Fuzzy Models to Predict Consumer Ratings for Biscuits Based on Digital Image Features

Valerie J. Davidson, Joanne Ryks, and Terrence Chu

***Abstract*—Fuzzy models to recognize consumer preferences were developed as part of an automated inspection system for biscuits. Digital images were used to estimate physical features of chocolate chip cookies including size, shape, baked dough color, and fraction of top surface area that was chocolate chips. Polls were conducted to determine consumer ratings of cookies. Four fuzzy models were developed to predict consumer ratings based on three of the features. There was substantial variation in consumer ratings in terms of individual opinions (30 panelists in each poll) as well as poll-to-poll differences (three calibration polls). Parameters for the inference system, including fuzzy values for cookie features and consumer ratings, were defined based on judgment and statistical analysis of data from the calibration polls. Two of the fuzzy models gave satisfactory estimates of average consumer ratings for two validation sets (44 cookies). One was a Mamdani inference system that was based on eight fuzzy values for consumer ratings. These were defined using rating distributions from calibration polls. The second model was a Sugeno inference system developed using the adaptive neurofuzzy inference system (ANFIS) algorithm (MatLab®Version 5.2, The MathWorks Inc., Natick, MA) with the calibration poll data.**

***Index Terms*—Consumer preferences, fuzzy models, image analysis.**

I. INTRODUCTION

WITH consumers' demand for high-quality products, quality assurance has become a major concern in all manufacturing environments, including the food industry. In food manufacturing, a substantial amount of product grading and quality assurance is performed by human inspectors. However, manual inspection tends to be laborious, tedious, and prone to inconsistency. To solve these difficulties, food manufacturers are interested in automated visual inspection for quality assessment, which is one of the fastest growing applications for machine vision, according to Gunasekaran [4].

At a low level of information processing, there are many advantages to automated inspection. Feature extraction (e.g., physical aspects such as size, area, and color) is consistent, unbiased, and quantitative. However, many food inspection operations also require a higher level of information processing. It is often necessary to integrate of a number of physical features to make an inference about overall quality that is consistent with expert graders' or consumers' judgments. The subjective and psychological biases in human decisions are more difficult to capture in computer algorithms to assess product quality.

A variety of classification methods have been used to integrate image features for assessment of food quality: statistical methods such as multiple regression estimators and discriminant analysis, knowledge-based systems based on consultation with expert graders, neural networks, and fuzzy logic tools. Croft, de Silva, and Kurianto [2] developed an automated grading system for herring roe. The prototype system had a hierarchical structure that processed extracted features such as shape and weight as fuzzy quality values and then used a set of fuzzy rules to assign an overall grade (five classes). Testing with industrially graded roe (215 pieces) produced a grading accuracy of 85%–95% and reasonable repeatability. Inspection of almonds, crackers, and corn kernels by machine vision was reported by Ding and Gunasekaran [3]. A back-propagation neural network was used to classify food shapes. Several minimum indeterminate zone classifiers were developed to reduce classifier training time. With machine vision, the accuracy and speed of food shape inspection was greatly improved. Paulus, De Busscher, and Schrevens [7] determined size, shape, and color of apples using digital images and developed a "tree-based" model to simulate quality assessment by human graders. In this paper, they also evaluated the consistency of human graders. Decision trees were a succession of binary rules to classify a particular variety. Rules were determined from a training set of classified apples. "Tree-based" models provided some evidence about the product features that influence the judgments of human graders. Misclassification was tested for two varieties and was about 29% for both test sets. However, human graders were not 100% consistent.

In some applications, there are no government-regulated grades for a food product, but it is important to mimic judgments of general consumers in product inspection. In these cases, consumer assessments are expected to vary between individuals in a group and over time. Fuzzy methods are useful for representing the variability that is inherent in consumers' opinions as well as flexible data aggregation techniques.

We have explored the use of fuzzy models to recognize consumer preferences based on physical features extracted from digital images of chocolate chip cookies. We chose chocolate chip cookies because we were working with a commercial bakery and could obtain products with a range of characteristics. Furthermore, certain cookie attributes such as the surface appearance of chocolate chips and baked color are strong influences on consumers' assessment of quality. Hence, chocolate chip cookies were an appropriate product to consider both quantitative and subjective criteria in an automated inspection system.

