A low-cost image analysis technique for seed size determination

S. Mandal, S. Roy and H. Tanna

A low-cost image analysis technique was developed using a flatbed document scanner and MATLAB software for determination of seed dimensions. The technique showed strong correlation between the image analysis and experimental data of length and width of corn, baby corn, pigeon pea, soybean and paddy seeds. The regression equation developed for five seeds and two parameters for each seed can be effectively used for seed size determination.

Basic physical properties of seeds are required in order to design the machines for handling, harvesting, transporting, cleaning, separating, packing and processing of seeds[1]. Determining seed dimensions by slide callipers is tedious and time-consuming. Screen sieving is widely used as a standard method of determining seed size distribution in grains. In a typical sieving practice, a set of screens with different hole sizes is used to separate seeds into various size categories. Round-hole screens are used for circular or approximately circular seeds (peas or soybeans), whereas slotted-hole screens are used for noncircular seeds (kidney beans or cranberry beans). Sieving (manual or mechanical) is inconsistent, laborious and time-consuming. It may potentially cause involuntary damage to the seed coat, which can adversely affect the visual appearance, storability and processing quality of the grains. A quick, efficient and non-destructive method for determining seed size profiles would greatly benefit grain industries. Machine vision or image analysis can be a faster, non-destructive alternative to the traditional sizing equipment currently used in the grain industry[2].

Digital image analysis technique has been developed and used to determine the physical dimensions of seeds and grains like wheat[3], rice[4], linseed[5], lentil[6], basil[7], soybean[8], rapeseed[9] and wild sage seed[7]. These techniques require digital camera, CCD camera, scanner, image analysis software, etc. which are expensive and need specialized laboratory set-up. Therefore, the objectives of the present study were to develop a low-cost method for quick determination of seed physical dimensions and to compare the seed dimensions determined by image analysis with manual method.

Materials and methods

Seed samples

Five types of seeds were used for the determination of physical dimension – corn, baby corn, paddy, soybean and pigeon pea. Seeds were collected from the seed lots of the Indian Council of Agricultural Research (ICAR) farm. Fifty seeds were randomly selected from the samples for determination of seed dimensions by manual and image analysis methods. A vernier calliper with least count of 0.001 mm was used to determine length and width of seeds manually.

Image analysis

A document scanner and computer are commonly available. MATLAB software is also being used for its vast applications in many fields. An HP scanjet (model # C7716A) document scanner was used to take images of the seeds and the images were analysed using the MATLAB 7.8.0 (The Math Works, Inc., Natick, MA, USA) software, which made it a low-cost technique. Seeds were spread over the scanner surface without touching to each other scanned with highest resolution. Care was taken to spread the seeds steadily on the scanner bed. It was done using a plyboard by pressing and shaking it gently on the seeds. This ensured that the seeds deposited with highest dimensions (length and width) on a horizontal plane. The lid of the scanner was covered by a black sheet to reduce the colour of the background of scanned images.

Each image was analysed using MATLAB software. The flow diagram of the process of image analysis for determination of physical dimensions is shown in Figure 1. The 'imread' function returns a three-layer array of RGB (red green blue) colour intensity values within a range 0–255. Preliminary examination of colour arrays of seed coats showed the highest intensity values of red colour array in all seeds. Therefore, the red colour array was taken for further analysis of the image which showed the highest difference in the intensity values of seed and background than green and blue. White background of the scanner lid produces an array of high intensity of all three colours as white represents the highest colour intensity (R,G,B = 255,255,255). As the highest intensity value could only be 255, the background must be of lower intensity so that the image of seeds can be separated from the background. Covering the scanner lid with black sheet helped reduce the colour

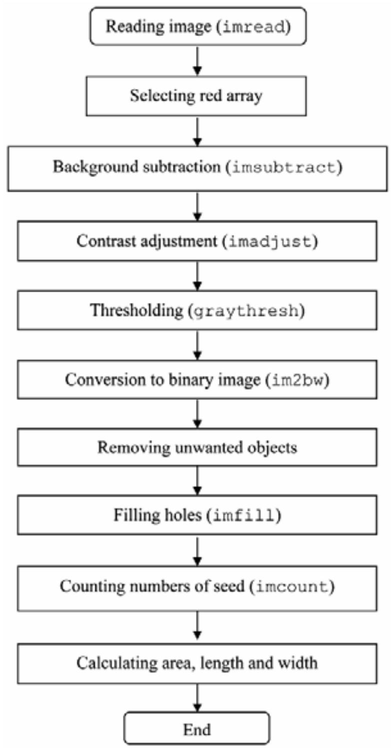

Figure 1. Flow diagram for measurement of dimension by image analysis.

