CAFEi2012 – 47

FEATURE EXTRACTION FOR HEAD AND BROKEN RICE DETECTION USING IMAGE PROCESSING TECHNIQUE

Siti Sharifah Bibi Hanibah, Siti Khairunniza Bejo*, Wan Ishak Wan Ismail, Aimrun Wayayok

Department of Biological and Agricultural Engineering, Faculty of Engineering
University of Putra Malaysia.
*Email: skbejo@eng.upm.edu.my

ABSTRACT

Rice (Oryza Sativa) is the most important staple food for a large part of human population, especially in Southeast Asia such as Malaysia and Indonesia. Rice has been graded based on three main components namely: grain composition, milling quality and defective parts. Rice grading is important to ensure only edible rice reaches the consumer standard. It also protects consumers from price manipulation. In this paper, a new approach of image processing technique has been developed to extract rice features. The features used were area, perimeter, minor axis length and major axis length of the rice. The rice images were first segmented automatically from its background by using Otsu's method. Morphological operation was later being applied to the segmented image in order to eliminate unwanted region(s). Results from the experiment have shown that area gave more consistent results of head and broken rice detection compared to the other features. It is due to the difference in surface coverage area of the rice. Meanwhile, minor axis length gave the worst results due to same value for both broken and head rice. The method give the overall accuracy of 98% when tested using 600 samples of rice image taken from six different percentage of broken.

***Keywords**: rice, Otsu, area, perimeter, minor axis length and major axis length*

INTRODUCTION

Rice grading is a process of removing the broken rice from head rice and sorting them into well defined fractions of different length. Optical inspection is the final quality control and enhancement step in the rice mill. To achieve a uniform and consistent quality in packed white rice, the various white rice fractions are blended according to the market requirement and trade standards. The marketable white rice is then packed and sent out to the customers either locally or for export.

Graded rice is classified into three grades namely standard, which refers to rice with high broken content (up to 45%) and may contain 100% damaged grain; premium which refers to rice with high broken content (up to 45%) but its damaged grain should not exceed 2% and super, which is further divided into three different grades, largely based on its broken content, namely Super Tempatan 15%, which content about 15% broken rice. Super Spesial Tempatan 10%, which contain about 10% broken rice and Super Spesial Tempatan 5%, which contain about 5% broken rice. Grain composition is one of the components used for rice grading. It refers to head rice and broken rice. Head rice consists of whole kernel or at least 8 part of kernel. Figure 1 shows the diagram of head and broken rice graded by Padiberas Nasional Berhad (BERNAS).

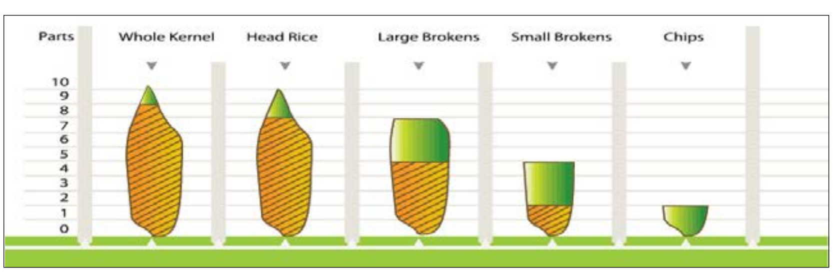

Fig. 1: Rice grain composition [1]

Today's image processing techniques has become an increasingly popular and cost effective method in rice grading. Yoshioka et al. [2] has evaluated the effectiveness of image information processing to measure and categorize chalkiness. It gave an accuracy rate of 90.2% in discriminating level of chalkiness. Aghayeghazvini et al. [3] has determined percentage of broken rice by using image analysis. This work indicates that the digital image processing technique can be used for estimating broken grains. Determination of the size and size distribution of rice and the amount of broken rice kernels using flatbed scanning and image analysis was developed by Dalen [4]. It was concluded that image analysis yields the same accuracy and better precision than manual time consuming method. A digital image analysis algorithm based on 7 color features namely mean value of red (R), mean value of green (G), mean value of blue (B), mean value of hue (H), mean value of saturation (S), mean value of Intensity (I), the standard deviation of H and morphological features such as area, length, width, major axis length, minor axis length and others were developed by Liu et al. [5] to identify rice seeds planted in Zhejiang province. These data sets were later classify using Neural Network and obtained overall of more than 70% of accuracy.

