

Original papers

Identifying rice grains using image analysis and sparse-representation-based classification

Tzu-Yi Kuo^a, Chia-Lin Chung^b, Szu-Yu Chen^b, Heng-An Lin^b, Yan-Fu Kuo^{a,*}^a Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan^b Department of Plant Pathology and Microbiology, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:

Received 23 February 2016

Received in revised form 11 June 2016

Accepted 16 July 2016

Keywords:

Cultivar identification

Sparse coding

Locality constraint

Machine vision

Machine learning

Image processing

ABSTRACT

Rice (*Oryza sativa* L.) is a major staple food worldwide, and is traded extensively. The objective of this study is to distinguish the rice grains of 30 varieties nondestructively using image processing and sparse-representation-based classification (SRC). SRC uses over-complete bases to capture the representative traits of rice grains. In the experiments, rice grain images were acquired by microscopy. The morphological, color, and textural traits of the grain body, sterile lemmas, and brush were quantified. An SRC classifier was subsequently developed to identify the varieties of the grains using the traits as the inputs. The proposed approach could discriminate rice grain varieties with an accuracy of 89.1%.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Rice (*Oryza sativa* L.) is an essential staple food for half of the global population. Rice grains of hundreds of varieties are cultivated and traded in considerable amounts worldwide. Most facilities share the same equipment for handling various products. Rice grains of different varieties can be mixed during the cultivation, harvesting, transporting, and processing, reducing the purity, quality, and value of the subsequent products. Despite the neglect, the introduction of impurities can also be intentional, in a dishonest manner. Hence, the demand for the nondestructive authentication of grain varieties is emerging. The grains of different varieties can vary in appearance (Fig. 1). Specialists traditionally identify grains manually based on their appearance. However, this process is time-consuming and subjective. This study proposes identifying the rice grains of 30 varieties (Appendix A) through image analysis and sparse-representation-based classification (SRC; Wright et al., 2009) techniques.

Various genetic marker-based methods have been applied for identifying rice grain varieties. Steele et al. (2008) selected insertion and deletion markers to distinguish Basmati rice grains from some other fragrant rice varieties. Cirillo et al. (2009) applied random amplified polymorphic DNA (RAPD) approach to fingerprint

rice grains of 13 Italian accessions. Becerra et al. (2015) determined the genetic variability of certain Chilean and foreign commercial rice cultivars using simple sequence repeat (SSR) markers. Another study reported using SSR markers to distinguish 36 varieties of rice grains from different countries (Chuang et al., 2011). Although these genetic marker-based methods are accurate, they are often too time-consuming or costly to be suitable for online applications.

Image-based approaches, by contrast, are nondestructive and rapid. They combine image analysis and machine learning techniques to achieve automatic inspection and evaluation. Image-based approaches have been applied for discriminating varieties of cereal grains using either one of the morphological, color, and textural traits, or a combination. Camelo-Méndez et al. (2012) characterized the rice grains of 9 Mexican cultivars by performing principle component analysis (PCA) and hierarchical analysis. Kong et al. (2013) classified the rice seeds of 4 accessions using a near-infrared hyperspectral imaging system and various machine learning algorithms. Mebatsion et al. (2013) distinguished between barley, oat, and rye using a least-squares classification approach. Another study applied multilayer perceptron and neuro-fuzzy classification networks for identifying 5 Iranian rice varieties (Pazoki et al., 2014). Although the results of these studies have been promising, they have included a relatively limited number of varieties for discrimination.

SRC is a machine learning algorithm that is suitable for solving high-dimensional problems. It encodes the representative characteristics of training samples as the atoms in a dictionary. When a

* Corresponding author at: Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan.
E-mail address: ykuo@ntu.edu.tw (Y.-F. Kuo).

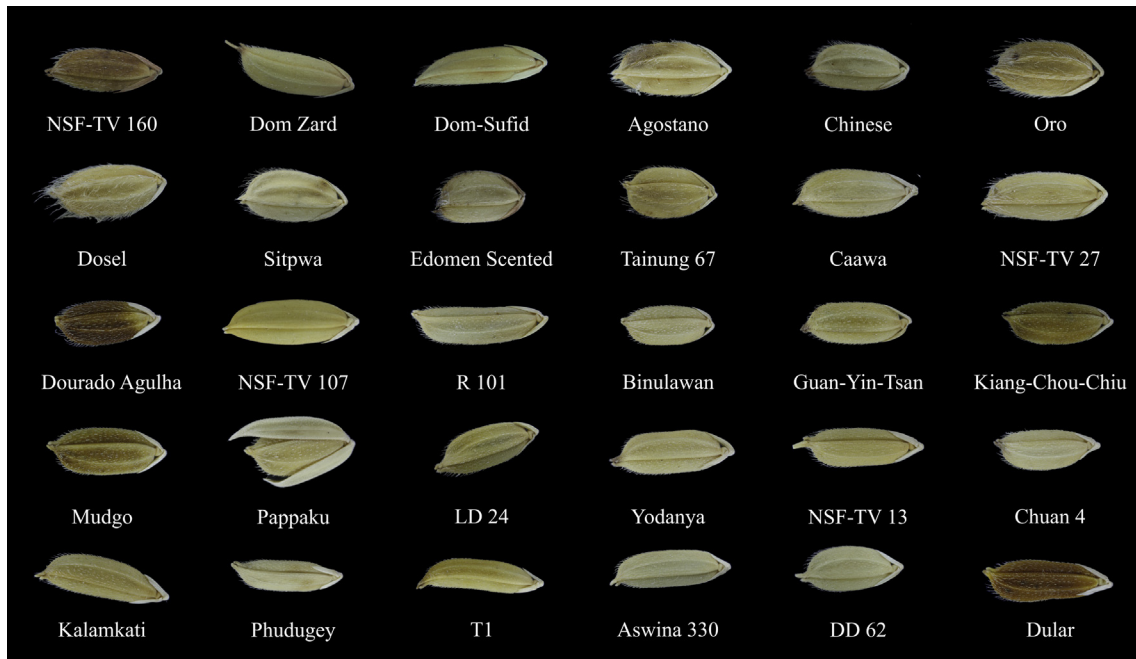


Fig. 1. Rice grains of 30 varieties.

query sample is provided, it is coded as a sparse combination of atoms. The query sample is then assigned to the class that yields the least coding errors. SRC is robust to noise and computationally inexpensive. The method has been used for denoising images detecting the flowers of various species (Yuan et al., 2012), recognizing field crop insects (Xie et al., 2015), and assisting with the diagnosis of Alzheimer's disease (Liu et al., 2012).

The current study aimed to differentiate the rice grains of 30 varieties using locality-constrained SRC (Wei et al., 2013). The following were the specific objectives of the study: (1) establish a microscopic imaging system for acquiring grain images; (2) quantify the morphological, color, and textural traits of the grain body and parts; and (3) develop a locality-constrained SRC classifier for identifying rice grain varieties.