intensity values of all three colours, as black represents zero intensity (R,G,B = 0,0,0). The average intensity value of the background was subtracted from the red colour array to eliminate the background from the image. This method gave better output than converting into greyscale image.

The 'imadjust' function was used to increase the contrast of the images. This function maps the intensity values in greyscale image to new values of adjusted image such that 1% of data is saturated at low and high intensities of the former image. This increased the contrast of the output image. After grey thresholding, images were converted to binary image to define the seeds on the background (Figure 2). It was performed by computing a global threshold (level) that can be used to convert an intensity image to a binary image with command 'im2bw'. The threshold level is a normalized intensity value that lies in the range [0, 1]. Function 'im2bw' converted the greyscale image to a binary image by replacing all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black)[10]. Any unwanted pixel was removed from the binary image. Holes on the seed image, if any, were filled to increase the total pixel count of the seeds and accuracy of analysis. A hole is a set of background pixels that cannot be reached by filling in the background from the edge of the image or an area of dark pixels surrounded by lighter pixels[11]. So, there will be no risk of filling the pixels in edge which may lengthen the size of seeds. Then by successive commands, the number of seeds in the image, area, length and width of each seed were calculated.

The next step in the process was to develop a scale. For this purpose, a coin was scanned and all the dimensions were calculated using the same procedure discussed above. Dimensions were also measured physically using a vernier calliper and a scale was developed by dividing the original dimensions by pixel values. This scale was used to determine area, length and width of the seeds by multiplying with pixel values.

Statistical analysis

Regression models were developed for length and width between electronic and manual measurements. Standard deviation, maximum and minimum values were calculated for each sample.

Results and discussion

The mean values, standard deviation and range of seed length, width and area are summarized in Table 1. The regression equations and linear regression lines with R^2 (coefficient of determination) values are presented in Figure 3. The mean dimensions in both methods were closely matched for all the seeds. The mean values of dimensions determined by the image analysis method exceeded the manual data within ±10%. Highest difference in mean values of manual and image analysis method was observed with width of baby corn (9.02%), whereas lowest was with length of paddy

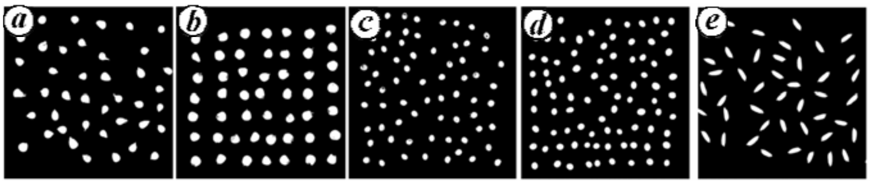

Figure 2 ***a–e***. Binary images of seeds. ***a***, Baby corn; ***b***, corn; ***c***, pigeon pea; ***d***, soybean; ***e***, paddy.

