Discrimination of Wheat and Nonwheat Components in Grain Samples by Image Analysis[1]

I. ZAYAS,[2] Y. POMERANZ,[3,4] and F. S. LAI[3]

ABSTRACT

Cereal Chem. 66(3):233-237

This study describes use of image analysis to discriminate between wheat and nonwheat and between weed seeds and stones in the nonwheat part of a grain sample. Two methods of pattern recognition were used in the discrimination: multivariate discriminant analysis to distinguish between wheat and nonwheat and among weed seeds, and development of a structural prototype to distinguish between wheat and nonwheat. The structural prototype method discriminated well between wheat and nonwheat. Weed seeds were differentiated by multivariate discriminant analysis, but identification of stones posed a problem with this method. Consequently, physical separation of stones prior to the image analysis program described in this paper may be necessary for satisfactory discrimination.

There are two basic approaches to evaluation of wheat in marketing channels: the Official United States Standards for Grain (USDA 1984) and Standards of the International Association of Cereal Science and Technology (ICC 1972). According to the official U.S. wheat standards, "dockage" is all material that can be removed from wheat by prescribed mechanical means (Zeleny 1971). Foreign material is all material, other than wheat, which is not separated in the determination of dockage. The U.S. grain standards specify the maximum amounts of certain types of foreign material (stones, glass, crotalaria seeds, or castor beans). Additional foreign materials of concern are weed seeds such as vetch, velvetleaf, chess, wild buckwheat, and foxtail (J. C. Halverson, FGIS-USDA, *personal communication*).

According to ICC standards (1972), *Gesamtbesatz* (total dockage) is classified into grain dockage (grain extraneous material) and black dockage (foreign extraneous material). The main objective of the *Besatz* test is the determination of the total amount of sound whole wheat kernels in a sample. The philosophy behind wheat evaluation by the U.S. Dapartment of Agriculture (USDA) and ICC systems differs. The U.S. dockage test, which is rapid and well-adapted to routine evaluation, has been criticized because it does not measure the total amount of nonmillable material. The ICC *Besatz* test, on the other hand, is time-consuming, somewhat subjective, and subject to errors because a small sample (250 g) that may not be representative of the total population is used for testing. The use of air-classification in combination with sifting and a trieur separator has met those difficulties to a limited extent only (Scott 1985, Seibel 1985).

We previously described the use of image analysis in discriminating among various types of cereal grains (Lai et al 1986, Zayas et al 1986) and among wheat classes and varieties (Zayas et al 1985, 1986). A digital image processing system was described by Sapirstein et al (1987) to determine composition of mixtures of wheat, oats, barley, and rye and Neuman et al (1987) used digital image analysis to classify wheat cultivars according to kernel type. Discrimination between varieties within classes gave inconclusive results with incorrect classification scores ranging from 4 to 85%. Travis and Draper (1985) used image processing and statistical analysis to obtain principal axis and confidence regions for 49 crop and weed species. Most species could be separated on the basis of a circularity shape factor (4π area/perimeter^2) and seed length. Included were only one wheat variety and two types of weed seeds of importance in wheat grading (wild buckwheat and green foxtail). Westerlind (1984) described a computerized separator as a mechanical aid for determination of seed purity (including wheat) on the basis of color, size, shape, and texture.

We now report on the use of techniques and software developed in our laboratories to distinguish between wheat and nonwheat and identify objectionable weed seeds.

This study was conducted in two parts, testing different approaches to the problem. In part I, a PDP11/03-L minicomputer (DEC, Digital Equipment Corp., West Concord, MA) served as a host computer for a Quantimet 720 image analysis system (Imanco-Cambridge, Monsey, NY). Both were used in the program ART30 and are described under Software. Images of wheat and nonwheat components were digitized; morphometrical data were extracted and stored on a computer disk. The program ART30, created for extraction of morphometrical parameters and discrimination of wheat varieties and based on multivariate discriminant analysis, was employed in this study for discrimination of wheat from nonwheat components of a sample.