Manuscript received October 27, 1999; revised August 24, 2000. This work was supported by the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA). The work of T. Chu in developing the image analysis system was supported by Natural Sciences and Engineering Research Council (NSERC).

The authors are with the School of Engineering, University of Guelph, Guelph, ON N1G 2W1 Canada (e-mail: vdavidso@uoguelph.ca).

Publisher Item Identifier S 1063-6706(01)01366-2.

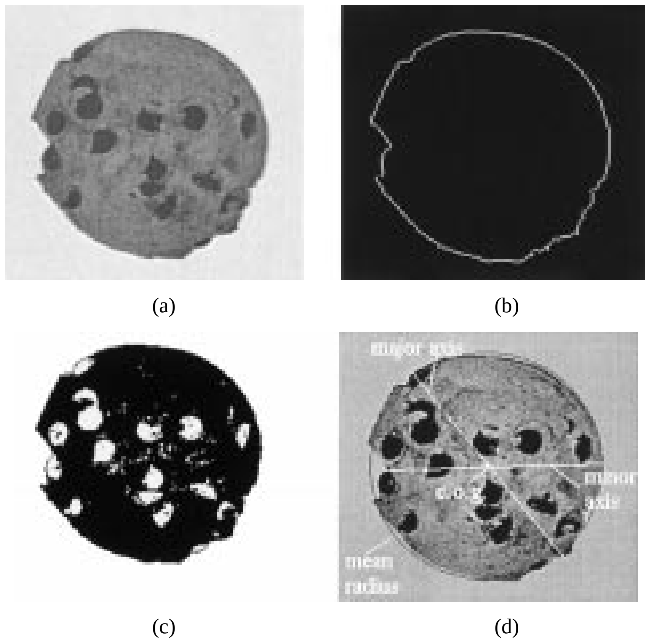

Fig. 1. Feature extraction: (a) digital image of a chocolate chip cookie; (b) perimeter definition; (c) chocolate chip definition; and (d) dimensions.

II. Methods

A. Quantitative and Fuzzy Values for Cookie Features

Digital images of typical chocolate chip cookies were obtained using a camera (JVC KY-F55BU, Elmwood Park, NJ) with three charge-coupled-device sensors (Red Green Blue) connected to a laboratory computer (200-MHz Pentium processor) with a Meteor RGB grabber board (Matrox Electronic Systems Ltd., Dorval, QC). Lighting was controlled by using two fluorescent fixtures (15 W) to illuminate the top surface of a cookie with minimal shadows. Lens adjustments were used to control the field of view and light transmission.

To measure physical attributes of the cookie from a digital image, a feature extractor was developed in C++ in a UNIX-based environment. The Image Vision library (Silicon Graphics, Inc.) was used for basic image processing functions. The feature extractor identified the cookie perimeter and cookie surface color values and quantified the major and minor axes, average diameter, and fraction of top surface area that was chocolate chips (Fig. 1). RGB color values were transformed to CIE (Commision Internationale de L'Eclairage) XYZ values and then to the $L^*a^*b^*$ color space, which is commonly used in the food industry [6]. A lightness (L^*) threshold was used to identify chocolate chips. Baked color was calculated based on L^* values for cookie surface pixels that were not identified as chocolate chips. Calibration procedures were developed for L^* (using an independent colorimeter) and dimensions. Details related to the acquisition of digital images and feature extraction are given in Chu [1].

Initially fuzzy values for four physical features (size, shape, visible chips, and dough lightness) were defined by the system designers. The number of levels and range for each fuzzy value were based on judgments about consumers' abilities to differentiate cookies, as well as normal production ranges for each characteristic. Membership functions for these fuzzy values are shown as solid lines in Fig. 2. Full membership in the 12 combinations of the 3 features defined archetypes or classes (e.g., a large, dark cookie with few visible chips). These 12 classes were the basis for designing consumer polls and preliminary data analysis.

