IDENTIFICATION OF TRASH TYPES AND CORRELATION BETWEEN AMS AND SWCGRL TRASH CONTENT IN GINNED COTTON

M. Siddaiah, M. A. Lieberman and S. E. Hughs
Southwestern Cotton Ginning Research Laboratory, USDA-ARS
Mesilla Park, NM
N. R. Prasad
New Mexico State University
Las Cruces, NM

Abstract

This paper discusses the identification of various trash types in cotton (non-lint material/foreign matter) using soft computing techniques, such as, Fuzzy Logic and Neural Network based approaches. Trash identification provides the basis for computing the trash content in ginned cotton. The effectiveness of a hybrid neuro-fuzzy structure, namely the Adaptive Network-Based Fuzzy Inference System, to classify trash types is compared with other techniques. A correlation between trash content computed by Agricultural Marketing Service and those computed by the Southwestern Cotton Ginning Research Laboratory is presented.

Introduction

The Southwestern Cotton Ginning Research Laboratory (SWCGRL) is part of the United States Department of Agriculture, Agricultural Research Service (USDA-ARS). SWCGRL is involved in various aspects of cotton ginning research. The industry needs a capability to classify cotton based on trash content in real time, i.e., at video rates, and to develop a system capable of defining trash types for possible use by textile processors and in grading cotton. SWCGRL is working in conjunction with USDA-Agricultural Marketing Service (AMS), and other ARS laboratories to develop such a system.

Traditionally, cotton grade has been based on four physical properties: color, trash, preparation, and extraneous material (USDA-AMS, 1999). The reflectance (R_d) and the degree of yellowness ($+b$) describe cotton color. Trash is a measure of the amount of non-lint material, such as leaf and bark, in cotton. The term preparation describes the smoothness of the sample, i.e., lack of lumps and twists. Cotton in its raw form contains lint and non-lint material. Lint is the cotton fiber, non-lint or foreign matter is essentially everything other than lint. The term ginning refers to the process of separating lint from cottonseeds. A commercial gin also cleans non-lint material from the fibers. This non-lint material can consist of bark, leaf, pepper, hulls, seed coat fragments, and motes left behind in the sample. Pepper refers to broken or crushed pieces of leaf, hulls are the outer coverings of the cotton boll, and motes are immature cottonseeds. Current techniques to identify these trash objects include classical statistics, Linear Vector Quantization (LVQ) (Lieberman, 1997), clustering algorithms, Neural Networks (NN's) (Lieberman, 1994) and Adaptive Network-Based fuzzy Inference System (ANFIS) (Siddaiah, 1999a).

Currently 15 to 19 million bales of cotton are produced in the United States each year; and 98% of these are classed by the USDA-AMS. Since each bale of cotton produced in the United States is classed in a classing office, an on-line automated system capable of identifying both trash types as well as assigning a class grade is of great importance to AMS. With the addition of trash identification, the criteria for grading ginned cotton can be improved. The current task includes identifying trash and categorizing it as bark1, bark2, leaf, or pepper. In our research, we differentiate bark in terms of bark1 and bark2. Stringy pieces of bark objects are referred to as bark1. Pieces of bark with no filaments are referred to as bark2. The objectives are (1) identify a segmentation technique to separate non-lint material from lint material; (2) identify features to recognize trash in cotton samples; (3) evaluate the performance of a recognition system to accurately identify the stated trash types; and (4) compute the trash content in ginned cotton.

This paper presents a methodology for identifying various types of trash objects using soft computing techniques and computation of trash content (by percent area) in cotton samples. Using area, solidity, and extent as features of the various trash types, it is possible to identify objects in cotton samples as bark1, bark2, leaf, and pepper. These features provide the necessary inputs to train the various classifier systems, examined in this paper, i.e., back propagation neural network, fuzzy clustering algorithm (fuzzy C-means), and the Adaptive Network-Based Fuzzy Inference System (ANFIS).

Approach to Classification of Trash Types

The problem of pattern classification is generally considered to be a high-end task for computer based image analysis. The complexities range from locating and recognizing isolated objects of known types, to recognizing objects in poorly defined classes, to much more open-ended problems of recognizing possible overlapping objects or classes and properly handling shadows, empty regions and poorly prepared samples.

