Available online at www.sciencedirect.com
SCIENCE DIRECT
Computers and Electronics in Agriculture 50 (2006) 121–134
Computers and electronics in agriculture
www.elsevier.com/locate/compag

Plant species identification using Elliptic Fourier leaf shape analysis

João Camargo Neto [a], George E. Meyer [b,*], David D. Jones [b], Ashok K. Samal [c]

[a] *Embrapa Information Tecnology, Av. André Tosello 209, Cidade Universitária "Zeferino Vaz", P.O. Box 6041, Barão Geraldo, 13083-886 Campinas, SP, Brazil*
[b] *Biological Systems Engineering, 250 L.W. Chase Hall, University of Nebraska, Lincoln, NE 68583-0726, USA*
[c] *Computer Science, University of Nebraska, Lincoln, NE 68583-0115, USA*

Received 29 April 2005; received in revised form 7 September 2005; accepted 21 September 2005

Abstract

Elliptic Fourier (EF) and discriminant analyses were used to identify young soybean (*Glycine max* (L.) *merrill*), sunflower (*Helianthus pumilus*), redroot pigweed (*Amaranthus retroflexus*) and velvetleaf (*Abutilon theophrasti Medicus*) plants, based on leaf shape. Chain encoded, Elliptic Fourier harmonic functions were generated based on leaf boundary. A complexity index of the leaf shape was computed using the variation between consecutive EF functions. Principle component analysis was used to select the Fourier coefficients with the best discriminatory power. Canonical discriminant analysis was used to develop species identification models based on leaf shapes extracted from plant color images during the second and third weeks after germination. The classification results showed that plant species during the third week were successfully identified with an average of correct classification rate of 89.4%. The discriminant model correctly classified on average: 77.9% of redroot pigweed, 93.8% of sunflower, 89.4% of velvetleaf and 96.5% of soybean. Using all of the leaves extracted from the second and the third weeks, the overall classification accuracy was 89.2%. The discriminant model correctly classified 76.4% of redroot pigweed, 93.6% of sunflower, 81.6% of velvetleaf, 91.5% of soybean leaf extracted from trifoliolate and 90.9% of soybean unifoliolate leaves. The Elliptic Fourier shape feature analysis could be an important and accurate tool for weed species identification and mapping.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Elliptic Fourier; Discriminant analysis; Leaves; Machine vision; Shape features

1. Introduction

Site-specific weed control and herbicide application using machine vision or remote sensing has been the focus of many recent studies. Spot control of weeds has the potential to reduce the amount of chemicals applied by as much as 80% for improved farm profitability and water quality (Timmermann et al., 2003). Weed scientists have traditionally used manual scouting for botanically identifying weed species and mapping populations (Manley et al., 2001). They have also considered automated methods for species identification (Medlin et al., 2000). Automated species identification would assist scientists and growers to track the growth and development of young weeds, which could help to determine when to apply herbicides. The best time to apply post-emergent herbicide is defined as the critical time when weeds must be controlled to prevent crop yield losses (Knezevic et al., 2002). The critical time depends

* Corresponding author. Tel.: +1 402 4723377; fax: +1 402 4726338.
E-mail address: gmeyer1@unl.edu (G.E. Meyer).

0168-1699/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2005.09.004

on crop cultivar, weed species, environmental conditions, plant density and timing of weed competition (Zimdahl, 1988).

Canopy structure and leaf shape have been key features for plant species identification by weed specialists. Machine vision and shape feature analysis is usually the final step of a process that includes identifying green plant regions of interest and then isolating individual plant canopy crowns (Meyer et al., 2004a). For accurate species identification, one needs to extract and analyze individual leaves or leaflets (Camargo Neto, 2004). Kincaid and Schneider (1983) first used normalized Fourier coefficients, a complexity index (based on frequencies of Fourier coefficients), and a dissection index (roundness) to approximate leaflet shape. Guyer et al. (1986) studied four classical shape features for leaves from eight species of plants grown in containers. Species identification errors were reported as low as 9%. In a follow-up study, Guyer et al. (1993) investigated 17 quantitative shape features to classify 8 different plant species (soybean, milkweed, dandelion, jimsonweed, velvetleaf, giant morning glory, ragweed and foxtail). A mean correct identification rate of 69% was reported using a leaflet database of only 40 images. Petry and Kuhbauch (1989) found that leaf shape parameters with five canonical indices were statistically different for several weed species. Franz et al. (1991a) identified whole and occluded leaves of velvetleaf, ivy leaf, morning glory, giant foxtail and soybean using a leaf shape curvature and Fourier–Mellin correlation method at two stages of growth. When testing 124 samples, 73.4% of the leaves were correctly identified, 11.3% were matched incorrectly and 15.3% were rejected because of insufficient data. Canopy shape features as opposed to individual leaf features were used to distinguish monocotyledonous from dicotyledonous weeds by Woebbecke et al. (1995). The first invariant moment and aspect features distinguished monocotyledon from dicotyledonous canopies with correct classification rates from 60 to 90%.

Franz et al. (1991b) also studied a set of leaf spectral features (mean, variance, skewness, kurtosis, computed from spectral wavebands of red, green, blue and near infrared) to discriminate leaf types of soybean unifoliolates, ivy leaf, morning glory cotyledons, velvetleaf cotyledons, foxtail, first true leaves of ivy, morning glory and first true leaves of velvetleaf. The species identification error rate was 24.2%, when leaf orientation was not considered. When leaf orientation was considered, the species error rate decreased to 18%. The method tended to fail for leaves rotated more than 30° perpendicular to the petiole axis. The method only worked reasonably well, if 80% of the leaf boundary was exposed. These results suggest looking for whole or fully exposed leaves in plant canopies. Manh et al. (2001) proposed a template method to locate edges of leaves based on a deformable skeleton and leaf color. They reported a correct identification rate of 84% for green foxtail using 600 individual leaves from 40 images. Chi et al. (2002) presented a leaf species identification model using Bezier curves to discriminate four types of vegetable seedlings (cabbage, Chinese mustard, lettuce and edible amaranthus). The Bezier curves were approximations of the leaflet boundary, with endpoints defined by the leaf base and apex. They reported a correct classification rate of 95.1%, when applying the Bezier coefficients to a back propagation neural network. It was found that this accuracy depended upon the leaf base and apex, which defined the Bezier endpoints.