2. Materials and methods

2.1. Sample preparation

The rice grains of 30 varieties were used in this study (Fig. 1; Appendix A). *Oryza sativa* L. consists of 5 genetic subpopulations (Garris et al., 2005): *indica*, *aus*, *aromatic*, *temperate japonica*, and *tropical japonica*. The varieties used in this study were selected from these subpopulations, including 7 from *indica*, 8 from *aus*, 3 from *aromatic*, 7 from *temperate japonica*, and 5 from *tropical japonica*. The grain samples were acquired from the Genetic Stocks *Oryza* germplasm collection (Agricultural Research Service, United States Department of Agriculture) and reproduced in a local greenhouse (Kaohsiung District Agricultural Research and Extension Station, Taiwan) in 2013. For the details of the rice varieties, refer to Zhao et al. (2011). After harvesting, the grains were dried to a moisture content of approximately 13% and refrigerated at 4 °C. Fifty grains of each variety were prepared.

2.2. Rice grain exterior

A grain body is typically composed of husk and sterile lemmas (Fig. 2). In certain varieties, brush covers the husk surface. The

morphological, textural, and color traits of these organs (i.e., husk, sterile lemmas, and brush) can be used for distinguishing between grain varieties. Because the awns of the grains typically shed during drying, their traits were not considered in this study.

2.3. Imaging system and image acquisition

Fig. 3 displayed the image acquisition system developed for acquiring the grain images. It comprised a digital camera (EOS 450D, Canon; Tokyo, Japan), a microscope (BXFM, Olympus; Tokyo, Japan), a 2X objective lens (PLN UIS2, Olympus; Tokyo, Japan), and a ring-shaped LED illuminator. The LED illuminator was placed 15 mm above the surface of the sample placement platform of the microscope. The system was enclosed in a dark chamber to prevent exposure to stray light. The barrel or pincushion distortion of the microscope system was measured using a checkerboard, and the distortion level was negligible. Before image acquisition, the system was calibrated using a standard color reference board (Color Checker Passport, X-rite; Grand Rapids, USA) to estimate the device-independent color parameters of the grains. The camera was set in manual mode for image acquisition with an ISO of 400 and a shutter of 1/30 s. The acquired images were saved in raw format, in which no adjustment (e.g., white balance) was applied.

2.4. Multifocus image fusion and background removal

Multifocus image fusion (Wang and Chang, 2011) was applied to improve the quality of the grain images (Figs. 4a–c). The lenses of an optical microscope typically have a limited field depth. Two micrographic images of the same rice grain, one focused at the grain center (Fig. 4a) and the other at the grain edge (Fig. 4b), were obtained. The 2 photographs were merged to obtain an image with all of the pixels in focus (Fig. 4c). The fusion involved matching, registration, and consolidation. In the matching process, the same characteristic points of the rice grain in the 2 photographs were identified using speeded up robust features (Bay et al., 2006). Next, the 2 images were registered using the identified characteristic points and the fast approximate nearest-neighbor search algorithm

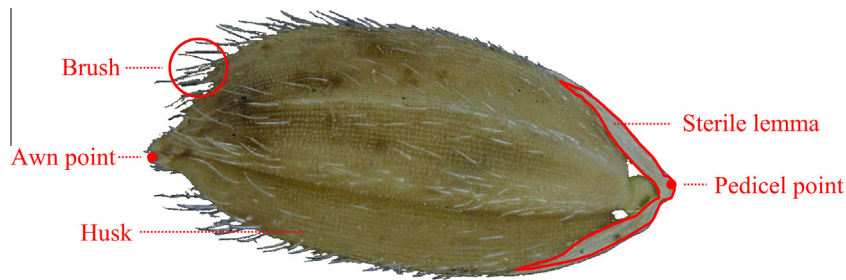


Fig. 2. Husk, sterile lemmas, and brush of a rice grain. The husk is the outermost layer. Sterile lemmas are 2 flowerless bracts connected to the pedicel. Brush is the hair on the husk, and is clearly visible in certain varieties.

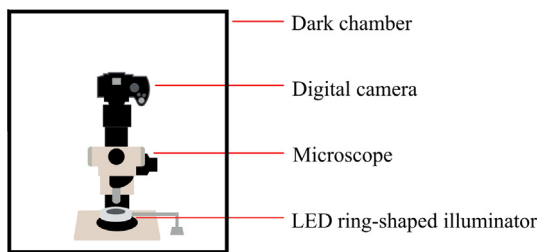


Fig. 3. Image acquisition system.

(Muja and Lowe, 2009). The 2 images were subsequently consolidated into a single image through Laplacian pyramid transformation (Jin et al., 2005). The fused image contained 1068×712 pixels at a resolution of approximately 95 pixels per millimeter (2413 pixels per inch). The image fusion algorithm was implemented in C language, and the program was developed using Xcode (Apple Inc.; Cupertino, CA, USA).

The rice grain in each image was segmented from the background (Fig. 4d). First, a k -means operation (Hartigan and Wong, 1979) was applied to the hue channel in the hue-saturation-value color space. The number of clusters in the k -means operation was set to 2. The algorithm labeled each pixel in the image “foreground” or “background.” Next, connected-component labeling (Dillencourt et al., 1992) was applied to identify the largest foreground object as the grain body. For certain varieties, the brush was recognized as a part of the grain body (Fig. 4d). This false recognition reduced the accuracy in estimating the grain perimeter. Therefore, morphological closing (Gonzalez and Woods, 2007) was performed to remove the brush outside the grain contour from the grain body (Fig. 4e). After grain body segmentation was complete, the awn and pedicel points were determined as the 2 points on the grain contour with the largest distance apart. The sterile lemmas and brush (Fig. 4f), which are usually associated with colors different from the color of the husk, were identified through color thresholding.

2.5. Trait quantification

To describe the characteristics of the rice grains, their traits were categorized into 4 groups: morphological traits, color traits, textural traits, and Fourier descriptors. Twelve morphological traits were calculated for the grain body: perimeter, surface area, length of major axis, length of minor axis, aspect ratio, arc ratio, standard deviation (SD) of radii, maximum radius, minimum radius, radius ratio, Haralick ratio, and thinness ratio (Fig. 5; Camelo-Méndez et al., 2012; Majumdar and Jayas, 2000). The major axis was defined as the line connecting the awn and pedicel points. Conversely, the minor axis was defined as the line perpendicular to the major axis with the longest segment intersecting the grain. The aspect ratio was the ratio of the major axis length to the minor axis length. The arc ratio was the ratio of the major arc length to the minor arc length. The major and minor arcs were the long and short contour segments, respectively, along the grain contour between the awn and pedicel points. The radii were the distances between the contour points and geometric centroid of the grain. The radius ratio was the ratio of the maximum radius to the minimum radius. The Haralick ratio was the ratio of the mean radius to the SD of the radii. The thinness ratio was the ratio of the area to the perimeter. A grain with a larger thinness ratio is more circular, whereas that with a smaller thinness ratio is pointier. The morphological traits of the sterile lemmas were also quantified, and included the lengths along the major arc (L1 in Fig. 5) and the minor arc (L2 in Fig. 5), the ratio of L1 to the major arc, the ratio of L2 to the minor arc, the area of the sterile lemmas, and area ratio. The area ratio was defined as the ratio of the sterile lemma area to the grain body area.

