EVOLUTION OF VIDEO COLOR/TRASH PROCESSING AT SOUTHWESTERN COTTON GINNING RESEARCH LABORATORY

Michael A. Lieberman
USDA-ARS, SWCGRL
Mesilla Park, NM
Murali Siddaiah
New Mexico State University
Las Cruces, NM

Abstract

The status of video cotton trash and cotton color measurement work at Southwestern Cotton Ginning Research Laboratory (SWCGRL) is presented. New accomplishments are in the areas of video measurement of absolute color in CIE XYZ and Hunter's $R_d ab$ color spaces and improvements of bark trash type identification.

Introduction

The need for this research is best seen in the context of how the cotton industry is measuring cotton color and trash content for marketing and process control. United States cotton is High Volume Instrument (HVI) graded before entering the cotton marketing system. But this grading happens after the fact. The cotton has already been ginned, baled, and is ready to enter the cotton market. Any chance for the ginner or farmer to effect lint quality has passed. A gin based color/trashmeter able to indicate possible USDA-Agricultural Marketing Service (AMS) HVI grade is expected to improve the ability of any gin to optimize control and to communicate seed cotton problems to the farmer. This system is not meant to replace the current AMS HVI color and trash measurement systems. The AMS system should be maintained. A method for relative measurements within a gin is being suggested. In addition, with a system that can identify bark, process control could decrease the bark component of the final product.

Zellweger Uster[†] sells a process control system (PCS) that currently uses color/trashmeter sub-systems derived from their HVI systems, (Anthony-90). Global sample color, total sample trash, and trash count are used as inputs to the PCS. In addition to these features, the proposed system measures lint-only color, identifies trash type, and can accumulate information about each type of trash. A single instrument is used to measure both color and trash.

But why the need to identify trash types? This instrument would give more information to the PCS. Knowing relative and total bark amount could permit the ginner to better choose the equipment mix for improving final product value. For instance, removing relatively more bark might decrease bark discounts.

If this system's "barky" calls correlate with AMS', then AMS might evaluate the system for classing office use. Textile mills could use the system to better categorize cotton going into a laydown.

Approach

The approach is divided into three parts, general, color measurement, and trash identification.

General

The measurement techniques should be transportable and reproducible, so problems like the differences between the early MCI® and Spinlab® systems do not re-occur. For this work, an MCI stand-alone color/trashmeter was modified to use (1) a Sony® m/n 930, 3-chip color camera with ½ CCDs, (2) quad fiber optic spotlights with defocusing lenses placed at 45 ° left and right, and (3) an Illumination Technologies® 3900 fiber optic light source. Two different Matrox® IM-1280 Framestores, IM-CLD frame grabbers, and computers were used as two systems (\\486 and \\brg). Acquired images were 640 by 480 pixels although only the central 512 by 480 pixels were used to make our predictions. The images have three planes, red, green, and blue (RGB).

Color

Since a high resolution 3-CCD color camera has been used for our trash identification work (Lieberman 95), the system could be designed to measure total sample and lint-only color in CIE LAB (L*a*b*) (CIE 86, Hunter 87) or Hunter's $R_d ab$ (Hunter 87) color spaces at little extra cost. There would then be only one instrument to maintain in a gin environment rather than two.

A system design parameter was instrument stability over more than 24 hours without recalibration. Thomasson (92) suggests a calibration interval of four hours with the current trashmeters. The stability of the system was evaluated by tracking mean image levels and level standard deviations every six-minutes over 2½ days. After the stability had been characterized, the absolute color test was begun.

Absolute Color Measurement

The first step in developing an absolute color measurement system was to obtain color-stable samples. These were measured by a method traceable to a national standard. These color-stable samples were then used to configure CIE XYZ and $R_d ab$ measuring systems, (Pratt 91; Hunter 87). Evaluation consisted of comparison between the measured R_d, *a*, and *b* values and the traceable values. Repeatability was evaluated by comparing multiple measurements on one imaging system to each other and then measurements on one system to the other system. Accuracy was evaluated by comparing measurements to the traceable values.

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:657-663 (1999)
National Cotton Council, Memphis TN

The color conversion test used 28 ceramic and 2 porcelain tiles. These were sent to Hunter's Laboratory and measured on instruments traceable to the NIST. (NIST performs the functions of the former National Bureau of Standards.) CIE XYZ and $R_d ab$ measurements from both a spectrophotometer and colorimeter were acquired, (Hunter Labs 96). On some of the tiles, we also have AMS R_d and $+b$ values.