Oide and Ninomiya (2000) used the Elliptic Fourier (EF) method to classify soybean varieties, using a normalized leaf shape. The EF method using a chain-coded, closed-contour, invariant to scale, translation and rotation was first introduced by Kuhl and Giardina (1982). EF has been used in recent studies to describe the shape of objects. Innes and Bates (1999) used an Elliptical Fourier descriptor to demonstrate an association between genotype and morphology of shells. Chen et al. (2000) used Elliptic Fourier descriptors to describing shape changes in the human mandible for male and female at different ages. McLellan and Endler (1998) compared several morphometric methods for describing complex shapes. They found that 20 harmonics of the Elliptic Fourier method accurately depicted shapes of *Acer saccharinum*, *Acer saccharum* and *Acer palmatum* leaves. Most methods previously investigated ignore leaf serration. Leaf serration or edgeness is an important morphologic feature used for identifying plant species. For example, the curvature functions developed by Franz et al. (1991b) were found generally inadequate where leaflet serration was quite pronounced.

The EF method apparently provides an excellent and accurate graphical rendition of the edges and shapes of individual segmented or isolated leaves. The main problem in using EF has been how to use its large coefficient matrix for object identification. Obviously, machine vision operations that precede leaf shape feature analysis may affect the usefulness of these methods for precision agriculture. We will assume that those machine vision operations are available, (Camargo Neto, 2004). In this study, we therefore focus on the EF shape feature method for weed and crop species identification. The objectives of this study were to (a) develop a size and rotation invariant shape feature method for identifying young plants such as soybean (*Glycine max* (L.) *merrill*) and common weeds and (b) to determine the best time after emergence to identify these plant species.

Table 1
Total number of hand extracted leaves by species

Plant species	Week		Total
	Second	Third	
Sunflower	44	81	125
Redroot pig weed	41	99	140
Soybean	45	58	103
Velvetleaf	49	93	142
Total	179	331	510

2. Materials and methods

2.1. Image acquisition

Color digital images were taken using a DC120 digital camera (Kodak Digital Science Rochester, NY) of individual plants of soybean (*G. max* (L.) *merrill*), sunflower (*Helianthus pumilus*), redroot pigweed (*Amaranthus retroflexus*) and velvetleaf (*Abutilon theophrasti Medicus*) under natural lighting conditions around noon, during the first 3 weeks after germination. The camera was set to automatic exposure and focus, mounted overhead or nadir at 1 m above the plant containers, and with light behind the camera, such that no shadows were imposed on the leaves. An image resolution of 1280 × 960 pixels was used, which translated into 2 pixels per mm. Images were downloaded using the Kodak DC120 TWAIN interface® and stored as *jpegs*. These images are part of the Hindman plant set of 681 images (Hindman, 2001; Meyer et al., 2004a,b).

Individual leaves and trifoliolate leaflets were accurately segmented from the canopy images, visually and by mouse with Adobe®Photoshop®5.0 LE. This resulted in one individual whole leaflet per sub image. In general, plant species with real leaves were quite recognizable by visual inspection of the photographs. A few whole leaves were slightly rotated away from the camera lens, but most appeared parallel to the lens. A total of 510 individual leaves or leaflets were hand extracted from the Hindman image set, for the second and third weeks after germination (Table 1).

Soybean leaf sub images were divided into two categories representing cotyledon (VC soybean) and first trifoliolate (V1 soybean) growth stages. During the second week, the soybean presented fully expanded unifoliolate leaves, as shown in Fig. 1a. By the third week, the first trifoliolate leaflet was fully expanded, as shown in Fig. 1b. No leaves were used from the first week after germination, since all newly emerged plants showed cotyledons with very similar shapes. Cotyledons or leaves of the first week after germination are not considered as the critical period for weed control decisions (Knezevic et al., 2002; Zimdahl, 1988).

2.2. Leaf shape features analysis

The edges of each leaflet in the sub images were binarized according to standard procedures reported by Camargo Neto (2004). Leaf shape was then modeled using normalized Elliptic Fourier descriptors of a closed-contour of the leaflet edge (Kuhl and Giardina, 1982). The closed-contour was defined with differential chain code, represented as a sequence of vectors of unit length and directional coding as shown by an example in Fig. 2. Individual leaf boundary

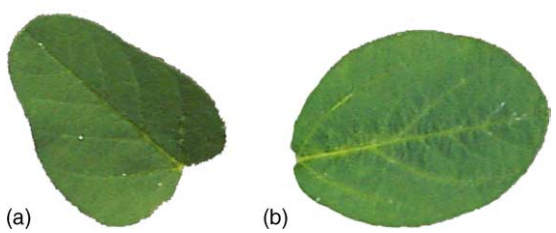

Fig. 1. (a) Unifoliolate soybean (*Glycine max* (*L.*) *merrill*) leaflet (VC, cotyledon stage) and (b) soybean trifoliolate leaflet (V1, first trifoliolate).

