2002 BELTWIDE COTTON CONFERENCES, ATLANTA, GA – JANUARY 8-12

COTTON COLOR CLASSIFICATION BY FUZZY LOGIC

B. Xu, D.S. Dale and Y. Huang
University of Texas at Austin
Austin, TX
M.D. Watson
Cotton Inc.
Raleigh, NC

Abstract

This paper describes the application of fuzzy logic to cotton color grading in an attempt to improve the accuracy of the machine grading for cotton colors. Color grades of cotton are a number of classes in the (R_d, b) color space. Adjacent color classes have blur and overlapping boundaries, making crisp-boundary methods ineffective for the cotton color classification. Fuzzy logic is specialized in dealing with uncertainty and imprecision in a decision-making process, and thus offers a new approach for grading cotton colors. In this paper, we will present the procedures of constructing a fuzzy inference system (FIS) using fuzzy logic to classify major classes of cotton colors, and the preliminary results to demonstrate the FIS's effectiveness in reducing machine-classer disagreements in color grading. The results from the FIS have shown great consistency for multiple years' cotton color data.

Introduction

The Nickerson-Hunter color diagram illustrates a partition of a two-dimensional color space (R_d, b) that defines the USDA color grades, and has been used by the colorimeter of high-volume-instruments (HVI) to grade cotton colors [(USDA, 1993)]. The color diagram consists of one set of linear lines nearly in the horizontal directions and another set of non-linear lines nearly in the vertical directions, forming the boundaries of various color grades. During the past several decades, the HVI color grading was not accepted by the industry, because it did not achieve a satisfactory agreement with visual grading [(Xu, et al., 2000), (Xu, et al., 1998)]. One of the major reasons attributing to the disagreement is that some of the color boundaries in the diagram, especially the one between white and light spotted color classes, do not properly separate neighboring color classes. This may be due to the fact that the diagram was established more than 30 years ago and cotton fiber colors had chronic shifts because of changes in seed varieties and environments. Another main reason for the disagreement is that the crisp, abrupt separations of color grades in the diagram do not reflect the clustering nature of cotton color classes, which often have blur boundaries. Neighboring classes always overlap to some extent. Thus, the belongingness of a sample point in an overlapping region is inherently ambiguous.

The above analysis can be further evidenced by the color data of 2,489 bales of cotton selected from the1996's crop. To facilitate the discussion, we focused on two major color classes, white and light spotted, which are also the two most disputable color grades [(Xu, et al., 2000)]. Figure 1 shows the distributions of the white and light spotted classes labeled by classers (o-- white, *-- light spotted). Both the white and light spotted classes seem to follow a two-dimensional Gaussian distribution. Note that the real boundaries separating three major color classes, white (W), light spotted (LS) and spotted (S), were also drawn on the R_d-b plane. Although the two classes have distinct populations, they overlap extensively and their intersection does not seem to coincide with the W-LS boundary. It is evident that the W-LS boundary does not provide a realistic separation between the white and light spotted classes. This mismatch brings a systematic error into the HVI's color grading.

Figure 2 shows the distributions of the white and light spotted classes classified by the HVI. The clear split between the white and light spotted classes arises from the crisp boundary used by the HVI. However, the W-LS separation by the HVI does not indicate the natural grouping of the cotton color data in these two classes. It is logic to consider that the two peaks of the distribution represent two separate populations in the color data as seen in Figure 1. But the HVI did not allocate these two populations properly. This is the reason why the HVI tends to grade cotton colors for the white class more likely than for the light spotted class.

In order to make the machine grading more realistically reflect the natural grouping of cotton colors, the Agriculture Marketing Service of the USDA and the cotton community agreed to adjust the boundaries of the Nickerson-Hunter color diagram. This measure effectively reduced the systematic bias in the color grading with the HVI colorimeters, and therefore officially adopted as official grading starting from year 2000. However, the modified color diagram does not deal with problems associated with blur, overlapping boundaries of color classes. In a previous paper, we presented how to use an artificial neural network to reduce the machine-classer disagreements [(Xu, et al., 2000)]. The neural network acts as a black-box classifier that does not use explicitly defined boundaries. In this paper, we will report an investigative work of applying fuzzy logic to eliminating the hard boundary problems in cotton color grading.