Nine color traits were acquired for the husk: red, green, and blue (RGB) parameters; hue, saturation, and value parameters; and Commission Internationale de l'Eclairage (CIE) L^* , a^* , and b^* (Hunter, 1975) parameters. The color traits were the mean color parameters of the pixels in the region of interest (ROI). The ROI was a rectangle centered at the grain centroid, with its edges parallel to the major and minor axes of the grain. The length and width of the ROI were 50% of the lengths of the major and minor

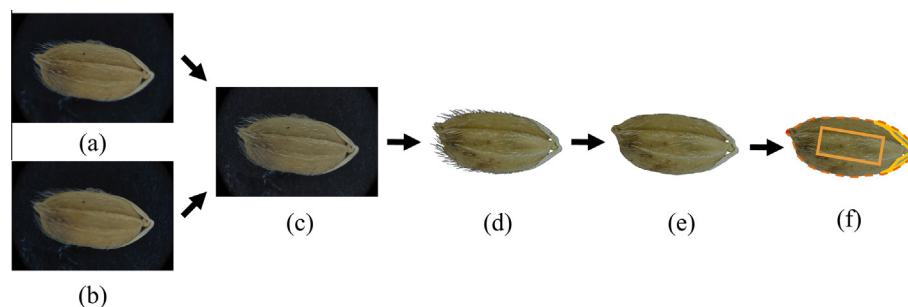


Fig. 4. Multifocus image fusion and background removal. Grain images with the (a) center and (b) edge in focus; (c) image fused from (a) and (b); (d) foreground image; (e) brush-eliminated image; and (f) sterile lemmas and region of interest.

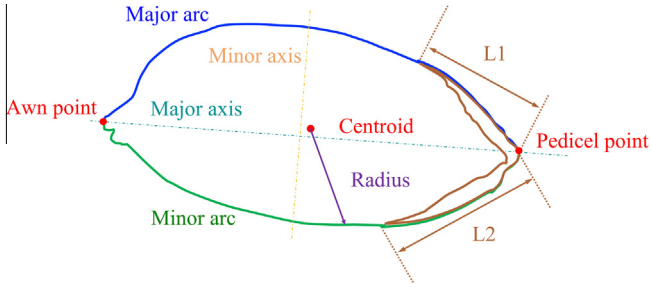


Fig. 5. Morphological traits of the grain body and sterile lemmas. L1 and L2 denote the lengths of the sterile lemmas along the major and minor arcs, respectively.

axes, respectively. The L^* , a^* , and b^* parameters were converted from the ROI RGB values using the transform functions obtained during calibration. These 9 color parameters were also quantified for the sterile lemmas using the entire area of the sterile lemmas as the ROI.

Seven textural traits were assessed for the husk: brush ratio and 6 gray-level co-occurrence matrix (GLCM; Haralick et al., 1973) traits. The brush ratio was defined as the percentage of the husk ROI covered with brush. Here the brush was identified using edge detection and color thresholding. The 6 GLCM traits included the mean, variance, uniformity, entropy, contrast, and correlation (Galloway, 1975). To quantify the GLCM traits, the ROI was first converted into a grayscale image with a 3-bit word length. The GLCM of this image was then calculated using a displacement vector with a direction parallel to the major axis and a distance of 1 pixel. The 6 traits were subsequently calculated from the GLCM following their definitions (Haralick et al., 1973).

Fourier descriptors were quantified for the grain body. Fourier descriptors are a set of sine and cosine harmonics at various frequencies for encoding the outline of an object (Rohlf and Archie, 1984). To calculate the descriptors, the grain contours were first represented as sequentially connected points (8-connected) in a Cartesian coordinate system. The coordinates of the connected points in each dimension were then converted into Fourier descriptors by using discrete Fourier transform (Harris, 1978). The first 20 low-frequency descriptors were retained for characterizing the grain shapes. The harmonics at high frequencies were disregarded because the grain images had a limited resolution. The high-frequency harmonics may depict noise, rather than the native structures of the grain contours. The Fourier power of the first 20 harmonics was 99.9% (Costa et al., 2009). The grain images were two-dimensional, with a sine and a cosine harmonic in each dimension. Thus, a total of 80 coefficients of the descriptors were collected as the traits.

2.6. Grain shape variation

The shape variations among the grains of the 30 varieties were examined. PCA was first conducted on the Fourier descriptors to obtain the principal components (PCs). Each PC was associated with a particular grain shape variation, and the PCs were arranged in descending order based on their percentage variance. The first few PCs accounted for a large proportion of the variance and could represent the major shape variations. Next, the shape variation associated with each PC was visualized by reconstructing the grain contours. In this process, the mean and SD of the PCs were calculated. Fourier descriptors were next derived through inverse PCA, with a specific PC value being manipulated while maintaining the mean values of the others. The manipulated PC values were set to be the mean and mean ± 2 SD (Williams et al., 2013). Grain contours were then reconstructed using the resulting Fourier descriptors and by conducting inverse Fourier transform.

2.7. Grain variety identification

Locality-constrained SRC was applied for identifying the rice grain varieties. The objective of SRC was to encode the grain traits as a sparse linear combination of a dictionary, which is composed of atoms that represent the essences of the grain traits. The dictionary was learned from the collected traits. The learning process is explained as follows. Consider that $N \in \mathbb{N}$ training samples (i.e., number of grain images) of the j th variety, where $j = 1 \dots J \in \mathbb{N}$ were gathered. Each training sample was associated with $M \in \mathbb{N}$ traits. Assume that the dictionary has $K \in \mathbb{N}$ atoms. Let $\mathbf{X}_j \in \mathbb{R}^{M \times N}$, $\mathbf{D}_j \in \mathbb{R}^{M \times K}$, and $\mathbf{A}_j \in \mathbb{R}^{K \times N}$ denote the matrix of the training samples, the dictionary to be trained, and the sparse coefficient matrix, respectively, of the j th variety. The dictionary was obtained by solving the following equation:

$$\begin{aligned} \mathbf{D}_j = \arg \min_{\mathbf{D}_j, \mathbf{A}_j} & \|\mathbf{X}_j - \mathbf{D}_j \mathbf{A}_j\|_2^2 + \lambda \sum_{i=1}^N \|\mathbf{p}_{ji} \odot \boldsymbol{\alpha}_{ji}\|_2^2 \\ \text{s.t. } & \mathbf{1}^T \boldsymbol{\alpha}_{ji} = 1, \quad i = 1, \dots, N \end{aligned} \quad (1)$$

where $\boldsymbol{\alpha}_{ji} \in \mathbb{R}^{K \times 1}$ is the i th column of \mathbf{A}_j , $\lambda \in \mathbb{R}$ is a regularization parameter that controls the sparsity of $\boldsymbol{\alpha}_{ji}$, $\mathbf{p}_{ji} \in \mathbb{R}^{K \times 1}$ is the locality adaptor, and the symbol \odot denotes element-wise multiplication. The locality adaptor \mathbf{p}_{ji} was defined as a vector consisting of the Euclidean distances between a training sample (i.e., a column of \mathbf{X}_j) and the columns of \mathbf{D}_j . This regularization formulation using the locality adaptor considered the underlying manifold structures of the grain traits (Wang et al., 2010). Hence, SRC could produce a dictionary that encodes the essential patterns of the grains. The dictionaries for all varieties were developed using the locality-sensitive dictionary learning algorithm proposed by Wei et al. (2013). The optimal regularization parameter λ and dictionary size K were determined by conducting a grid search and 10-fold cross-validation (Arlot and Celisse, 2010).

