CLASSIFICATION OF WEED SPECIES USING COLOR TEXTURE FEATURES AND DISCRIMINANT ANALYSIS

T. F. Burks, S. A. Shearer, F. A. Payne

ABSTRACT. *The environmental impact of herbicide utilization has stimulated research into new methods of weed control, such as selective herbicide application on highly infested crop areas. This research utilized the Color Co-occurrence Method (CCM) to determine whether traditional statistical discriminant analysis can be used to discriminate between six different classes of groundcover. The weed species evaluated were giant foxtail, crabgrass, common lambsquarter, velvetleaf, and ivyleaf morningglory, along with a soil image data set. The between species discriminant analysis showed that the CCM texture statistics procedure was able to classify between five weed species and soil with an accuracy of 93% using hue and saturation statistics, only. A significant accomplishment of this work was the elimination of the intensity texture features from the model, which reduces computational requirements by one-third.*

Keywords. *Machine vision, Image processing, Precision farming, Herbicide application, Leaf and plant identification.*

The environmental impact from herbicide utilization has been well-documented in recent years. As a result, efforts to protect groundwater quality by restricting the use of certain herbicides is likely to increase in the years to come. However, society is likewise concerned about maintaining cost-effective production of agricultural crops. According to Bridges and Anderson (1992), "Weeds pose one of the most important threats to our supplies of food and fiber. Losses in both yield and quality of crops due to weeds, as well as costs of weed control, constitute an enormous problem in all agricultural areas." The reduction in use of herbicides without a viable alternative will likely result in decreased production and thus higher unit production costs.

Chemical application technologies are being targeted to reduce herbicide use and maintain production levels. Selective herbicide application to weed-infested areas of the field, rather than the entire field, was suggested by Thompson et al. (1991). The goal of this research was to investigate the potential for using a machine vision system to identify localized crop areas which are threatened by weed competition, thereby enabling reduced reliance on chemical application in other areas of a field.

Article has been reviewed and approved for publication by the Information & Electrical Technologies Division of ASAE. Presented as ASAE Paper No. 98-3037.

The investigation in this Paper No. 97-05-172 is in connection with a project of the Kentucky Agricultural Experiment Station and is published with approval of the Director. The paper was originally prepared for the 1996 Society of Automotive Engineers, Summer meeting in Indianapolis, Indiana.

The authors are **Thomas F. Burks**, Post Doctoral Scholar, **Scott A. Shearer**, ***ASAE Member Engineer***, Associate Professor, and **Fred A. Payne**, ***ASAE Member Engineer***, Professor, Department of Biosystems and Agricultural Engineering, University of Kentucky, Lexington, Kentucky. **Corresponding author:** Thomas F. Burks, University of Kentucky, 128 Agricultural Engineering Bldg., Lexington, KY 40546-0276, phone: 606.257.3000, fax: 606.257.5671, e-mail: <tburks@bae.uky.edu>.

IMAGE PROCESSING TECHNIQUES

The identification of various plants and crops using image processing techniques has been addressed by several researchers. A weed detection system for Kansas wheat was developed using four color filters (Chaisattapagon and Zhang, 1991). Gray scale ratios were used to discriminate between three weed species common to wheat fields (kochia, redroot pigweed, and Russian thistle). The method was capable of identifying the weeds with respect to the crop, soil and sandstone, but did not classify individual species. Woebbecke et al. (1995a) used RGB color features to discriminate between weeds, soil, and residue background features. They found that color features could be used to segregate plant and background features, but were inadequate for separating plant species.

Image shape matching functions were used to match weed species of completely visible and partially occluded leafs (Franz et al., 1990). Classification of velvetleaf, ivyleaf morningglory, giant foxtail, and soybean seedlings were studied at growth stages of fully expanded cotyledons and fully developed first true leaves. Image classification was found to be independent of leaf scale and orientation. However, the method was found to be an inadequate descriptor of leaves with variable leaf serrations and for aggregate boundaries of multiple leaves. In a later work, Zhang and Chaisattapagon (1995) used four leaf shape features to classify weeds common to wheat fields (kochia, redroot pigweed, Russian thistle, and wild buckwheat). They found that eccentricity, compactness, and two invariant moments where sufficient to classify weed species using a multivariate discriminant analysis. Woebbecke et al. (1995b) conducted an extensive study on the use of shape features for the classification of 10 weed species, corn, and soybeans. They grouped the species according to monocotyledon (corn and two weed species) and dicotyledon (soybeans and eight weed species), and then further separated the dicotyledon into three random sub-groups. They found that the aspect ratio and the first invariant central moment were the most successful for discriminating between monocots and dicots, achieving a greater than 80% classification accuracy. However, they noted that "...young weed plants can be highly variable in canopy shape features..." and "... there is no guarantee that any particular shape feature will continue to work successfully as a plant classifier as the size of the plant increases". They also noted the difficulties associated with wind movement and plant canopy interaction. They conclude that the best time to separate monocots and dicots is between 14 and 23 days of development, which is a time favorable for post-emergence weed control.