Fuzzy logic uses the fuzzy set theory and approximate reasoning to deal with imprecision and ambiguity in decision-making [(Cox, 1999), (Jang and Gulley, 1997) (Hguyen and Walker, 1999), (Lin and Lee, 1996)]. It provides intuitive, flexible ways to create fuzzy inference systems for solving complex control and classification problems. For classification applications, fuzzy logic is a process of mapping an input space into an output space using membership functions and linguistically specified rules. In this study, we take the output of the HVI colorimeter, the R_d-b data, as the input, and color grades as the output. Our discussion in this paper will be limited to the classification for five major color classes, white (W), light spotted (LS), spotted (S), tinged (T) and yellow stained (YS). Figure 3 presents a schematic diagram of the fuzzy inference system (FIS) for cotton color grading [(Jang and Gulley, 1997)].

Methods

Fuzzy Sets and Memberhsip Functions

Elements in ordinary or crisp sets have full memberships in one set and zero memberships in others. A fuzzy set contains elements only with partial memberships ranges from 0 to 1 to describe uncertainty for classes that do not have sharply defined boundaries. For each input and output variable of an FIS, fuzzy sets are created by dividing its universe of discourse (entire space) into a number of sub-regions and are named in linguistic terms. Fuzzy sets' linguistic terms are useful in establishing fuzzy rules. In designing an FIS for cotton color grading, five fuzzy sets were selected for the input variable R_d and six for b. The fuzzy sets for R_d represent five levels of brightness varying from very low (I), low (II), median (III), high (IV) to very high (V), and the fuzzy sets for b represent six levels of yellowness ranging from very low (I) to extremely high (VI). Table I presents the ranges and other distributions parameters of the input fuzzy sets. Each fuzzy set overlaps with its adjacent fuzzy sets. The reason for adding one more fuzzy set for b is that b seems more critical than R_d in determining cotton major color classes (white, light-spotted, etc.). In general, the more intermediate levels are used, the higher accuracy the classification would be. But increasing the fuzzy sets will significantly increase the number of fuzzy rules in the next step. The final selection on the number of fuzzy sets and their range may be determined by trial and error. Since this FIS was designed to classify five major color classes, the output variable was split by five fuzzy sets named as white (1), light-spotted (2), spotted (3), tinged (4) and yellow stained (5). The range of the output variable was equally divided into five sections for the five fuzzy sets.

Once the fuzzy sets are chosen, a membership function for each set should be created. A membership function is a curve that maps an input element to a value between 0 and 1 showing its degree of belongingness to a fuzzy set. The curve can have different shapes, such as bell (Gaussian), sigmoid, triangle and trapezoid, for different types of fuzzy sets [(Cox, 1999), (Jang and Gulley, 1997)]. In this study, the Gaussian distribution curve was used to build the membership functions for the input fuzzy sets R_d and b:

$$\mu(x) = e^{-(x-m)^2/2\sigma^2}$$

where m and σ are the mean and the standard variation of one fuzzy set in x (R_d or b). Finding the right parameters for the functions is a major task, which may be selected arbitrarily and then tweaked by using a known set of input-output data. The m and σ values used in this FIS are included in Table I, and the membership functions are displayed in Figure 4. The extent of overlap between the membership functions of two adjacent sets indicates the nature of the unsharp boundary between two color classes.

For the simplicity of defuzzification, a triangular shape was used to construct the membership functions for the output fuzzy sets:

$$\mu(x) = \begin{cases} = 0, & x < a \text{ or } x \geq c \\ = (x-a)/(b-a), & a \leq x < b \\ = (c-x)/(c-b), & b \leq x < c \end{cases}$$

The shape and size of the triangular function depend on the values of a, b and c. To make the output clear and unbiased, the symmetric, non-overlapping and equal-size membership functions were used for all the output sets (Figure 5).