This test consisted of 8-measurement repetitions on 30 tiles. Four measurements were made with each of the two systems presented earlier (\\486 and \\brg). The tile was rotated 90° between acquisitions. Four-repetitions were first acquired on the imaging computer \\486. The next day, four repetitions were similarly acquired using imaging computer \\brg.

Trash

There are two components to the trash work, identification of the types of trash and measuring percent trash. For trash identification, eight training images (two each of pepper, leaf, bark, and bark2) were used. Bark and bark2 are both "bark", but they have different ranges of the measured features (described below). One might call bark: hairy-bark and bark2: solid bark. Operators identified the trash in each test image. A back propagation neural network (Lieberman 94) was used to train the system. Five actual cotton samples were used for testing.

Results and Evaluation

In any lighting system there are two obstacles to good measurements: spatial and temporal non-uniformity. Spatial non-uniformity refers to cases where lighting is brighter in one region than another. This has been addressed, in previous years, at this conference (Lieberman 95). Briefly, a reference image was acquired and used to define the spatial non-uniformity. The reference image was normalized; i.e., all pixels in each color plane were divided by the plane's mean value. Every new image was then divided by this normalized shape image. The improvement was demonstrated by evaluating the standard deviation in levels for the image. If lighting was uniform, the reference tile was uniform over its whole surface, all camera pixels were identical, and the optics caused no distortion then the standard deviation of level in each image plane should be zero. Ours is not a perfect world, hence, the need for the correction. The plots in Figure 1 and Figure 2 must be discussed before evaluating the improvement in standard deviation, (Figure 3). Many identification tasks using imaging rely on relative gray level or color level information, keyword *relative*. Making absolute color measurements, traceable to some standard, is a very different ball game.

Temporal non-uniformity is a problem and is where lighting level changes over time in either total amount or relative amount among the colors. Figure 1 through Figure 5 show measurements on an 80%-white reference tile. A tile image was acquired and processed every six minutes over 2½ days. The vertical line located near midnight Sunday shows a flawed image and its subsequent bad data. This was caused by the loss of one of two fields in one image frame. This condition is easily recognized and the system would have the image retaken. All of these figures show six traces versus time. These traces consist of two sets of three image planes of information. For example, The raw red, green, and blue (RGB) levels from the camera are shown in Figure 1. These RGB traces start at 160, 174, and 170, respectively. The corrected RGB signals start at 156, 158, and 160, respectively. Since the variation in raw RGB is so much larger than the corrected RGB, just the corrected RGB and target RGB are shown in Figure 2. Target values are the measurements produced by Hunter Laboratory or values derived from those numbers. The trend of the corrected curves is much flatter during the first 24-hours than the latter part of the time. This implies a limit of 24-hours between calibrations, but this should be looked at again, in the future.

Evaluating the effect of the correction is shown in Figure 3. In this figure, the standard deviation of the quarter million pixels in each image plane is plotted over time. The two sets of traces are raw-image standard deviations using the left axis, and corrected-image standard deviations using the right axis. Raw-image standard deviations for red, green, and blue are about 22.8, 24.2, and 25, respectively. The corrected values are 3.2, 3.1, and 4.5, respectively, an improvement of about a factor of 5.

The AMS R_d, $+b$ color space used for cotton color is meant to be identical to Hunter's $R_d ab$. Hunter's $R_d ab$ color space is defined in terms of the CIE XYZ color space (Hunter 87, CIE 86). The CIE XYZ color space can be defined in terms of various RGB color spaces. An XYZ value was computed for each pixel in an RGB image. The mean CIE XYZ image levels are shown in Figure 4. It can be seen that the Z signal has the most variation. This will be a factor in the $R_d ab$ plot. Besides the measured values, the Hunter values of X, Y, and Z are shown.

$R_d ab$ pixel values were computed from the XYZ pixels. Figure 5 shows mean level information in the Hunter's $R_d ab$ color space. The cotton classing system currently uses only the R_d and $+b$ components. Besides the measured values, the Hunter R_d, a, and b values are shown; R_d uses the left axis, a and b uses the right.

Over the first 32 hours of the test, the a and b components varied less than 0.25 units. However, the b signal shows a bias, which is being evaluated. AMS has color tolerances on their colorimeters of 1.0 and 0.5 respectively. R_d changes less than 0.2 peak to peak over 2½ days (Figure 5). The $+b$ change is less than 0.25 peak to peak most of the time and less than 0.35 for the whole period.