The reflectance of leaf surfaces in the near-infrared, red, and blue wavebands was used to classify leaves from soybeans, velvetleaf, ivyleaf morningglory, and foxtail (Franz et al., 1991). The image intensities in the three wavelengths were measured near the leaf periphery and were statistically evaluated for mean, variance, and skewness. They achieved a classification accuracy of 97% when leaf orientation was not a factor, and 75% when leaf orientation was considered. Their conclusion was that leaf orientation with respect to the light source was critical for reflectance measurement at these wavelengths. Wang et al. (1999) reported the development of a spectral-based weed sensor. They grouped nine weed species together and used discriminant analysis to classify between wheat, soil, and weeds achieving classification accuracies of 98.3%, 98.7%, and 64.3%, respectively.

Levine (1985) observed that textures represents the phenomenon of shift invariance in image structure, where visual perception of the image is not affected by position within the image pattern. Statistical texture patterns (stochastic) are random and usually occur naturally. These patterns may represent rough seas, grasses, crops and forest. The spatial gray-level dependencies method for statistical texture analysis was introduced by Julesz (1962). Julesz worked with first and second order statistics of gray scale images. Haralick and Shanmugam (1973) generalized the concept of using second order statistics into spatial gray-level dependencies matrices (SGDM). These matrices were later used by Haralick and Shanmugam (1974) to develop a set of 16 texture features associated with the SGDMs.

The use of color features in classical gray image texture analysis techniques was first reported by Shearer (1986). Shearer and Holmes (1990) reported accuracies of 91% for classifying different types of nursery stock by the color co-occurrence matrix (CCM) method. The traditional gray image texture features were expanded by Shearer (1986) to utilize HSI (hue, saturation and intensity) color features. As a result, the HSI color co-occurrence method consisted of three co-occurrence matrices, one each for the hue, saturation, and intensity color features. These matrices were used to generate the texture features suggested by Haralick and Shanmugam.

Meyer et al. (1998) used red, green, blue (RGB) true color to produce an excess green color feature for discriminating between four different species of weeds and soil regions with a 99% accuracy. The weed species studied were Shattercane, Green foxtail, Velvetleaf and Red Root Pigweed. They used the traditional gray scale co-occurrence matrix to generate four texture statistics: angular second moment, inertia, entropy, and local homogeneity. They observed classification accuracies of 93% for grasses and 85% for broadleaf categories when using angular second moment, inertia and local homogeneity. However, they were only able to achieve individual species classification accuracies of 30 to 77%. Tang et al. (1999) used a Gabor wavelets-based feature extraction method and neural networks to classify images into broadleaf and grass categories. They achieved 100% classification accuracies when testing 20 sample images from each of the two categories.

Color Co-occurrence Method

The image analysis technique selected for this study was the CCM method. This method has proven ability to discriminate between multiple canopy species as reported by Shearer and Holmes (1990), and appears to be insensitive to leaf scale and orientation. The use of color features in the visible light spectrum provide additional image characteristic features over the traditional gray-scale representation.

The CCM procedure consists of three primary mathematical processes. First the image is transformed from an RGB color representation to an HSI color representation. Each image pixel is transformed from the RGB value to a corresponding HSI pixel value. The intensity is calculated using the mean value of the three RGB values. The hue and saturation values are determined using geometrical transformation of the CIE chromacity diagram (Ohta, 1985). In this process, the CIE chromacity diagram represents a two-dimensional hue and saturation space (Wyszecki and Stiles, 1992). The pixel RGB values determine the chromacity coordinates on the hue and saturation space, which are then used to geometrically calculate the value of hue and saturation. This process has been documented by Shearer (1986).