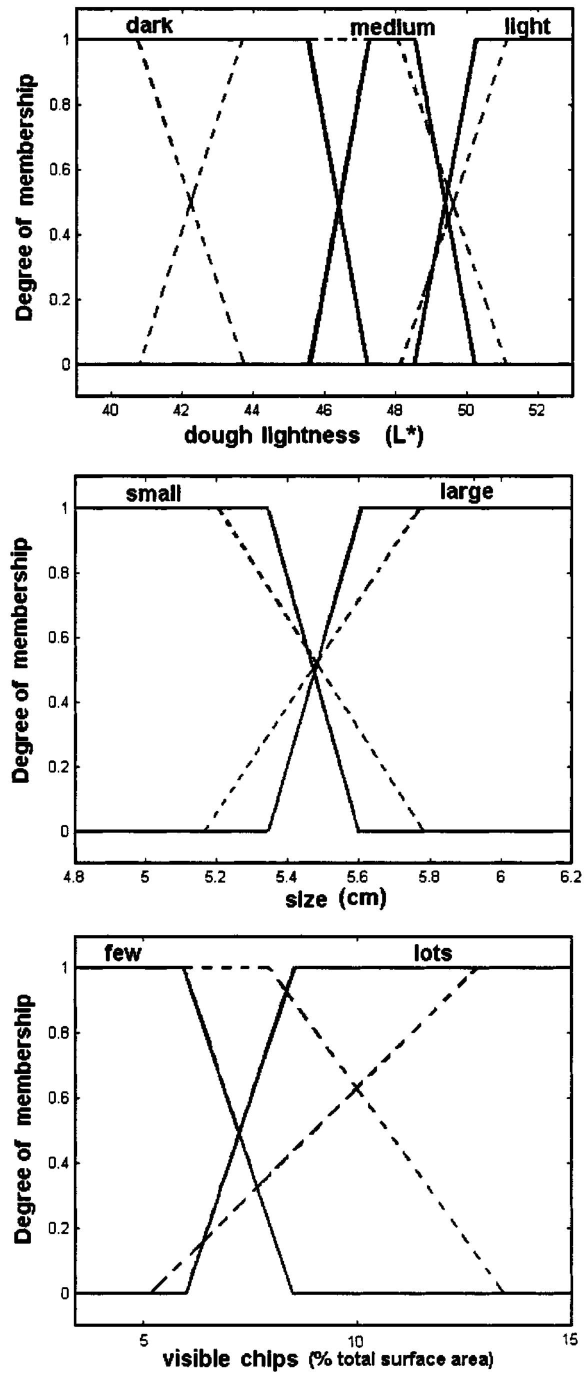

Fig. 2. Membership functions for cookie features (dough lightness, size, and visible chips). —— defined by system designers, - - - - - defined by ANFIS algorithm.

B. Consumer Evaluations

Chocolate chip cookies from a commercial bakery were evaluated by students, faculty, and staff in the School of Engineering in five separate polls. Three of the polls were conducted to develop models for consumer ratings based on cookie features and these are referred to as the calibration polls. Thirty panelists par-

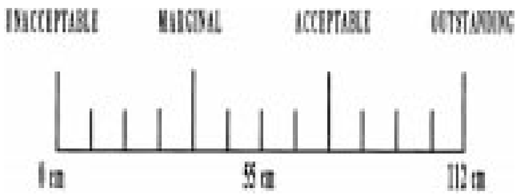

Fig. 3. Line scale used by consumers to rate biscuits.

ticipated in each poll and 17 panelists took part in all three polls. Two additional polls were conducted with 30 panelists to validate the models for consumer ratings. Cookies in the validation polls were different from those used in the first three polls, but with similar physical features.

In each of the first three polls, two sets of 12 chocolate chip cookies representing archetypes were evaluated by individual panelists. There were 24 (i.e., 2×12) cookies because two values of cookie shape (regular and irregular) were considered in the original set of archetypes. The sets were blocked to confound blocks with the crossed effect of size $\times$ shape $\times$ visible chips. In total, there were 44 observations in the calibration set since some cookies were used in more than one poll. Panelists were asked to place cookies on a line scale, as illustrated in Fig. 3. Panelists rated cookies based on individual preferences and were not required to use the entire scale. Linguistic labels (unacceptable, marginal, acceptable, and outstanding) were used to define approximate intervals on the line scale.