Sample Domain Description

Problem domain specifications have a major impact when designing a classification system. The population of all lint and non-lint material present in ginned cotton defines the

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 2:1549-1555 (2000)
National Cotton Council, Memphis TN

problem domain. For trash identification, this domain is reduced to a sample domain of prepared samples of trash placed on bleached cotton. Identification is limited to four common types of trash, namely, bark1, bark2, leaf, and pepper. These categories are not distinct, as they do not have well defined separating boundaries. For example, leaf and pepper types are actually a continuum with an arbitrary separation boundary. Test samples were cotton samples extracted randomly from bales. For trash content measurements, the sample domain was 18 AMS trash level check boxes.

Preprocessing

Images (640 x 480 pixels) were acquired using a 3-chip CCD Sony® color camera. The imaging hardware consists of a Matrox® IM-1280 imaging board and IM-CLD acquisition board. Image frame-grabber acquisition timing was adjusted so both horizontal and vertical pixel resolutions are 0.005 in. The acquired images are flat field corrected to remove spatial illumination nonuniformity. Trash objects were separated from the cotton lint background (segmented) using a simple threshold on the intensity plane in HLS color space to obtain a binary image where each trash object is identified. A unique, automatically chosen threshold was used for each image. A set of 3 features was measured with Visilog® image analysis software by NorPix, Inc. (formerly Noesis Vision) for each object in the segmented image.

Feature Hyperspace

For each object, a set of features is computed. These features form the basis for a component feature vector that defines a point in multi-dimensional feature space. The classifier partitions this multi-dimensional hyperspace into specific regions, where each region corresponds to a type of non-lint material. Each classification system forms a unique hypersurface separating the feature space into object regions. The image analysis software provide the following features:

i. Area A: Area of object in a binary image, the number of nonzero pixels in an object.
ii. Ferrets X_f and Y_f: Projection of an object along image edges.

From these basic features the following additional features were computed (Russ, 1994).

i. Convex Area: Area of a 32-sided irregular polygon bounding the object.
ii. Bounding box area: Product of the ferrets X_f and Y_f when object has been rotated with long axis horizontal or 45 degrees.
iii. Solidity: Ratio of Area to Convex Area.
iv. Extent: Ratio of Area to Bounding Box Area.
v. E^{dif}: Difference in extent at 0 degrees and 45 degrees.

Figure. 1 shows examples of the trash objects with the feature measurement values shown in Table 1. All trash objects with an area less than 10 pixels (2.5×10^{-4} in^2) were considered as noise and ignored. Pepper is the only trash type that can be distinguished from the others using one feature, namely "area". All objects with area less than or equal to 200 pixels (5×10^{-3} in^2) can be considered pepper. The convex area for bark1 objects is significantly larger than the actual area of the object. Solidity for bark1 objects are typically less than 0.5 in comparison to leaf and bark2 objects, which have solidity values closer to 1 as the convex area for leaf and bark2 objects are closer to their actual area. Area and solidity are independent of rotation in the image plane for the trash types. However, the extent for bark2 objects varies with the orientation of the objects. For a bark2 object, when the object edge is oriented horizontal to the image axis, the bounding box area is very close to the actual area of the object (Lieberman, 1999). This results in a value for extent that is closer to unity. However, when these objects are oriented at an angle of 45 degrees, the bounding box area is maximum resulting in values for extent that are lower. The difference between the extent at 0 and 45 degrees orientation (E^{dif}) provides a distinguishing characteristic between leaf and bark2 trash types (Siddaiah, 1999a).

Classification Methods

In this section, various classification techniques used to identify trash types in ginned cotton are presented. Different trash types were placed on bleached cotton and used to collect the training data. Images (640 x 480 pixels) were acquired and flat field corrected to improve spatial uniformity. The flat field corrected images are 512 x 480 pixels. Each acquired color image (RGB) is converted into the hue, luma (intensity), and saturation (HLS) color space. The intensity plane is used to obtain the threshold for separating non-lint material from lint. Automatic threshold values are determined using an entropy measure on the intensity plane of the image. The thresholded images are used to mask all lint from the cotton image. Any object that touches the boundary of the image is removed from the image. All remaining objects are numbered as blobs and transferred to a label image for the collection of data. Fig. 2 through 5 represent training samples for various trash types and their respective thresholded binary image. Area and ferrets are measured; convex area, bounding box area, extent, solidity, and E^{dif} are computed for all the blobs in the binary image.