Accuracy/Repeatability

At an initial level we can handle for temporal and spatial lighting variations. Next, the system's ability to measure 30 tiles in Hunter's $R_d ab$ color space within AMS tolerances was evaluated. Figure 6 shows the results of eight repetitions plotted around an AMS color chart. The most obvious aspects of the plot are points that have "missed" the AMS chart. That was done on purpose. Points outside of the nominal AMS region were included to improve the accuracy of our equations.

Each apparent single triangle actually consists of eight triangles, one per repetition. PROC CORR (SAS CORR) was used to acquire the correlations among Hunter' values of $R_d ab$, values for each repetition (8-each), the means of the four-repetitions from each image processing system and the mean of all 8 repetitions. All Pearson correlation coefficients were essentially the same. All correlations among the eight repetitions were above 0.999 with most above 0.9999. But, AMS does not use single repetitions for their measurements; they use the mean of four. The correlation among the Hunter values and the mean-of-four-runs from each of two different systems is shown below. Tables for R_d, a, b are located in Table I through Table III Points to notice are, the two systems correlate very well with one another (0.9993, 0.9994, and 0.9991) for R_d, a, and b respectively. The correlations between the target and the mean test sets are 0.99923 ±0.00020, 0.99608 ± 0.00032, 0.99400 ± .00065 for R_d, a, and b, respectively. The lower correlation for b than for R_d agrees with the fact that more of the error in Figure 6 is in the b direction.

There is a correlation among the different measures for $R_d ab$. Now we see if there is a predictive capability. Table IV and Table V show the coefficient of determination (r^2) for numerous cases were our data was compared to the target Hunter Laboratory values. All of the measurements presented until the end of the color section have a confidence level of 99.99%. The first data line of Table IV reports how well the Hunter $R_d ab$ values are predicted from using the 8-repetitions of 30 tiles as one data set. The next two lines report how well the mean of 4-repetitions for each system (called \\486 and \\brg) performed. Of course, one performed better than the group of 240 and one worst, but what is beneficial is that all three were essentially the same in their ability to predict Hunter's $R_d ab$. The fact that the mean of 8-repetitions and the set of 240 individual measurements is not surprising. It was mentioned in the discussion of Figure 6 that all of the eight repetitions were essentially equal. The last line of the table, where the mean of one system is compared with the mean of the other, confirms this consistency. Table V contains similar information for tiles that are located within the AMS color diagram. There was improvement in the regression coefficients for R_d and b, but a performance was not as good.

This says there is no simple correction and our accuracy still needs more work. There appears to be a trend that the error increases with $+b$.

Trash Identification

There are numerous methods to identify regions of a cotton image as trash (segment the image). Image levels were adjusted to imitate spatially uniform lighting and responses. Flat-field correction (Lieberman 95) was used to compensate for spatial effects caused by lighting, cell sensitivity and other problems. Correcting for spatial non-uniformity enables the use of thresholding, abet with different levels for segmenting each image. Changes of illumination level over time are less of a problem as our methods find a threshold unique for each image. Trash objects were segmented from lint background by thresholding. The results of this are shown in the leaf-training sample (Figure 8). Once objects are identified, simple features are acquired from the imaging hardware. After previous experimentation, area, shape factor (4p * area / perimeter^2), solidity, and extent (Russ 94) were chosen Figure 7. Solidity is the ratio of actual area to convex area. Convex is the technical term; rubber-band area seems more descriptive; the convex perimeter is the shape a rubber band takes around the trash object. Convex area is within that perimeter. Convex area versus actual area for a number of objets is shown in Figure 7. Convex area is constant, within digitization errors, when an object is rotated in an image. Since both area and convex area are constant, solidity is rotationally invariant. Extent is the ratio of actual area to bounding box area. If an object rotates in the image, the size of the bounding box changes drastically for objects like bark and bark2, but not for leaf. The minimum and maximum extent provides information that is useful for classification. An example of how these measures might be used for classification can be seen in Table VI. Solidity for bark is usually less then 0.5. Once bark is removed, extent can then be used to separate leaf from bark2.

One further comment, Figure 7 shows that the silhouette of a short bark2 would look like a leaf silhouette. Also, the silhouette of oblong leaf would look like a bark2 silhouette. This observation will be used later.