Placement of a cookie on the linguistic scale was converted to a numeric value by measuring distances (cm) from the left end of the scale to the left and right edges of the cookie. These values were averaged to define an individual consumer rating for each cookie. The final data set included each consumer's rating of individual cookies, with each poll and block (within a poll) identified. Cookie attributes were coded as fixed levels of the fuzzy values for size, shape, dough lightness, and visible chips. An analysis of variance was conducted to test significance of the following effects: poll, block, size, visible chips, shape, and dough lightness, as well as crossed effects between the image features (i.e., size $\times$ visible chips).

C. Fuzzy Models to Recognize Consumer Preferences

For comparison purposes, four fuzzy inference systems were developed to predict average consumer ratings based on three cookie features (size, dough lightness, and visible chips). These features were chosen based on significance in the analysis of variance. The models differed in two key areas: fuzzy values for cookie features and consumer ratings and the methods of defuzzification.

The four fuzzy models were based on 12 combinations of cookie features (i.e., $3 \times 2 \times 2$ levels). *A priori* definitions of feature values based on the system designers' judgment were used to develop three of the fuzzy models: two Mamdani (ModelM-cent and ModelM-mom) and a Sugeno (ModelS) inference systems as shown in Fig. 2. After the calibration polls were completed, an adaptive neural network technique was also used to define fuzzy values for cookie features. Numerical values for physical features, as extracted from digital images, and average consumer ratings were used as input–output data in an adaptive neuro-fuzzy inference system (ANFIS) from the Fuzzy Logic Toolbox of MatLab®(Version 5.2, The MathWorks Inc., Natick, MA). The ANFIS tool defined membership functions for inputs and 12 values for consumer ratings using a hybrid optimization algorithm to minimize the sum of squared residuals. Membership functions for the three cookie features, as defined by the ANFIS algorithm, are shown as dashed lines in Fig. 2 for direct comparison to the definitions selected by the system designers.

For each cookie archetype, the distributions of consumer ratings from the three polls were characterized by calculating statistical moments (first, second, and third moments). For the Mamdani inference systems, fuzzy values for consumer ratings (i.e., rule consequents) were defined to match as closely as possible to the poll distributions. A total of eight membership functions in Gaussian and exponential forms were defined to represent variance and skewness in distributions of consumer ratings, as well as recognizing poll-to-poll variation. The distributions and parameters for each Mamdani consumer rating are summarized in Table I and, as examples, four distributions (two Gaussian and two exponential) are shown in Fig. 4. Consumer ratings in ModelS were defined using mean values (i.e., first moments) from the calibration poll distributions. Five singleton values were considered adequate based on poll-to-poll variations (Table I). The ANFIS model defined 12 singleton values for consumer ratings since there were 12 unique rules (Table I).

Each inference system included twelve rules to define relationships between each set of three physical features and consumer ratings (Appendix). All rules were given equal weights. Different methods of defuzzification were used to compare estimates of consumer preferences. In the Sugeno inference systems (ModelS and ANFIS), a weighted average was used. In the Mamdani inference systems, two methods of defuzzification to numerical values for consumer rating were compared: centroid and mean-of-maximum [5].

III. Results and Discussion

Cookie features were extracted from digital images as quantitative, numeric values for dimensions, baked color, and coverage of chocolate chips. Separation of baked color, (i.e., dough color) from the dark chocolate chip areas was not done by the existing automated inspection system in the bakery. However accurate identification of chocolate chips in the cookie images was accomplished using an adaptive threshold that was adjusted as the baked color changed from light to dark [1]. One advantage of separating dough color from chips was an improved signal for feedback control of the baking oven. Furthermore, surface appearance of chocolate chips was clearly a factor in consumer preferences.

In food inspection systems, both numeric and fuzzy values of physical features are useful. A cookie feature such as major axis length is useful in numeric format because a maximum limit is imposed by package size. However, fuzzy values for physical features are appropriate for representing variation in consumer preferences. In this paper, fuzzy values for size, shape, baked color, and chocolate chip coverage were defined by two methods. The system designers made *a priori* choices

TABLE I
COMPARISON OF MEMBERSHIP FUNCTIONS FOR CONSUMER RATINGS FOR FOUR FUZZY MODELS

Model	Number of membership functions	Membership functions for consumer ratings
ModelM-cent & ModelM-mom	8	*G1[19,33] G2[23,44] G3[23,50] G4[25,71] G5[22,77] E1[0,22] E2[0,29] E3[0,49]
ModelS	5	*S1[22] S2[34] S3[47] S4[68] S5[81]
ANFIS	12	*S1[-6] S2[12] S3[22] S4[31] S5[42] S6[46] S7[50] S8[52] S9[63] S10[79] S11[84] S12[98]

