Image Texture Analysis for Discrimination of Mill Fractions of Hard and Soft Wheat

I. Y. ZAYAS[1,2] and J. L. STEELE[1]

ABSTRACT

Cereal Chem. 73(1):136–142

Digital image texture analysis was utilized to identify mill fractions from different mill streams and to assess wheat hardness differences. The study was conducted using a soft red winter wheat (Terra SR-87) and a hard red winter wheat (Thunderbird). Black and white images were acquired in a 256×256 pixel format to examine samples of coarse and fine mill fractions. Sixteen 64×64 pixel subimages per image were evaluated using texture analysis. Software was developed to calculate the image textural features used to develop the mill stream and hardness classification models. Several models based on image textural features were computed for different sets of subimages belonging to wheat of different hardness or mill stream. Recognition of hard wheat vs. soft wheat was achieved with 100% correct recognition rate for each mill fraction when a three-feature model was used for pairwise analysis. Different mill fractions of the same wheat, coarse vs. fine, were similarly discriminated with 100% accuracy for each pairwise comparison. All four mill fractions were successfully recognized with 100% correct recognition rate when a three feature model was used for four class analysis. The wheat class and mill fraction discrimination was achieved with <3 g (≈0.2 g/subimage) of material.

This study addressed the problem of distinguishing differences among wheat flour samples of different wheat cultivars and mill streams. Digital image analysis was first applied to grain in the early 1980s. Efforts to reduce subjective judgments in the wheat grading and classification process using new image analysis methods are reflected in several studies. At the U.S. Grain Marketing Research Laboratory (USGMRL), several image analysis studies were directed toward grain classification and grading and were primarily based on morphometrical descriptors of single kernels (Zayas et al 1985, 1986, 1990). Combinations of morphometrical descriptors using multivariate statistical analysis for pattern recognition showed potential for wheat cultivar and class recognition.

Wheat hardness and mill fraction differences were studied with an image texture technique applied to images of mill fractions. Distinguishing mill fractions of wheats of different hardness or mill fractions of different particle size in different mill streams is important for the milling industry and breeding of wheat. During 1988–1989, image texture analysis was applied to bulk samples of various mill fractions to evaluate the potential for discrimination of two wheat cultivars, one soft and one hard wheat, and also for discrimination of mill fraction or samples from different mill streams of the same wheat (Zayas et al 1989). The same image texture methodology was successfully applied to a much larger study of 17 wheat cultivars of crushed wheat kernels and successfully distinguished hard and soft cultivars (Zayas et al 1991). More detailed description of image texture analysis can be found in Gonzales et al (1987).

Inhouse, PC-based software was developed to study the image texture methodology in greater depth. The PC-based software was designed for study of various quantifications of the 0–255 gray scale or gray level sampling methods. A series of sampling methods were studied to determine recognition sensitivity of developed image texture methodology. The study objectives were to determine optimal gray level sampling procedures before computing image textural features for two purposes: 1) maximize discrimination of classes of wheat (hard and soft); and 2) maximize discrimination of mill fractions of wheat (coarse and fine).

[1]Electronics Engineer and Research Leader, U.S. Grain Marketing Research Laboratory, USDA, ARS, Manhattan, KS 66502.
[2]Corresponding author.

Publication no. C-1996-0115-04R.

MATERIAL AND METHODS

Image Analysis System

For image acquisition and analysis, a Kontron Image Processing System (IPS), (Kontron Bildanalyse, Munich, Germany) was used. A SUN-3/160C work station with a Unix operating system, 8Mb RAM memory, 1,068Mb disk storage, a 60Mb tape backup, and a high-resolution 19-in. color monitor served as the host system. The system included a standalone, microprogrammable 10 MIPS pipeline array processor. The images were acquired using a DAGE MTI-81 black and white Newvicon camera (DAGE-MTI, Inc., Michigan City, IN) with a resolution of 1,600 lines and a Mikro-Nikkor 55mm, f/2.8 lens with aperture setting of f/11. Images were then acquired in a 256×256 pixel format, 256 gray levels, and with a magnification resulting in 1 pixel = 0.08675 mm. Four 250-watt halogen lamps illuminated the field of view and provided a color temperature 3020 K. An analog-to-digital converter operated at 20 MHz and provided 8-bit gray tone resolution (256 gray levels). The acquired images were stored for subsequent processing.

Sample Preparation and Image Acquisition

Samples of two wheat cultivars, a soft red winter (SRW) wheat, Terra SR 87, and a hard red winter (HRW) wheat, Thunderbird, were tempered to 15% moisture content and milled on an Allis-Chalmers Experimental Mill (Norwood Works, OH). The experimental mill used two types of rolls: corrugated break rolls that open up the wheat kernel to remove endosperm from bran, and reduction rolls, which are usually smooth and reduce endosperm particles to flour. Samples (15 g) of the 4th break fraction, coarse and fine, were obtained for this study. Coarse bran was that portion of the 4th break that did not pass through a 26W sieve and fine bran was that portion passing through the 26W sieve and retained on a 54S sieve. Apertures for the 26W and 54S sieves were 0.787 mm (0.0310 in.) and 0.368 mm (0.0145 in.), respectively.