To handle larger sets of wheat kernels, weed seeds, etc., the data were transferred to a mainframe computer CMS-IBM/3780, and multivariate discriminant analysis (SAS 1986) was used to discriminate among several types of nonwheat components and wheat.

For part II, a wheat pattern prototype was developed and run in the Quantimet 720 system to discriminate between wheat and nonwheat components.

MATERIALS

Samples of wheat and foreign material (which included weed seeds and stones) were obtained from the Federal Grain Inspection Service of the USDA (FGIS). The weed seeds were vetch (*Vicia angustifolia*), velvetleaf (*Abutilon theophrasti*), chess (*Bromus secalinus*), crotalaria (*Crotalaria spectabilis*), castor bean (*Ricinus communis*), and wild buckwheat (*Polygonum convolvulus*) mixed with small seeds such as yellow foxtail (*Setaria pumila*), green foxtail (*Setaria viridis*), and others.

A composite sample was formed of wheat kernels of different classes: five kernels of Arkan (hard red winter [HRW]), six kernels of Arthur (soft red winter [SRW]), six kernels of Era (hard red spring [HRS]), six kernels of North Star (HRW), six kernels of Coker (SRW), and five kernels of NK-812 (HRW); it was used for the ART30 program described under Software. In addition, eight wheat samples (four Arkan and four Arthur sets of 30 kernels each) were used in the training process by the SAS-DISCRIM procedure.

METHODS

Part I: Hardware

Image acquisition system. The hardware and software are described in Zayas et al (1985). The input device was a Vidicon

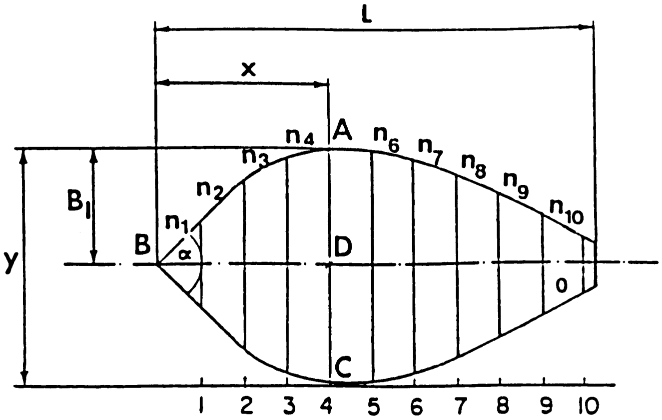
[1]Cooperative investigations of the U.S. Department of Agriculture, Agricultural Research Service, U.S. Grain Marketing Research Laboratory, Manhattan, KS 66502, and the Department of Chemical Engineering, Kansas State University.

[2]Research assistant, Department of Chemical Engineering, Kansas State University.
[3]Research chemist and research chemical engineer, respectively, U.S. Grain Marketing Research Laboratory.
[4]Present address: Research professor, Department of Food Science and Human Nutrition, Washington State University, Pullman, WA 99164-2032.

camera (9677 Thorn EMI) linked to the image analysis hardware, the computer interface, and the monitor. The host computer is PDPII/03-L with a 32K memory, dual hard disks RLO1K, and an LA36 Decwriter line printer. The six-bit digital value of each pixel (64 gray levels) can be transferred to a computer, stored, and analyzed; alternatively using Quantimet 720, desirable morphological measurements such as area, length, width, and perimeter can be taken and stored on the disk of the computer. The host is connected through a modem to an IBM-370 via a conversational monitor system interface.