* G [σ,c] = Gaussian (σ = standard deviation (cm) c = mean (cm))
E [b,θ] = exponential (b = value (cm) where f (x) = 1 θ = decay constant (cm^{-1}))
S [c] = singleton (c = mean (cm))

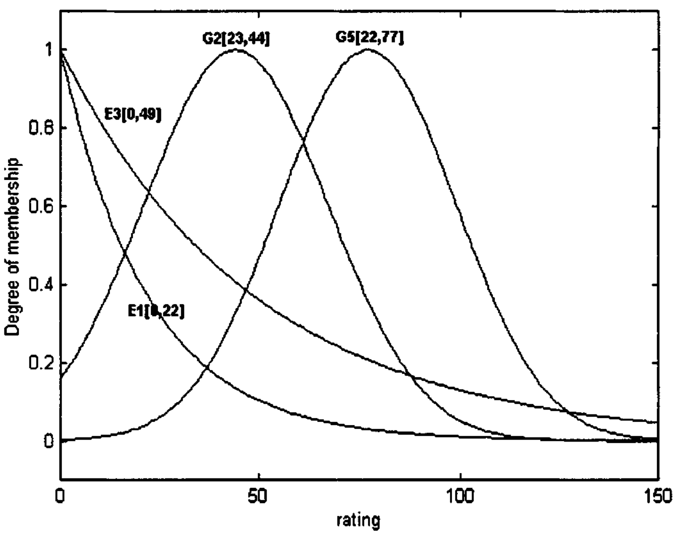

Fig. 4. Membership functions for two Gaussian (G2 and G5) and two exponential (E1 and E3) distributions for consumer ratings in Mamdani models.

for fuzzy values based on judgments about consumers' ability to differentiate. These fuzzy values were used to analyze sources of variation in the calibration poll data and to develop three of the fuzzy models for consumer preferences (ModelM-cent, ModelM-mom, ModelS).

The analysis of variance for the calibration polls was based on individual consumer ratings of each cookie, rather than average values for one or multiple polls. Not surprisingly, there was substantial variation in consumer ratings of a particular cookie. The model sum of squares in the ANOVA was significant relative to the error sum of squares however the correlation coefficient was low $(r^2 = 0.30)$ and the root-mean-square error (RMSE) was 26 cm. The following effects were not significant at the 95% confidence level: poll, block (within poll), shape, and the crossed effect of dough lightness and chip coverage. This statistical analysis was useful in identifying significant cookie features (size, baked color, and chocolate chips) as well as recognizing substantial variation in ratings among individual consumers. The type of distribution and distribution parameters for cookie ratings in the Mamdani systems were chosen to reflect variability in consumer responses.

The second approach to parameter estimation was to use the ANFIS algorithm to optimize the membership functions based on data from the calibration polls. The ANFIS membership functions are shown as dashed lines overlayed on the *a priori* definitions (solid lines) in Fig. 2 and there are substantial differences in terms of the extent of overlap of membership functions. In retrospect, the original choices were close to "crisp" classifications and did not reflect variations in consumer opinions. The ANFIS values, fitted from the calibration poll data, were better reflections of the range of consumer judgments of size and color as these factors influenced their ratings. The ANFIS membership functions for visible chips were quite different from the *a priori* definitions in terms of ranges and the overlap of "few" and "lots." The interval between 5–8% of the total area defines nonzero membership in the fuzzy value "lots," as well as membership grades of 1.0 in "few." Clearly, the visibility of chocolate chips influences consumer ratings and this bias is difficult to predict *a priori*. Data from calibration polls and a tool such as ANFIS were required in order to fit appropriate membership functions.

Four fuzzy models were developed to predict average consumer rating for a cookie based on three image attributes. Model predictions for the calibration cookies are compared in Table II on the basis of RMSE

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

where

y_i = observed value of average consumer rating
$\hat{y}_i$ = predicted value of average consumer rating.