For image acquisition and texture analysis, each of the four wheat fractions was poured onto black velvet paper, so that the paper was completely covered over an area that filled the camera field of view (70 × 70 mm). Four repetitions of each sample presentation were used. The depth of the layer was ≈2–3 mm and the material covering the field of view weighed ≈3 g.

Image Texture Analysis

Figure 1 is a typical representation of the images obtained. Each image (256×256) was subdivided into 16 subimages. Each subimage represented ≈0.2 g of material and a mill fraction image (3 g of wheat) contained 16 subimages (observations). For image texture analysis, inhouse software was written to determine various gray level ranges, compute co-occurrence matrices and several image textural features. The image textural features were computed from the co-occurrence matrices based on the Haralick's definitions (Haralick et al 1973). The image textural features characterize the spatial distribution of gray levels in an image or subimage region. The image textural features were then used in discriminant analysis to classify mill fractions by hardness and by mill stream.

Image texture can be described in terms of tonal primitives or regions which are spatially organized in patterns. Image texture implies smoothness, coarseness, fineness, granulation, irregularity, etc., which may relate to the spatial or regional gray scale patterns. Coarse textures are those for which the gray scale changes only slightly with distance, and fine textures are those for which the gray scale changes rapidly with distance. For example, gray scale changes occurred more rapidly for fine bran than for coarse bran and for the hard wheat compared to soft wheat (Fig. 1).

The inhouse software was used to preprocess the images to determine gray level ranges, reduce the number of gray levels from 256 to eight and to reduce computation time. The original 256 gray levels were mapped to eight gray levels, each having 12.5% of the original gray level distribution. To study gray level sampling procedures, the samples of different hardness and samples from different mill streams were organized into seven sets as shown in Table I. Each data set had different gray scale distributions which affected the limits of each of the eight gray level ranges. The software computed the subimage textural features based on the different ranges of gray levels, which were extracted from different sets of subimages. A block diagram of data organization to determine the different gray level sampling ranges is shown in Figure 2. The goal was to determine which gray level range (combination of samples) would give better recognition of hard-soft or coarse-fine mill fractions. For example, would the textural features extracted from the gray level ranges within coarse and within fine fractions for pooled hard and soft wheats (data set **wqbb5**), recognize a hard coarse and a soft coarse fraction at the same recognition rate?

The first gray level sampling data set (**wqbb1**) provided image textural features based on the individual subimage gray levels only. The second set (**wqbb2**) comprised image textural features based on ranges of gray values determined for each image. The image textural features for the third set (**wqbb3**) were based on gray levels extracted from each fraction of each wheat class; for example, the coarse fraction of soft wheat. The fourth set (**wqbb4**) of gray level ranges were determined from the images of coarse and fine wheat within each wheat class; for example, coarse and fine fractions of soft wheat. In the fifth set (**wqbb5**), textural features were extracted from images of the same mill stream; for example, coarse fractions of the hard and soft wheats. In the sixth set (**wqbb6**), textural features were extracted from all images, hard-soft and coarse-fine. In the seventh set (**wqbb7**), textural features from all images were based on eight fixed gray levels rather than equal percentage distributions.

The image texture statistics are represented in the spatial gray level co-occurrence matrices that represent the probabilities of two gray levels occurring in certain vector directions or in the same relative position. The co-occurrence matrices were determined using the indicated gray level ranges and a neighborhood window size of 9×9 pixels. This window size produced 40 spatial vector directions and co-occurrence matrices for any subimage under analysis. Only the upper right vector directions were used to determine the co-occurrence probabilities. The shifting distance was one pixel.

After preprocessing as described above, the software extracted tonal features such as statistical measures and mean and variance of eleven image textural features (angular second moment, etc.). These textural features are commonly used and described in the early work of Haralick et al (1973). The mean and variance of these features were computed for each subimage and gray level sampling procedure for each wheat class and mill stream fraction.

Abbreviations for the extracted features were defined as: 1) GAVG - mean gray value of subimage; 2) GVAR - variance of gray value of subimage; 3) ASMM - angular second moment mean; 4) CONTM - contrast mean; 5) CORM - correlation mean; 6) VARM - variance mean; 7) IDMM - inverse difference moment mean; 8) SAVGM - sum average mean; 9) SENTM - sum entropy mean; 10) ENTM - entropy mean; 11) DVARM - difference variance mean; 12) DENTM - difference entropy mean; 13) SVARM - sum variance mean.

The variances of the same features (3–13 above) over the 40 image planes were also determined and defined as: ASMV, CONTV, CORV, VARV, IDMV, SAVGV, SENTV, ENTV, DVARV, DENTV, SVARV.