For the experiments described under part I, a Nikkor 55-mm lens, F-2.8, with two 40-mm extension rings was used with the Vidicon camera, so that only one object would be in the field of view. A fiber-optic circular lamp mounted on the lens was used for illumination. Black Plasticine and black velvet paper were used as background. A dark background was used for light objects (wheat, chess, stones, yellow foxtail, castor beans); light background was used for dark objects (crotalaria, vetch, velvetleaf). Some objects, i.e., castor beans, were difficult to set for satisfactory detection because of surface reflectivity or a dark surface pattern. Calibration was performed with a preliminary shade correction setting. The Quantimet shading corrector measured and stored a pattern for a blank field. "Shading" is a result of nonuniformity of illumination. The stored shading pattern is used to correct the shading of subsequent fields. Shading correction is done before the image is digitized and objects are detected.

Setting a proper detection threshold is crucial for error-free measurements of an object in the field of view. The procedure generates a segmented, binary image in which ones correspond to objects of interest and zeros correspond to background.



1. WIDTH - Y ;
2. LENGTH - L ;
3. ASPECT RATIO - $\frac{L}{Y}$;
4. TANGENT α * = $\frac{AD}{BD}$;
5. LENGTH OF PARABOLIC SEGMENT (ABC) ;

$$L_{ABC} = \sqrt{4x^2+y^2} + \frac{y^2}{2x}\log_e\left[\frac{2x+\sqrt{4x^2+y^2}}{y}\right];$$

COEFFICIENT $n_i = \frac{L_i}{Y_i}$ $(i = 1 \ldots\ldots 10)$ **

6. COEFFICIENT $n_9 = \frac{L_9}{Y_9}$;
7. COEFFICIENT $n_5 = \frac{L_5}{Y_5}$;
8. COEFFICIENT $n_2 = \frac{L_2}{Y_2}$;
9. NORMALIZED GRAY (GR.) LEVEL - $\frac{\text{MAX GR. LEVEL} - \text{MIN GR. LEVEL}}{\text{AVG GR. LEVEL} - \text{MIN GR. LEVEL}}$;
10. PARALLELISM - $\frac{\text{\#COL. TOP * \# COL. BOTTOM}}{\text{MAX WIDTH}}$;

NOTE: *Angle is germ angle for wheat.
**10 coefficients can be entered; each coefficient is ratio of the length to the width at this point (n_i).

Fig. 1. Wheat size and shape descriptors used for multivariate discriminant analysis.

Part I: Software

The program ART30 (a newer version of ARTARK; Zayas et al 1986) was modified and used for discrimination of wheat from some nonwheat components. In the original version of ARTARK, based on multivariate statistical analysis (DEC, Scientific Subroutine Package), 10 variables were used. The program was modified to enter up to 30 variables. Those variables describe the size, shape, and gray level of the objects.

In the first step of the program, sets of different materials were run for calibration. In the second step is an option to run the previously stored experimental set of data. Program ART30 also has an option, "Read from the screen," in which the user can place an unknown object under the camera and after several seconds classify the object into one of the calibration sets.

The 32K memory of PDP11/03-L was a limiting factor in the experiments. To test larger numbers of variables and features, we transferred the data files into a mainframe CMS-IBM 370 computer. Under our conditions, the use of the mainframe computer from the beginning was not economical. In the mainframe computer, we used multivariate discriminant analysis (MDA), running the SAS (1986) software package, to discriminate nonwheat components.

The STEPDISC procedure (SAS) is helpful for exploratory analysis, because it facilitates the selection of independent variables that are likely to be good (but not necessarily the best possible) discriminators (SAS 1986). Using the STEPDISC procedure, nine variables (as in Fig. 1) were selected for MDA (Table I).

Part II: Hardware

The Quantimet 720 image analysis system was used for this series of experiments.