RMSEs for three of the models (ModelM-cent, ModelS, and ANFIS) were equivalent based on the calibration polls. All three models used a weighted average or center-of-gravity method of defuzzification compared to mean-of-maximum (ModelM-mom). The mean-of-maximum method was biased

TABLE II
PREDICTION RESULTS FOR CALIBRATION SET AND TWO VALIDATION SETS

	Calibration Set (44 observations)	Validation Set 1 (14 cookies)	Validation Set 2 (30 cookies)
Model	RMSE (cm)	RMSE (cm)	RMSE(cm)
ModelM-cent	10	20	16
ModelM-mom	17	25	26
ModelS	9	23	19
ANFIS	9	17	14

TABLE III

	Consumer rating*	
Rule	**ModelS**	**ModelM**
If lightness is dark & size is small & chips is few	unacceptable - S1[22]	E1[0,22]
If lightness is dark & size is small & chips is lots	unacceptable - S1[22]	E2[0,29]
If lightness is dark & size is large & chips is few	unacceptable - S1[22]	G1[19,33]
If lightness is dark & size is large & chips is lots	unacceptable - S1[22]	E3[0,49]
If lightness is medium & size is small & chips is few	almost unacceptable - S2[34]	G1[19,33]
If lightness is medium & size is small & chips is lots	acceptable - S4[68]	G4[25,71]
If lightness is medium & size is large & chips is few	marginal - S3[47]	G2[23,44]
If lightness is medium & size is large & chips is lots	almost outstanding - S5[81]	G5[22,77]
If lightness is light & size is small & chips is few	marginal - S3[47]	G3[23,50]
If lightness is light & size is small & chips is lots	acceptable - S4[68]	G5[22,77]
If lightness is light & size is large & chips is few	marginal - S3[47]	G3[23,50]
If lightness is light & size is large & chips is lots	almost outstanding - S5[81]	G5[22,77]

*Membership functions defined in Table I.

to the cookie archetype with highest membership. Given the range of ratings for different combinations of features, it was better to use "democratic" methods such as center-of-gravity or weighted average defuzzification.

Certain cookies were used in several polls, and there was considerable variation in average ratings over the three polls. This is not surprising for evaluations by amateur judges. Cookie ratings reflected individual preferences (e.g., a preference for chocolate) and were influenced by factors such as time of day and comparisons to other cookies in poll set. Hence, predictions should not be interpreted as a precise numerical value, but should be associated with a range that reflects variability in consumer opinions.

A summary of the prediction results for the two validation polls is also presented in Table II. All four models had higher RMSE values for the validation sets relative to the calibration set. The ModelM-cent and ANFIS models were comparable in terms of RMSE values. It may be tempting to use additional polling results to recalibrate the models by adjusting membership functions. However the RMSE values for the validation sets were comparable to or lower than standard deviations estimated from distributions of individual ratings for each cookie in the consumer polls. Over the validation sets, RMSE was somewhat larger for ModelS than ANFIS values. The five singleton values for consumer ratings for ModelS (Table I) were chosen based on the mean ratings in the calibration polls. The ANFIS model had more parameters (i.e., 12 singleton values of consumer ratings) that covered a wide range of ratings, including a negative value which was outside the limits that panelists used for evaluations. The ANFIS model predictions were better in terms of RMSE.

IV. CONCLUSION

Although validation sets in this paper were small, the results indicated reasonable performances of two fuzzy models (ModelM-cent and ANFIS) for predicting average consumer ratings. RMSE values were comparable to variation observed in distributions of individual ratings. The fuzzy inference systems were simple in terms of number of rules and inference methods.

TABLE IV

Rule	Consumer rating
If lightness is ANFIS dark & size is ANFIS small & chips is ANFIS few	-6
If lightness is ANFIS dark & size is ANFIS small & chips is ANFIS lots	22
If lightness is ANFIS dark & size is ANFIS large & chips is ANFIS few	12
If lightness is ANFIS dark & size is ANFIS large & chips is ANFIS lots	46
If lightness is ANFIS medium & size is ANFIS small & chips is ANFIS few	31
If lightness is ANFIS medium & size is ANFIS small & chips is ANFIS lots	63
If lightness is ANFIS medium & size is ANFIS large & chips is ANFIS few	42
If lightness is ANFIS medium & size is ANFIS large & chips is ANFIS lots	84
If lightness is ANFIS light & size is ANFIS small & chips is ANFIS few	52
If lightness is ANFIS light & size is ANFIS small & chips is ANFIS lots	79
If lightness is ANFIS light & size is ANFIS large & chips is ANFIS few	50
If lightness is ANFIS light & size is ANFIS large & chips is ANFIS lots	98