Three techniques were examined for their effectiveness in identifying the various trash types.

Fuzzy Clustering

A fuzzy clustering algorithm (Lin, 1996) is used on the data set to obtain the fuzzy center for each trash type. Based on fuzzy centers obtained from the training data, trash objects in two test images (Fig. 6 and 7) are classified. The minimum distance criteria was used to classify the trash types. The classification results for the two test images are summarized in the results section.

Back-Propagation Neural Network

A three-layer back-propagation neural network (Medsker, 1994) is trained using inputs area, solidity, and E^{dif}. The input-output training pairs of data consist of 212 patterns (features from the training images). Based on the weight vectors obtained from the training data, the trash objects in two test images (Fig. 6 and 7) are classified. The classification results for the two test images are summarized in the results section.

Adaptive Network-Based Fuzzy Inference System (ANFIS)

ANFIS is based upon a set of fuzzy rules originally proposed by Takagi, Sugeno, and Kang, commonly referred to as TSK rules. The consequents of these rules are linear combinations of their preconditions (Jang, 1996). The TSK fuzzy rules can be expressed as follows:

$$R^j: \text{IF } x_i \text{ is } A_i^j \text{ AND } x_2 \text{ is } A_2^j \text{ AND} \cdots \text{AND } x_n \text{ is } A_n^j,$$
$$\text{THEN } y = f_j = a_0^j + a_1^j x_1 + a_2^j x_2 + \cdots + a_n^j x_n,$$

where x_i is the input variable, y the output variable, A_i^j are linguistic terms of the precondition part, and with membership functions of $\mu_{A_i^j}(x_i)$, where $a_i^j \in R$ are coefficients of linear equations $f_j(x_1, x_2, \cdots, x_n)$ and $j = 1, 2, \cdots, M;\ i = 1, 2, \cdots, n.$

Consider the input variables area, solidity, and E^{dif}. The outputs of the network are fuzzy singletons. For the given input-output training pair, the premise parameters, namely the centers (means), and the widths of the fuzzy membership functions are updated using the hybrid learning algorithm. The generalized bell functions shown in Figs. 8 through 13 were used as membership functions.

The TSK rules for the given input-output data pairs can be expressed as shown in Table 2. The initial and final membership functions of the premise parameters are shown in Figures 8 through 13 (Siddaiah, 1999b). The classification results for the two test images(Figures 6 and 7) are summarized in the results section.

Trash Content

Trash Content is defined as the ratio of the total trash area to the image area, i.e.,

$$\text{Trash Content (\%)} = \frac{\text{Trash Area}}{\text{Image Area}} \times 100.$$

The trash content for 18 AMS box samples with various trash levels were computed. The box samples were rotated over the image plane and 10 imags (reps) of each sample were acquired. Six of the 18 samples were used to compute the ratio between AMS measurements and SWCGRL measurements. The mean value of these ratios were used to compute the trash content of the remaining 12 AMS samples.

Results

Classification

Classification results for the trash objects contained in test samples #1 and #2 using back propagation neural network, fuzzy C-means, and ANFIS are shown in Table 3 (Siddaiah, 1999c). Comparing classifier performances, it is seen that ANFIS provides far superior results. The neural network classifier appears to provide better results than the fuzzy C-means. The misclassification of a bark2 object as pepper objects in test sample #1 is due to segmentation (blobs 13 and 22 in Fig. 6). The ANFIS parameter optimization method used for training the fuzzy inference system yields better convergence compared to the back propagation neural network.