RESULTS AND DISCUSSION

Coarse and fine mill fractions of a soft red winter wheat (Terra SR-87) and a hard red winter wheat (Thunderbird) were evaluated

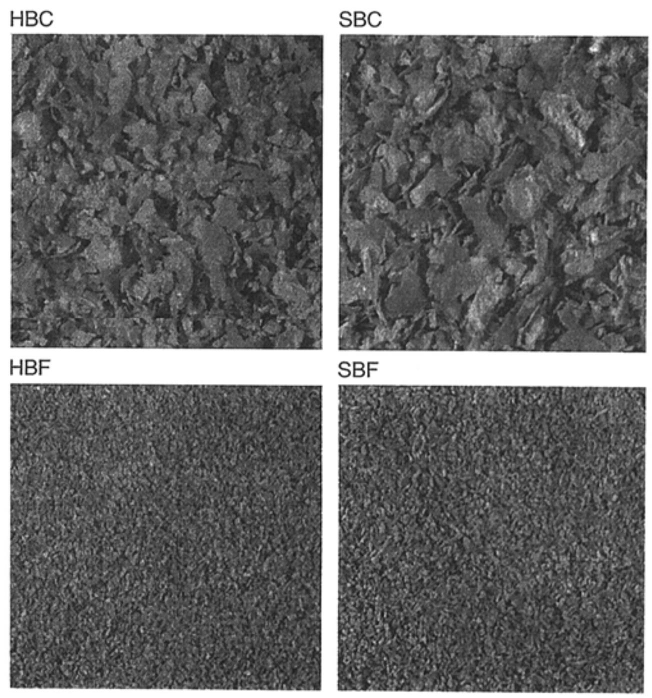

Fig. 1. Images of mill fraction samples. HBC = hard bran coarse, SBC = soft bran coarse, HBF = hard bran fine, SBF = soft bran fine.

TABLE I
Organization of Different Gray Level Sampling Data Sets

Data Set	Quantizing Within
WQBB1	Subimage $2 \times 2 \times 4 \times 16$
WQBB2	Image $2 \times 2 \times 4$
WQBB3	Class/Fraction 2×2
WQBB4	Class 2×1
WQBB5	Fraction 1×2
WQBB6	Data Set 1×1
WQBB7	(Gray+1)/32

using image textural features based on selected gray level ranges. The discrimination power of the textural features was evaluated using a statistical (SAS 1993) ranking procedure (STEPDISC) and by visual evaluation of graphical representations of sample clustering. The most effective features were used for further analysis. Several single textural features distinguished the coarse and fine mill fractions with a 100% correct recognition rate. The above was obtained for all subimages and the gray level sampling ranges represented by **wqbb1**, **wqbb2**, **wqbb3,** and **wqbb4**. Image textural features, CORM and VARM, successfully distinguished coarse from fine mill fractions (SBC vs. SBF) of soft wheat for the **wqbb4** sampling set, as shown in Figure 3. The same was true for coarse vs. fine fraction (HBC vs. HBF) of hard wheat and the **wqbb4** set. Hard fractions were not distinguished from soft fractions by any single feature from any data set, though clustering was observed for several single image textural features.

Because single textural features did not distinguish hard from soft wheat fractions, multivariate analysis was investigated. The CANDISC and DISCRIM (SAS 1993) procedures were used to perform the multivariate analysis and determine the best performing model. The CANDISC procedure computes canonical values that can be used to determine and visualize class separation. The DISCRIM procedures were used to determine correct recognition rates for calibration and test data sets. Discriminant analysis was used for pairwise comparison—a two class problem and for all individual fractions comparison—a four class problem. For more details on multivariate analysis see Morrison (1985) and Zayas et al (1990).

CANDISC and DISCRIM procedures were used to determine which of the gray level sampling procedures was more effective for class recognition. The 13 image textural features comprised the feature vector: GAVG, GVAR, ASMM, CONTM, CORM, VARM, IDMM, SAVGM, SENTM, ENTM, DVARM, DENTM, SVARM. Though all seven data sets resulted in high recognition rate for pairwise analysis of hard vs. soft fractions, some data sets demonstrated better recognition and clustering than others, as shown in the canonical plots in Figure 4. Results of pairwise discriminant analysis using 13 feature and three feature models for all gray level sampling data sets are shown in Table II. Recognition of hard vs. soft mill fractions was noticeably better for the gray level sampling data sets **wqbb2**, **wqbb3** and **wqbb4** (100% correct recognition). For the other sampling procedures, the hard vs. soft class fractions were discriminated with recognition rates less than 100%. Recognition of coarse vs. fine was 100% correct with a high degree of separation between clusters for all **wqbb1–wqbb7** data sets (Fig. 5 and Table II).