Once developed, the dictionaries \mathbf{D}_j were employed to classify the grains using the traits as the inputs. Let $\mathbf{y} \in \mathbb{R}^{M \times 1}$ be a query sample. The sparse coefficients $\boldsymbol{\alpha}^j \in \mathbb{R}^{K \times 1}$ were calculated for all varieties ($j = 1 \dots J$) to assemble the sparse reconstructions $\hat{\mathbf{y}} \equiv \mathbf{D}_j \boldsymbol{\alpha}^j$ of the query sample. The variety of the query sample was then determined as a variety associated with the dictionary that yields the minimum reconstruction error:

$$\begin{aligned} \text{cultivar}(\mathbf{y}) = \arg \min_j & \|\mathbf{y} - \mathbf{D}_j \boldsymbol{\alpha}^j\|_2^2 + \lambda \|\mathbf{p}_j \odot \boldsymbol{\alpha}^j\|_2^2 \\ \text{s.t. } & \mathbf{1}^T \boldsymbol{\alpha}^j = 1, j = 1, \dots, J \end{aligned} \quad (2)$$

where \mathbf{p}_j represents the Euclidean distances between the sample \mathbf{y} and the columns of \mathbf{D}_j . The dictionary training and variety identification were performed using MATLAB (The MathWorks; Natick, MA, USA).

3. Results and discussion

3.1. Morphological and textural traits

Figure 6 shows the boxplots of (a) area, (b) aspect ratio, (c) area ratio, (d) ratio of L2 to the minor arc, (e) brush ratio, and (f) GLCM entropy. The traits were color-coded by their subpopulations, where purple, blue, green, red, and yellow stand for *aromatic*, *temperate japonica*, *tropical japonica*, *indica*, and *aus*, respectively. Variations in the traits were observed. Some varieties were associated with larger areas (e.g., varieties NSF-TV 27, NSF-TV 107, and Caawa; Fig. 6a) or longer major axis (e.g., varieties Dom-Sufid, T1, and Dom Zard; Fig. 6b) compared with some other cultivars. In addition, the aspect ratio of subpopulation *temperate japonica* is significantly smaller than the aspect ratios of subpopulations

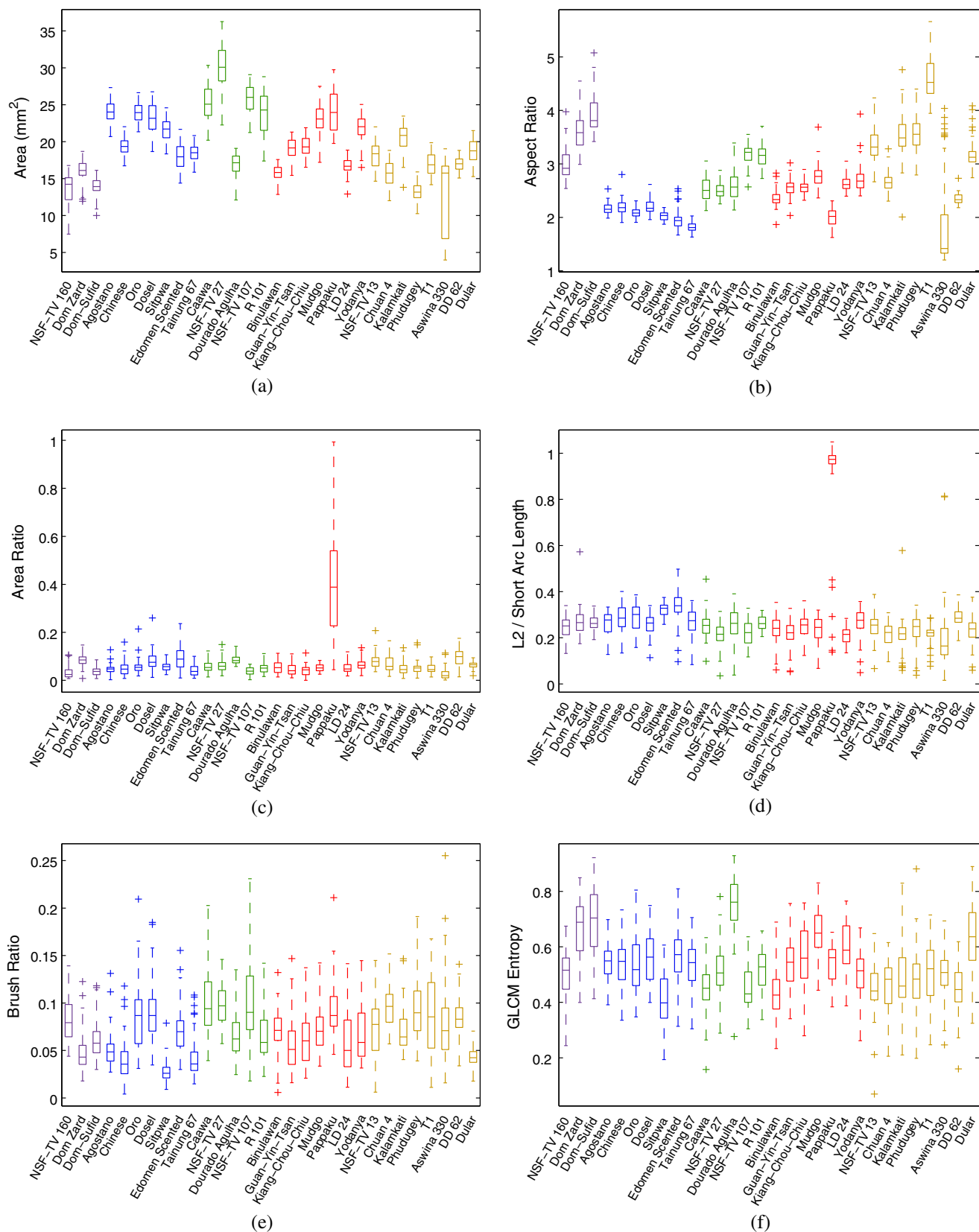


Fig. 6. Boxplots of the morphological and textural traits. The varieties were color-coded by their subpopulations, where purple, blue, green, red, and yellow stand for *aromatic*, *temperate japonica*, *tropical japonica*, *indica*, and *aus*, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

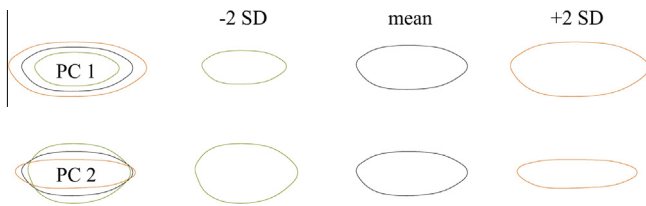


Fig. 7. Major grain shape variations. Each column shows the grain contours with altered PC values (mean – 2SD, mean, and mean + 2SD) as labeled. The left-hand column shows the 3 contours stacked together. That PC2 primarily corresponds with the roundness of the grain is clearly visible.