The bakery is currently evaluating on-line vision sensors to augment the existing at-line vision system and these fuzzy techniques would be useful for automating quality assurance decisions (i.e., accept/reject decisions prior to packaging). There is also potential to use baked color (L^*) in a fuzzy control system for the oven. Currently, the baker makes adjustments to oven conditions based on his perception of color changes, as well as feedback from the at-line vision system.

One of the ultimate objectives of this research is to use digital images for automated inspection of food products. Our work with bakery products serves to illustrate a common problem in machine vision for food inspection. Feature extraction must be combined with a higher level of information processing that is capable of integrating features in a way that consumers make judgments. In the case of chocolate chip cookies, consumers are influenced by feature combinations. For example, small cookies are judged to be acceptable if there are "lots" of chocolate chips, but not with "few" visible chips. If acceptable limits are set on individual features, decisions to accept or reject product will be conservative (i.e., reject small cookies with "lots" of chips) or excessive and carry the risk of accepting cookies that consumers reject. Ultimately, intelligent system design will require both designer judgments (e.g., relevant image features and the number of fuzzy values for features) and calibration tools. Neurofuzzy techniques such as ANFIS will be useful for fitting parameters for membership values based on well designed consumer surveys.

APPENDIX

Fuzzy rules for Mamdani (ModelM) and Sugeno (ModelS) inference systems (membership functions for lightness, size and visible chips defined *a priori*) are shown in Table III.

Fuzzy rules for ANFIS inference system (membership functions for lightness, size, and visible chips defined from consumer polls) are shown in Table IV.

REFERENCES

[1] T. Chu, "Quality control system for a bakery product using fuzzy logic to predict consumer preferences," M.Sc. degree thesis, University of Guelph, 1999.

[2] E. A. Croft, C. W. de Silva, and S. Kurnianto, "Sensor technology integration in an intelligent machine for herring roe grading," *IEEE/ASME Trans. Mechatron.*, vol. 1, pp. 204–215, 1996.

[3] K. Ding and S. Gunasekaran, "Shape feature extraction and classification of food material using computer vision," *Trans. ASAE*, vol. 37, pp. 1537–1545, 1994.

[4] S. Gunasekaran, "Computer vision technology for food quality assurance," *Trends Food Sci. Technol.*, vol. 7, pp. 245–256, 1996.

[5] J.-S. R. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice-Hall, 1997, ch. 4, pp. 75–76.

[6] P. P. Ling, V. N. Ruzhitsky, and T. Lee, "Measuring the color of food," *Chemtech*, vol. 26, no. 3, pp. 46–53, 1996.

[7] I. Paulus, De Busscher, and E. Schrevens, "Use of image analysis to investigate human quality classification of apples," *J. Agric. Eng. Res.*, vol. 68, pp. 341–353, 1997.

Valerie J. Davidson received the B.Eng. degree from McMaster University, Hamilton, ON, Canada, in 1975, the M.Sc. degree from University of Guelph, Guelph, ON, Canada, in 1977, and the Ph.D. degree from the University of Toronto, Toronto, ON, Canada, in 1983.

She has been a Faculty Member in the School of Engineering, University of Guelph, since 1988. Previously, she was with Cambrian Processes Ltd. (1977–1979), Griffith Laboratories (1983–1986), and Ryerson Polytechnical Institute (1986–1988).

Joanne Ryks received the B.Sc. degree from the University of Guelph, Guelph, ON, Canada, in 1988 and the B.Ed.degree from Lakehead University, Canada, in 1994.

She is a Research Assistant at the School of Engineering, University of Guelph.

Terrence Chu received the B.Sc. (engineering) and M.Sc. degrees from the University of Guelph, Guelph, ON, Canada, in 1996 and 1999, respectively.

He is currently a Manager with the Technical Consulting Eloqua Corporation, Toronto, ON, Canada.