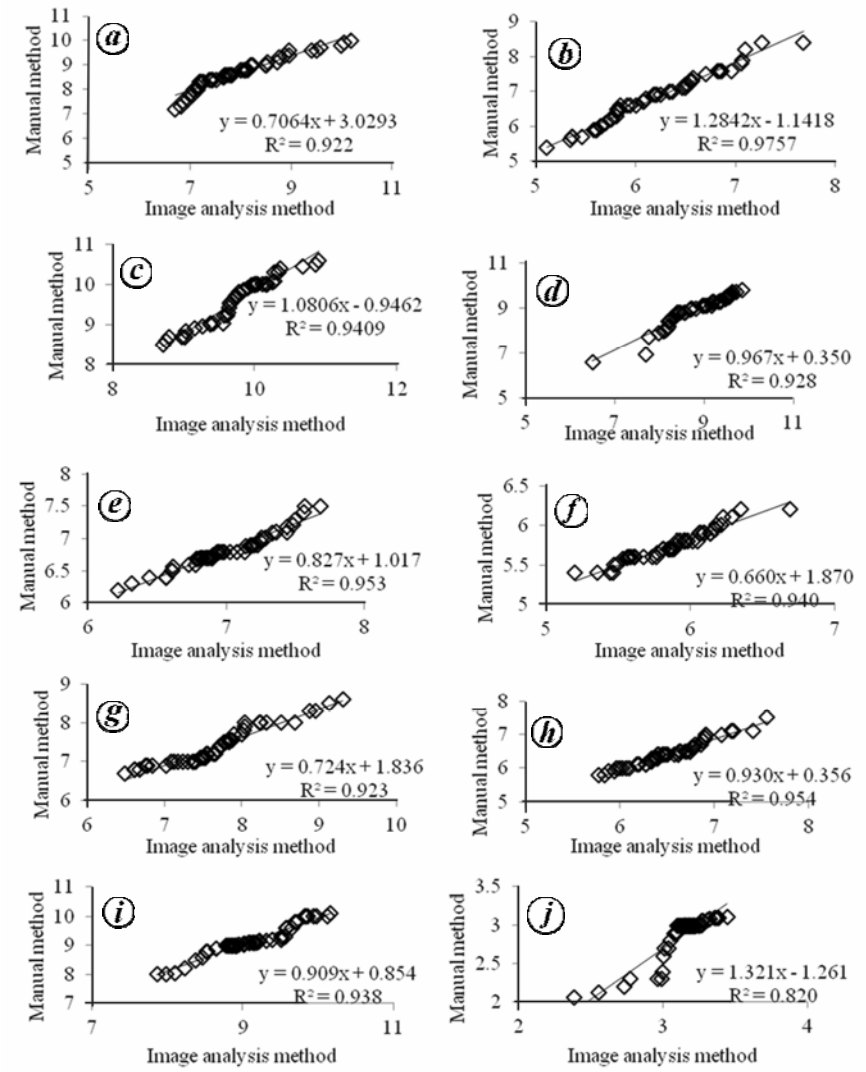

Figure 3. Regression analysis of manual and image analysis data of (***a***) baby corn length, (***b***) baby corn width, (***c***) corn length, (***d***) corn width, (***e***) pigeon pea length, (***f***) pigeon pea width, (***g***) soybean length, (***h***) soybean width, (***i***) paddy length and (***j***) paddy width.

Table 1. Dimension of seeds determined by manual and image analysis method

Seed	Geometrical features	Manual method Mean	Manual method Range	Manual method Standard deviation	Image analysis method Mean	Image analysis method Range	Image analysis method Standard deviation
Baby corn	Length (mm)	8.678	7.2–10.0	0.655	7.997	6.7–10.2	0.890
	Width (mm)	6.782	5.4–8.4	0.742	6.170	5.1–7.7	0.571
	Area (mm^2)	–	–	–	36.954	20.2–53.6	7.006
Corn	Length (mm)	9.632	8.5–10.6	0.571	9.789	8.7–10.9	0.513
	Width (mm)	8.938	6.6–9.8	0.694	8.872	6.5–9.8	0.691
	Area (mm^2)	–	–	–	67.555	49.9–83.1	7.809
Pigeon pea	Length (mm)	6.807	6.2–7.5	0.280	6.995	6.2–7.7	0.330
	Width (mm)	5.744	5.4–6.2	0.197	5.864	5.2–6.7	0.290
	Area (mm^2)	–	–	–	32.076	27.0–38.9	2.602
Soybean	Length (mm)	7.388	6.7–8.6	0.487	7.664	6.5–9.3	0.646
	Width (mm)	6.452	5.8–7.5	0.386	6.551	5.8–7.6	0.405
	Area (mm^2)	–	–	–	39.499	29.9–53.2	5.439
Paddy	Length (mm)	9.160	8.0–10.1	0.513	9.129	7.7–10.2	0.546
	Width (mm)	2.881	2.1–3.1	0.281	3.135	2.4–3.4	0.192
	Area (mm^2)	–	–	–	22.156	17.0–26.5	2.001