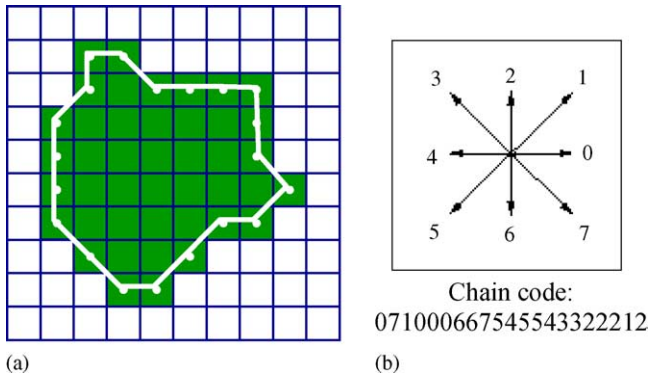

Fig. 2. (a) Binary pixel and shape pattern. (b) Differential chain code sequence starting in upper left hand corner (eight connected grids).

pixels were used to describe the chain code, starting with the upper left pixel of the contour and tracing the boundary clockwise. The chain code enumeration was completed when the original starting point was reached.

Using the chain code, a truncated Fourier series expansion of the closed-contour which is a projection on the x and y axes was obtained, given as:

$$x_N(t) = A_0 + \sum_{n=1}^{N} a_n \cos\left(\frac{2n\pi t}{T}\right) + b_n \sin\left(\frac{2n\pi t}{T}\right) \tag{1}$$

$$y_N(t) = C_0 + \sum_{n=1}^{N} c_n \cos\left(\frac{2n\pi t}{T}\right) + d_n \sin\left(\frac{2n\pi t}{T}\right) \tag{2}$$

where t is the step required to traverse 1 pixel along the closed-contour, such that $t_{p-1} < t < t_p$ for code values of p, within the range of $1 \leq p \leq k$ and k is the total number of codes describing the boundary contour. n is the number of Fourier harmonics required to generate the approximation of the boundary (each harmonic has four coefficients). T is the basic period of the chain code, or steps needed to traverse the entire contour, $T = t_k$. A_0 and C_0 are the bias coefficients, corresponding to a frequency of 0. These coefficients are related to image translation and N is the total number of EF harmonics needed to generate an accurate approximation of the boundary.

For each harmonic, the nth set of four harmonic coefficients a_n, b_n, c_n and d_n was defined as:

$$a_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{K} \frac{\Delta x_p}{\Delta t_p} \left[\cos\left(\frac{2n\pi t_p}{T}\right) - \cos\left(\frac{2n\pi t_{p-1}}{T}\right)\right], \tag{3}$$

$$b_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{K} \frac{\Delta x_p}{\Delta t_p} \left[\sin\left(\frac{2n\pi t_p}{T}\right) - \sin\left(\frac{2n\pi t_{p-1}}{T}\right)\right], \tag{4}$$

$$c_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{K} \frac{\Delta y_p}{\Delta t_p} \left[\cos\left(\frac{2n\pi t_p}{T}\right) - \cos\left(\frac{2n\pi t_{p-1}}{T}\right)\right], \tag{5}$$

$$d_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^{K} \frac{\Delta y_p}{\Delta t_p} \left[\sin\left(\frac{2n\pi t_p}{T}\right) - \sin\left(\frac{2n\pi t_{p-1}}{T}\right)\right], \tag{6}$$

where t_p is the number of steps required to traverse the first p components or links of the chain code. Δx_p and Δy_p are the spatial changes in the x and y projections of the chain code, respectively, at link p. These values may be 1, 0 or -1 depending on the orientation of link p. Δt_p is the step change required to traverse link p of the chain code.

Fig. 3 shows a non-normalized EF approximation to the shape of a velvetleaf using the set of coefficients a_n, b_n, c_n and d_n, generated with the 1st, 4th, 8th, 16th and 30th harmonics. The non-normalized and normalized EF coefficient values for the 16 sets of harmonics for the velvetleaf closed-contour are shown in Tables 2 and 3, respectively.

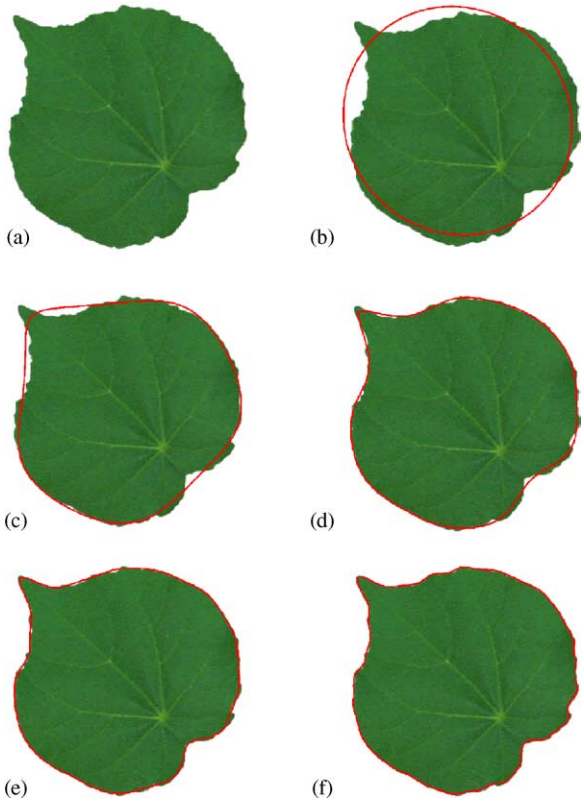

Fig. 3. Elliptic Fourier shape approximations for velvetleaf (*Abutilon theophrasti Medicus*) (a) original leaf image, (b) 1st EF harmonic, (c) 1st + 2nd + 3rd + 4th EF, (d) 1st + · · · + 8th EF harmonics, (e) 1st + · · · + 16th EF harmonics and (f) 1st + · · · + 30th EF harmonics.

