for each image using the Gray Scale Co-occurrence Matrix (GLCM) method. A Principal Component Analysis (PCA) was performed on the features data and using an optimal partitioning method three groups of images were identified. Examination of the images in each group revealed consistency between the visual features of the images and the groups obtained from the textural features, PCA and partitioning methods. Is interesting to note that an independent analysis[7] of the images using visual morphological features from the epidermis (type of epidermis, type of stomata, structure of guard cells, etc.) reveals the formation

of four groups. A future work may reveal the connection (if any) between the textural classification and this morphological classification.

5. Acknowledgments

This research has been sponsored by the National Security Agency (NSA) Grant NUM. H98230-07-1-0114. We would like to thank Dr. Elvia Meléndez-Ackerman from CREST-Center for Applied Tropical Ecology and Conservation (CATEC) for providing the images utilized in the analysis.

6. References

1. Abramoff, M.D., Magelhaes, P.J., Ram, S.J. "Image Processing with ImageJ", 2004.
2. Bharath Kumar, S.V. et al, Proc. of International Conference on Signal Processing and Communication, 2004 Biophotonics International, volume 11, issue 7, pp. 36-42, 2004.
3. Cabrera J., Texture Analyzer, http//rbs.info.nih.gov/ij/plugins/texture.html 2005
4. Carpenter, K.J., *Am. J. Botany*, 92, 2005
5. Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. Graphical Methods for Data Analysis, Belmont, CA: Wadsworth. 1983.
6. Haralick, R.M., Proceedings of the IEEE, 67, 1979.
7. Melendez-Ackerman, E. Private Communication, 2008.
8. Smith, L.I., A tutorial on Principal Component Analysis, http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf 2002
9. Venaples, W.N., Ripley, B.D., Modern Applied Statistics with S, Springer-Verlag, New York, 2002.