Once the RGB image was transformed into the three HSI pixel maps, each pixel map was used to generate a color co-occurrence matrix, resulting in three CCM matrices. That is, one CCM matrix for each of the HSI pixel maps. The color co-occurrence texture analysis method was developed through the use of spatial gray-level dependence matrices (SGDMs). These matrices measure the probability that a pixel at one particular gray-level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray-level (Shearer and Holmes, 1990). The SGDMs are represented by the function $P(i,j,d,\theta)$ where i represents the gray-level of location (x,y) in the image I(x,y), and j represents the gray-level of the pixel at a distance d from location (x,y) at an orientation angle of θ (Shearer, 1986). This is illustrated in figure 1c, where i is the row indicator and j is the column indicator in the SGDM matrix $P(i,j,d,\theta)$. The nearest neighbor mask is shown in figure 1a, where the reference pixel (x,y) is shown as an asterisk. All eight neighbors shown are one pixel distance from the reference pixel '*' and are numbered in a clockwise direction from one to eight. The neighbors at positions one and five are both considered to be at an orientation angle equal to 0°, while positions eight and four are considered to be at an angle of 45°. An example image matrix I(x,y), with a gray scale range of zero to three is shown in figure 1b. This process may be understood by looking for the number of occurrences of zero adjacent to three in I(x,y) in figure 1b. When looking at both the one and five neighbor positions as each row of the matrix is scanned (forward and backward looking), we find a total of two occurrences of zero adjacent to three. This corresponds to the two value at matrix location (0,3) in P(i,j,1,0), shown in figure 1c. The resulting SGDMs for other i and j values are shown in figure 1c. This study used a one pixel offset distance and a zero degree orientation angle (neighbors one and five).

The hue, saturation, and intensity CCM matrices are then used to generate the texture features described by Haralick and Shanmugam (1974). Shearer and Holmes (1990) reported a reduction in the 16 gray scale texture features through elimination of redundant variables. The resulting 11 texture feature equations are defined by Shearer and Holmes (1990) and Burks (1997). The same equations are used for each of the three CCM matrices, producing 11 texture features for each HSI component and thereby a total of 33 CCM texture statistics. The texture features are identified by a coded variable name where the first letter represents whether it is a hue (H), saturation (S) or intensity (I) feature and the number following represents one of the eleven texture features described in Shearer and Holmes (1990). As an example, the feature (I_7) is a measure of the entropy in the intensity CCM matrix which represents the amount of order in an image and is calculated by equation 1.

$$I_7 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \ln p(i,j) \qquad (1)$$

The p(i,j) matrix represents the normalized intensity co-occurrence matrix, and N_g is the total number of intensity levels. The equation for normalizing the co-occurrence matrix is given in equation 2, where P(i,j,1,0) is the intensity co-occurrence matrix.

$$p(i,j) = \frac{p(i,j,1,0)}{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j,1,0)} \qquad (2)$$

A physical representation of entropy (uncertainty) may be visualized by comparing a checkerboard-like image to an image where one half is black and the other half is white. The latter image is highly ordered having all pixels of the same intensity segregated into two distinct pixels groups which gives greater certainty of the pixel value of the adjacent pixels. The checkerboard image has a lower amount of order due to intermixing of black and white squares which results in a greater level of uncertainty of neighboring pixel values. The lower order image would therefore have more uncertainty and thus a higher entropy measure.

Objective

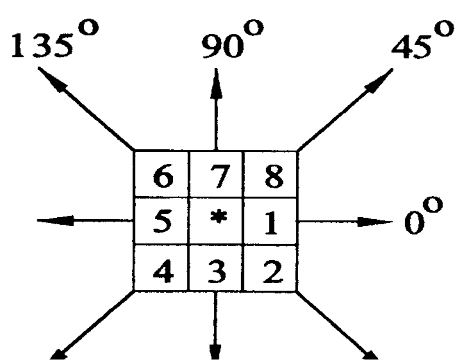
The objectives of this study were to:

- Determine if HSI color features can be used to discriminant between soil and plant matter (weeds).
- Generate CCM texture feature data models using six classes of groundcover: giant foxtail, crabgrass, velvetleaf, lambsquarter, ivyleaf morningglory and soil, and then use stepwise discriminant analysis techniques to identify combinations of the CCM texture feature variables which have the highest classification accuracy with the least number of texture variables.
- Train a discriminant classifier to identify weeds using the models generated in the preceding objective, and conduct classification tests with each model to determine their potential for classifying weed species.

Experimental Methods

Plant Specimen Preparation

Five weed species common to Kentucky rowcrops were selected for this study. The weed species investigated were ivyleaf morningglory *(Ipomoea hederaceae)*, giant foxtail *(Setaria faberi)*, large crabgrass *(Digitaria sanguinalis)*, velvetleaf *(Abutilon theophrasti)*, and common lambsquarter *(Chenopodium album)*. Each of the five species were grown from seed in a sterilized plant bed with average germination times of approximately 7 to 10 days. The weed species were planted in individual 0.6 m × 3.0 m plots and allowed to grow under normal ambient conditions until they reached an appropriate maturity level. The



(a)

$$I(x,y) = \begin{bmatrix} 0 & 0 & 3 & 1 \\ 2 & 1 & 0 & 2 \\ 3 & 2 & 0 & 3 \\ 1 & 2 & 1 & 3 \end{bmatrix}$$

(b)

$$P(i,j,1,0^{o}) = \begin{bmatrix} 2 & 1 & 2 & 2 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ 2 & 2 & 1 & 0 \end{bmatrix}$$

(c)

Figure 1–Gray-level dependence example: (a) nearest neighbor diagram, (b) gray-level image, and (c) SGDM for different orientations.