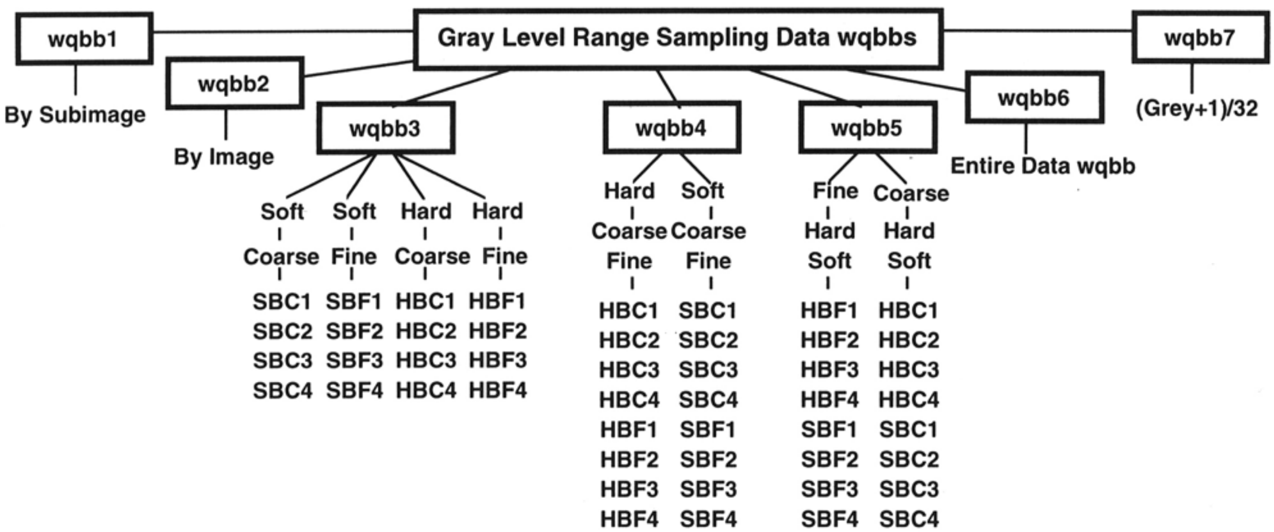

Fig. 2. Block diagram of data organization for gray level sampling. Mill fractions: HBC = hard bran coarse, SBC = soft bran coarse, HBF = hard bran fine, SBF = soft bran fine.

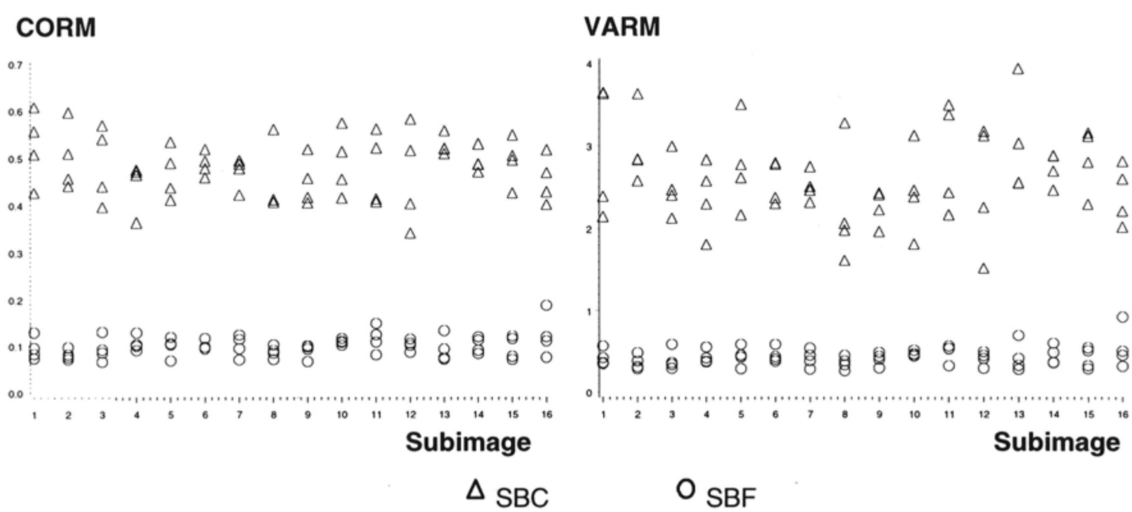

Fig. 3. Plots of individual image textural features (correlation mean [CORM] and variance mean [VARM]) for pairwise mill fraction identification of soft bran coarse(SBC) vs. soft bran fine (SBF) using gray level sampling set **wqbb4**.

Results for a three-feature model are also shown in Table II. Correct recognition rates of 100% were achieved with a three-feature model consisting of CORM, GAVG and SAVGM for data sets **wqbb3** and **wqbb4**. These three features were ranked by STEPDISK as the most powerful. The three-feature model recognition rates were lower than 100%, from 63 to 92%, for some gray level sampling data sets, but not for **wqbb4** and **wqbb3**.

The **wqbb4** data set showed the best performance for distinguishing mill fractions with the least number of features in the model. The next best sampling data set, **wqbb3**, showed 100% recognition of each individual mill fractions for pairwise analysis with a minimum nine feature model. Further analysis was done using data set **wqbb4**. The canonical coefficients for the three feature model and pairwise analysis for **wqbb4** are shown in Table III. Table IV shows minimum, maximum, mean, and standard deviation of all canonical values for pairwise analysis of mill fractions. In Figure 4, the clusters of SBC subimages were separated from the SBF cluster of subimages by CAN1. The SBC boundaries are defined from CAN1 = 1.67 to 7.05, and for the SBF subimages, CAN1 = - 5.20 to –2.56. Figure 5 illustrated perfect mill fraction separation for all data sets.