Fuzzification

Fuzzification is a step to determine the degree to which an input data belongs to each of the appropriate fuzzy sets via the membership functions. For a given input point (R_{d0}, b_0), the memberships of all the fuzzy sets are calculated, and only the fuzzy sets with non-zero memberships are forwarded to the next steps. In Figure 4, an example of determining the relevant fuzzy sets was shown for an input data (R_{d0}, b_0)=(67.5, 9.0). R_{d0} belongs to the medium (III) and high (IV) sets of R_d with the memberships being 0.51 and 0.52, while b_0 belongs to the low (II) and medium (III) sets of b with the memberships being 0.20 and 0.82. There are four combinations with the selected fuzzy sets:

$$[\mu_{III}(R_{d0}), \mu_{II}(b_0)] = [0.51, 0.20],$$
$$[\mu_{III}(R_{d0}), \mu_{III}(b_0)] = [0.51, 0.82],$$
$$[\mu_{IV}(R_{d0}), \mu_{II}(b_0)] = [0.52, 0.20],$$
$$[\mu_{IV}(R_{d0}), \mu_{III}(b_0)] = [0.52, 0.82].$$

These four combinations will be evaluated by fuzzy rules to determine the output fuzzy sets and the weight of each rule influencing the output.

Fuzzy Rules

In an FIS, fuzzy rules provide qualitative reasoning that links input fuzzy sets with output fuzzy sets. They are a collection of linguistic rules of the form [(Nguyen and Walker, 1999)]:

R_i: **If** $\underline{R_d \text{ is } A_i}$ AND $\underline{b \text{ is } B_i}$, **then** $\underline{color \text{ is } C_i}$, i=1, 2, …, k

where A_i, B_i and C_i are the fuzzy sets for the inputs R_d and b and the output color in the ith rule R_i, and k is the number of the rules. The values of A_i and B_i are the linguistic terms such as very low (I) and very high (V), and the values of C_i are the linguistic terms such as white (1) and light spotted (2). An example of such a rule may be given as follows:

If R_d is *very high* (V) AND b is *very low* (I), **then** *color* is *white* (1).

The **if**-part of the rule is called the antecedent, and the **then**-part of the rule is called the consequent. Since the antecedent in this FIS always involves two conditions (one for R_d and one for b), fuzzy operators are needed to specify the relationships of the fuzzy sets in the antecedent. AND (intersection), OR (union) and NOT (complement) are the three common fuzzy operators. Because R_d and b should be simultaneously observed in selecting a color class, the fuzzy operator in all the antecedents must be AND. For two fuzzy sets A and B, the fuzzy AND is defined as [(Cox, 1999), (Nguyen and Walker, 1999)]:

$$A \text{ AND } B: min\{\mu_A(x), \mu_B(x)\}.$$

Fuzzy AND aggregates two membership functions by outputting the minimum value at a given input x (R_d or b). The result of fuzzy AND serves as a weight showing the influence of this rule on the fuzzy set in the consequent.

The fuzzy rules should be established based on both visual grading experience and the basic relationships between color data and color grades in the HVI color diagram. Since there are five fuzzy sets in R_d and six fuzzy sets in b, there are 30 possible combinations in the antecedents when only fuzzy AND is applied. Thus, the maximum number of the fuzzy rules that can be established is 30. To simplify the fuzzy rule expressions, we designed a chart that illustrates (Figure 6). In the chart, the five sets of R_d is arranged vertically and the six sets of b is arranged horizontally. A knot (a black dot) between a horizontal line and a vertical line indicates an antecedent formed by the connecting fuzzy sets from R_d and b, and leads to a consequent that in turn gives an output fuzzy set. The aggregated membership of the antecedent is then used as a weight factor to modify the size and shape of the membership function of the output fuzzy set in a way of either truncation or scaling. Truncation is done by chopping-off the triangular output function, while scaling is done by compressing the function. In this FIS, the truncation operation was used. If the membership function of an output fuzzy set is $\mu(x)$ and the weight generated from the antecedent is w, the truncated functions are:

$$\mu^T(x) = max\{\mu(x), w\}.$$

As given previously, the input fuzzy sets containing sample (R_{d0}, b_0) have four combinations, which satisfy the four following rules:

R_1: **If** R_d is *median* (III) AND b is *low* (II), **then** *color* is *white* (1);
R_2: **If** R_d is *median* (III) AND b is *median* (III), **then** *color* is *spotted* (3);
R_3: **If** R_d is *high* (IV) AND b is *low* (II), **then** *color* is *white* (1);
R_4: **If** R_d is *median* (IV) AND b is *median* (II), **then** *color* is *light spotted* (2).