Training

A back propagation NN was used to train the system on sample images (Lieberman 94; Lieberman 97). The NN generated weights were incorporated into the imaging system. The trash type of each trash object is computed as a function of the relevant weights and features. There were two sample images for each trash type, pepper, leaf, bark, and bark2. Again referring to Figure 8 showing a raw cotton image and segmented cotton image, each black region of the segmented image represents one trash object. Figure 9 shows a segmented and numbered image. These numbers were software generated. The operator used numbers while identifying trash type. Note that not all trash objects are numbered. There are three reasons for not numbering an object common to all training images. 1) Objects with an area less than 10-pixels were considered noise and removed from the database. (Optics and camera timing were adjusted so each pixel was 0.127 mm (0.005 in.), 10 pixels have an area of 0.16 mm² (250μ in²).) Object "northwest" of object 11 is an example of this. 2) Object 42 is large enough, but it is made from two joined objects. 3) For training, only objects of one type were retained in a given image's database. We consider pepper and leaf to differ only by size. A trigger point of 200 pixels was used to separate these for training; 200 pixels have an area of 3.23 mm² (0.005 in²). In the pepper images, objects greater than the trigger point were removed from the database. In the leaf images, objects less then the trigger point were removed from the database.

Testing

The system was trained and self-tested to identify features to be used. Once the best set of features was found, the system was tested on five actual cotton samples. Figure 10 shows test sample 5. Figure 11 shows the same sample with the objects numbered and with unnumbered objects deleted. Objects deleted from the database were not numbered. The classification results of 245 trash pieces contained in five test images is shown in Table VII. There were 223 classified properly (Table VIII). The 22 mis-classified objects will be evaluated. These are the Type I errors where a known object is mis-classified (Table IX). There was a trivial amount of pepper mis-classified. Leaf on the other hand has 50% mis-classification into two categories: pepper and bark2.

In previous years, at this conference, it had been shown that hierarchical clustering could help. If an area threshold is used to remove most pepper before NN classification (Lieberman 97), leaf called pepper error will be minimized. In the best system, it is expected the pepper called leaf and leaf called pepper errors to be equal. Leaf was called bark2 (19%) also bark2 was called leaf (50%); more work is needed here. As pointed out previously, shape related parameters might not be sufficient for leaf and bark2 separation or bark2 called pepper error. If a threshold reduces the leaf called pepper error, then this error should decrease also.

Figure 10 /Figure 11 results can be evaluated to show some of these errors (Table VII). Objects 87, 88, 91, and 94 appear as pieces of leaf or pepper in Figure 11. However, they are all part of a piece of bark (Figure 10). Object 45 is called stick, but the human observer believes it is leaf.

Conclusion

The technique to measure color works as an indicator of color not as an accurate absolute measurement. The method for trash identification works well for bark and pepper. The overlap between leaf and pepper is understood and should be corrected. The overlap of bark2 with leaf and pepper requires further evaluation. Using color might be a tool to help differentiate these classes

Future Research

Color

These measurements related to the Hunter Laboratory colorimeter $R_d ab$ color space. We plan to see how well we can predict AMS color. Currently, we need to get some calibration tiles measured in AMS color space. This preliminary test should be repeated using two different imaging systems (different lighting geometry, different camera types). The next step would then be to apply the $R_d ab$ measuring system to cotton samples to produce measurements equivalent to current AMS R_d and $+b$. A parallel step is to get the lint-only color. The color and range of color would be evaluated as factors in identifying "difficult to identify" trash. Some researchers have found lint-only color does not give statistically better color information than sample color, (Thomasson, 93). That research was done with a different configuration and will be repeated using our technique.

Trash

Now that trash can be identified, a method to measure AMS percent trash should be developed. The prior probability of pepper versus leaf, bark, and bark2 should be more representative of our test samples. Our sample had 211 pepper and 32 all other types. Eventually, AMS percent trash should be computed as the ratio of area of all trash objects to the image area. The minimum area threshold and/or trash type might be a factor to enable AMS system results to be duplicated. Bark discount might be computed by relating bark count and area to all trash area for each color region. The classes of trash that we can identify could be expanded to include such things such as grass, hairy seed coat fragments (scf) (outer surface of scf), smooth scf (the inner side of scf), pea vine, grease, man-made material, and other trash items of interest. Detection of light spot could be evaluated. Since bark can be identified, a method to predict bark discount should be developed. The use of color to help differentiate among bark2, leaf and pepper should be evaluated.

Footnotes

† Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products that may be suitable.

Acknowledgments

Thanks to student Brian Baker and the technical support group at the laboratory.

References

Anthony, W. S., 1990. Computerized gin process control. Applied Engineering in Agriculture, 12-18.

Commission Internationale de L'Éclairage (CIE), 1986. CIE Publication No. 15.2 (1986), Colorimetry, 2nd edition. Central Bureau of the CIE.