TABLE I
Summary Ranking of Object Features by Stepwise Discriminant Analysis

Model Variables[a]	Squared Partial Correlation (R^2)	*F* Statistic	Probability *F*	Wilks' Lambda	Average Squared Canonical Correlation
2	0.924	1,135.371	0.0001	0.0763	0.185
4	0.607	144.459	0.0001	0.0300	0.305
6	0.587	132.841	0.0001	0.0124	0.398
1	0.509	96.776	0.0001	0.0061	0.489
7	0.312	42.207	0.0001	0.0042	0.535
8	0.269	34.057	0.0001	0.0031	0.557
10	0.231	27.803	0.0001	0.0024	0.583
9	0.128	13.359	0.0001	0.0005	0.707
3	0.023	2.032	0.0723	0.0003	0.746

[a]Size and shape descriptor variables as shown in Fig. 1.

Part II: Software

The objective of this part was to discriminate wheat from nonwheat without identifying nonwheat components. Simple shape descriptors were used to create a wheat pattern structured prototype that could be compared by matching against a stored structured prototype.

The number of objects in the field of view was determined by the size of the object and optical setting. Only two castor beans fit the size of the field of view but 15 seeds of wild buckwheat filled the screen. Objects were oriented along their longest axis parallel to the bottom line of the screen. A detection threshold for overlayed binary image was set for each field, for each group of objects of interest. The final output was acceptance or rejection of objects in the field of view.

The Quantimet 720 image analysis system allows creating patterns and writing subroutines in Quantimet facilities. Morphometrical derivative parameters of the basic parameters (area, perimeter, length, width, and Feret's diameters 0, 45, 90, and 135°) were combined to create a wheat pattern prototype on the basis of six shape factors: 1) area, from 1,000 to 21,000 pixels; 2) $(2\sqrt{\text{area}/3.14})/(\text{perimeter}/3.14)$, from 0.6 to 0.9; 3) perimeter/convex perimeter, from 1 to 1.3; 4) convex perimeter/length, from 2 to 2.5; 5) Feret 0/Feret 45, from 0.4 to 0.7; 6) Feret 90/Feret 135, from 1.1 to 1.4.

Shape factor 1 eliminates objects smaller or larger than wheat kernels. Shape factor 2 represents the relationship between equivalent diameter and normalized perimeter and distinguishes among objects that differ in concavities and smooth outlines. Shape factor 3 depicts the ratio between perimeter and convex perimeter and differentiates among objects with the same length of perimeter but different symmetry. Shape factors 4, 5, and 6 differentiate among objects with similar width and length but with different degrees of symmetry. Those six shape factors (and the appropriate ranges) were selected on the basis of visual analysis and previous experience; their validity was confirmed by a series of tests.

The wheat pattern was formed as a preset of classification limits; only objects that passed all limiting criteria were recognized as wheat. Counts (on a monitor or in print) specified the number of objects recognized as wheat kernels (Quantimet 720 manual). The wheat pattern was tested against several sets of nongrain samples.

RESULTS AND DISCUSSION

Part I

Table II summarizes the results of MDA-SAS (DISCRIM procedure) on four different weed seeds: wild buckwheat, chess, velvetleaf, and vetch in combination with wheat. This MDA rule classified all the wheat samples correctly. Weed seeds in the calibration data set were classified correctly, except for two chess seeds and one velvetleaf. For test data there were the following misclassifications: one wild buckwheat seed out of 30, four chess seeds out of 30, and seven vetch seeds out of 38.

When stones were included in MDA-SAS, the results summarized in Table III were obtained. Both calibration and test data showed correct identification of wheat by the MDA rule. The main discrepancy was for stones, with six out of 30 misclassified in the calibration data and 22 out of 59 misclassified in the test data. Chess and vetch were generally identified correctly in the calibration data but not in the test data.

In general, the incorrect classifications in Table II are comparable to those in Table III, except for the stones in the latter. The main concern is misclassification of irregularly shaped stones as wheat; 12 out of 59 stones were misclassified as wheat for test data (Table III). In addition, some chess weeds were identified as wheat and some vetch weeds were misclassified as wild buckwheat. The latter is probably of less concern than misclassification of chess as wheat.