The four output fuzzy sets were circled in Figure 6. The weights of the rules on the four outputs are 0.20, 0.51, 0.20 and 0.52, respectively. The truncated membership functions of the output fuzzy sets were presented in Figure7.

Defuzzification

After all the fuzzy rule evaluations are done, the FIS needs to output a crisp member to represent the classification result (color classes) for the input data. This step is called defuzzification. As seen in the (R_{d0}, b_0) example, one input data may generate several weighted output fuzzy sets. The multiple sets need to be aggregated into a single set in preparation for the defuzzification. If those output fuzzy sets are different, the aggregation can be done simply by placing all the truncated functions together to form the final fuzzy set. If two of the output fuzzy sets are identical, they can be combined by using fuzzy OR, which is defined as [(Cox, 1999), (Nguyen and Walker, 1999)]:

$$A \text{ OR } B: max\{\mu_A(x), \mu_B(x)\}.$$

Fuzzy OR gives the maximum value of the two membership functions at any given point. For example, sample (R_{d0}, b_0) has two 'white', one 'light spotted' and one spotted output functions. After the fuzzy OR operation, the two trapezoidal 'white' functions merge into one so that the agrregated curve becomes the one shown in Figure 8.

The most popular method for defuzzification is the centroid calculation, which returns a grade weighted by the areas under the aggregated output functions. Let a_1, a_2, … a_n be the areas of the truncated triangular areas under the aggregated function, and c_1, c_2, … c_n be the coordinates of their centers on the x-axis. The centroid of the aggregated area is given by [(Jang and Gulley, 1997), (Lin and Lee, 1996)]:

$$G = \sum_{i=1}^{n} a_i c_i \Big/ \sum_{i=1}^{n} a_i$$

The location of the centroid indicates the color class to be designated to the input data. For sample (R_{d0}, b_0), G is 17.1, which falls in the light spotted set (see Figure 8). The location of this centriod inside the identified fuzzy set can suggest a finer rating.

Experiment

The constructed FIS was tested using cotton samples in 1996, 1997 and 1998 crop years. Since the majority of the U.S. Upland cottons are 'white' and 'light spotted' and the most disputable grading occurs between these two classes, samples only in these two classes were selected for the experiment. There were totally 2489 samples from the 1996's crop, 1375 samples from the 1997's crop and 658 samples from the 1998's crop in the selection. The samples were firstly graded by official cotton classers, and then measured by the HVI colorimeters. The (R_d, b) data were converted into color grades using the HVI color diagram and the FIS. We used the classers' grades as a reference to check the consistency of the HVI and FIS data, since the classers' grades were the official grades at that time and they were coincided with the natural grouping of the color data. Table II presents the data showing the disagreements among the testing methods. % disagreement was calculated by dividing the disagreement count with the number of the tested samples. For the white class, 54.1% of the samples in 1996's crop, 61.5% in 1997's, and 51.2% in 1998's were graded differently by the classers and the HVI. The amount of disagreements between the classers and the FIS has decreased considerably, and the FIS results are consistent for consecutive years. The classer-FIS disagreement much more evenly spread between the white and light spotted classes, suggesting that there is no systematic bias in the FIS's grading.

Figure 9 shows the distributions of the white and light spotted color classes of the 1996 and 1997 samples classified by the FIS. It can be seen that the two color classes in both years' data were reasonably separated. Based on the two newly separated data clusters, a best-fit polynomial curve was calculated (curve 2 in the figure). This curve provides a new boundary that more properly reflects the segregation between the white and light spotted color classes. The distribution for the 1998's crop was not included because the number of the available samples from that year was not sufficient for drawing an effective distribution.