Trash Content

The trash content for 6 AMS box sample images were computed to obtain the ratios of the AMS to SWCGRL measures. Fig. 14 and Fig. 15 represent the images of the AMS trash boxes with the lowest and highest trash content of the six box samples. The mean value of the ratio is used to compute the trash content and the correlation coefficient between the AMS and SWCGRL trash content of the remaining 12 AMS trash boxes. Figs. 16, 17, and 18 are three images of the 12 samples shown for illustration. Table 4 represents the SWGGRL and AMS measures for the six samples. A total of 10 reps of each sample were used to compute the trash content. Table 5 shows the mean values of the computed percent trash and the AMS percent trash. The correlation coefficient between AMS values and SWCGRL values is 0.9988.

In the identification of trash types in cotton samples, the trash objects lying on the periphery (edges) of the images are removed from the segmented image since any such object may result in misclassification. For instance, if a bark2 object is on the periphery, it may appear as a leaf object and result in misclassification. However, in the computation of trash content it is necessary to include all objects in the camera's field of view (FOV) to measure the total trash area. In the computation of trash content in this research effort, all pixels in the FOV were used to measure the trash area. All objects with area less than 10 pixels were considered as noise and removed from the images before the computation of trash content.

In Table 4, the ratios for all samples but #6 are within ±1 standard deviation of the mean of all 6 samples. The reason for this discrepancy is that not all the trash objects in sample #6 surface can be acquired when rotating an image, resulting in a lower trash area. This is evident in Fig. 15. To correct this, all trash in an AMS trash box would have to be within a 2.5-in. circle.

Conclusions

Results indicate superior performance in the classification compared to previously employed methods. The classification rates are superior with the ANFIS and back propagation neural networks. The proposed approaches can be used in developing a classification system capable of identifying various trash types in real time. The computed percent trash is consistent with AMS percent trash measurements.

Footnote

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.

Acknowledgment

This research effort was partially funded by Dr. David McAlister III, Research Leader, USDA-ARS, CQRS, Clemson, South Carolina.

References

U.S. Department of Agriculture, Agricultural Marketing Service, The classification of Cotton, Agricultural Handbook 566, Washington, D.C. (Revised Jan. 1999).

M. A. Lieberman and R. B. Patil, "Evaluation of learning vector quantization to classify cotton trash," Optical Engineering, Vol. 36 No. 3, pp 914-921, March 1997.

M. A. Lieberman and R. B. Patil, "Clustering and neural networks to categorize cotton trash," Optical Engineering, Vol. 33 No. 5, pp 1642-1653, May 1994.

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

M. A. Lieberman and M. Siddaiah, "Evolution of video color/trash processing at the Southwestern Cotton Ginning Research Laboratory," Beltwide Cotton Production Conferences, Orlando, Florida, 1999.

C. T. Lin, C. S. G. Lee, *Neural Fuzzy Systems*, Prentice Hall, New Jersey, 1996.

L. R. Medsker, *Hybrid Neural Networks and Expert systems*, Kulwer Academic Press, Norwell, Massachusetts (1994).

J. S. R Jang, C. T. Sun, E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, New Jersey, 1996.

M. Siddaiah, M. A. Lieberman, S. H. Hughs, and N. R. Prasad, "Computation of Trash Content in Ginned Cotton using Soft Computing Techniques", IEEE MWSCAS'99-42nd IEEE Midwest Symposium on Circuits and Systems, (accepted for publication), New Mexico State University, Las Cruces, NM, August 8-11, 1999.

M. Siddaiah, M. A. Lieberman, S. E. Hughs, and N. R. Prasad, "A soft computing approach to classification of trash in ginned cotton," Proceedings of the Eighth International Fuzzy Systems Association World Congress, Taiwan, August 1999, Vol. 1, pp. 151-155.

M. Siddaiah, M. A. Lieberman, S. E. Hughs, and N. R. Prasad, "Identification of Trash Types in Ginned Cotton using Neuro Fuzzy Techniques," 1999 IEEE International Fuzzy Systems Conference Proceedings, FUZZ-IEEE'99, South Korea, August 1999, Vol. II, pp. 738-743.

Table 1. Feature Measurements.

Trash type	Bark1	Bark2	Leaf
Area	1526	1578	581
Convex area	3242	1734	651
Bounding box area	4879	3990	858
Solidity	0.4707	0.9100	0.8925
Extent_0	0.3287	0.7599	0.6952
Extent_45	0.1728	0.2777	0.6439
E^{dif}	0.1559	0.4822	0.0513

Table 2. TSK Fuzzy Rules.