Table 1. Weed species listing

Speci-men	Species Description	Avg. Leaf* Count	Avg. Height (cm)	Min. Height (cm)	Max. Height (cm)	Range of Plant Density† (%)	Mean Plant Density‡ (%)
1	Crabgrass	4.8	5.6	3.8	7.6	86.8 to 100	95.9 ± 6.7
2	Foxtail	4.8	8.6	6.4	10.2	87.5 to 99.9	96.3 ± 6.6
3	Lambsquarter	9.6	5.2	3.2	7.6	53.4 to 99.6	86.4 ± 25.1
4	Morning glory	3.0	5.5	3.8	6.4	91.3 to 100	98.3 ± 4.1
5	Velvetleaf	4.0	4.7	3.8	5.7	84.4 to 100	97.2 ± 5.9
6	Soil	-	-	-	-	0.0	0.0

* The average leaf count per plant with in the species maturity range.
† The plant density is listed in terms of percent plant matter versus soil with in the specimen sub-image. The range is listed as the minimum and maximum values for the species.
‡ The mean plant density is given for each species with the range listed as the mean value plus or minus two standard deviations.

general plant data for each of the six classes of observations are presented in table 1. The average leaf count, average height, minimum height, maximum height, and plant density are provided for each class of observation. This study focused on a plant maturity which ranged in average height from 4.7 cm to 8.6 cm as shown in table 1. Once the weed species reached maturity, an image acquisition system was employed to capture 40 raw images of each specie. The individual 512 × 400 pixel raw images covered approximately 23 cm × 30 cm of ground surface area, which included both plant matter and soil. The plant density within the sub-image was determined by pixel counting using a bi-modal threshold based on the pixel hue value.

It was observed during earlier investigations, that the combination of windy conditions with partly cloudy skies produced rapid changes in ambient light conditions and created significant leaf movement. In an attempt to minimize the influence of changes in ambient conditions, a diffuse off-white cotton cover was placed over the top of the image acquisition system and supplemental lighting was used to compensate for the effects of cloud cover and angle of sun. Additionally, a nylon canvas was used as a wind break around the windward sides and base of the image acquisition system to prevent excessive motion in the plant leaves.

Image Acquisition System

A field imaging system was used to collect the digital RGB images. The assembly was a wheel mounted unit which provided a mobile self contained digital acquisition system. Four, 250 W Sylvania Superflood BCA lights (5000°K) with diffuser covers were suspended at each of the four corners of a 1.0 m × 1.0 m frame assembly. The lights were adjusted to equal heights above the ground (approximately 1.5 m) and then directed toward the center of the illumination area to eliminate shadows under the camera mounting assembly. A JVC, model TK-870U RGB color video camera with three CCDs was mounted on a height adjustable assembly suspended at the center of the unit at a fixed height of approximately 0.9 m above ground level. The camera output was connected to the video input of an ATT, TARGA 24 image capture board, which was installed in a IBM compatible 386-16 MHZ PC. The acquired image resolution was 512 × 400 pixels. In addition, a Mitsubishi analog video monitor was connected to the TARGA 24 analog output and used for live image monitoring.

A computer software program termed GCVIS (Ground Cover VISion) was developed to drive the imaging system, generate RGB image files, create a HSI color model of the image, generate color co-occurrence matrices, and calculate CCM statistics for the specified matrices. The image acquisition and CCM texture analysis method is illustrated in a flowchart format in figure 2. A complete description of GCVIS can be found in Burks (1997).

Image Processing and CCM Statistics

Six classes of field observations were captured in 512 × 400 pixel digital images which included both soil and plant matter. Digitized, 128 × 128 pixel gray scale, images for each class of observation are presented in figure 3. The full size raw images contained a high percentage of soil matter in some species. To select isolated class specimens from within the raw image, smaller sub-images were manually extracted for each species using a 64 × 64 pixel window template (approximate 3.8 cm × 3.8 cm canopy field of view). This size sub-image provides adequate resolution for a wide range of plant maturity. A sub-image was extracted by placing a window template over the desired region of the displayed raw image and then executing the extract routine which would automatically create a new image file containing the 64 × 64 pixel RGB image data. Each species' image data set consisted of 40 sub-images. Sub-images were selected from the raw images to acquire the highest percentage of plant matter. The percentage of plant matter in the sub-image was determined after the sub-image extraction process, ranging from 53% to 100% for sub-images used in this report.