Both ART30 and DISCRIM (SAS) in the mainframe computer identified wheat correctly. The procedure CANDISC (SAS) performs a canonical discriminant analysis that gives linear combinations of the variables that best identify differences between classes (i.e., wheat vs. nonwheat). Values of the first two

TABLE II
Identity of Discriminated Wheat Kernels and Weed Seeds[a] by Multivariate Discriminant Analysis

Identity of Observed Objects	No. of Observed Objects	Wheat	Wild Buckwheat	Chess	Velvetleaf	Vetch
Calibration data						
Wheat	236	236	···	···	···	···
Wild buckwheat	60	···	60	···	···	···
Chess	30	2	···	28	···	···
Velvetleaf	30	···	1	···	29	···
Vetch	20	···	···	···	···	20
Test data						
Wheat	34	34	···	···	···	···
Wild buckwheat	30	···	29	···	1	···
Chess	30	4	···	26	···	···
Vetch	38	···	7	···	···	31

[a]Without stones.

TABLE III
Identity of Discriminated Wheat Kernels, Stones, and Weed Seeds by Multivariate Discriminant Analysis

Identity of Observed Objects	No. of Observed Objects	Wheat	Wild Buckwheat	Chess	Velvetleaf	Vetch	Stones
Calibration data							
Wheat	236	236	···	···	···	···	···
Stones	30	1	4	···	1	···	24
Wild buckwheat	59	···	56	···	···	···	3
Chess	30	2	···	28	···	···	···
Velvetleaf	30	···	1	···	29	···	···
Vetch	20	···	···	···	···	20	···
Test data							
Wheat	34	34	···	···	···	···	···
Stones	59	12	5	1	4	···	37
Wild buckwheat	31	···	28	···	1	···	2
Chess	30	4	···	26	···	···	···
Vetch	38	···	5	···	···	33	···

canonical functions were computed for each sample data point and plotted in Figure 2. The plot illustrates a nearly perfect separation of the wheat and different weed seeds by the canonical functions (Fig. 2).

In light of the above misclassifications of the various nonwheat objects, we used MDA-SAS to distinguish between wheat and nonwheat only. The results are summarized in Table IV and illustrated in Figure 3. In practically all cases wheat was identified

