Seed classification using machine vision

P. SHATADAL, D.S. JAYAS, J.L. HEHN and N.R. BULLEY

Department of Biosystems Engineering, 438 Engineering Building, University of Manitoba, Winnipeg, MB, Canada R3T 5V6.
Received 16 August 1994; accepted 17 February 1995.

Shatadal, P., Jayas, D.S., Hehn, J.L. and Bulley, N.R. 1995. **Seed classification using machine vision.** Can. Agric. Eng. **37**:163-167. This paper reports the results of applying digital image analysis in conjunction with statistical pattern recognition to measure the size and shape features of various seed types and to classify them into the primary grain, small seed, and large seed categories. The seed types used in each category were: hard red spring (HRS) wheat and barley as primary grains; canola, brown mustard, yellow mustard, oriental mustard, and flaxseed as small seeds; and 'Laird' lentils, 'Eston' lentils, pea beans, green peas, black beans, and buckwheat as large seeds. The objective of the study was to assess the classification success in identifying HRS wheat and barley from other small and large seeds using morphological features. Orientation of the kernels for camera viewing was random. The kernels were, however, positioned manually in a non-touching manner. Hard red spring wheat and barley were correctly identified from all other seed types with more than 99% accuracy. Small and large seed categories were successfully discriminated from each other. Within each of the small and large seed groups, however, the classification was poor with up to 54.7% misclassification in small seed group and up to 30.3% misclassification in the large seed group. Canola yielded the worst classification with only 45.3% of canola seeds correctly discriminated from other small seeds.

Cet article contient les résultats d'application de l'analyse d'images digitales en conjonction avec la reconnaissance statistique de formes, pour mesurer la grandeur et les caractéristiques de la forme de différents types de graines, et pour les classifier dans le catégories suivantes: graines primaires, petites graines et grosses graines. Les types de graines utilisés pour chaque catégorie étaient: blé de force roux du printemps et orge pour les graines primaires; canola, moutarde noire, moutarde jaune, moutarde chinoise, et lin pour les petites graines; et lentilles 'Laird', lentilles 'Eston', fèves à pois, pois verts, fèves noires, et sarrasin pour les grosses graines. L'objectif de l'étude était d'établir le taux de succès de la classification en différenciant le blé de force roux du printemps et l'orge des autres petites et grosses graines, en utilisant des caractéristiques morphologiques. L'orientation des grains par rapport à la caméra était aléatoire. Les graines étaient toutefois disposées manuellement de façon à éviter qu' elles se touchent. Le blé de force roux du printemps et l'orge ont été correctement identifiés par rapport à tous les autres types de graine, avec une exactitude de plus de 99%. Les catégories de petites et grosses graines ont été discriminées les unes des autres avec succès. Toutefois, à l'intérieur de chacun des groupes de petites et grosses graines, la classification a été moins bonne; le taux de mauvaise classification a grimpé jusqu' à 54.7% dans le groupe de petites graines, et a atteint 30.3% pour les grosses graines. Les pires taux de classification ont été obtenus avec le canola, pour lequel seulement 45.3% des graines ont été discriminées correctement des autres graines.

INTRODUCTION

There is a growing interest in the grain industry for on-line monitoring of grain at terminal elevators. Rapid evaluation of the contents of a grain sample can be used in deriving optimum cleaning strategies, in making appropriate binning decisions, and for complete automation of certain terminal operations such as railcar unloading. Such monitoring would result in increased throughput and enhanced recovery of salvageable grains. Machine vision technology holds promise for developing an on-line grain monitoring system. Preliminary investigations demonstrate that machine vision can be successfully used in identification of contrasting grains (Sapirstein and Bushuk 1989; Shatadal et al. 1994). The scope of the earlier reported investigations (Brogan and Edison 1974; Sapirstein et al. 1987; Sapirstein and Bushuk 1989; Hehn and Sokhansanj 1990; Shatadal et al. 1994) was limited because only a few grain species were considered in any one study. A sample of grain may contain seeds of several species. The machine vision system should be robust enough to identify the primary grain from other seeds present in a sample. The positive identification of the primary grain and determination of the fractions of small and large seeds in a sample is important in automating the controls of grain cleaning machinery.