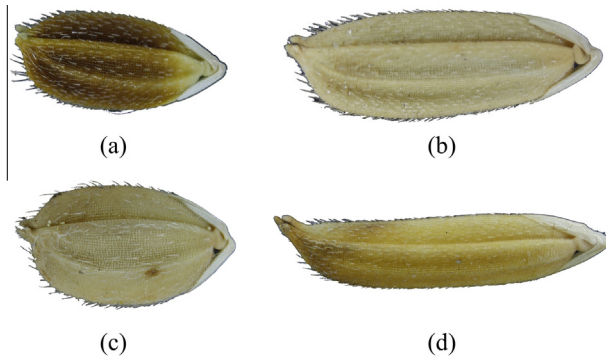


Fig. 8. Grains of varieties (a) Dourado Agulha, (b) R101, (c) Tainung 67, and (d) T1. The grains are dissimilar in size and roundness.

aromatic and *tropical japonica* ($P < 0.001$). Particularly, the sterile lemmas of variety Pappaku were larger in size and longer in length (Figs. 1, 6c and d). As for the textural traits, the brush ratios of varieties Dosel, Dourado Agulha, and Caawa were evidently larger than the brush ratios of varieties Sitpwa, Dular, and Tainung 67 (Fig. 6e). The GLCM entropy also varied from one variety to another. The GLCM entropy of varieties Dom-Zard, Dom-Sufid, and Dourado Agulha were greater than the GLCM entropy of some other varieties (Fig. 6f). Note that some traits of the same varieties may vary at a considerable degree (e.g., the aspect ratio of variety Aswina

330, Fig. 6b; the area ratio of variety Pappaku and Edomen Scented, Fig. 6c). These within-variety variations increase the challenge of grain variety identification.

3.2. Major grain shape variations

The grain shape variations among the 30 varieties were determined. The PCA results revealed that the first 2 PCs accounted for 98.6% and 0.5%, respectively, of the total variance. Each remaining PC accounted less than 0.3%. Hence, only the shape variations associated with the first 2 PCs were devised. Fig. 7 illustrates the grain contour variations. Major variations were observed in grain length, width, and roundness. The figure indicates that PC1 primarily corresponded with the grain size. The grains of large PC1 values were greater in volume compared with those of small PC1 values. PC2 principally corresponded with roundness. The grains of large PC2 values were relatively slender, whereas those of small PC2 values were more spherical.

Fig. 8 shows sample grain images of varieties (a) Dourado Agulha, (b) R101, (c) Tainung 67, and (d) T1 (Appendix A). The grain contours vary considerably from one variety to another. The grains of Dourado Agulha and R101 resembled the contours reconstructed using extreme PC1 values (–2 SD and +2 SD, respectively) in Fig. 7. Also, the grains of Tainung 67 and T1 resembled the contours reconstructed using extreme PC2 values (–2 SD and +2 SD, respectively) in Fig. 7. These observations indicated that the morphological traits and Fourier descriptors can be used to differentiate the grains of certain varieties effectively.

3.3. Grain color discrepancies

The color discrepancies among the grains of the 30 varieties were examined. In the analysis, PCA was conducted to summarize the color distribution for the 50 grains of each variety in the CIE $L^*a^*b^*$ color space (Fig. 9). The color distribution was presented using an ellipsoid. The ellipsoid was centered at the mean L^* , a^* , and b^* color parameters of a variety. The principal axes of the ellipsoid were set parallel to the first 3 PCs, and were 2 SDs of the PC scores in length. Fig. 8(a) and (b) display the ellipsoids

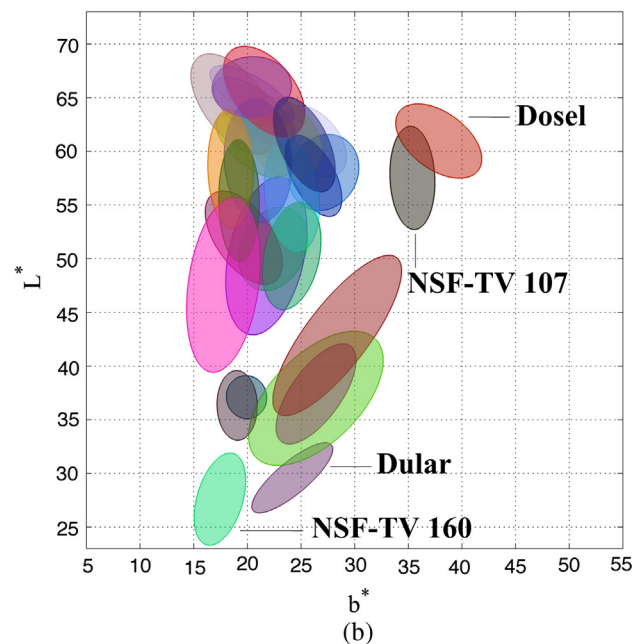
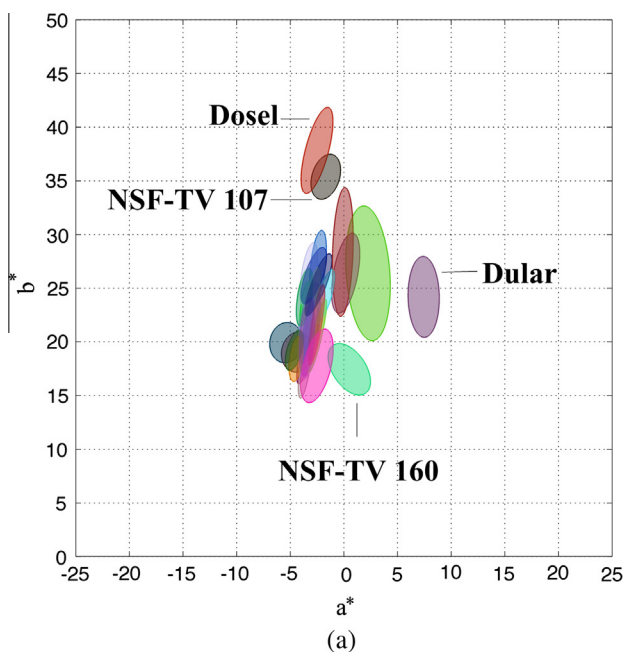


Fig. 9. Grain color distribution of the 30 varieties on the (a) chromaticity and (b) chromaticity-lightness planes. The ellipsoids represent the color ranges of the varieties, and are pseudo-colored. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



Fig. 10. Grains of varieties (a) Dosel, (b) NSF-TV 107, (c) Dular, and (d) NSF-TV 160. The grains are dissimilar in color. Traits such as the brush coverage and sterile lemma color considerably differ among varieties. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the 30 varieties projected onto the chromaticity ($a^* - b^*$) and chromaticity-lightness ($b^* - L^*$) planes, respectively. The ellipsoids were pseudo-colored for illustration purposes. The figures show that the grains were associated with a considerable variation in lightness (L^*) and yellow-blue channel (b^*). The colors of certain varieties (e.g., Dosel, NSF-TV 107, Dular, and NSF-TV 160) were apparently distinct from those of certain other varieties.