Table 2
Sample non-normalized Elliptic Fourier coefficients for velvetleaf[a]

Harmonic number	Fourier coefficient value			
	a	*b*	*c*	*d*
1	10.5216	185.9722	184.3727	−20.4650
2	−10.5959	4.1591	−10.6389	0.4225
3	−4.3226	7.8789	−2.5729	6.7795
4	8.4663	0.3244	−5.5601	4.4589
5	2.2916	−3.3906	1.8603	1.5773
6	−2.8998	−4.5531	2.9643	4.5055
7	−0.6946	0.3600	2.1388	−1.2771
8	−2.3432	2.1866	1.8762	−1.8403
9	0.5892	−0.2586	0.0889	−0.1551
10	1.1117	0.9109	−1.0565	−1.5187
11	−0.3360	0.1537	0.2449	0.3746
12	0.9690	−1.1446	−1.7818	0.7122
13	−0.2377	−0.0413	0.1890	0.1710
14	−1.2316	−0.4128	0.5999	0.7302
15	0.2016	0.0895	0.1991	−0.4502
16	0.4839	0.4086	0.3383	−0.4756

[a] First 16 harmonics out of a possible 30.

Table 3
Sample normalized Elliptic Fourier coefficients for velvetleaf[a]

Harmonic number	Fourier coefficient value			
	a	*b*	*c*	*d*
1	1.0000[b]	0.0000	0.0000	−0.9472
2	−0.0134	−0.0062	0.0030	−0.0803
3	−0.0043	0.0086	−0.0010	0.0597
4	−0.0531	−0.0115	0.0008	−0.0203
5	0.0187	−0.0012	0.0086	0.0140
6	−0.0395	−0.0062	0.0015	−0.0001
7	0.0107	−0.0053	−0.0060	0.0032
8	−0.0212	−0.0045	−0.0012	−0.0005
9	0.0002	−0.0020	−0.0007	−0.0028
10	−0.0106	−0.0056	0.0010	0.0024
11	−0.0015	−0.0018	−0.0013	0.0016
12	−0.0080	−0.0092	−0.0009	0.0035
13	0.0017	−0.0006	−0.0003	−0.0004
14	−0.0063	−0.0050	−0.0022	0.0011
15	0.0016	−0.0010	−0.0020	−0.0006
16	−0.0033	−0.0006	−0.0008	0.0030

[a] First 16 harmonics out of a possible 30.
[b] Corrected for leaf orientation and size using the first set of harmonic coefficients (see Fig. 4).

Shape descriptions need to be invariant with rotation, size, translation and starting point on the contour, if they are going to be useful for any size leaf. These are also important if the leaf shape is used as a template for texture analysis. Leaf venation and texture will be also subject to leaf orientation, size and scan direction (Camargo Neto, 2004). The normalization of sample velvet leaf EF coefficients shown in Table 2 was achieved by applying a spatial rotation factor θ_1, which is defined as the angle between the starting point and the first semi-major axis or the same counterclockwise rotation about the contour. θ_1 was therefore defined as:

$$\theta_1 = \frac{1}{2}\arctan\left[\frac{2(a_1b_1 + c_1d_1)}{a_1^2 + c_1^2 - b_1^2 - d_1^2}\right] \quad 0 \le \theta_1 \le 2\pi \tag{7}$$

where a_1, b_1, c_1 and d_1 are the first set of harmonic coefficients.

A rotational transformation matrix using θ_1 (radians) was then applied to all of the harmonic coefficients to provide a new set for the standard leaf orientation, which is given as:

$$\begin{bmatrix} a_n^* & c_n^* \\ b_n^* & d_n^* \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} a_n & c_n \\ b_n & d_n \end{bmatrix} \tag{8}$$

After a starting point correction, a rotation invariant normalization operation was next performed. The semi-major axis of the first harmonic was rotated by angle ψ_1 until it was parallel to the positive x-axis of the first quadrant. The rotation angle ψ_1 was defined using the elliptic locus coordinates of the first point, given as:

$$x_1^*(t^*) = a_1^* \cos\left(\frac{2\pi t^*}{T}\right) + b_1^* \sin\left(\frac{2\pi t^*}{T}\right), \tag{9}$$

$$y_1^*(t^*) = c_1^* \cos\left(\frac{2\pi t^*}{T}\right) + d_1^* \sin\left(\frac{2\pi t^*}{T}\right) \tag{10}$$

Considering that the first harmonic phasor can be aligned with the semi-major axis at the first point, $t^* = 0$, the rotational angle ψ_1 can be obtained as:

$$\psi_1 = \arctan\left[\frac{y_1^*(0)}{x_1^*(0)}\right] = \arctan\left(\frac{c_1^*}{a_1^*}\right), \quad \text{where } 0 \le \psi_1 < 2\pi \tag{11}$$

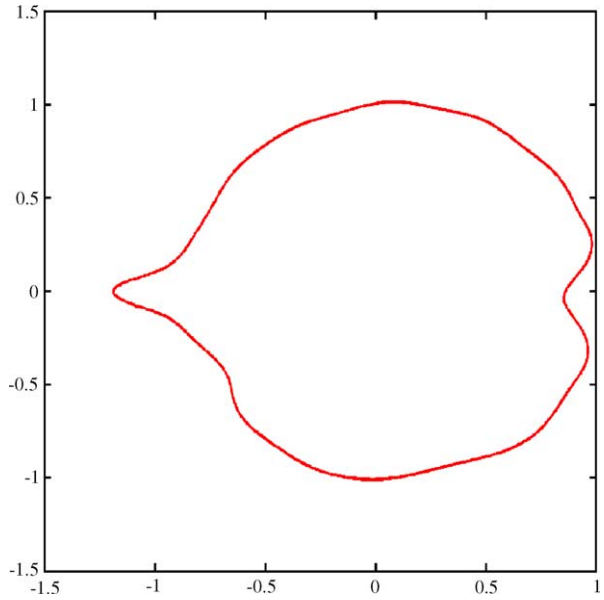

Fig. 4. Final normalized and rotated Elliptic Fourier leaflet boundary for velvetleaf.

