Discrimination Between Wheat Classes and Varieties by Image Analysis[1]

I. ZAYAS,[2] F. S. LAI,[3] and Y. POMERANZ,[3]

Cereal Chem. 63(1):52–56

ABSTRACT

Image analysis was used on-line to discriminate variables of grain morphology to differentiate among individual kernels of four hard red winter (HRW) and four soft red winter wheat cultivars, and between two HRW and two hard red spring wheat cultivars. Individual kernels and sets of kernels were differentiated using software developed in our laboratories. The main objective was to differentiate rapidly, using a minimum of critical and simple parameters. Differentiation was based on the development of canonical equations for comparing kernels on the basis of six parameters: width, length, length ratio, tangent, sine, and length of arc of parabolic segment. The program was verified by SAS multivariate discriminant analysis using the same variables extracted directly from the image of a kernel. The method was also used for discrimination of mixtures of wheat varieties from the three classes. For mixtures of three varieties of either HRW or of soft red winter wheat, the average percentages of correctly classified kernels were 85% for training set and 83% for experimental set samples. For mixtures of two varieties of either HRW or of hard red spring wheat, the average percentages were 78% for calibration samples and 77% for test samples.

Differentiation among wheats of various classes and varieties by the Federal Grain Inspection Service of the U.S. Department of Agriculture is based primarily on grain morphology (Anonymous 1957, 1984). We have recently reported on the use of image analysis to distinguish between morphological characteristics of the soft red winter (SRW) cultivar Arthur (C.I. #14425) and the hard red winter (HRW) cultivar Arkan (C.I. #47577) (Zayas et al 1985). Arkan, a cross between Arthur and the HRW cultivar Sage (C.I. #17277), has some of the morphological characteristics of its SRW parent, yet it possesses the milling and baking characteristics of its HRW parent (Martin et al 1983). The ambiguous nature of this variety has presented difficulties in classification according to conventional grading methods. In addition, problems have been encountered by grain inspectors in differentiating by visual inspection between HRW and hard red spring (HRS) wheats on the basis of grain morphology. Image analysis is an attempt to address these problems by focusing on other more fundamental aspects of variation in grain varieties.

The purpose of using discriminant analysis in this project was to establish a standard for placing observations into one of two classes (for example, HRW versus SRW wheat) by determining those variables that best characterize structural (morphological) differences between the two classes. The variables available in this study included width, length, length ratio, width ratio, tangent, sine, length of a parabolic segment, minimum grey level, maximum grey level, and average grey level. Of these, we chose to use all but the factors of color (grey levels) and width ratio in our analysis. The color factors were deleted because of the difficulties presented in interpreting the meaning of color variations. Also, because the width ratio is influenced by a factor known to be irrelevant, i.e., the position of kernel growth on either the left or right hand side of a wheat head, this variable was rejected.

This report describes the use of image analysis in differentiating between several SRW and HRW cultivars and several HRW and HRS cultivars. In addition, differentiation between the three major wheat classes by image analysis was studied.

[1]Cooperative investigations, U.S. Grain Marketing Research Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Manhattan, KS 66502, and Department of Chemical Engineering, Kansas State University, Manhattan 66506. Contribution 86-82-J, Department of Chemical Engineering, Kansas Agricultural Experiment Station, Manhattan 66506.

Mention of firm names or trade products does not constitute endorsement by the U.S. Department of Agriculture over others not mentioned.

[2]Research assistant, Department of Chemical Engineering, Kansas State University.

[3]Research chemical engineer and research chemist, respectively, U.S. Grain Marketing Research Laboratory.

MATERIALS

The HRW wheat varieties included NK 812, Centurk (vitreous and nonvitreous), TAM 101, Roughrider, and Arkan (two samples); the SRW wheat varieties were Coker 745, Southern Bell, Double Crop, and Arthur (two samples); the HRS wheats were Oslo and Era. All samples were obtained from the USDA Federal Grain Inspection Service. Broken, shrunken, immature, and very small kernels (below about 20 mg) were separated and not examined. Only sound kernels were used; damaged kernels (scabby, moldy, sick, black point, etc.) were rejected.