The performance of the chosen model was tested by training and validation by testing with unknown observations. The data set, 64 observations (subimages) per each mill fraction, was split

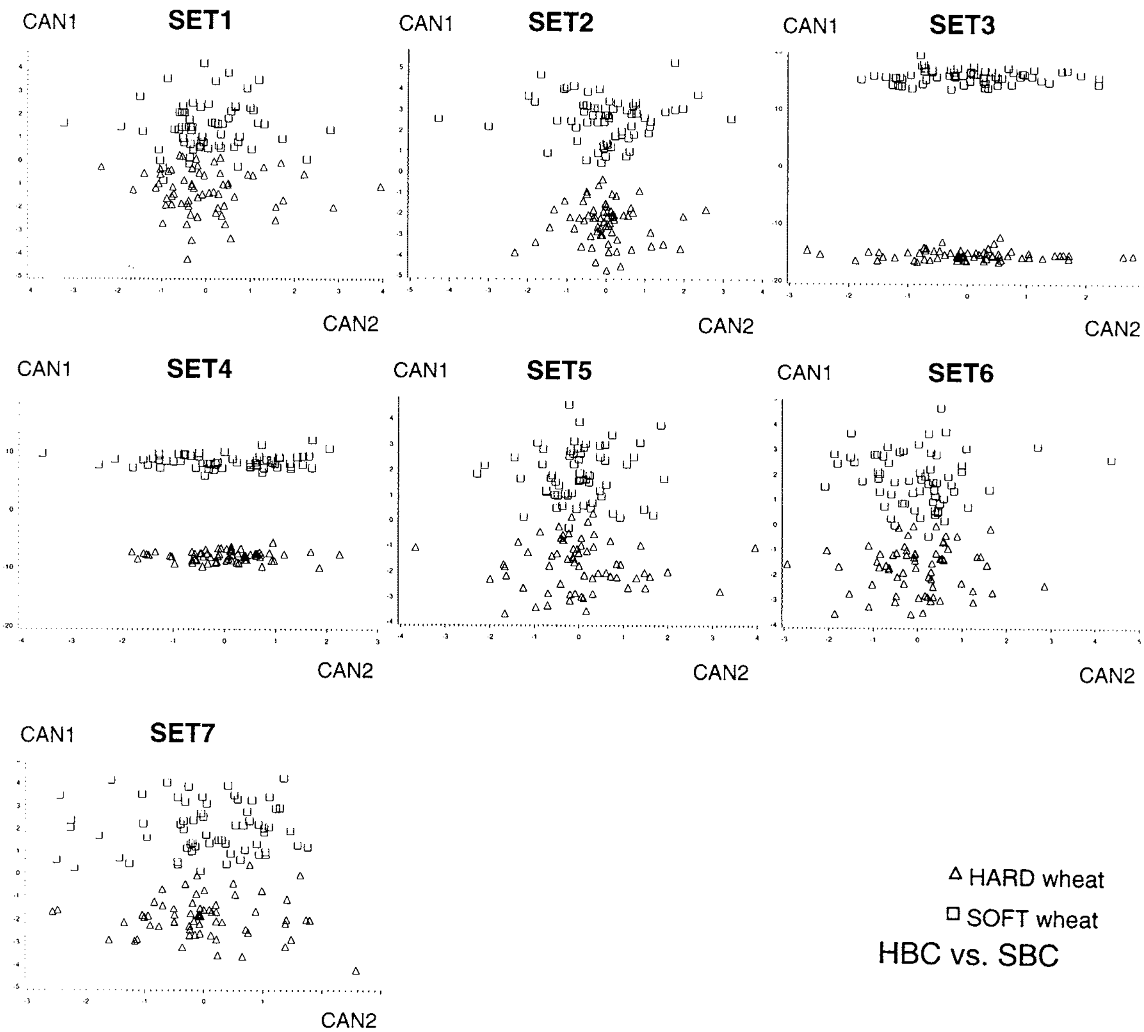

Fig. 4. Canonical pairwise classification of different gray level sampling sets by hardness in the coarse fraction (a hard bran [HBC] vs. soft bran coarse [SBC]) for 13-feature model.

TABLE II
Pairwise Classification (%) of Coarse vs. Fine and Hard vs. Soft Mill Fractions with Different Gray Level Sampling Data Sets and Different Models

	WQBB1		WQBB2		WQBB3		WQBB4		WQBB5		WQBB6		WQBB7	
MF[a]	13f[b]	3f[c]	13f	3f	13f	3f	13f	3f	13f	3f	13f	3f	13f	3f
HBC	100	100	100	100	100	100	100	100	100	100	100	100	100	100
HBF	100	100	100	100	100	100	100	100	100	100	100	100	100	100
HBC	94	69	100	63	100	100	100	100	97	92	100	92	98	63
SBC	92	100	100	100	100	100	100	100	98	90	97	89	100	100
HBF	100	78	100	100	100	100	100	100	100	77	100	77	100	83
SBF	92	84	100	89	100	100	100	100	98	83	98	88	98	83
SBC	100	100	100	100	100	100	100	100	100	100	100	100	100	100
SBF	100	100	100	100	100	100	100	100	100	100	100	100	100	100

[a] Mill fractions: HBC = hard bran coarse, SBC = soft bran coarse, HBF = hard bran fine, SBF = soft bran fine.
[b] 13f = 13-feature model (see Fig. 6).
[c] 3f = three-feature model (see Fig. 6).

randomly into two sets, one for calibration and the remainder for validation testing. The four class analysis results for calibration and validation showed 100% recognition of all fractions when a three-feature (CORM, GAVGM, SAVGM) model was used (Table V). A two-feature model (CORM, GAVG) showed lower four class recognition rates, 66–100%, for the **wqbb4** gray level sampling data set.