Size invariance was accomplished by dividing each coefficient by the magnitude E^* of the semi-major axis, defined as:

$$E^*(0) = (x_1^*(0)^2 + y_1^*(0))^{1/2} = (a_1^{*2} + c_1^{*2})^{1/2} \tag{12}$$

Setting the bias term A_0 and C_0 to zero made the EF coefficient set invariant to translation. The final normalized and rotated EF boundary approximation for the closed-contour of the velvetleaf is shown in Fig. 4. Values for the normalized EF coefficients for the 16 sets of harmonics are also shown in Table 3. (Note that the first three normalized Elliptic Fourier coefficients will be always equal to 1, 0 and 0, respectively).

The number of sets of EF coefficients increases as the number of harmonics increases. For each harmonic set, four coefficients are always generated. For discrimination proposes, the first three coefficients may be ignored. The total number of coefficients was computed to be:

$$Nh = 4h - 3 \tag{13}$$

where h is the harmonic number.

The variation between successive EF boundary approximations was used to measure morphological edge differences or serration along the leaf boundary. For example, leaves extracted from redroot pigweed had more serrations along the leaf edge than did sunflower leaves. This is a typical feature of pigweed leaves. The total variation of Elliptic Fourier boundary approximation (TVH) was computed for the nth harmonic as:

$$\text{TVH} = \sum_{k=2}^{n} \max(\text{sqrt}((x_{k-1} - x_k)^2 + (y_{k-1} - y_k)^2)) \tag{14}$$

where n is the total number of harmonics, x_j and y_j are the Elliptic Fourier boundary approximations for points; x and y along the leaf edge using the jth harmonic. TVH was computed for each leaf sub image and included with the EF coefficients as inputs to the subsequent species discrimination analysis, described below.

Leaf chain coding and Elliptic Fourier processing was performed, using MATLAB script (Version 6.1 with functions from the Image Processing Toolbox (The MathWorks, Inc., 2000). Processing script was also written with *UICONTROL* functions to provide a Windows's graphics user interface to assist operation. All computations were all performed using WINDOWS 2000 Professional® on a Dell 3.2 GHz, Pentium IV computer with one gigabyte of memory (http://www.dell.com/).

2.3. Statistical analyses

Normalized Elliptic Fourier coefficients were generated for the first 30 harmonics for all leaf samples and species. Next, stepwise discriminant analysis, PROC STEPDISC (Principle Component Analysis) from SAS® (SAS Institute, Cary, NC) was performed using all EF coefficients to produce reduced set with the best discriminatory power. Canonical analysis, PROC CANDISC, from SAS® was next performed using the reduced sets of coefficients to conduct a visual interpretation of leaf-species group differences as cluster plots. Canonical analysis created linear species discrimination functions that combined all input features used and presented them as new canonical variables. Statistical analyses were performed for the leaf data set described above to determine the best week to identify plant species, based on leaf shape. The second and third weeks were also analyzed together, because it is possible to have weeds with mixed stages of growth.

SAS species classification models were derived using training leaf images, which was then applied to a test images. To scientifically evaluate classification accuracy, the leaf image set was divided into groups for a five-fold cross-validation. For each experiment run, two files were held out as testing data and the remaining three were used as training data. The cross-validation results were averaged to give a final plant species classification mean.

3. Results and discussion

In this study, at least one fully exposed leaflet was found in the Hindman plant images used. The leaflets extracted from the images appeared to be visually parallel to the camera lens plane. It was therefore assumed that orientation angle was relatively invariant. The leaf subimages were carefully extracted using Photoshop and then analyzed for shape using EF analysis. Misclassifications may have occurred as a result in slight variations in leaf plane orientation, which would have distorted the leaf shape. Otherwise, normalization with size invariance criteria and the first harmonic coefficient should have taken care of the second but larger orientation angle.

3.1. Second week analyses

Initial training and testing results using all Elliptic Fourier coefficients obtained from the 30 harmonics, along with total variation of the EF boundary approximation (TVH) are shown in numerical form in Tables 4 and 5. Error rates for this first SAS species discrimination model using all coefficients were fairly high ranging from 10.9 to 55.5%. This

Table 4
Training data results using all Elliptic Fourier coefficients and cross-validation—second week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Pigweed	44.5	23.8	11.6	20.2	100.0
Sunflower	14.9	81.7	3.5	0.0	100.0
Soybean (VC)	9.5	5.6	80.5	4.5	100.0
Velvetleaf	20.9	6.1	12.3	60.7	100.0
Error rates (%)	55.5	18.3	19.5	39.4	32.9

Table 5
Test data classification results using all Elliptic Fourier coefficients and cross-validation—second week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Pigweed	46.7	16.7	12.2	24.4	100.0
Sunflower	6.7	89.1	4.2	0.0	100.0
Soybean (VC)	4.7	8.9	75.2	11.2	100.0
Velvetleaf	16.0	8.2	16.0	59.8	100.0
Error rates (%)	53.3	10.9	24.9	40.2	32.1

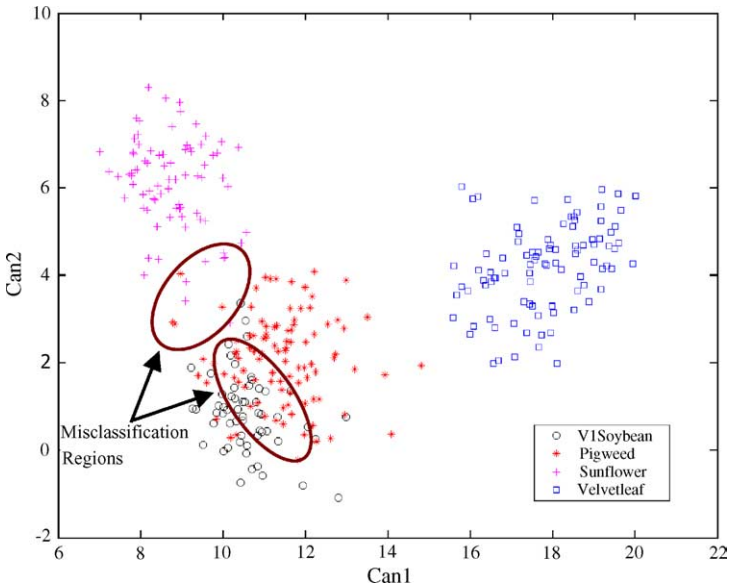

Fig. 5. Canonical analysis for the third week leaf image sets (V1 soybeans with trifoliolates).