The plant density within the sub-image was determined by pixel counting using a threshold based on the pixel hue value. It was found that pixel hue can accurately

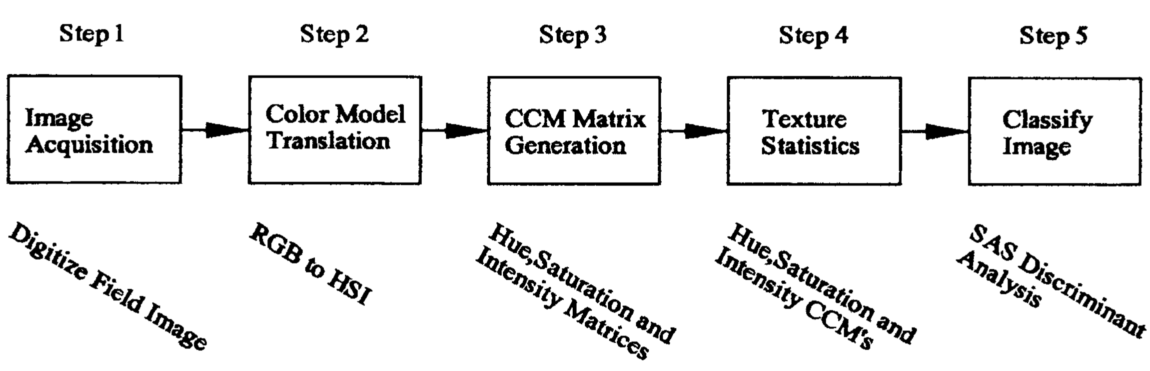


Figure 2–Image acquisition and CCM texture analysis flowchart.

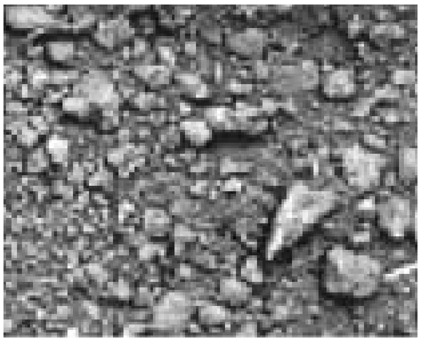
(a) Crabgrass

(b) Foxtail

(c) Lambsquarter

(d) Morningglory

(e) Velvetleaf



(f) Soil

Figure 3–Typical weed species image (approximately 8 cm × 8 cm coverage). Images were digitized from backup photographic slides, and do not represent actual digitized images used in the investigation.

discriminate between soil and plant material consisting of the five weed species under investigation. However, it could not discriminant between species. The mean hue value for soil sub-images was 26.6 (0 to 63 scale) with a standard deviation of 3.5, while the mean hue value for all plant species observed was 39.1 with a standard deviation of 7.1 (these plant sub-images contained small areas of soil). When the histograms of the soil and plant hue values were plotted and overlayed, a very distinct bi-modal effect was observed and a threshold of 34 provided good separation for the two classes of observations, (Burks, 1997). As indicated in table 1, all soil images showed a 0% plant matter content as would have been expected. Even though the hue based bi-modal classification procedure has shown potential for segregating plant from soil, for the purpose of this study it was decided to include soil in the discriminant analysis to ensure that the CCM method could identify various weed species images from soil images. This is especially important considering that the image specimens had up to 47% soil surface in the sub-image.

The HSI model was generated for each sub-image using the RGB image file created during sub-image extraction. These RGB files contained pixel data scaled from 0 to 255, for each of the respective colors. However, the size of the CCM matrices is dependent upon the range of the image scale. RGB images scaled from 0 to 255 required 197 Kbytes of memory to store the three 256×256 CCM matrices. This greatly increases the CPU time required to process the images. As a result, the RGB values were re-scaled to 0 to 63 reducing the storage requirements by a factor of 16. The HSI conversion was performed and the resulting HSI pixel maps were scaled to values between 0 and 63.

Thirty-three CCM texture statistics were generated for each of the 240 observations using the sub-image HSI pixel maps. Each HSI pixel map was evaluated using the CCM approach to generate three nearest neighbor maps, one each for hue, saturation and intensity. The approach used in selecting the SGDM orientation angle and offset distance for a specific application varies depending upon the texture pattern being considered and the spatial resolution available. Some researchers have limited their investigations to a one pixel offset and evaluate the system response to change in the orientation angle (0°, 45°, 90°, and 135°) (Shearer, 1986). The 0° orientation angle and one pixel offset distance were arbitrarily selected for this study, based on prior experience. The three SGDM nearest neighbor maps were then evaluated by 11 texture statistics yielding a 33 variable statistic set for each observation. The 240 observation CCM texture statistics output set was written into a common file which was formatted for export to SAS statistical package (SAS, 1985) for analysis.