Fig. 10 shows sample grain images of varieties (a) Dosel, (b) NSF-TV 107, (c) Dular, and (d) NSF-TV 160. The grain colors vary

considerably from one variety to another. The Dosel and NSF-TV 107 grains were evidently brighter than the Dular and NSF-TV 160 grains (Fig. 10b). The Dosel grain was associated with a stronger yellow compared with the NSF-TV 160 grain (Fig. 10a). These observations indicated that the color traits can be used to differentiate the grains of certain varieties effectively.

3.4. Classification performance

The dictionaries for the SRC classifier were developed using the 12 morphological traits, 9 color traits, 7 textural traits, and 20 Fourier descriptors quantified from the 1500 grains. The classifier was then evaluated using 10-fold cross-validation. In the cross-validation, the original samples were randomly partitioned into 10 groups. Nine groups were used as training data for developing the model, and the remaining group was retained as validation data for testing the classifier. The process was repeated for 10 times, with each of the group used once as the validation data. The averaged accuracies were then presented. Fig. 11 shows the classification accuracies of the 30 varieties in a confusion matrix. In the matrix, each row represents the varieties in an actual class, and each column represents the varieties in a predicted class. The color of each entry indicates the percentage of the prediction from the SRC classifier. Overall, the SRC classifier attained an averaged accuracy of 89.1% and a SD of 7.0%. The classification accuracies were not statistically different among the subpopulations.

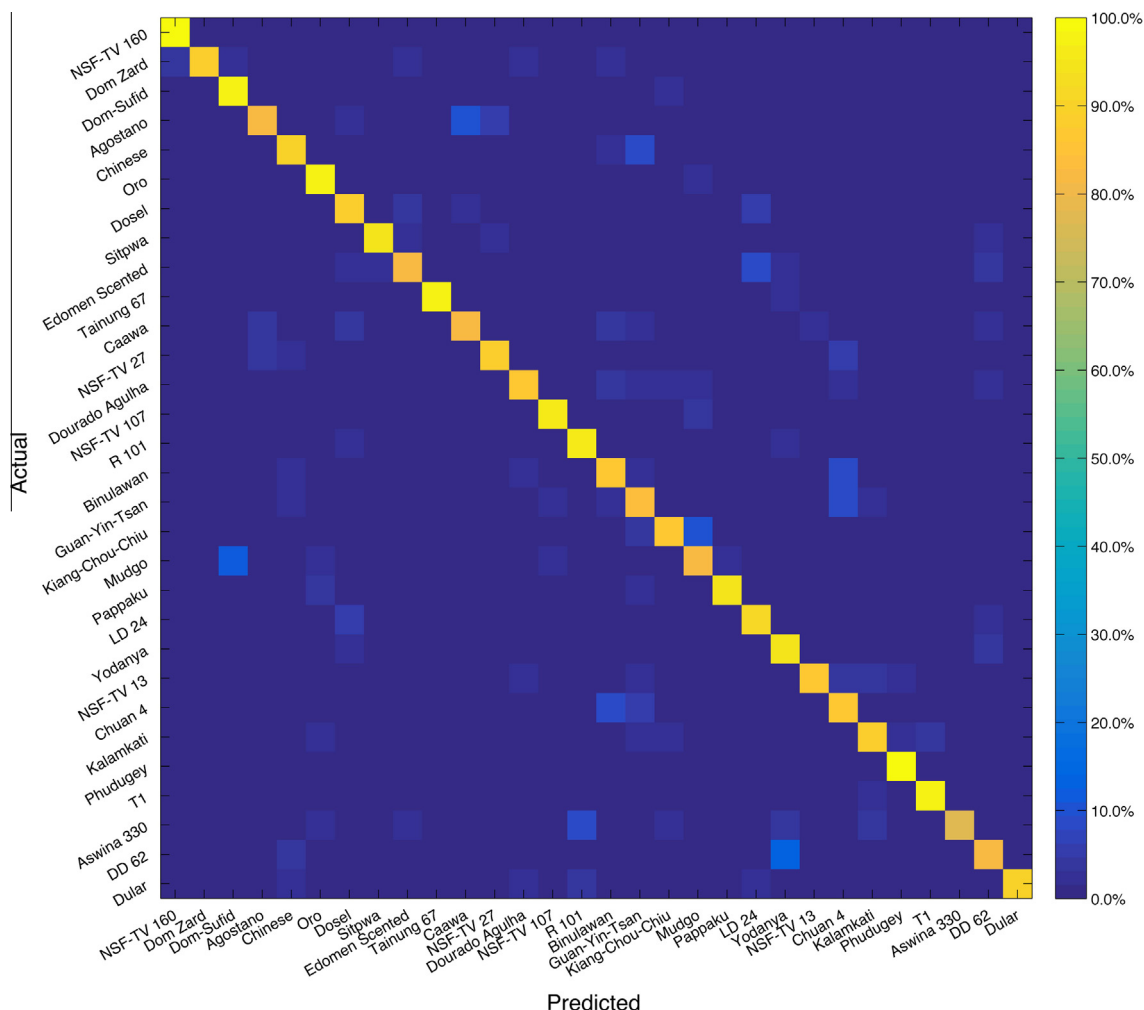


Fig. 11. Classification accuracies for the 30 varieties. Each row represents the grains in an actual class, and each column of the matrix represents the grains in a predicted class.