METHODS

The basic diagram of the on-line discriminant image analysis is given in Figure 1. Data in the form of video signals were sent to a processor via a vidicon scanner and macroviewer. Before an image was detected and digitized, the shade corrector measured and stored the shading pattern for a blank field as a reference that was used to correct the shading of subsequent fields. The analyzer could make field-specific and feature-specific measurements. The field-specific parameters available included area, parameter, count, horizontal and vertical projection, and feret diameter at 0, 45, 90, and 135 degrees.

In this study, we used the feature-specific parameters of length, length ratio, width, length of parabolic segment, and sine and tangent of angle (Fig. 2) to characterize individual kernels. The measurement field is usually restricted by the standard line frame to 800 × 625 picture points, leaving a surrounding guard region to eliminate edge errors. We used a rectangle frame of 650 × 350 picture points to obtain high resolution. A typical image of a wheat kernel, in the form of a matrix of alphanumeric characters, displayed by a PDP-11 monitor (host computer), is given in Figure 3.

During the initial step of converting an image into digital representation, the accuracy of detection of the superimposed bright-up (detected) image over the original one on the screen of the image analyzer was manually controlled by the operator. A Digital Equipment Corporation PDP-11-03 computer stored data on a dual hard disk. An LA36 Decwriter served as the computer terminal and also provided hard copy printouts. In addition, a CRT monitor permitted software control of the image analyzer. We developed a software program, ARTARK, to perform the discriminant analysis. After data from each of the six variables were provided to the computer, the ARTARK program presented four options: 1) build a set, 2) run discriminant analysis on a training set, 3) run discriminant analysis on an experimental set, and 4) read an image from the screen and perform a discriminant analysis.

Option 1

In this option, a set was formed from stored kernel images. The set depended on the objectives of the analysis to differentiate among kernels from two or more varieties or kernels from two or more classes. The extracted variables, such as length, width, tangent of kernel tip, etc., were stored as a compressed image data file.

Option 2

This option allowed the operator to perform a discriminant analysis of the calibration set using an SSP (scientific subroutine package) from Digital Equipment Corp. This subroutine also enabled the operator to select which variables to include in the discriminant analysis. In this study, we selected the six variables indicated before. We also calculated the means for each variable selected and the pooled dispersion (covariance) matrix for every group. Finally, a linear function of indices to classify each group was established.

Option 3

Using this option, the operator could perform a discriminant analysis of the stored unknown experimental set by referring to the function calculated in Option 2 on a training set.

Option 4

With this option, the operator could read an image from the screen and perform a discriminant analysis in real time as outlined in Option 3.

The software calculation procedures are outlined below:

1. Inputs: N image files in two or more groups. Outputs: one file of compressed image data.
2. Inputs: file of compressed image data (from step 1). Outputs: discriminant functions and the percentage of correct identifications of training sets.
3. Inputs: discriminant functions (from step 2) and file of compressed data. Outputs: the percentage of correct identifications of experimental sets.
4. Inputs: discriminant functions (from step 2) and image of the unknown kernel from the screen. Outputs: identification of wheat kernel as discriminated by functions from step 2.