Rule	If area is	And solidity is	And E^{dif} is	Then trash type is
R^1	Small	Small	Small	Pepper
R^2	Small	Small	Large	Pepper
R^3	Small	Medium	Small	Pepper
R^4	Small	Medium	Large	Pepper
R^5	Small	Large	Small	Pepper
R^6	Small	Large	Large	Pepper
R^7	Large	Small	Small	Bark1
R^8	Large	Small	Large	Bark1
R^9	Large	Medium	Small	Bark1
R^{10}	Large	Medium	Large	Leaf
R^{11}	Large	Large	Small	Bark2
R^{12}	Large	Large	Large	Bark2

Table 3. Summary of classification results.

Trash types Classifier		Bark1	Bark2	Leaf	Pepper	Correct(%)
Neural Network	Bark1	4	0	0	0	
	Bark2	1	0	0	1	93.33
	Leaf	0	1	5	3	
	Pepper	0	0	1	89	
Fuzzy C - Means	Bark1	0	1	2	1	
	Bark2	0	1	0	1	86.67
	Leaf	0	0	0	9	
	Pepper	0	0	0	90	
ANFIS	Bark1	4	0	0	0	
	Bark2	0	1	0	1	98.10
	Leaf	0	1	8	0	
	Pepper	0	0	0	90	

Table 4. Ratio of AMS to SWCGRL trash content.

Sample No.	AMS Values	SWCGRL Values	Ratio AMS to SWCGRL Values
#1	0.2716	0.6258	0.4345
#2	0.4209	1.0284	0.4196
#3	0.9006	2.1106	0.4275
#4	1.4160	2.9832	0.4754
#5	1.7370	3.4957	0.4970
#6	2.8687	5.5055	0.5212
		Mean	0.4625
		Std. Dev.	0.0416

Table 5. Mean values of AMS and SWCGRL trash content.

Sample No.	AMS Values	SWCGRL Values
#7	0.1010	0.1014
#8	0.2903	0.2921
#9	0.4428	0.4798
#10	0.6791	0.6923
#11	1.1573	1.1019
#12	1.9060	1.8245
#13	0.1066	0.1069
#14	0.2727	0.2753
#15	0.4344	0.4428
#16	0.7009	0.6507
#17	1.1077	1.0171
#18	0.4422	0.4476
Correlation coefficient		**0.99875**

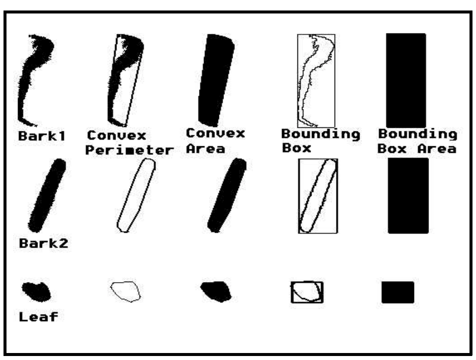

Figure 1. Trash types with feature measurements.

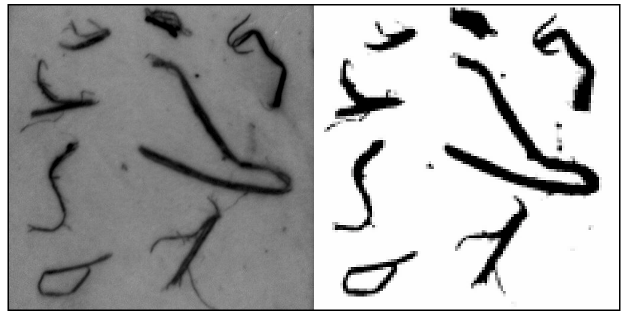

Figure 2. Bark1 training sample.

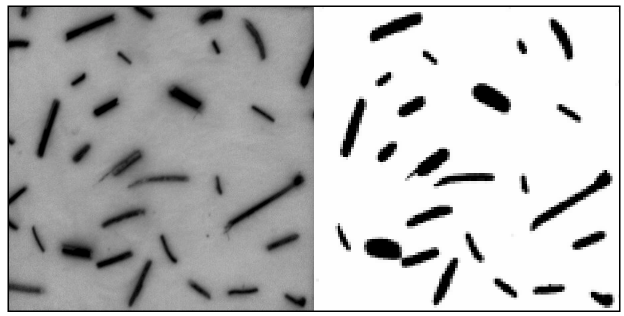

Figure 3. Bark2 training sample.