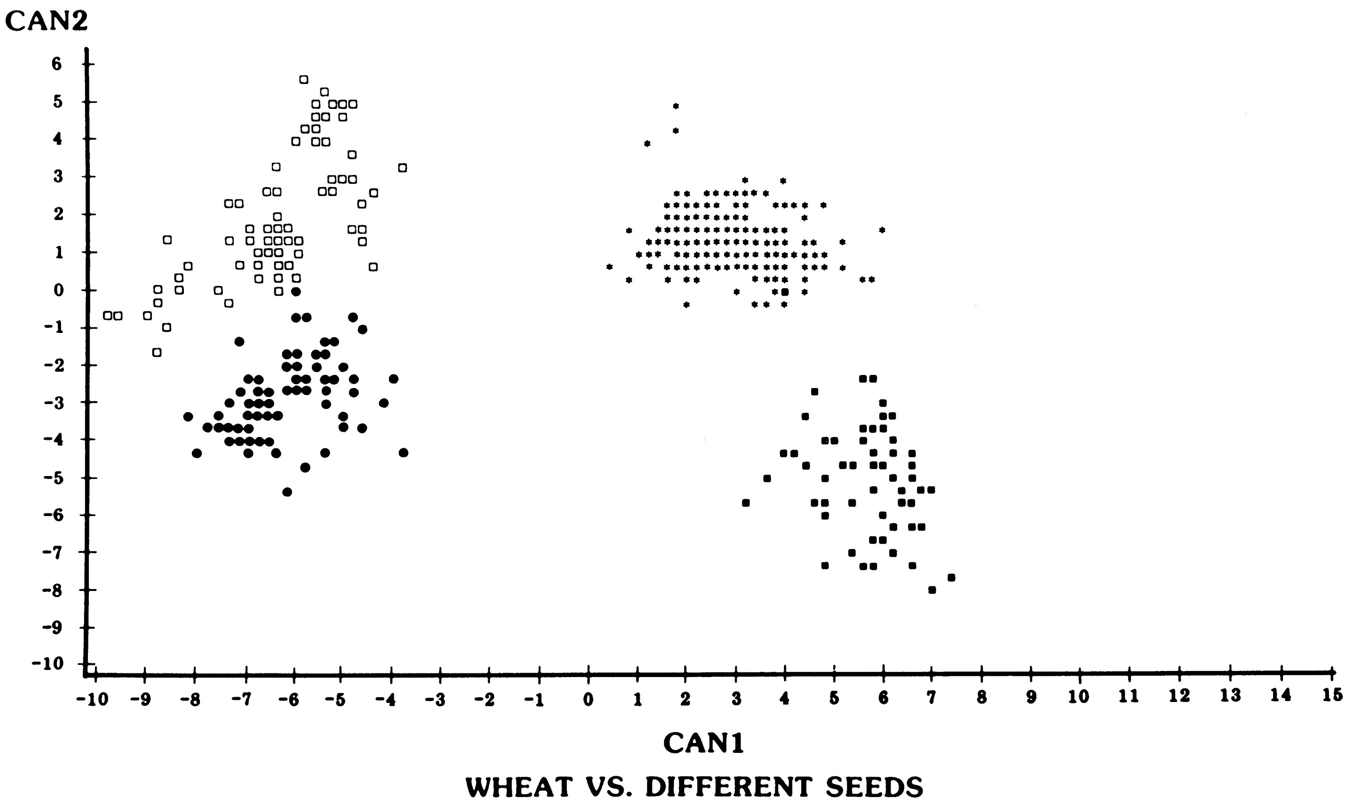


Fig. 2. Graph of canonical (**CAN**) functions from multivariate discriminant analysis for discrimination by image analysis between wheat and different weed seeds (calibration data): wheat (*), chess (■), wild buckwheat (•), and vetch (□). (195 observations are hidden.)