Table 6
Training data classification results using all Elliptic Fourier coefficients and cross-validation—second week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Pigweed	87.8	5.5	3.1	3.7	100.0
Sunflower	1.1	98.9	0.0	0.0	100.0
Soybean (VC)	1.1	0.6	97.7	0.6	100.0
Velvetleaf	4.6	0.0	0.5	94.9	100.0
Error rates (%)	12.2	1.1	2.2	5.1	5

model had a relatively low overall correct classification performance of 32%. The model when applied to the test data, correctly identified 89% of sunflower leaves, 46% of redroot pigweed leaves, 75% of soybean leaves and 59% of the velvetleaf. The low classification performance may be due to variations in leaflet orientation relative to the plane of the camera lens and subsequent shape distortion. Also, the EF method should have theoretically found a distinguishing leaflet edge characteristic (first harmonic) that would be located at the same position on similar leaves. Moreover, SAS canonical plotting in Fig. 5 does not always show all of the data points where misclassifications occurred.

Using the SAS STEPDISC procedure, screening of the coefficients was performed using principle component analysis. Overall classification accuracy was then improved, as shown in Tables 6 and 7. Classification error rates of

Table 7
Testing data classification results using selected Elliptic Fourier coefficients and cross-validation—second week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Pigweed	68.3	7.5	7.2	16.9	100.0
Sunflower	4.7	95.3	0.0	0.0	100.0
Soybean (VC)	6.9	2.2	86.1	4.7	100.0
Velvetleaf	16.4	0.0	10.0	73.6	100.0
Error rates (%)	31.7	4.7	13.8	26.4	19.2

Table 8
Training data classification results using all Elliptic Fourier coefficients and cross-validation—third week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (V1)	Velvetleaf	Total
Pigweed	70.5	9.6	14.9	5.1	100.0
Sunflower	7.7	89.2	2.8	0.3	100.0
Soybean (V1)	3.4	0.9	95.7	0.0	100.0
Velvetleaf	9.0	1.9	2.7	86.0	100.0
Error rates (%)	29.6	10.8	4.3	14.0	16.2

Table 9
Testing data classification results using all Elliptic Fourier coefficients and cross-validation—third week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (V1)	Velvetleaf	Total
Pigweed	76.8	6.1	13.1	4.1	100.0
Sunflower	6.3	90.0	3.6	0.0	100.0
Soybean (V1)	3.5	0.0	96.5	0.0	100.0
Velvetleaf	8.6	2.2	4.3	85.0	100.0
Error rates (%)	23.2	10.0	3.5	15.0	14.2

5% with cross-validation for the training data set and 19.2% for the testing data set were obtained, respectively. The species discrimination model correctly identified 95.3% of sunflower, 68.3% of redroot, 86.1% of soybean and 73.6% of velvetleaf species. The number of false classifications was also reduced. Fig. 5 shows the canonical classification results for selected Elliptic Fourier coefficients.

3.2. Third week analyses

Tables 8 and 9 present the training and testing classification rates with all leaflet images and a full set of EF harmonic coefficients for the second and third weeks. The new classification model had error rates for the training and test data at 16.2 and 14.2%, respectively. The model correctly identified 76.8% of redroot pigweed, 90% of sunflower, 96.5% of soybean and 85% of velvetleaf. When selected EF coefficients were used, overall classification accuracy for species identification was again improved, as shown in Tables 10 and 11. The overall classification error rates for the training and test data sets were 6 and 11.7%, respectively. This discriminant model correctly identified 77.9% of redroot pigweed, 93.8% of sunflower, 96.5% of soybean and 89.4% of velvetleaf species.

Fig. 5 shows the canonical discrimination results for species classification during the third week. Trifoliate soybean leaflets are sometimes confused with pigweed leaves, and those are circled on the plots. Those misclassifications are probably due to the similar rounded shape of the trifoliate soybean leaflet that was present in the third week after germination. Also noted in Fig. 5 is a small region where pigweed and sunflower clusters overlap. The performance of

Table 10
Training data results using selected Elliptic Fourier coefficients and cross-validation—third week

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (V1)	Velvetleaf	Total
Pigweed	87.9	2.8	7.6	1.8	100.0
Sunflower	3.4	95.7	0.9	0.0	100.0
Soybean (V1)	2.6	0. 0	97.4	0.0	100.0
Velvetleaf	2.7	0.0	0.3	97.0	100.0
Error rates (%)	12.1	4.3	2.6	3.0	6

Table 11
Testing data classification results using selected Elliptic Fourier coefficients and cross validation—third week after emergence

Species	Mean percentage of leaves classified as				
	Pigweed	Sunflower	Soybean (V1)	Velvetleaf	Total
Pigweed	77.9	6.1	11.0	5.1	100.0
Sunflower	5.0	93.8	1.3	0.0	100.0
Soybean (V1)	3.5	0.0	96.5	0.0	100.0
Velvetleaf	7.5	1.1	2.1	89.4	100.0
Error rates (%)	22.1	6.3	3.4	10.6	11.7

Table 12
Training data results using all Elliptic Fourier coefficients and one-out, cross-validation—second and third weeks after emergence

Species	Mean percentage of leaves classified as					
	Soybean (V1)	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Soybean (V1)	96.0	3.1	0.6	0.3	0.0	100.0
Pigweed	12.1	74.8	6.3	1.7	5.2	100.0
Sunflower	0.9	4.3	94.0	0.9	0.0	100.0
Soybean (VC)	5.9	1.0	0.3	92.2	0.7	100.0
Velvetleaf	2.1	13.0	0.1	3.8	80.3	100.0
Error rates (%)	5.6	28.3	8.4	7.3	17.9	15.9

classification for the third week with an average correct species identification of 88.3% was significantly better than the second week with an average correct species identification of 80.8%. This brings up an important question, as to why the EF was not applied to the entire soybean trifoliate. In reality that discrimination would be potentially much better. Unfortunately, some leaflets on soybean trifoliolate leaves were occluded in the images. The EF method should theoretically work with almost any shape of a simple or compound, fully exposed, and presented leaf image (McLellan and Endler, 1998).