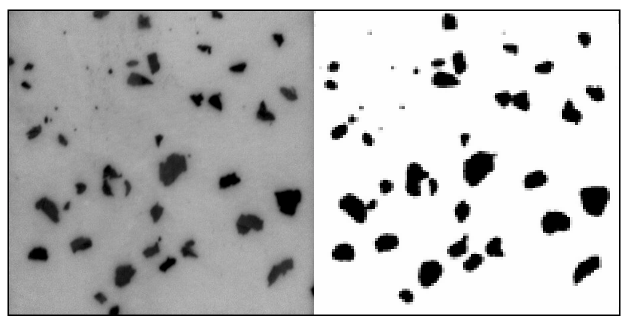

Figure 4. Leaf training sample.

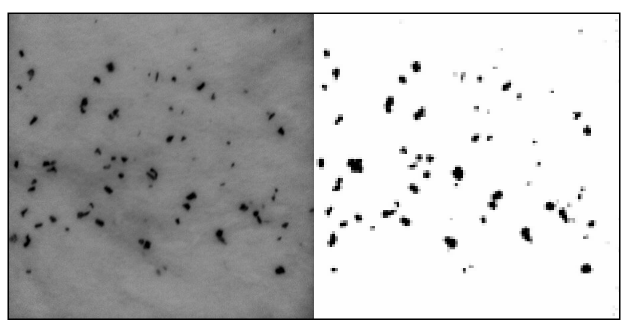

Figure 5. Pepper training sample.

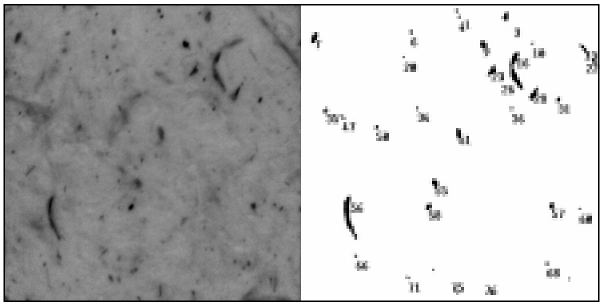

Figure 6. Cotton test sample #1.

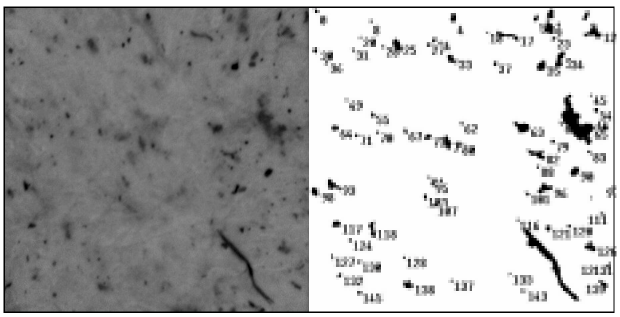

Figure 7. Cotton test sample #2.

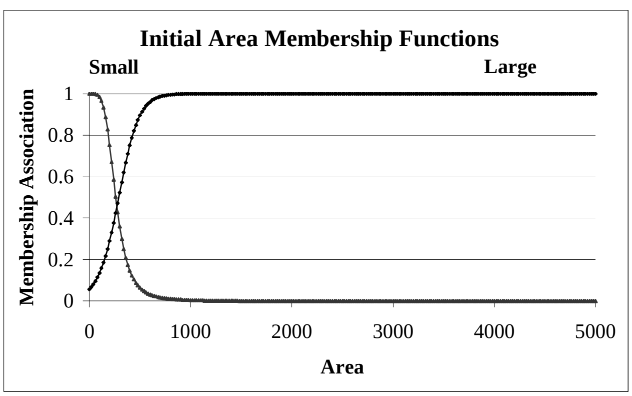

Figure 8. Input variable-Area.