Sakai, Yonekawa and Matsuzaki [6] analyze the shapes of Japonica, Indica and Javanica rice types with three different polishing methods. Area, perimeter, maximum length, maximum width, compactness and elongation were measured. This result shows the rice varieties were distinguishable at a probability level of 95.4% with combined or single dimensions. Honda et al. [7] applied an image analysis technique for automatic selection of regenerated rice. When the RGB values for each pixel were input into a fuzzy neural network (FNN), the shoot, callus, and medium regions were identified. The identification correctness of the FNN model was 95%.

Lin, Chen and He [8] developed a method based on flatbed scanning and image processing. These included image segmentation, denoising and thresholding to obtain the binary images. Bwboundaries function of Matlab R2009a [9] was used to trace region boundaries; later velocity representation method was developed for pattern recognition based on the contour characteristic of the rice kernels. High recognition rates for three rice varieties were reached by this method with 96.7% for Thailand rice, 98.73% for Pearl rice and 97.14% for Changlixiang rice respectively. Therefore broken rice are produced as a result of milling process, as known consumers prefer as few broken rice kernel as possible and this may also one of the factor influencing the price of rice, therefore the study to determine the broken rice with the aid of image processing is even more important nowadays.

In this paper, a new approach of image processing technique has been developed to extract rice features. The features used were area, perimeter, minor axis length and major axis length of the rice. The rice images were first segmented automatically from its background by using Otsu's method. Morphological operation was later being applied to the segmented image in order to eliminate unwanted regions.

MATERIALS AND METHODS

Image acquisition

In this study, the MR219 rice variety was used. It was obtained from BERNAS, Kedah. Images of rice were acquired using scanner (model Canoscan LIDE 20). The rice samples were placed on the glass plate of the scanner separately and covered with a black piece of board during image acquisition. All of the rice samples were randomly separated into training set and testing set. For training set, 100 samples of each broken rice and head rice were used for determining the milled rice characteristic. Meanwhile in testing set, 100 rice samples were scanned for different percentage of broken rice, namely 0%, 5%, 10%, 15% and 20%. Therefore, a total of 600 scanned images were used in testing datasets. Each sample was weighed 1 gram i.e. if 20% broken rice sample were to produce, 0.8 gram of head rice and 0.2 gram of broken rice were mixed and scanned.

Image segmentation

Otsu method has been used to automatically segment the rice image with the background. It will produce the binary image of rice samples. After that, a connected component analysis and morphological operations has been applied to remove unwanted noises which are defined as region with number of pixels less than 100 pixels.

Otsu's method

Otsu's method will search for the threshold that minimizes the intra class variance (within class variance), defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \qquad (1)$$

Weights ω_i are the probabilities of the two classes separated by a threshold t and σ_i^2 variances of these classes.

Otsu shows that minimizing the intra class variance is the same as maximizing inter class variance.

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = \omega_1(t)\omega_2(t)\left[\mu_1(t) - \mu_2(t)\right]^2 \qquad (2)$$

This is expressed in terms of class probabilities ω_i and class means μ_i.

The class probability $\omega_1(t)$ is computed from the histogram as t:

$$\omega_1(t) = \Sigma_0^t p(i) \qquad (3)$$

While the class mean $\mu_1(t)$ is:

$$\mu_1(t) = \Sigma_0^t p(i)\, x(i) \qquad (4)$$

Where $x(i)$ is the value at the center of the i th histogram bin. Similarly, $w_2(t)$ and μ_t can be computed on the right hand side of the histogram for bins greater than t.

The class probabilities and class means can be computed iteratively. This idea yields an effective algorithm. This operation is performed using *graythresh* function available in Matlab R2007a (The Math Works, Natick, USA).

Feature extraction

Physical rice properties have been used during rice grading. Head rice is identified as the rice with whole kernel or at least 8 part of kernel (Figure 1). Therefore, in this study physical properties of the rice will be used as the features. These include:

(a) Area
Area is defined as the actual number of pixels in the region.

(b) Perimeter
Perimeter is defined as the distance around the boundary of the region.

(c) Minor axis length
Minor axis length is defined as the length in pixels of the minor axis of the ellipse that has the same normalized second central moments as the region.

(d) Major axis length
Major axis length is defined as the length in pixels of the major axis of the ellipse that has the same normalized second central moments as the region.

regionprops function in Matlab was used to find all of these properties of each tested images. The databases of these properties for broken and head rice were built. The mean for each property were then calculated. The next stage is to identify ideal threshold value of the rice to remove the unwanted region in testing datasets, resulting image with only head rice left. The successful of this experiment was later tested with validation datasets.