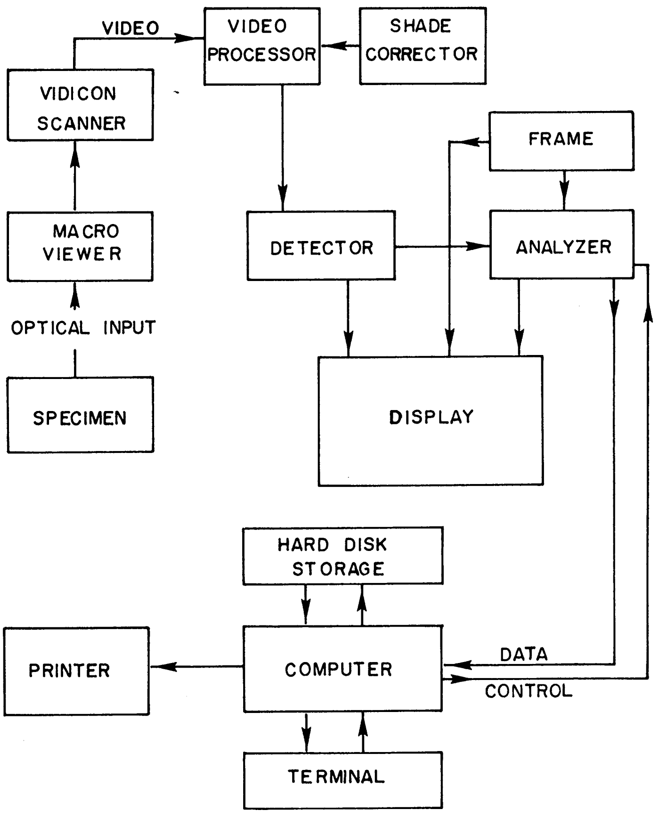

Fig. 1. Block diagram of an on-line discriminant image analyzer.

In the previous study (Zayas et al 1985) we used multivariate discriminant analysis based on nine parameters (from two orthogonal projections) to differentiate between Arthur and Arkan. We have modified the previous method in order to speed up the discrimination process and to obtain measurements on unknowns in real time (i.e., within several seconds). In the modified process, each kernel was analyzed on the image transfer mode of Quantimet 720 at a magnification of 23× (lense f = 75 mm and length of extension rings 82 = mm). The digitized image of the kernel was then transferred directly to the PDP-11 computer for image processing by software created for that purpose.

The first step involved building a training set or an experimental set using previously stored images. Once training sets were created for the varieties or classes that were to be compared, the user could enter the "Load Training Set" step. When this was done, the program requested the user to decide which variables were to be used in the discriminant analysis. Results of the analysis for the training set then appeared on the screen and/or in print as desired, and were available for classification of unknown kernels. To identify the class of an unknown kernel, it was placed under the scanning head, and a keyboard command signaled the computer to print out the name of the variety (out of the two being compared) into which the unknown kernel fit best, based on all introduced parameters.

In addition, results for a stored set of images of unknown kernels could be recalled and identified in one step. This method required only seconds to complete, as the training and unknown

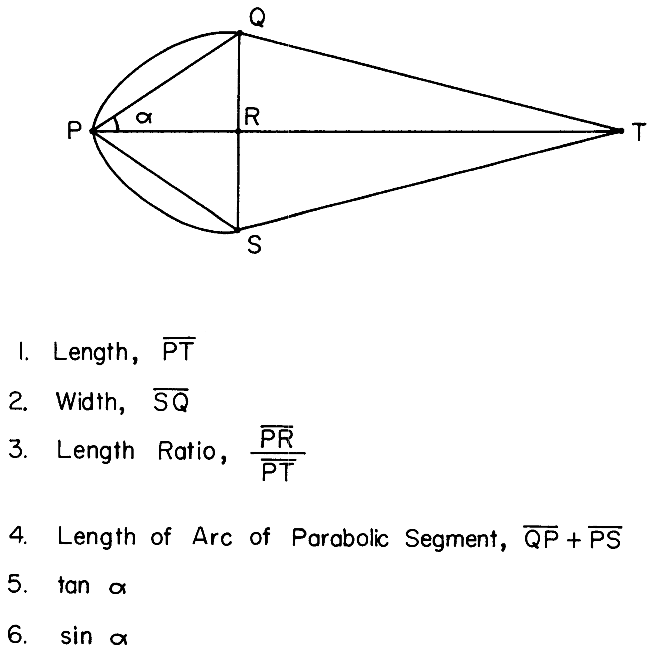

Fig. 2. Typical parameters of a wheat kernel used in the analysis.