The effectiveness of different numbers of features in a model is shown in the plots of canonical functions for data set **wqbb4** with 10 features: CORM, GAVG, SAVGM, ASMV, IDMV, IDMM, DENTV, CORV, SVARM and ENTM; and with three features: CORM, GAVGM and SAVGM (Fig. 6). Even though the three-feature model achieved 100% recognition of all fractions, the 10-feature model demonstrated much tighter clusters with larger separation distances. Coarse and fine fractions were distinguished by CAN1. Separation of hard and soft fractions was predominant by the value of CAN2. For the procedure used, CAN2 was indicative of hardness and CAN1 was indicative of particle size.

Pairwise analysis is important for feedback and automatic control of mill operations. Training of a computer controlled mill system using samples of hard vs. soft wheat or coarse vs. fine fractions would allow mill operation assessment. Nevertheless, recognition of all fractions is of importance. The potential of the software to indicate changes in hardness of a wheat mill stream is important for feedback for an automated mill. Consistency of mill fraction or mill stream output is also important. The capability of the procedure to indicate differences in mill fractions (different particle size), showed a potential for feedback and milling process automation. Once the computer is trained to recognize certain fractions, the software will give indication of conformity of material from the same stream compared to that for the samples used for training the machine. The use of image textural features to differ hard vs. soft could also be of practical value to breeders because class differentiation was possible on very small samples (≈0.2 g). The second study (Zayas et al 1991) was done, using the same methodology, for identification 17 wheat cultivars from six wheat classes. Crushed wheat kernels, which looked similar to coarse bran fraction, were used in that study. The result of the study of crushed wheat proved effectiveness of developed methodology for distinguishing wheat cultivars by hardness.

SUMMARY

An exploratory study was conducted to develop an objective method for assessing mill fractions of hard and soft wheat using digital image texture analysis and statistical multivariate discrimination techniques. A black and white camera was used to acquire digital images of two mill fractions, coarse and fine, of two wheat cultivars, a HRW wheat (Thunderbird) and a SRW wheat (Terra SR-87). Seven data sets organized by different gray level sampling procedures were studied. The best coarse-fine and hard-soft separations were achieved with gray level sampling