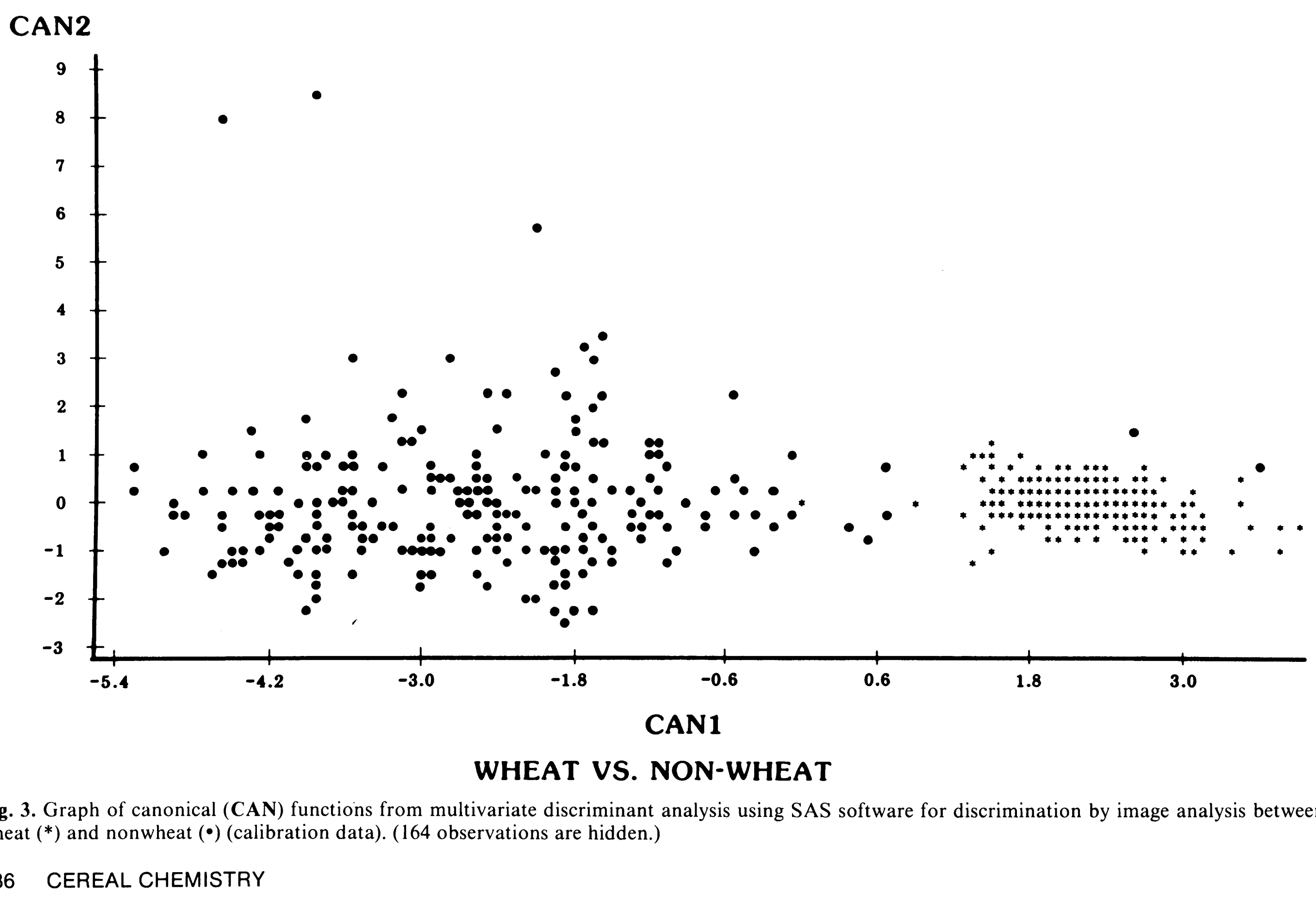


Fig. 3. Graph of canonical (**CAN**) functions from multivariate discriminant analysis using SAS software for discrimination by image analysis between wheat (*) and nonwheat (•) (calibration data). (164 observations are hidden.)

TABLE IV
Identity and Number of Wheat and Nonwheat Objects Identified by Multivariate Discriminant Analysis

Identity of Observed Objects	No. of Observed Objects	Wheat	Nonwheat
Without stones			
Calibration data			
Wheat	236	236	···
Nonwheat	140	7	133
Test data			
Wheat	34	33	1
Nonwheat	99	12	87
With stones			
Calibration data			
Wheat	236	236	···
Nonwheat	169	11	158
Test data			
Wheat	34	33	1
Nonwheat	158	23	135

correctly. Some of the nonwheat objects, however, were classified as wheat. The number of misclassified nonwheat objects (23 out of 158 nonwheat objects) was particularly high for test data containing stones.

Part II

Discrimination by a comparison with the structural wheat pattern yielded the results summarized in Table V. Discrimination was correct, except for two out of 50 samples of stones and one out of 50 seeds of chess. The approach used in part II was that a pattern was established for wheat and conformity to that pattern of nonwheat objects was tested. No attempt was made to identify the nonwheat objects (as in part I). In principle it was a wheat versus nonwheat discrimination. The relatively simple algorithm is attractive, as discrimination between wheat and nonwheat can be done in less than 10 sec. New, commercial image analyzers can complete the discrimination in real time (1/30 sec). This method can serve as a preliminary screening, in the ICC grading system, to determine nongrain components in a sample.

CONCLUSIONS

A practical scheme for use of image analysis for discrimination and identification of wheat and nonwheat components in grain samples could involve the following steps: 1) run a structural prototype wheat pattern to discriminate wheat from nonwheat components; 2) use the information from the first step in combination with a mechanical device (i.e., conveyor with rejecting plunger) to separate wheat from nonwheat; and 3) use multivariate discriminant analysis to identify and determine nonwheat components.