Hunter R. S., R. W. Harold, 1987. The measurement of appearance, 2nd edition, chapters 7, 8. John Wiley & Sons.

Hunter Associates Laboratory, Inc, 1996. Report on contracted tile measurements

Lieberman, M. A., R. B. Patil, 1994. Clustering and neural networks to categorize cotton trash. Optical Engineering. 1642-1653.

Lieberman, M. A. , S. N. Brennan, M. B. Zaidi, 1995. Update on video cotton trash measurements at Southwestern Cotton Ginning Research Laboratory, Proceedings of the Beltwide Cotton Conferences. 1260-1262

Lieberman, M. A., R. B. Patil, 1997. Evaluation of learning vector quantization to classify cotton trash. Optical Engineering. 914-921.

Pratt, W. K., 1991. Digital Image Processing, 2nd edition. John Wiley & Sons.

Russ, J. C., 1994. The image processing handbook, 2nd edition. CRC Press.

SAS Institute, Inc, 1987 SAS/Stat Guide, Proc CORR.

SAS Institute, Inc, 1994. SAS Procedures Guide, Proc REG.

Thomasson, J. A. 1993 Foreign matter effects on cotton color measurement: Determination and correction. Transactions of the ASAE, 663-669.

Thomasson, J. A. 1992. Recalibration interval for color/trashmeters in cotton gins. Applied Engineering in Agriculture, 8(2):147-157.

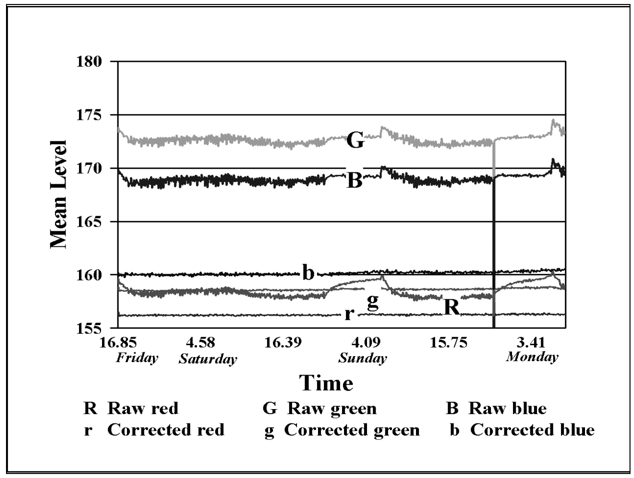

Figure 1. Mean raw and corrected levels.

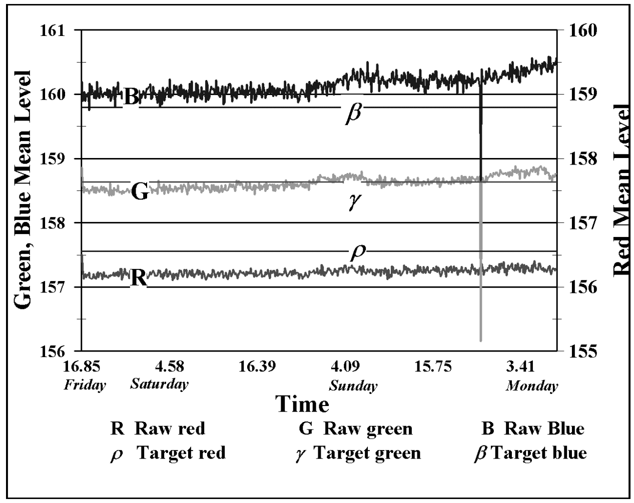

Figure 2. Corrected and target mean levels.

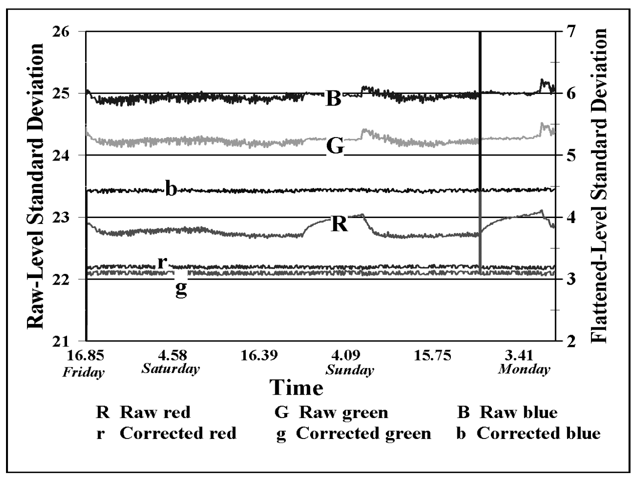

Figure 3. Standard deviation of RGB image planes.