(0.33%). Length of baby corn in image analysis was 7.84% lower than manual method, length of corn was 1.63% higher and width was 0.74% lower. Length and width of both pigeon pea and soybean were higher by 2.76%, 3.74% and 2.09%, 1.53% respectively. Width of paddy in image analysis method was 8.82% higher than manual method. However, these differences were much lower than those reported earlier[12] as 10% for medium graded crushed limestone aggregates and overestimation by 20% for pea gravel[2]. Deviations in regression coefficient of the equations developed were low in most of the parameters but high in width of baby corn, pigeon pea and paddy. This may be due to disagreement between manual and image analysis methods of few seeds. This error can be avoided by colouring the seeds with red, which will remove any possible black spots from the seed coat.

The standard deviation values were also comparable in both methods, although these values were higher than those reported for green pea, yellow pea, chick pea and soybean[2]. Difference in standard deviation values in the two methods was highest for length of baby corn (0.235) and lowest for width of corn (0.003). The correlations between manual and image analysis method were significant and ranged from 0.820 to 0.975, but these values were less than those reported for basil and wild sage seeds[1,7]. The lower correlation for width of paddy may be due to the small value and error that may have occurred while taking measurements manually.

ANOVA (analysis of variance) test showed statistically significant difference in length and width of all seeds between the two methods. But correlation between the two methods was highly significant, which confirms the suitability of image analysis method for determination of physical dimensions. The regression equation developed for the five seeds and two parameters for each seed can be effectively used. Also the time required to determine seed size physically (≈ 1.5 h for 50 seeds) is much higher than the image analysis method (≈5 min).

Conclusions

Image analysis tool to determine seed size is an established technique that is vastly being used in various applications. The algorithm developed in MATLAB software effectively determines the dimensions of five types of seeds from scanned images. The method is quick and low cost as there are no costly equipment and software. Strong correlation between manual and image analysis data was found for all the seeds studied. The regression equations developed can be used further to determine length and width of baby corn, corn, pigeon pea, soybean and paddy utilizing the values determined by image analysis.

1. Razavi, S. M. A., Bostan, A. and Razaie, M., *J. Food Process. Eng.*, 2010, **33**, 51–64.
2. Shahin, M. A. and Symons, S. J., *Can. Biosyst. Eng.*, 2005, **47**, 3.49–3.55.
3. Shouche, S. P., Rastogi, R., Bhagwat, S. G. and Sainis, J. K., *Comput. Electron. Agric.*, 2001, **33**, 55–76.
4. Sakai, N., Yonekawa, S., Matsuzaki, A. and Morishima, H., *J. Food Eng.*, 1996, **27**, 397–407.
5. Keefe, P. D., *Plant Var. Seeds*, 1999, **12**, 79–90.
6. Shahin, M. A. and Symons, S. J., American Society of Agricultural Engineers, Paper No. 01-3058, 2001.
7. Razavi, S. M. A., Bostan, A. and Rahbari, R., *Int. J. Food Prop.*, 2010, **13**, 308–316.
8. Shahin, M. A., Symons, S. J. and Poysa, V. W., *Biosyst. Eng.*, 2006, **94**, 191–198.
9. Tańska, M., Rotkiewicz, D., Kozirok, W. and Konopka, I., *Food Res. Int.*, 2005, **38**, 741–750.
10. Otsu, N., *IEEE Trans. Syst. Man Cyb.*, 1979, **9**, 62–66.
11. Soille, P., *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, 1999, pp. 173–174.
12. Maerz, N. H., Palangio, T. C. and Franklin, J. A., In Proceedings of the International Journal of Blasting and Fragmentation. Workshop on Measurement of Blast Fragmentation, Montreal, QC, Canada, 23–24 August 1996, pp. 91–99.

Received 10 April 2012; revised accepted 8 November 2012

S. Mandal is in the Division of Agricultural Engineering, ICAR Research Complex for NEH Region, Umiam 793 103, India; S. Roy is in the National Bureau of Plant Genetic Resource, Regional Station, Umiam 793 103, India; H. Tanna is in the College of Agricultural Engineering and Technology, Godhra 389 001, India.*
**e-mail: smandal2604@gmail.com*