RESULTS AND DISCUSSION

Feature extraction

Table 1 and 2 shows the value of correlation coefficient of rice features based on 100 samples of broken rice and head rice during training datasets. The correlation coefficient measures the degree to which two things vary together or oppositely. The maximum positive correlation is 1.00. The meaning of zero or near zero correlation is simply that the two things are vary separately. For the head rice (table 1), area is significantly correlated with perimeter, major axis length and minor axis length at the value of correlation coefficient greater than 0.6 in all conditions. Meanwhile, perimeter and major axis length does not significantly correlated with minor axis length; and minor axis length does not significantly correlated with area and major axis length. The area is also give significant correlation when compared to the other parameters in a broken rice with the value correlation greater than 0.6 (table 2) in all conditions. Therefore, area is chosen as the parameter to be used to develop the broken rice detection algorithm because it is highly correlated with other features that were studied in this research.

Table 1: Correlation coefficient table for head rice

Correlations (Spreadsheet1)
Marked correlations are significant at p < .05000
N=100 (Casewise deletion of missing data)

Variable	Means	Std.Dev.	area	perimeter	major	minor
area	1830.730	142.3511	1.000000	0.829783	0.659978	0.815169
perimeter	207.025	9.3935	0.829783	1.000000	0.841272	0.445795
major	90.211	3.8058	0.659978	0.841272	1.000000	0.116292
minor	26.419	1.6069	0.015169	0.445795	0.116292	1.000000

Correlation coefficient

Table 2: Correlation coefficient table for broken rice

Correlations (Spreadsheet2)
Marked correlations are significant at p < .05000
N=100 (Casewise deletion of missing data)

Variable	Means	Std.Dev.	area	perimeter	major	minor
area	1257.520	177.5105	1.000000	0.944623	0.884363	0.613185
perimeter	155.012	14.1149	0.944623	1.000000	0.918727	0.445736
major	61.685	6.7334	0.884363	0.918727	1.000000	0.178854
minor	26.615	1.7657	0.613185	0.445736	0.178854	1.000000

Correlation coefficient

Image segmentation

Figure 2 shows results of image segmentation. In this figure, 1 gram of rice sample image which consist of 20% of broken content was scanned. The image was thresholded using Otsu's method. Later, the connected component analysis was performed to its binary thresholded image to produce region of interest and thus remove unwanted region. It takes an average of 3 to 4 seconds to complete the iteration to produce final binary image that consist only head rice. The method is capable to detect broken rice with high accuracy.