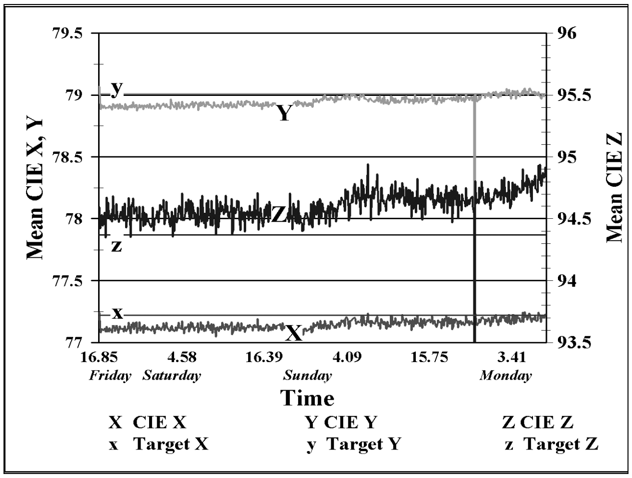

Figure 4. Mean CIE XYZ levels.

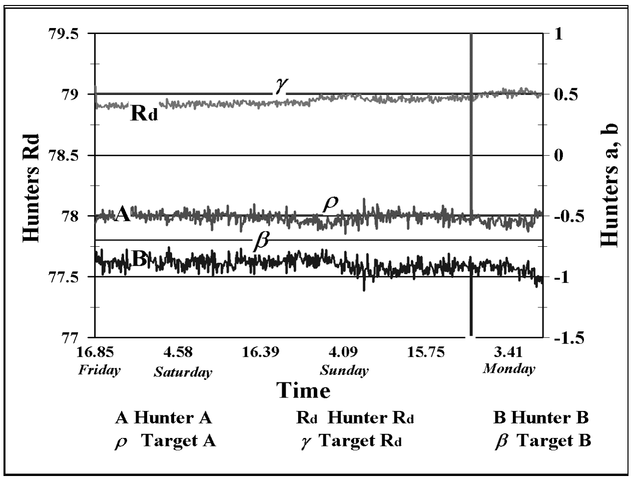

Figure 5. Hunter's $R_d ab$ mean levels.

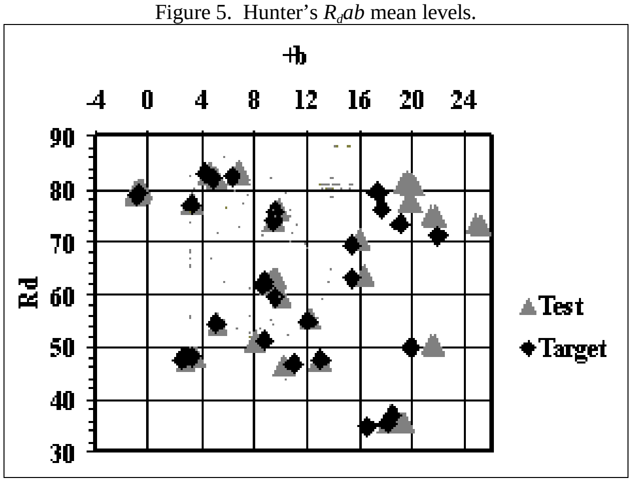

Figure 6. Evaluation of tile $R_d ab$.

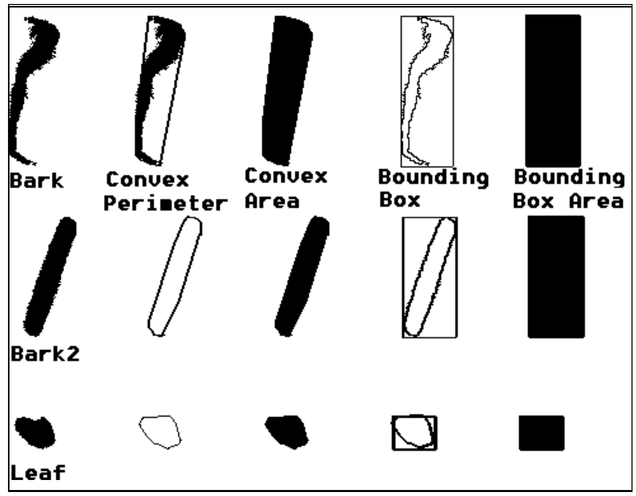

Figure 7. Examples of solidity and extent.