Statistical Analysis

The SAS STEPDISC procedure was used to reduce the number of variables for discriminant analysis (SAS, 1985). STEPDISC evaluated the 33 CCM texture variable set using the stepwise test procedure for entering and removing variables from the model. The Stepwise procedure begins with no entries in the model. At each step of the process, if the variable within the model which contributes least to the model, as determined by the Wilks' lambda method, does not pass the test to stay it is removed from the model. The variable outside the model which contributes most to the model and passes the test to be admitted is added. When no more steps can be taken the model is reduced to its final form. A test significance level of 0.0001 was used for both the SLS (test for variable to stay) and the SLE (test to enter) variables. Earlier research by Shearer (1986) had shown that plant canopy texture features can be represented by a multi-variate normal distribution. A similar assumption was made for this data set.

Five different reduced models were created using STEPDISC on various combinations of the three HSI/CCM texture statistic data sets. The training data set consisted of six classes with 20 images per class. Each of the unreduced models used different combinations of the original 33 texture features per image. The un-reduced variable sets consisted of 11 hue texture features, 22 hue and saturation texture features, all 33 HSI texture features, 11 saturation texture features, and 11 intensity texture features. Model design was motivated by the desire to minimize the computational requirements of the image processing algorithms, while maintaining high classification accuracy.

The SAS DISCRIM procedure evaluated the ability of the six models to discriminate between the six classes of groundcover. Five models were used from the previous STEPDISC variable reduction study, and a sixth model was added which was an unreduced data set containing all 33 variables. The discriminant function is established using a measure of the generalized squared distance between a specific test image texture variable input set and the class texture variable means, with an additional criteria being the posterior probability of the classification groups (Rao, 1973; SAS, 1985). Each test observation is placed in the class for which it has the smallest generalized square distance between the test observation and the selected class, or the largest posterior probability of being in the selected class (SAS, 1985). The DISCRIM procedure utilized a likelihood ratio test for homogeneity of the within-group covariance matrices, at a 0.1 test significance level. A training data set and a test data set, each consisting of six classes with 20 images per class, were created for each of the STEPDISC reduced models described above. The training set was used to train DISCRIM for classifying weed species and the test data set was used to evaluate the different model's classification accuracy. A more complete description of the experimental method used in this study can be found in Burks (1997).

Results and Discussion

The field observations from six weed classes were individually digitized and saved as RGB data files. The velvetleaf, crabgrass, ivyleaf morningglory, giant foxtail, lambsquarter, and soil samples provided a total of 240 sub-images (40/species), which were divided equally into training and test data sets. These images provided the RGB data used to generate the HSI color model data and the CCM texture statistics features.

The mean plant matter content for all weed species images combined was 94.8%, when using hue in a bimodal threshold classifier. Lambsquarter images exhibited the greatest variation with a mean of 86.4% and a standard deviation of 12.5%, as shown in table 1. The range of plant matter coverage for lambsquarter was 53.4% to 99.6%. Considering the lambsquarter classification accuracies that were observed, these results indicate that the CCM method is robust in the presence of a wide range of plant matter content.

Five CCM texture statistic models were analyzed using the SAS STEPDISC procedure to determine a reduced variable set with the greatest discriminant power. The STEPDISC results for each model are shown in table 2. The number of variables in the original model is dependent upon the number of HSI color features used. For instance, model 3 had 33 variables in the original model since it had hue, saturation and intensity texture statistics, while model 2 had 22 variables in the original model. The STEPDISC selected variables are listed in the order in which they were added to the model, and are therefore listed from highest to lowest significance. In the case of

Table 2. STEPDISC test variable sets

Model	Color Features	STEPDISC Reduced Variable List*
1	Hue	H2,H6,H7,H4,H1,H8,H10
2	Hue, Saturation	S2,H10,H2,H4,S6,H9,S9,H6,H1,H5,H11
3	HSI	S2,H10,H2,I9,S6,H9,S9,H4,H6,I5
4	Saturation	S1,S2,S6,S9
5	Intensity	I7,I2,I8,I6,I11

* The variables listed are coded so that variables prefixed by "H" correspond to hue CCM statistics, "S" corresponds to saturation CCM statistics, and "I" corresponds to intensity CCM statistics. The number indicates the texture statistic, as follows: (1) 2nd Moment, (2) Mean Intensity, (3) Variance, (4) Correlation, (5) Product moment, (6) Inverse difference, (7) Entropy, (8) Sum entropy, (9) Difference entropy, (10) Information correlation measure no. 1, and (11) Information correlation measure no. 2.

model 3, the number of CCM texture statistics selected to classify the images were reduced from the original 33 variables to 10 variables. The CCM texture statistic variable names are encoded using a single letter and a numeric extension. The footnotes to table 2 describe the interpretation of these statistics.Model 2 reduced from an original 22 variables to 11 variables using STEPDISC. However, since Model 2 requires only hue and saturation CCM texture statistics, the computational requirements will be less then Model 3 which has 10 variables, but requires that all three CCM matrices be generated. The elimination of one of the HSI color features is a significant reduction in computational requirements.