3.3. Second and third weeks combined

Assuming that leaves extracted from the soybean canopy could have both unifoliolate and trifoliolate shapes, the leaf data sets for the second and the third weeks were combined to simulate this situation. Tables 12 and 13 present the training and testing classification rates with all coefficients for both weeks. The overall error rate for the discriminant model was 15.9% for training data and 14.8% for the testing data. This discriminant model correctly identified 76.4% of redroot pigweed, 93.6% of sunflower, 80.3% of velvetleaf, 91.5% of soybean leaf extract from trifoliolate and 90.9% of soybean unifoliolate leaves. When selected Fourier coefficients were used, the training classification error rate was reduced to 9.2% and the testing error rate to 11.6%, as shown in Tables 14 and 15. The reduced set model

Table 13
Testing data classification results using all Elliptic Fourier coefficients and cross-validation—second and third weeks after emergence

Species	Mean percentage of leaves classified as					
	Soybean (V1)	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Soybean (V1)	91.5	6.7	0.00	1.8	0.0	100.0
Pigweed	12.1	76.4	6.5	1.4	3.6	100.0
Sunflower	1.6	4.1	93.6	0.8	0.0	100.0
Soybean (VC)	5.5	3.6	0.0	90.9	0.0	100.0
Velvetleaf	2.1	13.7	0.1	3.8	80.3	100.0
Error rates (%)	8.4	23.6	6.4	9.1	18.4	14.8

Table 14
Training data results using selected Elliptic Fourier coefficients and cross-validation—second and third weeks after emergence

Species	Mean percentage of leaves classified as					
	Soybean (V1)	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Soybean (V1)	97.0	2.2	0.0	0.9	0.0	100.0
Pigweed	6.6	82.0	6.6	1.4	3.4	100.0
Sunflower	0.4	4.4	94.2	1.0	0.0	100.0
Soybean (VC)	1.4	1.8	0.5	94.5	1.8	100.0
Velvetleaf	0.5	5.5	0.0	1.4	92.6	100.0
Error rates (%)	3.0	18.0	5.8	5.5	7.4	9.2

Table 15
Testing data results using selected Elliptic Fourier coefficients and cross-validation—second and third weeks after emergence

Species	Mean percentage of leaves classified as					
	Soybean (V1)	Pigweed	Sunflower	Soybean (VC)	Velvetleaf	Total
Soybean (V1)	94.9	3.3	0.0	1.8	0.0	100.0
Pigweed	7.1	80.8	7.2	1.4	3.6	100.0
Sunflower	0.0	4.9	94.3	0.8	0.0	100.0
Soybean (VC)	5.8	3.6	0.0	88.6	2.0	100.0
Velvetleaf	0.7	7.8	0.7	2.8	88.0	100.0
Error rates (%)	5.2	19.2	5.7	11.5	12.0	11.6

correctly identified 80.8% of redroot pigweed, 94.3% of sunflower, 88% of velvetleaf, 94.9% of soybean leaf extract from trifoliolate and 88.6% of soybean unifoliolate leaves. The average of correct species identification using the data sets for both weeks was 88.4%, similar to the third week alone. The combined sets from weeks two and three obviously provided a larger and better database for species classification. The canonical discrimination variables, CAN1 and CAN2 are plotted in Fig. 6. One may observe species cluster overlapping regions for pigweed and trifoliolate soybean leaves (V1 soybean) than unifoliolate soybean leaves (VC soybean).

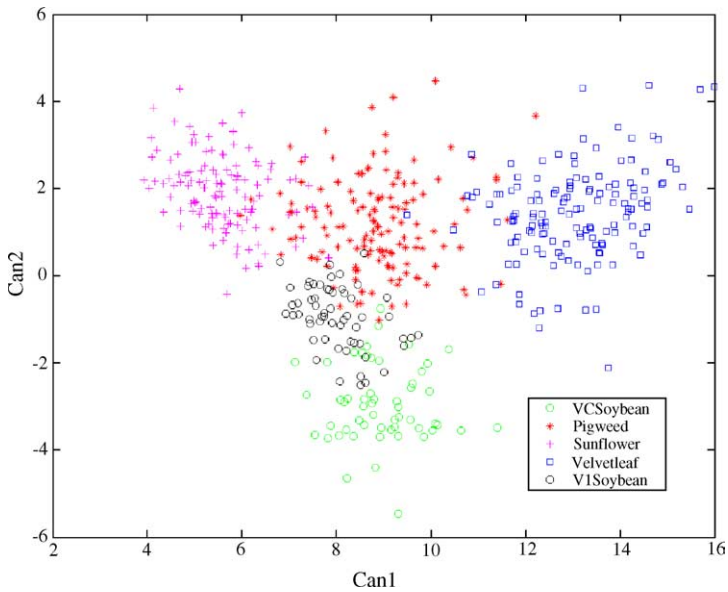

Fig. 6. Canonical analysis for combined second and third week leaf image sets.

4. Summary and conclusions

Plant species can be accurately identified using Elliptic Fourier shape features of whole, extracted leaflets from plant canopies. The EF method combined with principle component analysis and linear discriminant models performed very well. Older plants of the third week had more mature leaves and provided the best leaf images for identifying plant species, demonstrated with an 88.3% classification rate. As leaves develop, their shapes become an important species trademark. The redroot pigweed plant had the lowest correctly identified rate during both weeks. Redroot pigweed is misclassified with soybean during the third week, because of a similar rounded leaf shape to the trifoliolate soybean leaflet at this growth stage. Some lesser misclassification with sunflower also occurred. By combining the leaflet images for the second and third weeks, an overall species identification accuracy of approximately 88.4% was obtained. Future EF studies should consider improved timing, background lighting, improved camera work and application to identifying compound leaves.