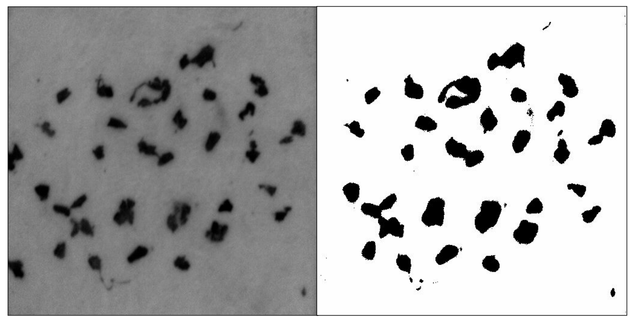

Figure 8. Leaf training sample, raw and segmented images.

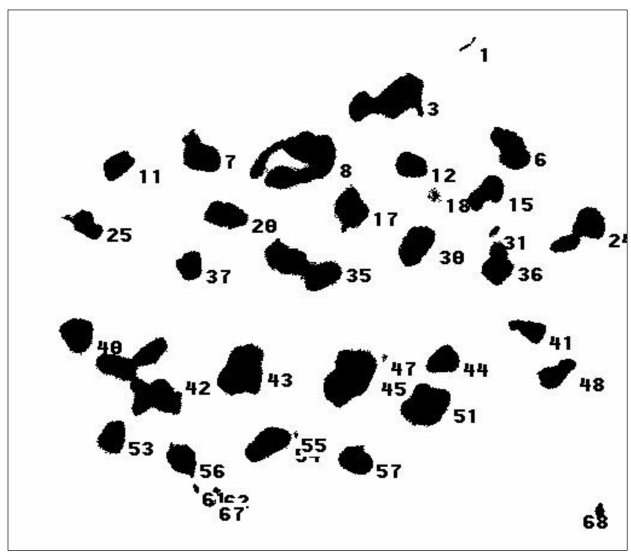

Figure 9. Leaf image segmented/numbered.

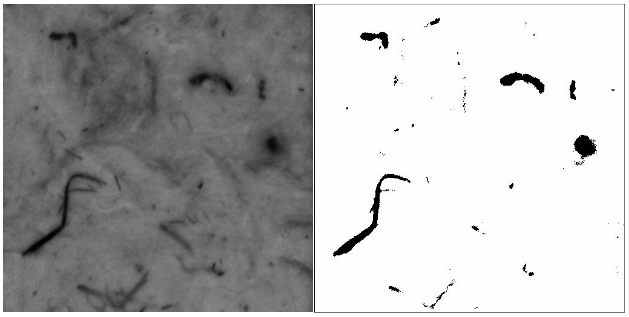

Figure 10. Test sample 5, raw and segmented images.

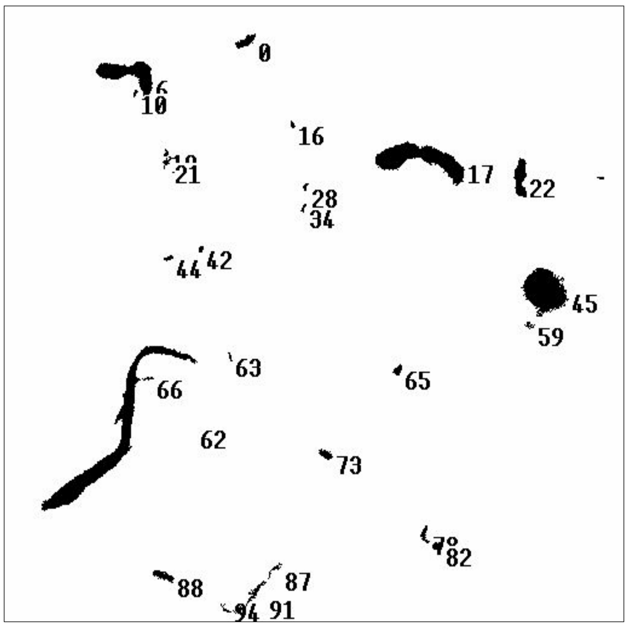

Figure 11. Test sample 5, trash numbered.

Table I. Hunter's R_d correlation results.