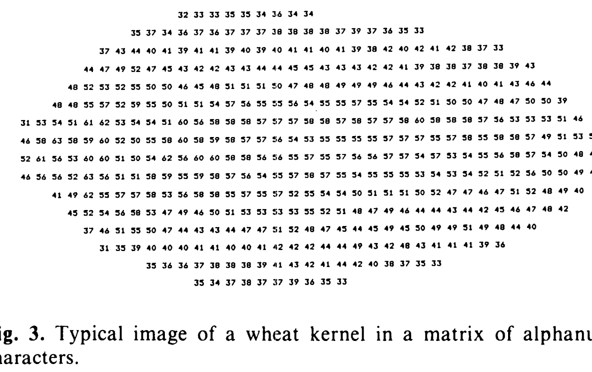

Fig. 3. Typical image of a wheat kernel in a matrix of alphanumeric characters.

experimental sets had already been stored. The process of storage was also relatively simple and rapid compared to the one described previously (Zayas et al 1985). As in the previous method, setting detection levels during image storage and measuring in the immediate mode were done carefully; settings included shade correction and calibration measurements.

All training files consisted of about 30 kernels for each pair of two varieties to be compared. After canonical equations were developed for each variety, individual kernels from each training set were tested against those equations. Then, 10 more kernels (test samples) were tested individually against the equations in real time directly from the screen. The same set was tested also as a stored set of images of unknown kernels.

TABLE I
Results of Discrimination by Image Analysis Between Wheat Varieties and Classes

Wheat Variety or Class[a]	Calibration Samples: No. Tested	Calibration Samples: No. Identified Correctly	Test Samples: No. Tested	Test Samples: No. Identified Correctly
Coker 745 (SRW)	29	29	10	10
NK 812 (HRW)	30	29	10	10
Double Crop (SRW)	30	30	10	10
TAM 101 (HRW)	29	28	10	10
Roughrider (HRW)	30	30	10	10
Oslo (HRS)	29	28	9	9
North Star (HRW)	30	29	10	10
Era (HRS)	30	29	10	10
Southern Bell (SRW)	30	25	10	10
Centurk (vitreous) (HRW)	30	28	10	10
Centurk (vitreous) (HRW)	30	29	10	6
Centurk (nonvitreous) (HRW)	30	26	10	8
Southern Bell (SRW)	30	25	10	8
Centurk (combined) (HRW)	60	49	20	16
Arkan (HRW)	30	30	10	10
Arthur (SRW)	27	27	10	10
Arkan (HRW)	30	30	10	10
Arthur (SRW)	30	28	9	9
SRW (Coker 745, Southern Bell, Double Crop)	89	72	30	25
HRW (NK 812, Centurk vitreous, TAM 101)	89	78	30	25
HRS (Oslo, Era)	59	42	19	12
HRW (Roughrider, North Star)	60	51	20	18

[a]HRS, hard red spring; HRW, hard red winter; SRW, soft red winter.

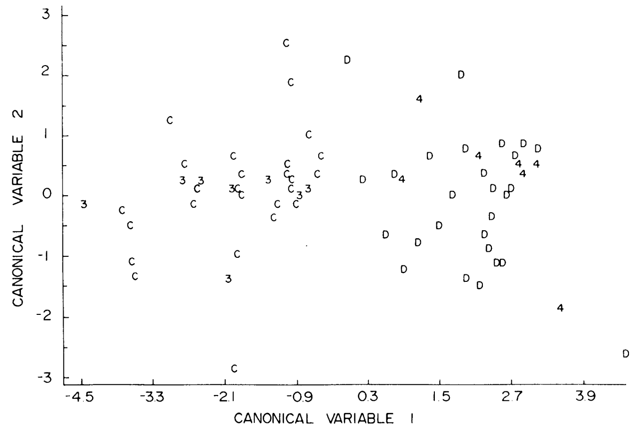

Fig. 5. Results of discrimination by image analysis between winter wheat varieties. Letters denote calibration samples and numbers denote test samples. Discrimination between hard red wheat Roughrider (**C** and **3**) and hard red spring Oslo (**D** and **4**). (10 observations are hidden.)