The training data set for each model was presented to SAS DISCRIM to train the Generalized Squared Distance classifier. Then, the test data set for the corresponding model was evaluated using DISCRIM's classification test. The process was repeated for all six models. The results presented in table 3 illustrate the classification accuracy of Model 2, which had an average classification accuracy of 93%. The lowest per species classification accuracy was 90% and soil was classified with 100% accuracy. There were 22 samples classified into the foxtail class even though there were only 20 samples in the original set. Two of the foxtail samples were classified outside the species, while two morningglory images, one crabgrass and one lambsquarter were mis-classified into foxtail. If crabgrass and foxtail were considered as a grasses category, and lambsquarter, morningglory, and velvetleaf were considered as a broadleaf category, Model 2 classified grasses with an 98% accuracy and broadleaf with an 95% accuracy.

The results of the six DISCRIM tests are given in table 4. Model 6, which had all 33 texture variables, exhibited 92% average classification accuracy. Likewise, Model 3 (STEPDISC reduced version of Model 6) achieved a 92% overall classification accuracy, showing that there is no advantage in having all CCM texture variables in the model. The models which used only one HSI color feature showed a loss of classification accuracy as might be expected. The lowest performer was Model 5 at 73%. Model 5 utilized the intensity HSI color feature, which is the equivalent of traditional gray-scale. This demonstrates the superior weed species discrimination capability of CCM texture descriptors over traditional gray-scale texture descriptors. Interestingly, Model 2 utilized hue and saturation features and was able to classify the observations with an average accuracy of 93%. This model results in a significant reduction in CPU time since the intensity CCM matrices would not be required.

Table 3. Classification results for Model 2 using PROC DISCRIM

From Species	Number of Observations Classified into Species: Crab-grass	Fox-tail	Lambs-quarter	Morning glory	Velvet-leaf	Soil	Classification Accuracy (%)
Crabgrass	19	1	0	0	0	0	95
Foxtail	1	18	1	0	0	0	90
Lambsquarter	0	1	18	1	0	0	90
Morning glory	0	2	0	18	0	0	90
Velvetleaf	0	0	0	1	19	0	95
Soil	0	0	0	0	0	20	100
Total count	20	22	19	20	20	19	93.3
Percent	16.7	18.3	15.8	16.7	16.7	15.8	
Prior percent	16.7	16.7	16.7	16.7	16.7	16.7	

Table 4. PROC DISCRIM classification accuracy in percentage

DISCRIM Model	Color* Feature	Crab-Grass	Fox-tail	Lambs-quarter	Morning glory	Velvet-leaf	Soil	Classification Accuracy
1	H	90	95	50	90	95	100	86.7
2	HS	95	90	90	90	95	100	93.3
3	HSI	95	85	90	95	85	100	91.7
4	S	70	80	90	80	85	100	84.2
5	I	60	65	80	75	55	100	72.5
6	ALL	90	85	100	90	85	100	91.7

* The listed color features refer to the base STEPDISC model variable set. For Instance, Model 1 used only hue texture features. In this column, H represents hue, S represents saturation and I represents intensity. HS refers to hue and saturation and HSI refers to all three features. The entry ALL refers to a Model which was not reduced using STEPDISC.

SUMMARY AND CONCLUSION

A set of 240 test images from six different classes of ground cover were digitized. The image classes included: giant foxtail, crabgrass, velvetleaf, lambsquarter, ivyleaf morningglory, and soil. An image acquisition program named GCVIS was used to extract 40 digital sub-images per species and to generate the texture feature data for each sub-image. It was found that a hue-based threshold classifier was capable of classifying between plant and soil pixels with a high degree of accuracy (100% accurate for soil pixels). SAS procedures were used to determine the discriminant powers of the texture features and their classification accuracy. The use of SAS procedure STEPDISC proved beneficial for reducing the number of variables in the data models. Model two was reduced from 22 to 11 variables, while model three was reduced from 33 to 10 variables.

The color co-occurrence method was shown to have a overall classification accuracy of 93% when using the 11 texture features selected for Model 2. The within class classification accuracy was 90% or higher for all six classes. These results show improvement over Meyer et al. (1998) which had individual classification accuracies as low as 30 to 77%. If crabgrass and foxtail were considered as a grass category, and lambsquarter, morningglory and velvetleaf were considered as a broadleaf category, Model 2 classified grasses with an 98% accuracy and broadleaf with an 95% accuracy for an overall accuracy of 97% (when considering grass, broadleaf and soil classes). The classification accuracy of this method for identifying grass and broadleaf are slightly below Tang et al. (1999), who achieved 100% classification. However, the ability to identify individual species may provide valuable insights for farmers, weed scientists, and agronomists who desire to map specific weed concentrations within a particular field. Additionally, Model 2 is unique, because it eliminates computational requirements for the intensity CCM matrix. The elimination of the intensity HSI color component may provide color texture statistic stability under varying ambient light conditions, since intensity is the measure of image brightness. If hue and saturation statistics are stable under varying ambient intensity, the effect of cloud cover variations can be eliminated. The CCM procedure has shown potential robustness under variations in plant matter content with in the sub-image, which will be beneficial under field conditions.