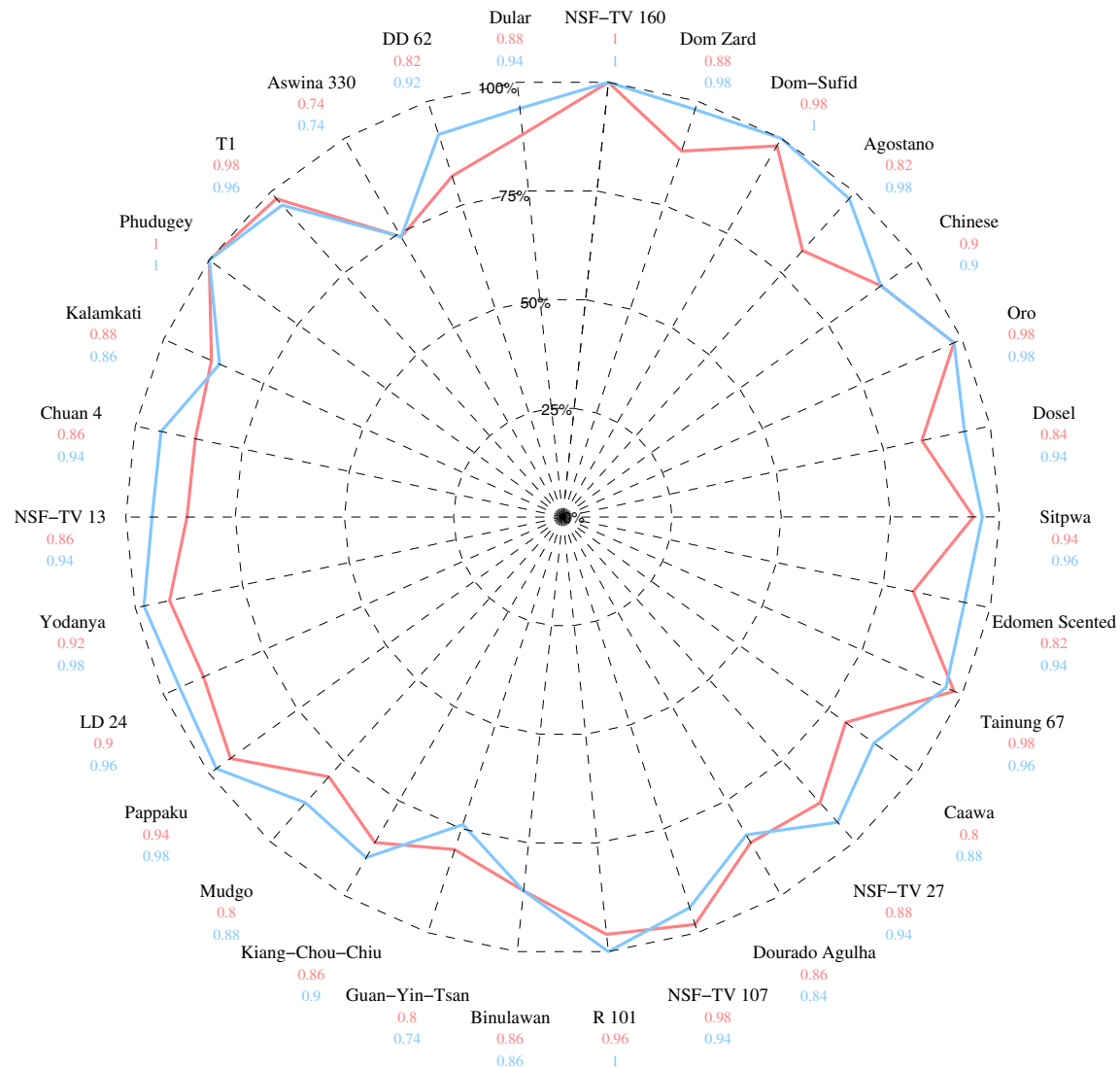


Fig. 12. Performance comparison of the SRC and SVM classifiers. The accuracies of the SRC and SVM classifiers were color-coded in blue and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The performance of the SRC classifier were compared with the performance of support vector machine (SVM) classifiers. Soft-margin SVM classifiers and radial basis function kernels were used. The optimal margin and kernel parameters for the SVM classifiers were determined using grid search and 10-fold cross-validation. The SVM classifiers were developed using an open-source library LIBSVM (Chang and Lin, 2011) and all the quantified traits. The SVM classifiers achieved an averaged accuracy of 92.8% and a SD of 6.8%. Fig. 12 shows the performance comparison of the SRC and SVM classifiers. The performance of the SRC classifier did not significantly differ from the performance of the SVM classifiers ($P < 0.0001$). Although the SVM classifiers achieved a slightly better performance compared with the SRC classifier, the SVM classifiers are known to be completely heuristic. They merely provide the function of classification, but the insight of the discrimination is unavailable. In contrast, the SRC classifier constructed a dictionary for each variety. The dictionaries contained essential trait information of the varieties. In addition, SVM classifiers are known to be binary. Thirty classifiers were required to classify 30 varieties of rice based on one-versus-all scheme.

One limitation in using the image-based approach for identifying the varieties of rice grains is that grains require distinct morphological and color traits. The SRC classifier achieved a

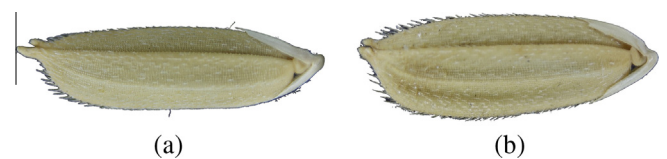


Fig. 13. Grains of varieties (a) Aswina 330 and (b) R101. The grains are similar in appearance.

relatively mediocre accuracy in classifying certain varieties, because the grain appearances of these varieties were similar to those of certain other varieties. For example, the classification accuracy for variety Aswina 330 was 74.0%. This was because that the appearance of variety Aswina 330 resembled the appearance of variety R101 (Fig. 13). A certain amount of Aswina 330 grains were misclassified as R101 grains owing to the similar appearances.

4. Conclusion















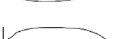








For this study, we nondestructively distinguished the rice grains of 30 varieties through image analysis and SRC techniques. Morphological and color variations among the rice grains of different varieties were observed. This prompted the use of

image-based approaches for differentiating the rice grains. In the proposed approach, we acquired the rice grain images through microscopy at a resolution of approximately 95 pixels per millimeter. The high resolution enabled observations of fine details of the rice grains. The morphological, textural, and color traits of the grains were quantified, and an SRC classifier was then developed to predict the varieties of the grains using the traits as the inputs. The classifier achieved an overall accuracy of 89.1%.








Acknowledgments

We thank Dr. Chih-Wen Wu and Ms. Fang-Yu Chang (Crop Improvement Section, Kaohsiung District Agricultural Research and Extension Station, Taiwan) for the help with seed reproduction. This research was supported by the Ministry of Science and Technology of Taiwan, grant MOST 104-2311-B-002-019-MY3.