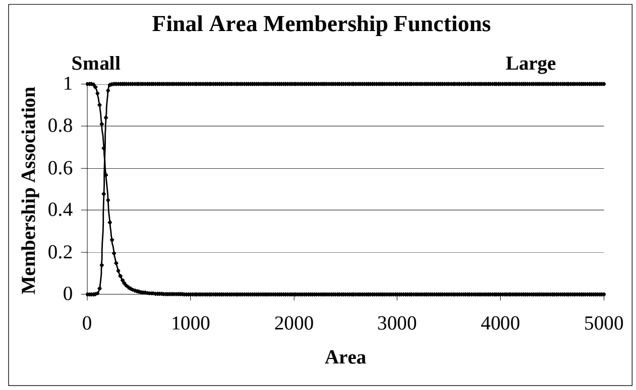

Figure 9. Input variable-Area.

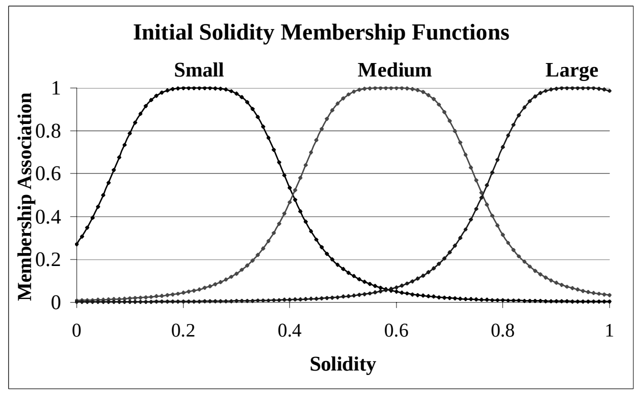

Figure 10. Input variable-Solidity.

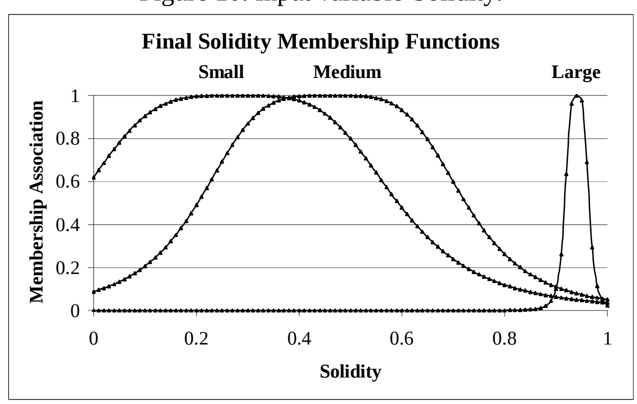

Figure 11. Input variable-Solidity.

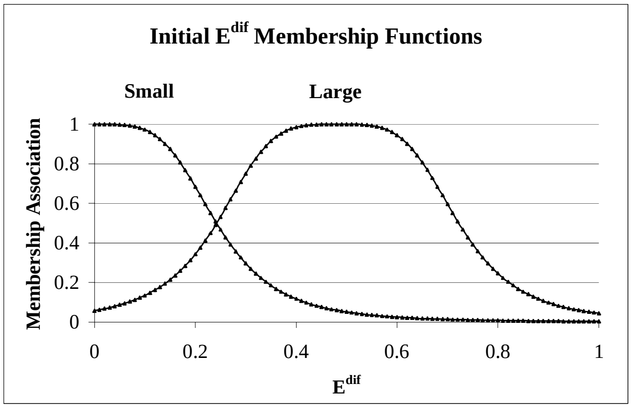

Figure 12. Input variable-E^{dif}.

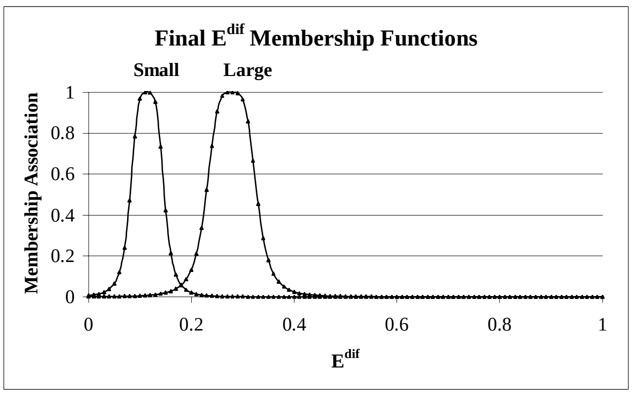

Figure 13. Input variable-E^{dif}.