Summary

This paper presents the investigative work of using fuzzy logic to construct a fuzzy inference system for classifying the white and light spotted cottons based on their (R_d, b) color measurements. The performance of the FIS depends primarily on the selections of input and output fuzzy sets, the design of the membership functions and the establishment of fuzzy rules that guide the input-output relationships. These parameters may be initially selected based on an analysis of a small set of data, and then tweaked by trial and error with more data. The classers' knowledge on color grading can be incorporated into the fuzzy rules to enhance the overall agreement with visual grading. The FIS is able to perform consistent classifications for the samples from different years. Although the generalization of the FIS could not tested more thoroughly in this study due to the lack of samples and other resources, the preliminary results has shown great potential for being a more reliable way of grading cotton colors.

Acknowledgement

We are grateful for the financial support provided by Cotton Incorporated and Texas Technology Development and Transfer Program for this research. We would like to thank Mr. James Knowlton, AMS of USDA, Memphis, TN, for the samples and comments he provided for the research.

References

Cox, E., "Fuzzy Systems Handbook, 2nd Edition", Academic Press, Chestnut Hill, MA, 1999.

Jang, Roger J. S. and Gulley, Ned "MATLAB/Fuzzy Logic Toolbox," MathWorks, Inc. Natick, MA, 1997.

Hguyen, H.T. and Walker, E.A., "A First Course in Fuzzy Logic, 2nd Edition," Chapman & Hall/CRC, Boca Raton, FL, 1999.

Lin, Chin-Teng and Lee, George C. S. "Neural Fuzzy Systems," Prentice Hall, Upper Saddle River, NJ, 1996.

United States Department of Agriculture, "The Classification of Cotton," Agricultural Handbook, 1993.

Xu, B., Su, J., Dale, D. and Watson, M.D., Cotton Color Grading by Neural Network, *Textile Research Journal*, **70**, 430-436, 2000.

Xu, B., Fang, C. and Watson, M.D., Investigating New Factors in Cotton Color Grading, *Textile Research Journal*, **68**, 779-787, 1998.

Table I. Parameters for the Fuzzy Sets of the Input Variables.

	R_d			b		
Fuzzy set	**Range**	***m***	**σ**	**Range**	***m***	**σ**
Very low (I)	40-52.5	46.5	3.5	4-7	4.0	1.00
Low (II)	45-65	55	3	4-11	7.2	1.00
Medium (III)	54-75	64	3	7-12.5	9.5	0.80
High (IV)	60-82.5	71.5	3.5	9-17	12.4	1.19
Very high (V)	67.5-87.5	77.5	3.5	11-18	15.1	1.19
Extremely high (VI)				14-18	18.0	1.19

Table II. Disagreements Between Different Grading Methods.

Color Class	**White**			**Light Spotted**		
Crop Year	**1996**	**1997**	**1998**	**1996**	**1997**	**1998**
By Classer (C)	1240 (49.8%)	416 (30.3%)	464 (70.5%)	1249 (50.2%)	959 (69.7%)	194 (29.5%)
By HVI	2350 (94.4%)	1200 (87.3%)	628 (95.4%)	139 (5.6%)	175 (12.7%)	30 (4.6%)
By FIS	1161 (46.6%)	355 (25.8%)	447 (68.0%)	1328 (53.4%)	1020 (74.2%)	211 (32.0%)
C-HVI Disagreement	1347 (54.1%)	846 (61.5%)	337 (51.2%)	9 (0.4%)	14 (1.0%)	3 (0.5%)
C-FIS Disagreement	154 (6.2%)	87 (6.2%)	40 (6.1%)	234 (9.4%)	144 (10.5%)	72 (10.9%)

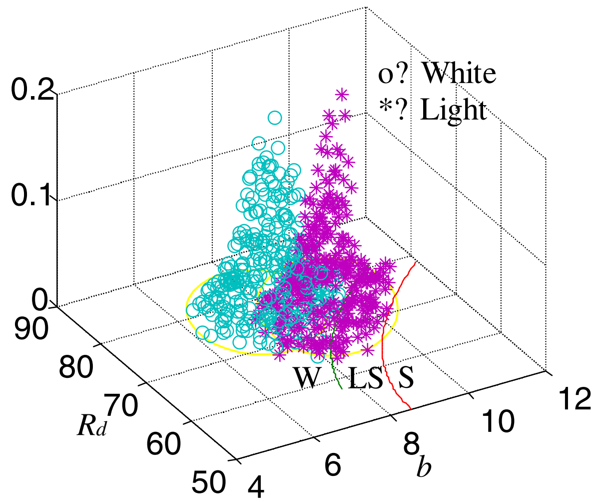

Figure 1. Distributions of White and Light Spotted Colors Classified by Classers.