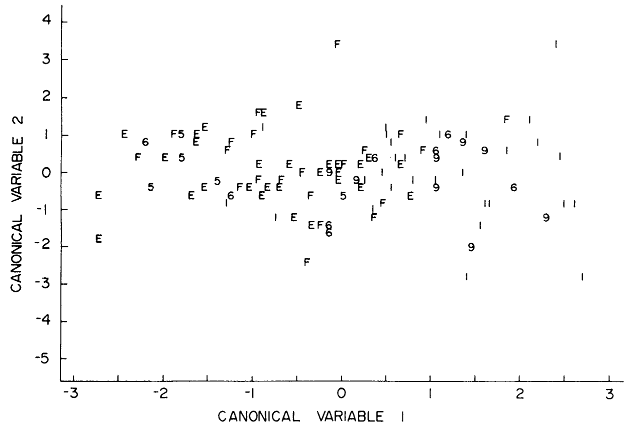

Fig. 6. Results of discrimination by image analysis between winter wheat varieties. Letters denote calibration samples and numbers denote test samples. Discrimination between a mixture of vitreous hard red wheat Centurk (**E** and **5**) and nonvitreous hard red wheat Centurk (**F** and **6**) and soft red wheat Southern Bell (**1** and **9**). (Six observations are hidden.)

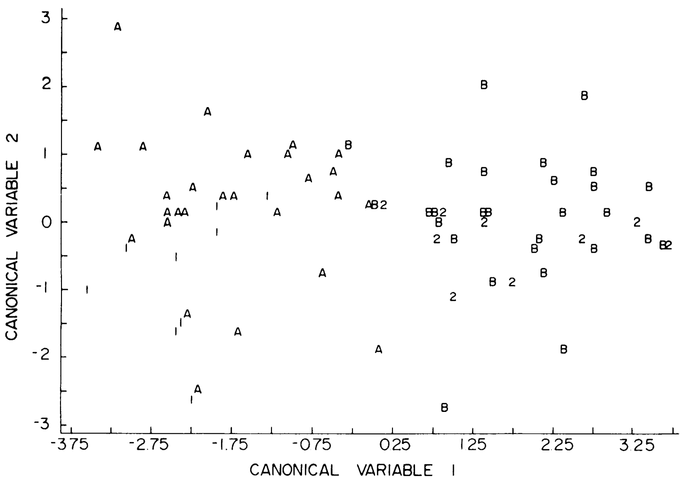

Fig. 4. Results of discrimination by image analysis between winter wheat varieties. Letters denote calibration samples and numbers denote test samples. Discrimination between soft red wheat Coker 745 (**A** and **1**) and hard red wheat NK 812 (**B** and **2**). (Six observations are hidden.)

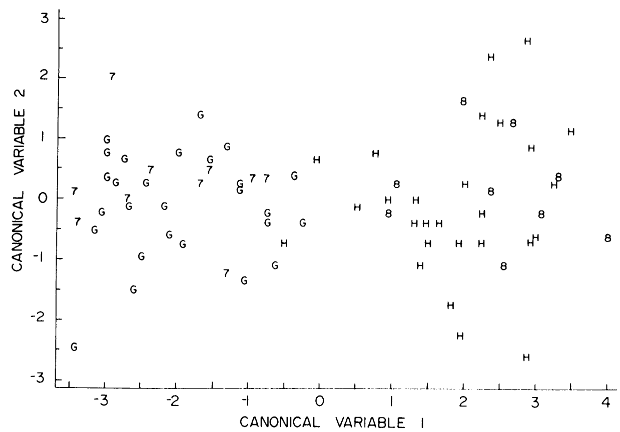

Fig. 7. Results of discrimination by image analysis between winter wheat varieties. Letters denote calibration samples and numbers denote test samples. Discrimination between hard red wheat Arkan (**G** and **7**) and soft red wheat Arthur (**H** and **8**). (44 observations are hidden.)