Leaf orientation is important in the EF analysis. To our knowledge, this has not been treated in previous plant species imaging studies. Two angles are needed to describe a leaf plane in three-dimensional space. One of the leaf angles is hard to control or adapt to, but sunlit leaves of many species can present themselves heliotropically toward a light source, such that one could select the best camera angles for full leaf exposure at the top of the canopy. The leaf angle in the plane of the canopy is apparently taken care of by the first EF harmonic, and that is a critical angle for rotationally invariant leaf texture or venation analysis. Additional studies regarding leaf orientation relative to the camera lens might help to resolve classification errors. Future studies are also needed to determine minimal digital image resolutions needed to maintain the highest species discrimination performance.

Acknowledgments

The Agricultural Research Division, University of Nebraska-Lincoln, has approved this article as Journal Series No. 14512. This work was supported in part by EMBRAPA, Campinas, Brasil and Nebraska ARD funds. Mention of specific trade names is for reference only and not to imply exclusion of others that may be suitable.

References

Camargo Neto, J., 2004. A Combined Statistical—Soft Computing Approach for Classification and Mapping Weed Species in Minimum Tillage Systems. Unpublished Ph.D. Dissertation. University of Nebraska, Lincoln, NE, 117 pp.

Chen, S.Y.Y., Lestrel, P.E., Kerr, W.J.S., McColl, J.H., 2000. Describing shape changes in the human mandible using Elliptic Fourier functions. Eur. J. Orthod. 22, 205–216.

Chi, Y.T., Chien, C.F., Lin, T.T., 2002. Leaf shape modeling and analysis using geometric descriptors derived from Bezier curves. Trans. ASAE 46 (1), 175–185.

Franz, E., Gebhardt, M.R., Unklesbay, K.B., 1991a. Shape description of completely visible and partially occluded leaves for identifying plants in digital images. Trans. ASAE 34 (2), 1991.

Franz, E., Gebhardt, M.R., Unklesbay, K.B., 1991b. The use of local spectral properties of leaves as an aid for identify weed seedlings in digital images. Trans. ASAE 34 (2), 1991.

Guyer, D.E., Miles, G.E., Schreiber, M.M., Mitchell, O.R., Vanderbilt, V.C., 1986. Machine vision and image processing for plant identification. Trans. ASAE 29 (6), 1500–1507.

Guyer, D.E., Miles, G.E., Gaultney, L.D., Schreiber, M.M., 1993. Application of machine vision to shape analysis in leaf and plant identification. Trans. ASAE 31 (1), 163–171.

Hindman, T.W. 2001. A fuzzy logic approach for plant image segmentation and species identification in color images. Unpublished Ph.D. Dissertation. University of Nebraska, Lincoln.

Innes, D.J., Bates, J.A., 1999. Morphological variation on *Mytilus edulis* and *Mytilus trossulus* in eastern Newfoundland. Mar. Biol. 133, 691–699.

Kincaid, D.T., Schneider, R.B., 1983. Quantification of leaf shape with a microcomputer and Fourier transform. Can. J. Bot. 61, 2333–2342.

Knezevic, S.Z., Evans, S.P., Blankenship, E.E., Van-Acker, R.C., Lindquist, J.L., 2002. Critical period for weed control: the concept and data analysis. Weed Sci. 50, 773–786.

Kuhl, F.P., Giardina, C.R., 1982. Elliptic Fourier features of a closed contour. Comput. Graphic Image Process. 18, 236–258.

Manh, A.G., Rabatel, G., Assemat, L., Aldon, M.J., 2001. Weed leaf image segmentation by deformable templates. J. Agric. Eng. Res. 80 (2), 139–146.

Manley, B.S., Wilson, H.P., Hines, T.E., 2001. Weed management and crop rotations influence populations of several broadleaf weeds. Weed Sci. 49, 106–122.

McLellan, T., Endler, J.A., 1998. The relative success of some methods for measuring and describing the shape of complex objects. Syst. Biol. 47 (2), 264–281.

Medlin, C.R., Shaw, D.R., Gerard, P.D., LaMastus, F.E., 2000. Using remote sensing to detect weed infestations in Glycine Max. Weed Sci. 48, 393–398.

Meyer, G.E., Camargo-Neto, J., Jones, D.D., Hindman, T.W., 2004a. Intensified fuzzy cluster for determining plant, soil, and residue regions of interest from color images. Electron. Agric. (Elsevier) 43 (3), 161–180.

Meyer, G.E., Hindman, T.W., Jones, D.D., Mortensen, D.A., 2004b. Digital camera operation and fuzzy logic classification of plant, soil, and residue color images. Eng. Agric. 20 (4), 519–529.

Oide, M., Ninomiya, S., 2000. Discrimination of soybean leaflet shape by neural networks with image input. Comput. Electron. Agric. 29, 59–72.

Petry, W., Kuhbauch, W., 1989. Automatisierte unterscheidung von unkrauten nach formparametern mit hilfe der quantitativen bild analyse. J. Agronomy Crop Sci. (Berlin) 163, 345–351.

Timmermann, C., Gerhards, R., Kuhbauch, W., 2003. The economic impact of site-specific weed control. Precision Agric. 4, 249–260.

Woebbecke, D.M., G.E. Meyer, B.K.V., Mortensen, D.A., 1995. Shape features for identifying young weeds using image analysis. Trans. ASAE 38 (1), 271–281.

Zimdahl, R.L., 1988. The concept and application of the critical weed-free period. In: Altieri, M.A., Liebman, M. (Eds.), Weed Management in Agroecosystems, Ecological Approaches. CRC Press, Boca Raton, FL, pp. 145–155.