It is recommended that additional research be conducted to ensure that these algorithms are stable under varying ambient light conditions. Likewise, sensitivity analysis should be conducted on the effects of change in sub-image size and percentage of plant matter in the sub-image. Additionally, sensitivity analysis should be conducted to determine if the transformation from 8 bit to 5 bit RGB prior to HSI conversion significantly affects the classification results.

REFERENCES

Bridges, D. C., and R. L. Anderson. 1992. *Crop Loses Due to Weeds in the United States*, 1-25.Champaign Ill.: Weed Science Society of America.

Burks, T. F. 1997. Color image texture analysis and neural network classification of weed species. Ph.D. thesis. Lexington, Ky.: University of Kentucky, Biological and Agricultural Engineering.

Chaisattapagon, C., and N. Q. Zhang. 1991. Weed detection using machine vision. ASAE Paper No. 91-3508. St. Joseph, Mich.: ASAE.

Franz, E., M. R. Gebhardt, and K. B. Unklesbay. 1991. The use of local spectral properties of leaves as an aid for identifying weed seedling in digital images. *Transactions of the ASAE* 34(20): 682-687.

_____. 1990.Shape description of completely visible and partially occluded leaves for identifying plants in digital images. ASAE Paper No. 90-7040. St. Joseph, Mich.: ASAE.

Haralick R. M., and K. Shanmugam. 1973. Computer Classification of Reservoir Sandstone. *IEEE Trans. Geosci. Electr.* Ge-11(14): 171-177.

Haralick, R. M., and K. Shanmugam. 1974. Combined Spectral and Spatial Processing of ERTS Imagery Data. *J. Remote Sens. Environ.* 3: 3-13.

Julesz, B. 1962. Visual pattern discrimination. *IRE Trans. Info. Theory* 8(2): 84-92.

Levine, M. D. 1985. *Vision in Man and Machine*. New York, N.Y.: McGraw-Hill Book Co.

Meyer, G. E., T. Mehta, M. F. Kocher, D. A. Mortensen, and A. Samal. 1998. Textural imaging and discriminant analysis for distinguishing weeds for spot spraying. *Transactions of the ASAE* 41(4): 1189-1197.

Ohta, Y. 1985. *Knowledge-based Interpretation of Outdoor Natural Color Scenes.* Marshfield, Mass.: Pitman Publishing Inc.

Rao, C. R. 1973. *Linear Statistical Inference and Its Application*. New York, N.Y.: John Wiley & Sons.

SAS. 1985. *SAS User's Guide;Statistics*, 5th Ed. Cary, N.C.: SAS Institute, Inc.

Shearer, S. A. 1986. Plant identification using color co-occurrence matrices derived from digitized images. Ph.D. thesis. Columbus, Ohio: Ohio State University, Agricultural Engineering.

Shearer, S. A., and R. G. Holmes. 1990. Plant identification using color co-occurrence matrices. *Transactions of the ASAE* 33(6): 2037-2044.

Tang, L., L. F. Tian, B. L. Steward, and J. F. Reid. 1999. Texture-based weed classification using gabor wavelets and neural network for real-time selective herbicide applications. ASAE Paper No. 99-3036. St. Joseph, Mich.: ASAE.

Thompson, J. F., J. V. Stafford, and P. C. H. Miller. 1991. Potential for automatic weed detection and selective herbicide application. *Crop Prod.* 10(4): 254-259.

Wang, N., N. Zhang, and Y. Sun. 1999. Development of a spectral-based weed sensor. ASAE Paper No. 99-3037. St. Joseph, Mich.: ASAE.

Woebbecke, D. M., G. E. Meyer, K. Von Bargen, and D. A. Mortensen. 1995a. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE* 38(1): 259-270.

_____. 1995b. Shape features for identifying young weeds using image analysis. *Transactions of the ASAE* 38(1): 259-270.

Wyszecki, G., and W. S. Stiles. 1992. *Color Science; Concepts and Methods, Quantitative Data and Formulae*, 2nd Ed.: 117-137. New York, N.Y.: John Wiley & Sons.

Zhang, N., and C. Chaisattapagon. 1995. Effective criteria for weed identification in wheat fields using machine vision. *Transactions of the ASAE* 38(3): 965-974.