RESULTS AND DISCUSSION

The results of the discrimination for pairs of varieties and classes are compared in Table I and in Figures 4–9. At this stage, our main interest was a rapid determination of variety or class identity of a single kernel or group of kernels that may represent either a pure variety or a mixture of different varieties or classes. The same criteria were used to develop canonical equations for comparison among any two varieties or classes.

We used the SAS discriminant analysis programs DISCRIM and CANDISC to perform the analysis as another method to verify results of the program ARTARK, using the same variables. Observations were classified in DISCRIM by assuming a multivariate normal distribution within each class and were assumed to have equal covariance matrices. The classification system was based on the pooled covariance matrices. A discriminant model was developed from each training set in order to segregate observations into either one of two classes. The model was generated by a measure of generalized squared distances (Rao 1973) from x to group t.

$$D_t^2(x) = g_1(x, t) + g_2(t),$$

where

$D_t^2(x)$ = a generalized squared distance,
x = a vector containing the variables of an observation,
t = a subscript to distinguish the groups,
$g_1(x, t) = (x-m_t)' S^{-1} (x-m_t)$, and
S = pooled covariance matrix.

$$g_2(t) \begin{cases} = \ln(q_t) \text{ if the prior properties are not all equal,} \\ = 0 \text{ if the prior properties are all equal.} \end{cases}$$

q_t = the prior probability for group (t).

The posterior probability of an observation x belonging to group t is

$$P_t(x) = \frac{\text{Exp}\,[-0.5 D_t^2(x)]}{\Sigma_u \text{Exp}\,[-0.5\, D_u^2(x)]}.$$

An observation was classified in group (u) if setting t=u produced the smallest value of $D^2_t(x)$ or the largest value of $p_t(x)$.

After establishing the above criteria for a training set, we tested an independent sample with the criteria and classified the test data into one of two categories representing two wheat varieties.

We used FORTRAN for software developed in our laboratories to establish a comprehensive discriminant analysis of a single kernel via a macroviewer in an immediate mode or of a stored set in a programming mode. This was set on-line with the image analyzer, which in turn was controlled by a PDP-11 computer.

In general, the percentage of properly identified kernels for any pair of varieties was almost identical for both the calibration and test sets. The discriminant analysis (both SAS and software developed in our laboratories) differentiated well between SRW and HRW varieties and between HRW and HRS varieties tested. Some differences were established by discriminant functions used in this study between vitreous and nonvitreous kernels of the Centurk variety. Consequently, the combined nonvitreous and vitreous Centurk kernels were variable and presented problems when compared to Southern Bell. No such difficulties were encountered for differentiation between kernels of Arthur (SRW) and Arkan (HRW) wheats, irrespective of their source and visual appearance. Consistent differentiation between those two wheat cultivars by visual examination is extremely difficult (Zayas et al 1985).

For mixtures of three varieties of either HRW wheat or of SRW wheat (based on the results of SAS discriminant analysis), the average percentages of correctly classified kernels were 85% for calibration samples and 83% for test samples. For mixtures of two varieties of either HRW or HRS wheats, the average percentages were 78% for calibration samples and 77% for test samples.