This study was conducted as part of a project to develop a machine vision system for monitoring wheat and barley at terminal elevators. The study is also relevant in the context of developing an automated grain grading system. The objective of the study was to determine the discriminating ability of the morphological features (i) for identification of hard red spring (HRS) wheat or barley from various small and large seeds, and (ii) for classification of different small and large seeds. The small seeds used in the study were canola, brown mustard, yellow mustard, oriental mustard, and flaxseed. The large seeds were 'Laird' lentils, 'Eston' lentils, black beans, pea beans, green peas, and buckwheat.

MATERIALS AND METHODS

Vision hardware

A 3 chip CCD (charge coupled device) color camera (Model DXC-3000A, SONY) was used to acquire images. A zoom lens (VCL-1012 BY) of 10-120 mm focal length was fitted in the camera. The camera was mounted on a stand (Model m3, Bencher Inc., Chicago, IL) which provided easy vertical movement and a stable support for the camera. The camera was connected to a camera control unit (Model CCU-M3, SONY). The iris or aperture control could be set to manual or automatic mode. The iris control was set to manual mode to achieve repeatability in the experiments. The automatic gain control of the camera was disabled.

The red (R), green (G), and blue (B) video signals from the camera control unit were converted to a 24 bit color digital image by a frame grabber board (Model DT 2871, Data Translation Inc., Molboro, MA). The frame grabber board was installed in an expansion slot of a personal computer (Model 80386, UNISYS). The frame grabber could convert the R, G, B color signals to hue (H), saturation (S), and intensity (I) signals in real time. The programs to control the frame grabber were written in C programming language using the Aurora subroutine library (Data Translation Inc., Molboro, MA). Only the intensity buffer was saved for further analysis. Image analysis was carried out on a workstation (model SPARCSTATION 2, Sun Microsystems, Inc., Mountain View, CA) using the development environment of Khoros software (Khoral Research Inc., Albuquerque, NM; Hehn et al. 1993).

Sample illumination

Uniform diffused front-lighting was used in all of the experiments. The method of illumination used is described by Batchelor (1985). The voltage to the illumination source (8 incandescent 25 W bulbs) was controlled by a voltage regulator (Sola Canada Inc.). A variac was used to control the illumination level by changing the voltage to the lamps. The illumination level was calibrated by repeatedly digitizing a small and fixed region of interest on the image of a Kodak gray card and simultaneously changing the voltage to the lamps until the mean pixel value of the region hit a pre-selected target value of 110. It was observed that hue, saturation, and intensity values did not change with voltage if the voltage was above 110 V. Voltage used in the tests ranged from 114 to 115 V.

Grain samples

Dry grain samples of hard red spring (HRS) wheat (Grade 1 Canada Western Red Spring), barley (Special Select Malt barley), and small and large seeds used in the experiments were obtained from the Grain Inspection Division of the Canadian Grain Commission, Winnipeg, MB. Samples of 'Laird' lentils, 'Eston' lentils, black beans, pea beans, green peas, canola, brown mustard, yellow mustard, oriental mustard, and flaxseed had not been previously graded and contained impurities. In the tests, however, only visibly sound kernels were used. Sub-samples of each grain type to be used in imaging were obtained by dividing the larger sample with a Boerner divider. The sub-samples were placed manually on a viewing plate. The kernel orientation was not controlled. Kernels were, however, not allowed to touch one another to facilitate their segmentation from the background.

Feature measurement

The kernel images were labelled for identification of each region in the image. The boundary tracking algorithm (Rosenfeld and Kak 1982) was used to find the contour of each kernel. Eight connected perimeter, as defined by Haralick and Shapiro (1992), was calculated. The length of a kernel was estimated by the Euclidean distance between the corresponding crossing points of the boundary on its major principal axis. The width of a kernel was estimated by the width of the minimum enclosing rectangle of the kernel region. Minimum and maximum radii were calculated as the minimum and maximum distances between points on the boundary and the centroid. Area was estimated by counting the number of pixels in a kernel region. The features were calibrated by digitizing a quarter coin and relating its physical dimensions to its dimensions in pixels. Rectangular aspect ratio (ratio of length to width), thinness ratio (ratio of square of perimeter to area), radius ratio (ratio of maximum to minimum radii), area ratio (ratio of area to product of length and width), and H-ratio (ratio of mean to standard deviation of all the radii) were calculated to quantify the shape of the kernels.