Source	Hunter R_d	System \\486 R_d	System \\brg R_d
Hunter R_d	1.00000	0.99938	0.99903
	0.0	0.0001	0.0001
System \\486 R_d	0.99938	1.00000	0.99993
	0. 0001	0.0	0.0001
System \\brg R_d	0.99903	0.99993	1.00000
	0. 0001	0.0001	0.0

Table II. Hunter's *a* correlation results.

Source	Hunter *a*	System \\486 *a*	System \\brg *a*
Hunter *a*	1.00000	0.97574	0.97641
	0.0	0.0001	0.0001
System \\486 *a*	0.97574	1.00000	0.99994
	0.0001	0.0	0.0001
System \\brg *a*	0.97641	0.99994	1.00000
	0. 0001	0.0001	0.0

Table III. Hunter's *b* correlation results.

Source	Hunter *b*	System 1 *b*	System 2 *b*
Hunter *b*	1.00000	0.99334	0.99465
	0.0	0.0001	0.0001
System \\486 *b*	0.99334	1.00000	0.99991
	0.0001	0.0	0.0001
System \\brg *b*	0.99465	0.99991	1.00000
	0.0001	0.0001	0.0

Table IV. Coefficients of determination for the 30 tile regressions.

Hunters $R_d ab$ versus	R_d	a	b
240 points (30 tiles x 8 repetitions)	0.9999	0.9402	0.9964
Mean of 4 \\486 repetitions	0.9999	0.9391	0.9959
Mean of 4 \\brg repetitions	0.9999	0.9414	0.9969
Mean 8 repetitions	0.9999	0.9403	0.9964
Predicting each other			
Mean \\486 versus mean \\brg	1.000	.9999	0.9999

Table V. Coefficients of determination for the 13 tile regressions.

Hunters $R_d ab$ versus	Rd	a	b
104 points (13 tiles x 8 repetitions)	1.0000	0.9147	0.9981
Mean of 4 \\486 repetitions	1.0000	0.9171	0.9979
Mean of 4 \\brg repetitions	1.0000	0.9125	0.9983
Mean 8 repetitions	1.0000	0.9149	0.9981
Predicting each other			
Mean \\486 versus mean \\brg	1.0000	0.9998	1.0000

Table VI. Examples of solidity and extent.

Trash type	Area	Convex area	Bounding box area	Solidity	Extent
bark	1526	3242	4879	0.4707	0.3128
bark2	1578	1734	3990	0.9100	0.3955
leaf	581	651	858	0.8925	0.6772

Table VII. Trash identification results.

		Classified			
		Pepper	Leaf	Bark	Bark2
Total	Known	210	17	10	8
211	Pepper	204	5	0	2
16	Leaf	5	7	1	3
10	Bark	0	1	9	0
8	Bark2	1	4	0	3

Table VIII. Identification: percent accuracy.

	Pepper	Leaf	Bark	Bark2
Known trash identified correctly	96.68	43.75	90.00	37.50
Identified trash that is correct	97.14	41.18	90.00	37.50

Table IX. Trash identification Type I errors.

		Classified			
Total	Known	Pepper	Leaf	Bark	Bark2
211	Pepper	-	2.37	0	0.95
16	Leaf	31.25	-	6.25	18.75
10	Bark	0	10	-	0
8	Bark2	12.5	50	0	-

Table X. Classification results

Pattern	Blob_ID	Actual Trash Type	Classified 0 Degrees	Classified 0, 30, 60 Degrees
1	0	Pepper	Pepper	Pepper
2	6	Bark	Leaf	Leaf
3	10	Pepper	Pepper	Pepper
4	16	Pepper	Pepper	Pepper
5	17	Bark	Stick	Bark
6	19	Pepper	Pepper	Pepper
7	21	Pepper	Pepper	Pepper
8	22	Stick	Leaf	Leaf
9	28	Pepper	Pepper	Pepper
10	34	Pepper	Pepper	Pepper
11	42	Pepper	Pepper	Pepper
12	44	Pepper	Pepper	Pepper
13	45	Leaf	Stick	Stick
14	59	Pepper	Leaf	Pepper
15	62	Bark	Bark	Bark
16	63	Pepper	Pepper	Pepper
17	65	Pepper	Pepper	Pepper
18	66	Pepper	Pepper	Pepper
19	73	Pepper	Pepper	Pepper
20	78	Pepper	Pepper	Pepper
21	82	Pepper	Pepper	Pepper
22	87	Stick(Broken)	Leaf	Stick
23	88	Stick(Broken)	Pepper	Pepper
24	91	Stick(Broken)	Pepper	Stick
25	94	Stick(Broken)	Leaf	Pepper