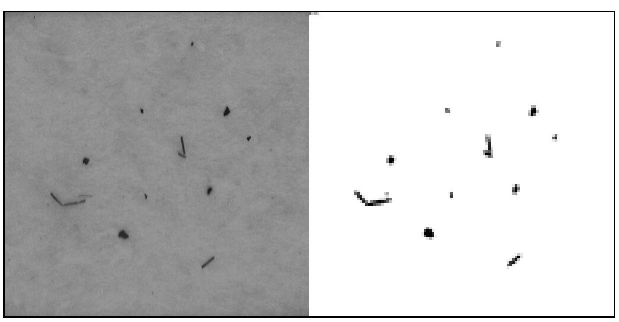

Figure 14. AMS box sample #1 (Trash Content = 0.27).

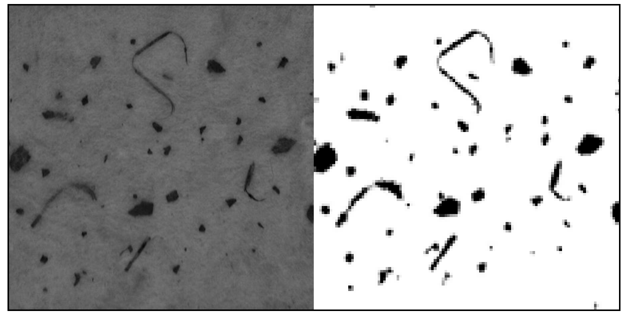

Figure 15. AMS box sample #6 (Trash Content = 2.87).

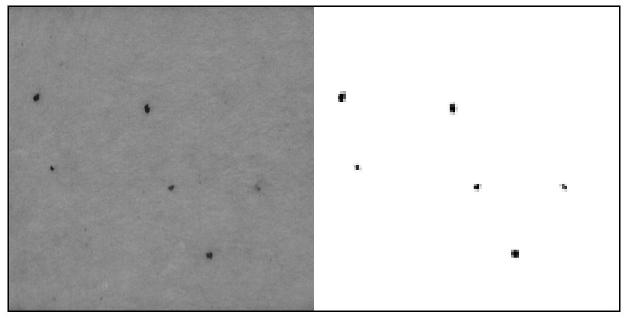

Figure 16. AMS box sample #7 (Trash Content = 0.1).

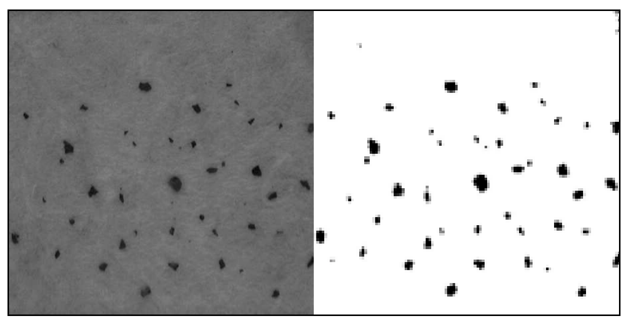

Figure 17. AMS box sample #11 (Trash Content = 1.15).

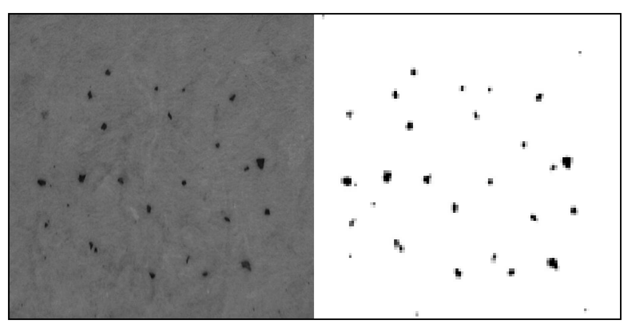

Figure 18. AMS box sample #18 (Trash Content = 0.44).