RESULTS AND DISCUSSION

Identification of HRS wheat and barley from small and large seeds

The basic data (mean, standard deviation, and range) on area, perimeter, length, and width of all the grains are given in Tables I - VII. To assess the pattern classification capability of the morphological features, procedure DISCRIM of SAS (1990) was used to classify the kernels into the categories of HRS wheat, barley, small seeds, and large seeds. All eleven extracted features (viz. length, width, area, perimeter, maximum and minimum radii, rectangular aspect ratio, thinness

Table I: Physical dimensions of hard red spring wheat and barley

	HRS wheat (n = 1001)*			Barley (n = 999)		
Feature	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Area (mm^2)	15.0	2.2	8.9 - 22.0	23.0	2.9	12.0 - 32.0
Perimeter (mm)	14.6	1.1	11.6 - 18.4	20.5	1.7	15.3 - 27.8
Length (mm)	5.3	0.4	4.0 - 6.8	8.3	0.8	6.0 - 11.9
Width (mm)	3.2	0.4	2.1 - 4.3	3.4	0.3	2.3 - 4.2

* n refers to sample size.

Table II: Physical dimensions of canola and brown mustard

	Canola (n = 992)*			Brown mustard (n = 1000)		
Feature	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Area (mm^2)	2.4	0.5	1.2 - 4.2	1.9	0.3	1.0 - 3.2
Perimeter (mm)	5.1	0.6	3.3 - 7.0	4.5	0.5	3.1 - 6.3
Length (mm)	1.6	0.2	1.1 - 2.2	1.5	0.2	1.0 - 2.1
Width (mm)	1.5	0.2	0.8 - 2.1	1.3	0.2	0.8 - 1.9

* n refers to sample size.

Table III: Physical dimensions of yellow and oriental mustards

Feature	Yellow mustard (n = 997)*			Oriental mustard (n = 1000)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Area (mm^2)	3.9	0.6	2.5 - 6.3	2.4	0.3	1.2 - 3.9
Perimeter (mm)	6.8	0.6	5.4 - 8.7	5.1	0.4	3.6 - 6.8
Length (mm)	2.2	0.2	1.6 - 3.0	1.7	0.2	1.1 - 2.2
Width (mm)	1.9	0.2	1.2 - 2.5	1.5	0.2	0.8 - 2.0

* n refers to sample size.

Table IV: Physical dimensions of 'Laird' and 'Eston' lentils

Feature	'Laird' lentils (n = 1000)*			'Eston' lentils (n = 999)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Area (mm^2)	36.0	3.6	21.0 - 46.0	16.0	2.0	10.0 - 24.0
Perimeter (mm)	21.8	1.2	16.7 - 24.9	14.5	1.0	11.2 - 17.8
Length (mm)	6.7	0.4	5.2 - 7.7	4.5	0.3	3.5 - 5.7
Width (mm)	6.4	0.4	5.1 - 7.5	4.3	0.3	3.2 - 5.3

* n refers to sample size.

Table V: Physical dimensions of pea beans and green peas

Feature	Pea beans (n = 1002)*			Green peas (n = 1000)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Area (mm^2)	40.0	4.6	26.0 - 56.0	35.0	3.3	22.0 - 50.0
Perimeter (mm)	23.4	1.4	18.8 - 28.0	21.6	1.0	17.3 - 26.1
Length (mm)	7.8	0.5	6.3 - 9.9	6.6	0.3	5.3 - 8.0
Width (mm)	6.1	0.5	4.4 - 7.9	6.4	0.4	4.8 - 7.9

* n refers to sample size.