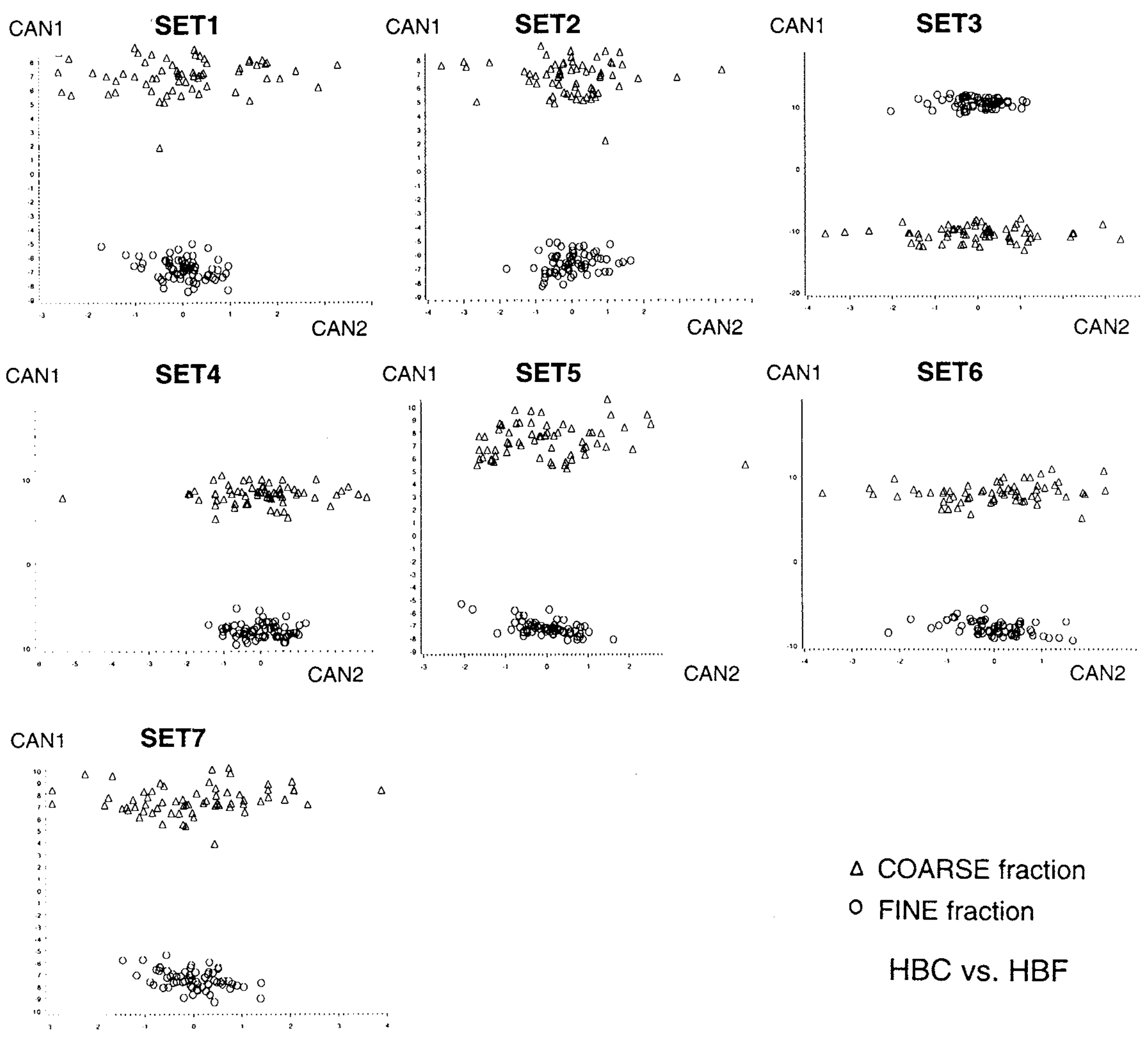

Fig. 5. Canonical pairwise classification of different gray level sampling sets by mill fraction in the hard class (hard bran coarse [HBC] vs. hard bran fine [HBF]) and a 13-feature model.

based on two fractions combined coarse and fine within each hard and soft variety (**wqbb4**). Some single-image textural features, like correlation mean, contained sufficient information to differentiate coarse and fine mill fractions for some gray level sampling procedures. Recognition of hard vs. soft wheat was achieved with 100% correct recognition rate for each fraction studied when discriminant analysis was used for pairwise distinction with three image textural feature (CORM, GAVG, SAVGM) model. Coarse and fine mill fractions of the same wheat were similarly discriminated with 100% accuracy for each pairwise comparison. The three-feature model successfully recognized the four mill fractions studied with a correct recognition rate of 100%.

TABLE III
Canonical Function Coefficients for Pairwise Discrimination of Wheat Mill Fractions[a] for a Three-Feature Model[b] and Data Set wqbb4

	Features	Can1	Can2
HBC vs. SBC			
	CORM	−1.17	−0.22
	GAVG	0.51	−0.07
	SAVGM	−3.48	0.99
HBC vs. HBF			
	CORM	26.82	2.33
	GAVG	−0.19	0.13
	SAVGM	0.55	−0.45
HBF vs. SBF			
	CORM	16.04	43.03
	GAVG	3.61	0.10
	SAVGM	−15.38	0.71
SBC vs. SBF			
	CORM	21.27	−5.72
	GAVG	−0.20	0.18
	SAVGM	1.28	0.42

[a] HBF = hard bran fine, SBF = soft bran fine, HBC = hard bran coarse, SBC = soft bran coarse.
[b] GAVG = gray level mean, SAVGM = sum average mean, CORM = correlation mean.

TABLE IV
Canonical Values for Pairwise Classification Using a Three-Feature Model and Data Set wqbb4

Bran Fraction[a]	Features	Minimum	Maximum	Mean	SD
HBC	Can1	−6.05	−0.80	−2.90	1.09
	Can2	−2.45	1.74	0.00	1.29
SBC	Can1	0.99	4.79	2.90	0.89
	Can2	−1.24	1.27	0.00	0.58
HBC	Can1	1.92	9.02	4.81	1.34
	Can2	−2.76	1.08	0.00	1.41
HBF	Can1	−5.80	−3.94	−4.81	0.44
	Can2	−0.14	0.12	0.00	0.06
SBC	Can1	1.67	7.05	4.34	1.33
	Can2	−2.51	2.74	0.00	1.21
SBF	Can1	−5.20	−2.56	−4.34	0.49
	Can2	−1.56	1.64	0.00	0.73
HBF	Can1	−19.86	−15.84	−17.70	0.87
	Can2	−2.42	2.14	0.00	0.98
SBF	Can1	14.46	19.73	17.70	1.12
	Can2	−1.92	3.12	0.00	1.02

[a] HBF = hard bran fine, SBF = soft bran fine, HBC = hard bran coarse, SBC = soft bran coarse.

TABLE V
Discriminating Four Classes (% of HBC, HBF, SBC, and SBF)[a] Using the wqbb4 Data Set and Two- and Three-Feature Models for Calibration and Validation for Set wqbb4

	Two-Feature Model[b]				Three-Feature Model[b]			
	HBC	HBF	SBC	SBF	HBC	HBF	SBC	SBF
Calibration	78	78	100	72	100	100	100	100
Validation	66	72	100	72	100	100	100	100

[a] HBF = hard bran fine, SBF = soft bran fine, HBC = hard bran coarse, SBC = soft bran coarse.
[b] GAVG = gray level mean, CORM = correlation mean.
[c] SAVGM = sum average mean; GAVG and CORM.