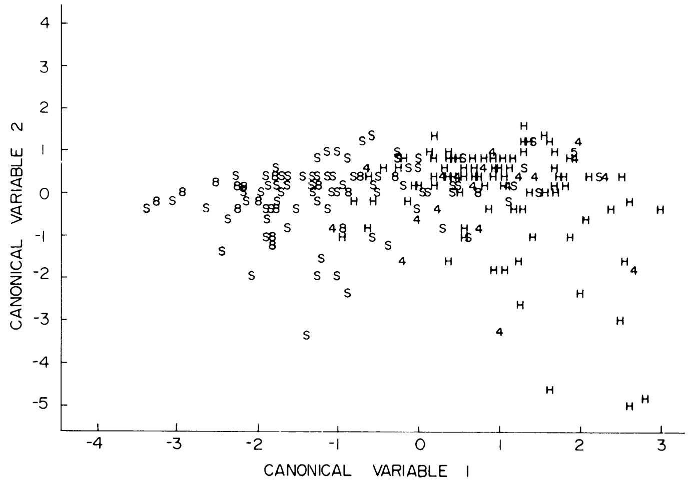

Fig. 8. Results of discrimination by image analysis between winter wheat classes. Letters denote calibration samples and numbers denote test samples. Discrimination between soft red wheat cultivars Coker 745, Southern Bell, and Double Crop (**S** and **8**) and hard red wheat cultivars NK 812, Centurk vitreous, and Tam 101 (**H** and **4**). (23 observations are hidden.)

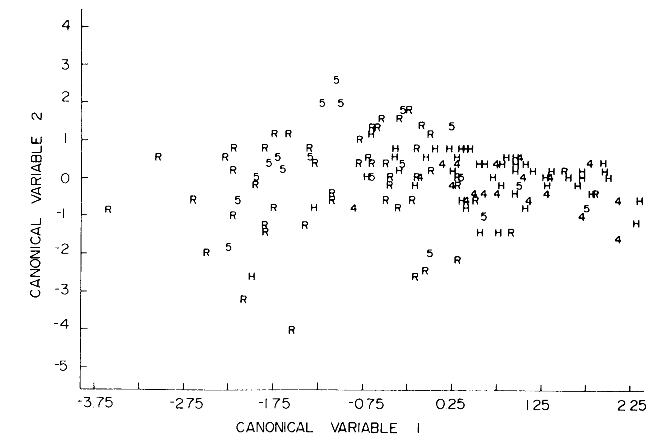

Fig. 9. Results of discrimination by image analysis between winter wheat classes. Letters denote calibration samples and numbers denote test samples. Discrimination between hard red spring cultivars Oslo and Era (**R** and **5**) and hard red wheat cultivars Roughrider and North Star (**H** and **4**). (19 observations are hidden.)

To identify kernels from various classes, additional parameters should be included for better characterization. A study of such parameters is underway in our laboratories. These methods are applicable to characterization of whole, sound, and well-developed kernels. Results are unaffected by source of material, as shown by the results for the two pairs of Arthur and Arkan samples. Results are affected, however, by visible damage (i.e., mold, scab, sick, black point wheat). For all practical purposes, it is of little consequence in routine grading to determine the precise class or variety of such damaged kernels. In addition, image analysis could well be used in grading cereal grains to detect the percentage of damaged kernels in a sample and to determine their conformity with grain standards. Such studies are underway in our laboratories.

ACKNOWLEDGMENT

We thank D. McAnarney for developing the software used in this study.

LITERATURE CITED

ANONYMOUS 1957. Page 10 in: Grain Grading Primer. Misc. Publ. No. 740, U.S. Department of Agriculture: Washington, DC.
ANONYMOUS 1984. New ways to class wheat. FGIS Update. October

1984, p. 3. Federal Grain Inspection Service, U.S. Department of Agriculture: Washington, DC.
MARTIN, T. J., BOCKUS, W. W., BROWDER, L. E., FINNEY, K. F., HATCHETT, J. H., and WETZEL, D. L. 1983. Registration of Arkan wheat. Crop Sci. 23:1221.
RAO, C. RADHAKRISHNA. 1973. Linear Statistical Inference and its Applications. John Wiley: New York, NY.
ZAYAS, I., POMERANZ, Y., and LAI, F. S. 1985. Discrimination between Arthur and Arkan wheats by image analysis. Cereal Chem. 62(6):478.

[Received July 2, 1985. Accepted October 11, 1985.]