In light of the misclassification of stones, physical separation of stones and other irregularly shaped inorganic material by a trieur separator or by air-classification in combination with sieving as recommended by Seibel (1985) and Scott (1985) is desirable. Alternatively, image analysis can be combined with one of the new methods of pattern recognition for discrimination of high-density inorganic material such as stones.

Image analysis in combination with other methods (i.e., sieving, separation by air) has a potential for discrimination between wheat and nonwheat components and among weeds in a sample using either the USDA or ICC grading systems. The combination seems to present an alternative to manual separation. At present, one of the limitations of the proposed method is the requirement to manually orient the kernels. The other limitation is the requirement to measure separated objects. The latter can be overcome by the now commercially available erosion dilation, image analysis processing technique. Finally, the design of an appropriate single-kernel feeder is a challenging requirement before the system can be implemented for practical use.

TABLE V
Discrimination of Wheat and Nonwheat by Structural Pattern

Set	No. of Observed Objects	No. of Correct Discriminations
Wheat	50	50
Wild buckwheat	50	50
Yellow foxtail	50	50
Chess	50	49
Vetch	30	30
Crotalaria	50	50
Stones	50	48
Glass	50	50
Castor beans	6	6

ACKNOWLEDGMENTS

D. Johnson, Department of Statistics, Kansas State University, is thanked for help with the analysis of data.

LITERATURE CITED

ICC. 1972. Standard Methods of the International Association for Cereal Chemistry. Verlag Moritz Schafer: Detmold, West Germany.

IMANCO. 1978. Instruction Manual–Quantimet 720. Imanco: Monsey, NY.

LAI, F. S., ZAYAS, I., and POMERANZ, Y. 1986. Application of pattern recognition techniques in the analysis of cereal grains. Cereal Chem. 63:168.

NEUMAN, M., SAPIRSTEIN, H. D., SHWEDYK, E., and BUSHUK, W. 1987. Discrimination of wheat class and variety by digital image analysis of whole grain samples. J. Cereal Sci. 6:125.

SAPIRSTEIN, H. D., NEUMAN, M., WRIGHT, E. H., SHWEDYK, E., and BUSHUK, W. 1987. An instrumental system for cereal grain classification using digital image analysis. J. Cereal Sci. 6:3.

SAS INSTITUTE. 1986. SAS User's Guide. The Institute: Cary, NC.

SCOTT, G. 1985. Definition of broken kernels and principle of the mechanical "Besatz" determination. pp. 39-42. In: Analyses as Practical Tools in the Cereal Field. K. M. Fjell, ed. Norwegian Grain Corporation: Oslo.

SEIBEL, W. 1985. Importance of Besatz for the cereal processing industry. Pages 33–37 in: Analyses as Practical Tools in the Cereal Field. K. M. Fjell, ed. Norwegian Grain Corporation: Oslo.

TRAVIS, A. J., and DRAPER, S. R. 1985. A computer based system for the recognition of seed shape. Seed Sci. Technol. 13:813.

USDA-FEDERAL GRAIN INSPECTION SERVICE. 1984. Official Grain Standards. U.S. Government Printing Office: Washington, DC.

WESTERLIND, E. 1984. Computerized separator as a mechanical aid for determination by number of other seeds in cereal species. Meddelands Statens Utsadeskontrol 59:83. (Original not seen; from a summary and technical publication by E. Westerlind, Kamas Westrup Co., Vellinge, Sweden).

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

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

ZELENY, L. 1971. Criteria of wheat quality. Page 26 in: Wheat Chemistry and Technology. Y. Pomeranz, ed. Am. Assoc. Cereal Chem.: St. Paul, MN.

[Received October 3, 1986. Revision received January 23, 1989. Accepted January 24, 1989.]