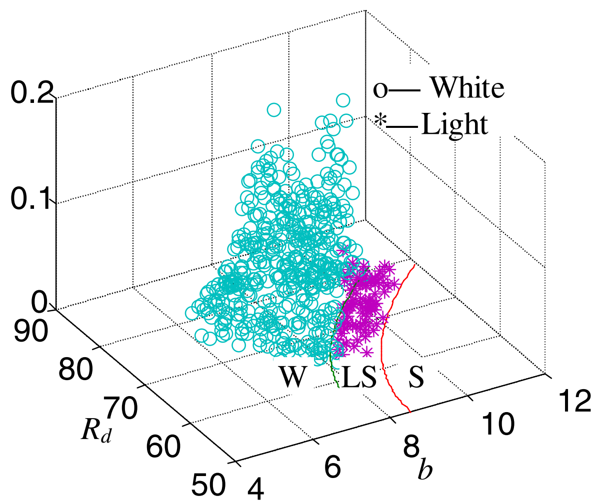

Figure 2. Distributions of White and Light Spotted Colors Classified by the HVI.

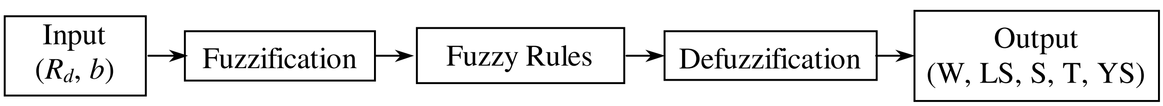

Figure 3. Fuzzy Inference System for Cotton Color Grading.

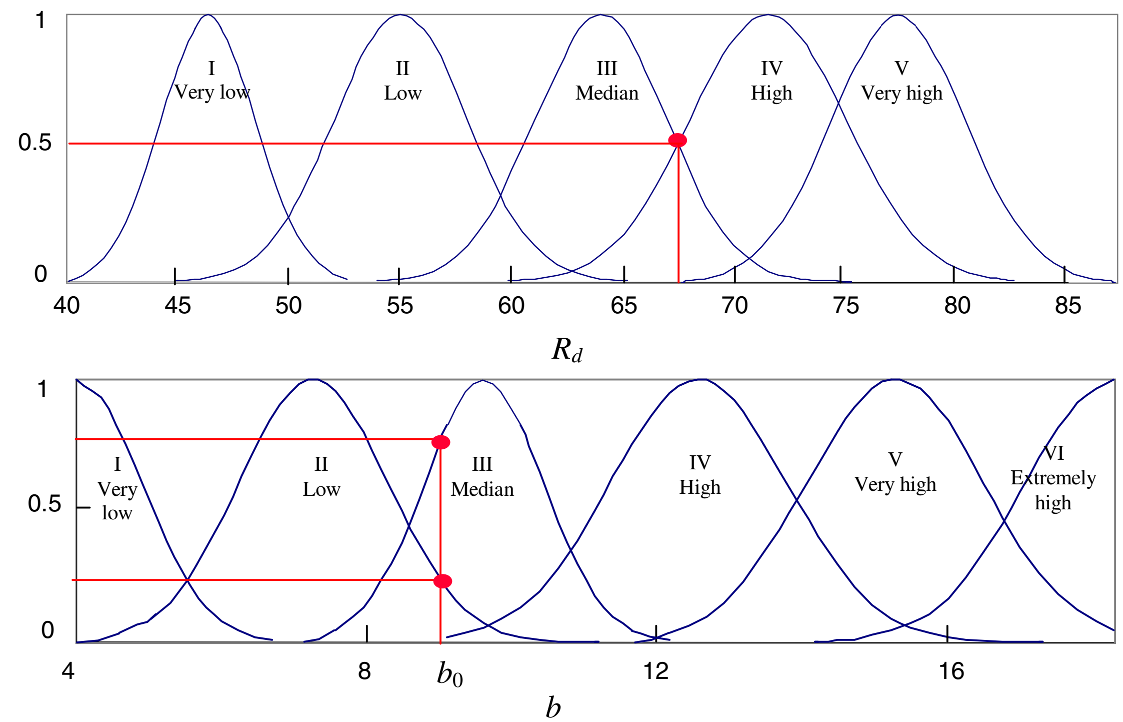

Figure 4. Membership Functions of Input Fuzzy Sets.