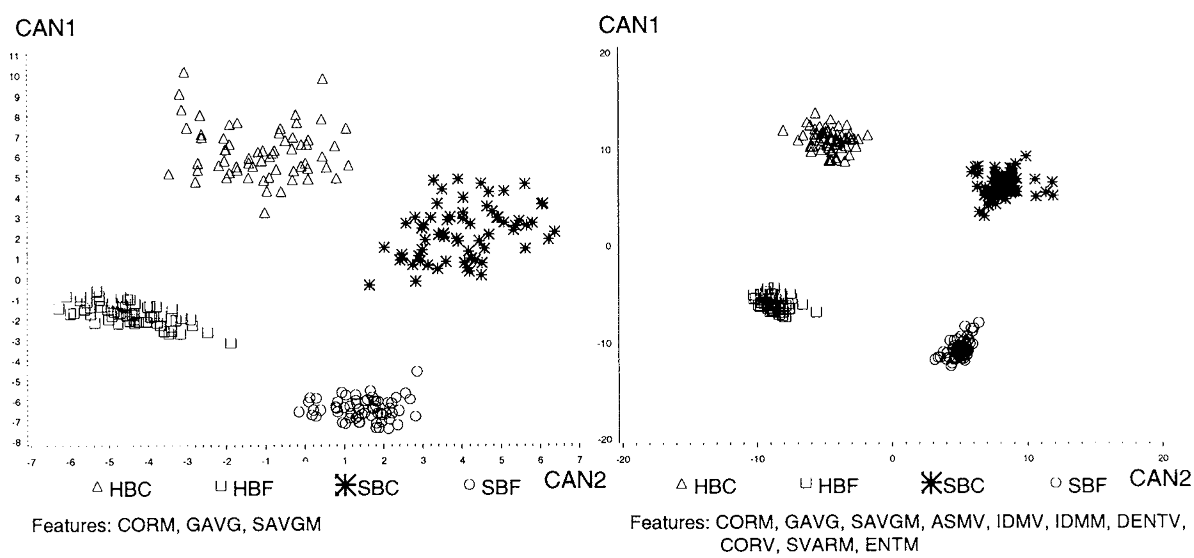

Fig. 6. Comparison of a three-feature model (correlation mean [CORM], mean gray value of subimage [GAVG], sum average mean [SAVGM]) and a 10-feature model (correlation mean [CORM], sum average mean [SAVGM], angular second moment variance [ASMV], inverse difference moment variance [IDMV], inverse difference moment mean [IDMM], difference entropy variance [DENTV], correlation variance [CORV], sum variance mean [SVARM], and entropy mean [ENTM]) for data set **wqbb4** and canonical discrimination of four mill fractions (hard bran coarse [HBC], hard bran fine [HBF], soft bran coarse [SBC], and soft bran fine [SBF]).

The method and software developed used conventional image techniques and can be used in several ways to assess mill fractions for wheat class or mill performance. The method developed has potential in grading systems, end-use quality assessment and automated mill control. Because the method uses a very small sample, it could be utilized for early generation evaluation by breeders.

ACKNOWLEDGMENTS

We thank L. Bolte for sample preparation, D. Walker for software development, and K. Buck for assistance.

LITERATURE CITED

GONZALEZ, R., and WINTZ, G. P. 1987. Digital Image Processing. 2nd ed. Addison-Wesley: Reading, MA.

HARALICK, R. M., SHANMUGAM, K., and DINSTEIN, I. 1973. Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics, 3(6):610-621. IEEE: New York.

MORRISON, D. F. 1985. Multivariate Statistical Methods. McGraw-Hill: New York.

SAS. 1993. User's Guide. SAS Institute Inc.: Cary, NC.

ZAYAS, I. Y., POMERANZ, Y., and LAI, F. S. 1985. Discrimination between Arthur and Arkan wheats by image analysis. Cereal Chem. 62:478.

ZAYAS, I. Y., LAI, F. S., and POMERANZ, Y. 1986. Discrimination between wheat classes and cultivars by image analysis. Cereal Chem. 63:52.

ZAYAS, I. Y., STEELE,. J. L., DEMPSTER, R. E., and BOLTE, L. C. 1989. Image analysis for texture pattern recognition of hard and soft wheat brans. Cereal Foods World 34:751.

ZAYAS, I. Y., CONVERSE, H., and STEELE, J. L. 1990. Discrimination of whole from broken corn kernels with image analysis. Trans. ASAE 33(5):1642-1646.

ZAYAS, I. Y., MARTIN, C. R., STEELE, J. L., and DEMPSTER, R. E. 1991. Image texture of crushed wheat., Machine Vision Architectures, Integration and Applications. Proc. SPIE 1615:203-215.

[Received March 17, 1995. Accepted September 15, 1995.]