Appendix A

No.	Name	NSFTV ID ^a	Subpopulation	Origin	Contour ^b
1	NSF-TV 160	160	<i>aromatic</i>	Iran	
2	Dom Zard	191	<i>aromatic</i>	Iran	
3	Dom-Sufid	640	<i>aromatic</i>	Iran	
4	NSF-TV 13	13	<i>aus</i>	Pakistan	
5	Chuan 4	33	<i>aus</i>	Taiwan	
6	Kalamkati	81	<i>aus</i>	India	
7	Phudugey	131	<i>aus</i>	Bhutan	
8	T1	152	<i>aus</i>	India	
9	Aswina 330	312	<i>aus</i>	Bangladesh	
10	DD 62	316	<i>aus</i>	Bangladesh	
11	Dular	651	<i>aus</i>	India	
12	Binulawan	17	<i>indica</i>	Philippines	
13	Guan-Yin-Tsan	61	<i>indica</i>	China	
14	Kiang-Chou-Chiu	90	<i>indica</i>	Taiwan	
15	Mudgo	110	<i>indica</i>	India	
16	Pappaku	126	<i>indica</i>	Taiwan	
17	LD 24	298	<i>indica</i>	Sri Lanka	
18	Yodanya	339	<i>indica</i>	Myanmar	
19	Agostano	1	<i>temperate japonica</i>	Italy	
20	Chinese	31	<i>temperate japonica</i>	China	
21	Oro	118	<i>temperate japonica</i>	Guinea	
22	Dosel	296	<i>temperate japonica</i>	Spain	
23	Sitpwa	338	<i>temperate japonica</i>	Myanmar	

Appendix A (continued)

No.	Name	NSFTV ID ^a	Subpopulation	Origin	Contour ^b
24	Edomen Scented	363	<i>temperate japonica</i>	Japan	
25	Tainung 67	641	<i>temperate japonica</i>	Taiwan	
26	Caawa	22	<i>tropical japonica</i>	Taiwan	
27	NSF-TV 27	27	<i>tropical japonica</i>	Pakistan	
28	Dourado Agulha	46	<i>tropical japonica</i>	Brazil	
29	NSF-TV 107	107	<i>tropical japonica</i>	Bangladesh	
30	R 101	310	<i>tropical japonica</i>	Zaire	

^a Accession identification number of the “Exploring the Genetic Basis of Transgressive Variation in Rice” project, National Science Foundation.

^b Mean contours of the 50 grains of the same variety reconstructed using Fourier descriptors.

References

- Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. *Statist. Surveys* 4, 40–79.
- Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up Robust Features. *Computer vision–ECCV 2006*. Springer, 404–417.
- Becerra, V., Paredes, M., Gutiérrez, E., Rojo, C., 2015. Genetic diversity, identification, and certification of Chilean rice varieties using molecular markers. *Chilean J. Agric. Res.* 75, 267–274.
- Camelo-Méndez, G.A., Camacho-Díaz, B.H., del Villar-Martínez, A.A., Arenas-Ocampo, M.L., Bello-Pérez, L.A., Jiménez-Aparicio, A.R., 2012. Digital image analysis of diverse Mexican rice cultivars. *J. Sci. Food Agric.* 92, 2709–2714.
- Chang, C., Lin, C., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3), 27.
- Chuang, H., Lur, H., Hwu, K., Chang, M., 2011. Authentication of domestic Taiwan rice varieties based on fingerprinting analysis of microsatellite DNA markers. *Botanical Stud.* 52, 393–405.
- Cirillo, A., Del Gaudio, S., Di Bernardo, G., Galderisi, U., Cascino, A., Cipollaro, M., 2009. Molecular characterization of Italian rice cultivars. *Eur. Food Res. Technol.* 228, 875–881.
- Costa, C., Menesatti, P., Paglia, G., Pallottino, F., Aguzzi, J., Rimatori, V., Russo, G., Recupero, S., Recupero, G.R., 2009. Quantitative evaluation of Tarocco sweet orange fruit shape using optoelectronic elliptic Fourier based analysis. *Postharvest Biol. Technol.* 54, 38–47.
- Dillencourt, M.B., Samet, H., Tamminen, M., 1992. A general approach to connected-component labeling for arbitrary image representations. *J. ACM* 39, 253–280.
- Galloway, M.M., 1975. Texture analysis using gray level run lengths. *Comput. Graph. Image Process.* 4, 172–179.
- Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S., McCouch, S., 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169, 1631–1638.
- Gonzalez, R.C., Woods, R.E., 2007. *Digital Image Processing*, third ed. Prentice Hall.
- Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. *Systems, Man Cybernet., IEEE Trans. Syst.*, 610–621.
- Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* 66, 51–83.
- Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. *Appl. Stat.*, 100–108.
- Hunter, A., 1975. The loss of community: an empirical test through replication. *Am. Sociol. Rev.*, 537–552.
- Jin, H., Yang, X., Jiao, L., Liu, F., 2005. Image Enhancement Via Fusion Based on Laplacian Pyramid Directional Filter Banks. *Image Analysis and Recognition*. Springer, pp. 239–246.
- Kong, W., Zhang, C., Liu, F., Nie, P., He, Y., 2013. Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. *Sensors* 13 (7), 8916–8927.
- Liu, M., Zhang, D., Shen, D., 2012. Initiative, The Alzheimer's Disease Neuroimaging Initiative, 2012. Ensemble sparse classification of Alzheimer's disease. *NeuroImage* 60, pp. 1106–1116.
- Majumdar, S., Jayas, D., 2000. Classification of cereal grains using machine vision: IV. Combined morphology, color, and texture models. *Trans. ASAE* 43 (6), 1689–1694.
- Mebastion, H., Paliwal, J., Jayas, D., 2013. Automatic classification of non-touching cereal grains in digital images using limited morphological and color features. *Comput. Electron. Agric.* 90, 99–105.
- Muja, M., Lowe, D.G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP 2* (1), 331–340.
- Pazoki, A., Farokhi, F., Pazoki, Z., 2014. Classification of rice grain varieties using two Artificial Neural Networks (MLP and Neuro-Fuzzy). *J. Anim. Plant Sci.* 24, 336–343.
- Rohlf, F.J., Archie, J.W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). *Syst. Biol.* 33, 302–317.
- Steele, K.A., Ogden, R., McEwing, R., Briggs, H., Gorham, J., 2008. InDel markers distinguish Basmati from other fragrant rice varieties. *Field Crops Res.* 105, 81–87.
- Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification. In: *Computer Vision and Pattern Recognition, 2010 IEEE Conference*, San Francisco, CA, pp. 3360–3367.
- Wang, W., Chang, F., 2011. A multi-focus image fusion method based on Laplacian pyramid. *J. Comput.* 6, 2559–2566.
- Wei, C., Chao, Y., Yeh, Y., Wang, Y., 2013. Locality-sensitive dictionary learning for sparse representation based classification. *Pattern Recogn.* 46, 1277–1287.
- Williams, K., Munkvold, J., Sorrells, M., 2013. Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.). *Euphytica* 190 (1), 99–116.
- Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence. IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2), 210–227.
- Xie, C., Zhang, J., Li, R., Li, J., Hong, P., Xia, J., Chen, P., 2015. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. *Comput. Electron. Agric.* 119, 123–132.
- Yuan, X., Liu, X., Yan, S., 2012. Visual classification with multitask joint sparse representation. *Image Processing. IEEE Transactions on Image Processing* 21 (10), 4349–4360.
- Zhao, K., Tung, C., Eizenga, G., Wright, M., Ali, M., Price, A., McClung, A., 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2, 467.