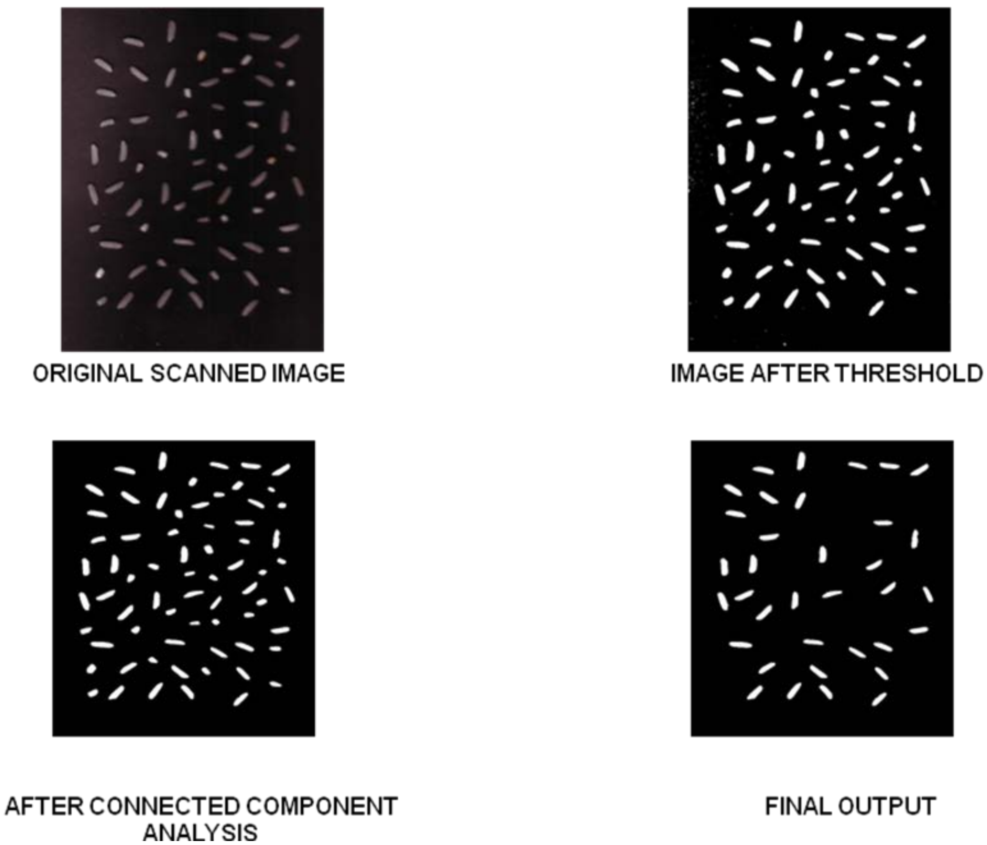

Fig. 2: Image segmentation using Otsu's method and connected component analysis

During a testing stage, 100 rice samples were scanned in different percentage of broken rice, namely 0%, 5%, 10%, 15% and 20%. Based on the results shown in table 4, 0% broken rice has 97% of success, meanwhile 1%,10% and 15% broken rice give 98% of success and finally 5% and 20% broken rice give 99% of success. In overall, the testing datasets give 98% of success in broken rice detection (calculated from the mean).

Table 4: Mean percentage of success in testing datasets

Samples	Percentage of success
0% broken rice	97
1% broken rice	98
5% broken rice	99
10% broken rice	98
15% broken rice	98
20% broken rice	99

Mean	98

CONCLUSIONS

In this study, four different types of rice image features had been analyzed for head and broken rice detection. Area has been identified as the most appropriate features when compared to perimeter, minor axis length and major axis length. A personal computer with a scanner was used to obtain the rice image. The scanned rice image was thresholded using Otsu's method, producing a binary image. Later, a connected component analysis was performed to produce regions of interest. Head rice is identified as a region with the area greater than 100 pixels. This method is fast and easy to be used. It only takes 3 to 4 seconds to complete the process of detection. It successfully detected 600 samples of broken rice taken from 6 different percentage of broken, with 1 gram each with the overall percentage accuracy of 98%. In conclusion, the method which is developed using MATLAB software can be used to detect broken rice images using its area as the feature.

ACKNOWLEDGEMENT

This work was financially supported by RUGS.

REFERENCES

[1] http://www.bernas .com.my (obtained on June 13, 2012)

[2] Yosuke Yoshioka, Hiroyoshi Iwata, Minako Tabata, Seishi Ninomiya and Ryo Ohsawa, 2007. Chakiness in rice: Potential for evaluation with image analysis. Crop Science Society of America, 47: 2113-2120.

[3] H. Aghayeghazvini, A. Afzal, M. Heidarisoltanabadi, S. Malek, L. Mollabashi, 2009. Determining percentage of broken rice using image analisys. Computer and Computing Technologies in Agriculture, 2:1019-1027.

[4] G.V. Dalen, 2003. Determination of the size distribution and percentage of broken kernels of rice using flatbed scanning and image analysis. Food Research International, 37:51-58.

[5] Liu Zhao Yan, Cheng Fang, Ying Yi Bin, Rao Xiu Qin, 2005. Identification of rice seed varieties using neural network. Journal of Zhejiang Universitiy Science, 6B(11): 1095-1100.

[6] N. Sakai, S. Yonekawa, A. Matsuzaki and H. Morishima, 1999. Two dimensional image analysis of the shape of rice and its application to separating varieties. Journal of Food Engineering, 27: 397-407.

[7] Hiroyuki Honda, Naoto Takikawa, Hideki Noguchi, Taizo Hanai and Takeshi Kobayashi, 1997. Image analysis associated with a fuzzy neural network and estimation of shoot length of regenerated rice callus. Journal of Fermentation and Bioengineering, 84: 342–347.

[8] Ping Lin, Yongming Chen and Yong He, 2010. Identification of Broken Rice Kernels Using Image Analysis Techniques Combined with Velocity Representation Method. Food and Bioprocess Technolog. An International Journal.

[9] http://www.mathworks.com (obtained on July 4, 2012)