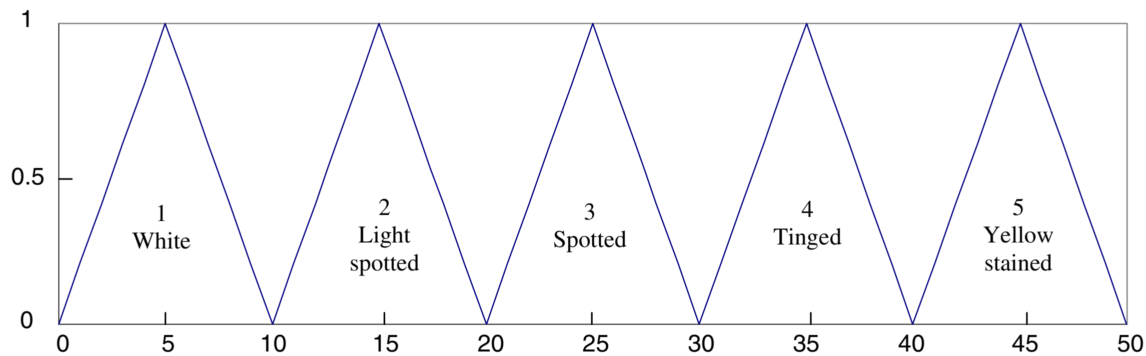

Figure 5. Triangular Membership Functions of Output Fuzzy Sets.