Table VI: Physical dimensions of black beans and buckwheat

Feature	Black beans (n = 1000)*			Buckwheat (n = 999)		
	Mean	Standard deviation	Range	Mean	Standard deviation	Range
Area (mm^2)	45.0	5.7	29.0 - 67.0	19.0	2.9	10.0 - 28.0
Perimeter (mm)	25.3	1.7	20.1 - 30.8	16.5	1.5	17.0 - 21.0
Length (mm)	8.4	0.6	6.4 - 10.8	5.6	0.7	3.6 - 8.0
Width (mm)	6.6	0.6	5.5 - 8.6	4.2	0.4	2.9 - 5.5

* n refers to sample size.

ratio, radius ratio, area ratio, and H-ratio) were used in the classification. A non-parametric probability density estimation (viz. k-nearest neighbor) was used. The classification was based on the Bayes decision rule which classifies an entity (represented by its pattern vector) to a class for which the entity has a maximum a posteriori probability (Hand 1981; Duda and Hart 1973). The classification was considered invalid if the corresponding maximum a posteriori probability was less than 0.5. The confusion matrix for a cross validation or leave-one-out classification for a four-way discrimination into HRS wheat, barley, small seeds, and large seeds is given in Table VIII.

Hard red spring wheat and barley were correctly identified from all of the small and large seeds included in the study with more than 99% accuracy (Table VIII). There were, however, misclassifications of 28 buckwheat kernels (large seeds) as HRS wheat and three buckwheat kernels as barley. Also, 0.2% of HRS wheat kernels were misclassified as buckwheat and another 0.2% as flaxseed (small seeds). These misclassifications may be acceptable from a grain monitoring viewpoint. Grey-level and color dependent features should be tested to determine whether their addition in the pattern vector can eliminate the misclassification among HRS wheat, buckwheat, and flaxseed.

The performance of a grain cleaning unit can be assessed by finding out the fraction of small and large seeds in the grain before and after cleaning. The results shown in Table VIII demonstrate that machine vision can be successfully used to calculate small and large seed fractions in wheat and barley grain samples. The study, however, used highest grade samples of HRS wheat and barley. Higher variability of morphological features in lower grade samples of HRS wheat and barley can result in increased misclassification than that observed in this study.

Classification among small and large seeds

The confusion matrix for a leave-one-out classification among canola, brown mustard, yellow mustard, oriental mustard, and flaxseed is given in Table IX. Except for the flaxseed, there were substantial misclassifications among small seeds (Table IX). Only 45.3% of canola seeds were correctly identified. Hehn and Sokhansanj (1990) reported 99% correct classification between

Table VII: Physical dimensions of flaxseed

Feature	Flaxseed (n = 998)* Mean	Standard deviation	Range
Area (mm^2)	6.4	0.5	4.5 - 8.6
Perimeter (mm)	10.1	0.4	7.9 - 11.5
Length (mm)	3.9	0.2	2.4 - 4.4
Width (mm)	1.8	0.1	1.3 - 2.3

* n refers to sample size.

Table VIII: Identification of HRS wheat and barley from small and large seeds

Categories (to)→ (from)↓	HRS wheat	Barley	Small seeds	Large seeds
HRS wheat (n=1001)*	995 (99.4%)	1	2	2
Barley (n=999)	4	995 (99.6%)	0	0
Small seeds (n=4987)	0	0	4987 (100.0%)	0
Large seeds (n=6000)	28	3	0	5969 (99.5%)

* n refers to sample size.

canola and mustard using morphological features. Their result is in strong contrast to those presented in Table IX. Hehn and Sokhansanj (1990) used wild mustard which is noticeably smaller than domestic mustard classes used in this study. The contrasting results of the two studies show the importance of including several seed types in a classification study for reaching a reliable conclusion.

Table X shows the confusion matrix for a leave-one-out classification into the categories of buckwheat, 'Laird' lentils, 'Eston' lentils, pea beans, green peas, and black beans. Only 'Laird' lentils and buckwheat were correctly identified (Table X). Misclassification between pea beans and black beans implies the lack of discriminating ability in morphological features and stresses the need for color or reflectance-dependent features.