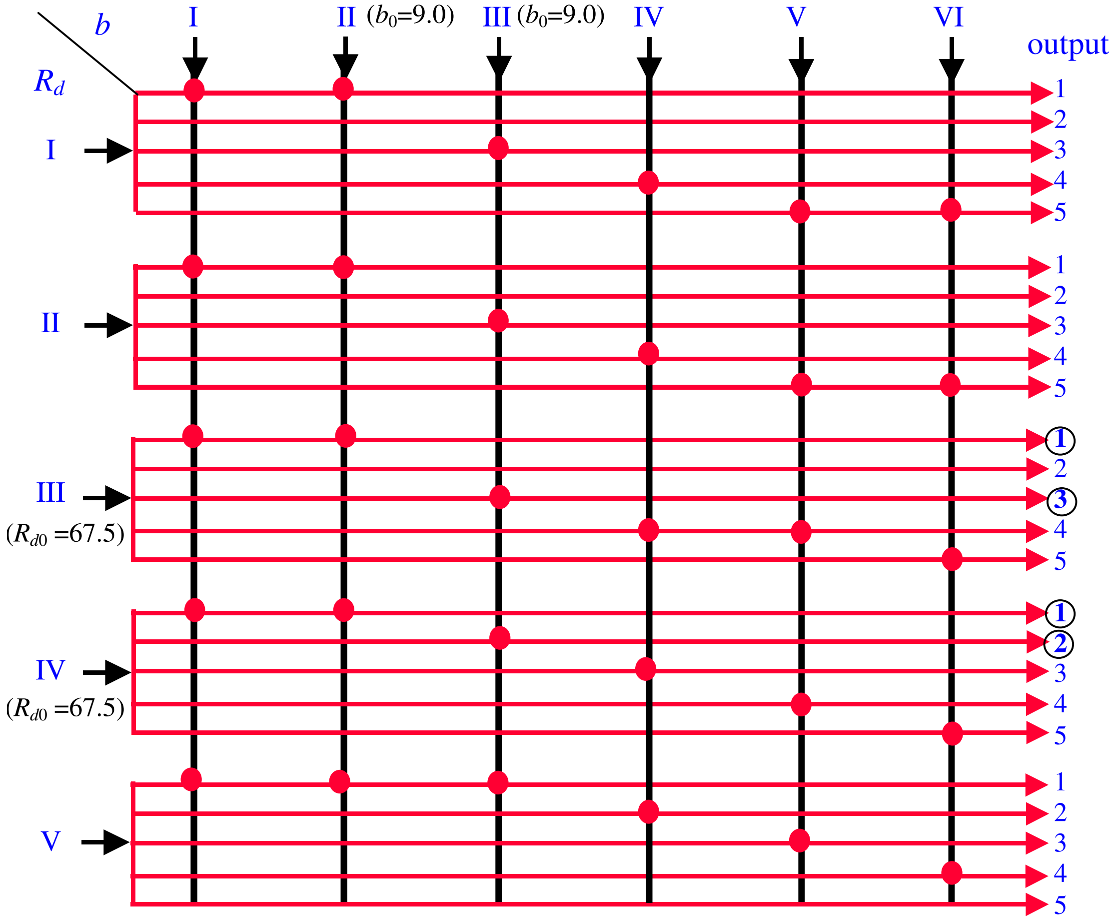

Figure 6. Fuzzy Rules.

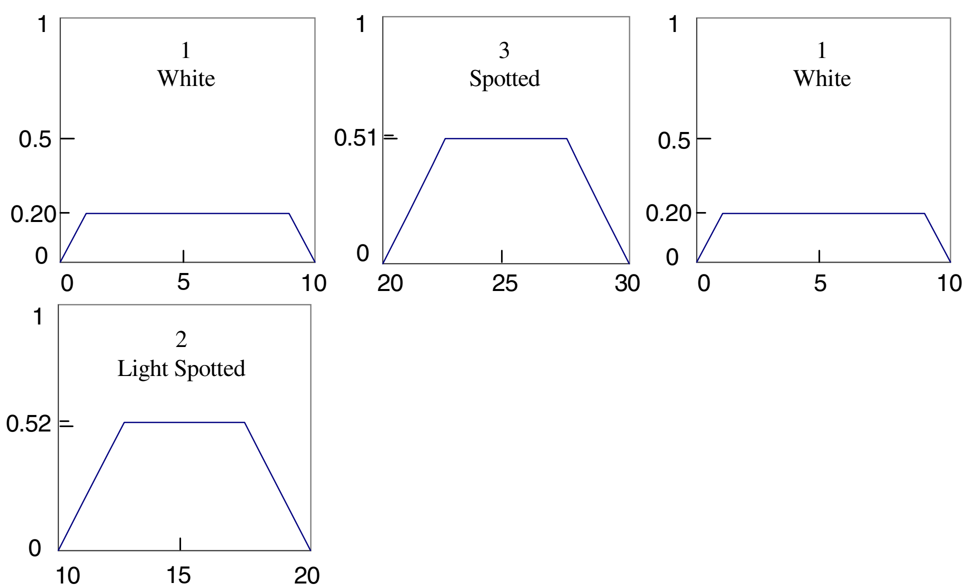

Figure 7. The Weighted Membership Functions of Output Fuzzy Sets.

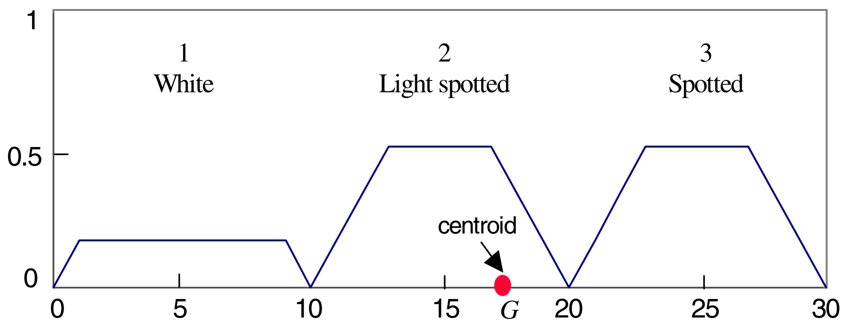

Figure 8. Aggregation of the Weighted Output Functions.