CONCLUSIONS

1. Hard red spring wheat and barley were identified from large (buckwheat, pea beans, black beans, lentils, green peas) and small (canola, mustard, flaxseed) seeds with more than 99% accuracy using morphological features.
2. There were substantial misclassifications (up to 54.7% for small seed group and up to 30.3% for large seed group) among seeds within each of the small and large seed groups. Flaxseed in the small seed group and 'Eston' lentils and buckwheat in the large seed group were, however, correctly identified.

Table IX: Classification among small seeds

Categories (to)→ (from)↓	Canola	Brown mustard	Oriental mustard	Yellow mustard	Flaxseed
Canola (n=992)*	449 (45.3%)	257	228	55	0
Brown mustard (n=1000)	140	733 (73.3%)	114	35	0
Oriental mustard (n=1000)	139	127	687 (68.7%)	35	0
Yellow mustard (n=997)	39	0	36	922 (92.5%)	0
Flaxseed (n=998)	0	0	0	1	997 (99.9%)

* n refers to sample size.

Table X: Classification among large seeds

Categories (to)→ (from)↓	Buckwheat	'Laird' lentils	'Eston' lentils	Pea beans	Green peas	Black beans
Buckwheat (n=999)*	993 (99.4%)	0	6	0	0	0
'Laird' lentils (n=1000)	0	697 (69.7%)	3	2	298	0
'Eston' lentils (n=999)	0	1	997 (99.8%)	0	1	0
Pea beans (n=1002)	0	1	0	880 (87.8%)	3	118
Green peas (n=1000)	0	283	2	9	701 (70.1%)	5
Black Beans (n=1000)	0	0	0	218	3	772 (77.2%)

* n refers to sample size.

ACKNOWLEDGEMENTS

The authors thank Prince Rupert Grain Ltd., Natural Sciences and Engineering Research Council of Canada, Agriculture and Agri-Food Canada, University of Manitoba Graduate Fellowship Committee, and Manitoba Department of Agriculture for financial support of this study. The authors also thank Mr. A. Verma for his technical assistance.

REFERENCES

Batchelor, B.G. 1985. Lighting and viewing techniques. In *Automated Visual Inspection*, eds. B.G. Batchelor, D.A. Hill and D.C. Hodgson, 103-179. Kempston, UK: IFS (Publications) Ltd.

Brogan, W.L. and A.R. Edison. 1974. Automatic classifications of grains via pattern recognition techniques. *Pattern Recognition* 6:97-103.

Duda, R.O. and P.E. Hart. 1973. *Pattern Classification and Scene Analysis*. New York, NY: John Wiley and Sons, Inc.

Hand, D.J. 1981. *Discrimination and Classification*. New York, NY: John Wiley and Sons, Inc.

Haralick, R.M. and L. Shapiro. 1992. *Computer and Robot Vision*. Reading, MA: Addison-Wesley Publishing Co.

Hehn, J.L., P. Shatadal, D.S. Jayas and N.R. Bulley. 1993. Image processing with "Khoros" - application to grain grading. ASAE Paper No. 93-3030. St. Joseph, MI: ASAE.

Hehn, J.L. and S. Sokhansanj. 1990. Canola and mustard seed identification using macintosh based imaging system. ASAE Paper No. 90-3534. St. Joseph, MI: ASAE.

Rosenfeld, A. and A.C. Kak. 1982. *Digital Picture Processing*, Vol. 2, 2nd ed. New York, NY: Academic Press.

Sapirstein, H.D. and W. Bushuk. 1989. Quantitative determination of foreign material and vitreosity in wheat by digital image analysis. In *International Association of Cereal Chemistry 1989 (ICC'89) Symposium on Wheat End-Use Properties*, ed. H. Salovaara, 453-474. Department of Food Chemistry and Technology, University of Helsinki, Lahti, Finland.

Sapirstein, H.D., M. Neuman, E.H. Wright, E. Shwedyk and W. Bushuk. 1987. An instrumental system for cereal grain classification using digital image analysis. *Journal of Cereal Science* 6:3-14.

SAS. 1990. *SAS User's Guide: Statistics*. Raleigh, NC: Statistical Analysis System Inc.

Shatadal, P., D.S. Jayas and N.R. Bulley. 1994. Segmentation of connected cereal grains. ASAE Paper No. 94-3022. St. Joseph, MI: ASAE.