# Spatial and Spectral Features Utilization on a HyperSpectral Imaging System for Rice Seed Varietal Purity Inspection

Hai Vu*, Christos Tachtatzis†, Paul Murray†, David Harle†,
Trung Kien Dao*, Thi-Lan Le*, Ivan Andonovic†, Stephen Marshall †
*International Research Institute MICA, Hanoi University of Science and Technology
† Dept. of Electronic and Electrical Engineering, University of Strathclyde
Email: hai.vu@mica.edu.vn

***Abstract*—A conventional method to inspect the varietal purity of rice seeds is based on human visual inspection where a random sample is drawn from a batch. This is a tedious, laborious, time consuming and extremely inefficient task. This paper presents an automatic rice seed inspection method using Hyperspectral imaging and machine learning, to automatically detect unwanted seeds from other varieties which may be contained in a batch. Hyperspectral image data from Near-infrared (NIR) camera are acquired for six common rice seed varieties. The results of applying two classifiers are presented, a Support Vector Machine (SVM) and a Random Forest (RF), where each consists of six one-versus-rest binary classifiers. The results show that combining spectral and shape-based features derived from the rice seeds, increase precision of the multi-label classification to 84% compared 74% when only visual features are used.**

## I. Introduction

Ensuring rice seed quality is a significant challenge for the large rice export nations such as India, Thailand, US and Vietnam. Rice seed impurities can impact on the yield by introducing weeds and off-types into the crop making it susceptible to disease. The consequences are not limited to a decrease in yield but also to the grade and price of the produce. The responsibility lies with rice seed producers to ensure high quality seed and a critical procedure is the batch screening and inspection. Conventional methods to inspect seeds, as shown in Figure 1(a), rely on extracting a sample from a batch. The inspection is performed visually to assess the grain properties, such as shape, length, width and size. This task is tedious, laborious, time consuming and requires experienced personnel.

Recently, the cost and size of Hyperspectral Imaging (HSI) Systems has reduced significantly. This technology has emerged proves to be a useful tool in food sciences and applications. Such systems provide spatial and textural information like other traditional cameras with the added advantage of high resolution spectral signatures for each pixel in the image data acquired. In this paper, we investigate the benefits of analysing the extracted features taken from a HSI system to solve issues of rice seed varietal purity inspection. We deploy an automatic inspection method which combines hyperspectral imaging and machine learning techniques to automatically detect seeds which are erroneously contained within a batch when they actually belong to a completely different species. In this study, the purity of six common rice seed varieties are examined, as shown in Figure 1(b).


Fig. 1. (a) A conventional way (human visual) to inspect purity of rice seed samples. (b) Six common rice seed varieties examined in this study.

Automatic rice seed inspection systems that employ machine vision and address this challenge have been shown in previous works [1]–[3]. Commonly, shape descriptors of the seed samples are extracted through image processing and vision-based approaches. The challenge in comparing and quantifying performance between these approaches, is that each one has been evaluated on different rice seed varieties. It is therefore unclear if the differences in performance come from better feature descriptors or if this is due to varying inter-class/intra-class variations among the examined species. In this study, a HSI system provides both spatial and spectral information about the seed samples. Therefore, the inspection techniques that utilize both types of feature is investigated. We formulate the purity inspection problem as six one-versus-rest binary classifiers. The binary classifiers are built using SVM and a RF techniques and both approaches are compared. While the spatial features measure physical properties of rice seed, the mean spectrum of all pixels in a seed sample can be used to infer chemical properties of the species. The use of discriminant analysis techniques and the combinations of both types of features provide significant benefits and potential in HSI offering great advantages for the development of a machine vision system for rice seed quality assessments.

The remainder of paper is organized as follows. Section II briefly describes related techniques for rice and rice seed quality assessments using vision-based/HSI systems. Section III describes the device configuration, data acquisitions and correction procedures. Section IV presents the proposed feature

978-1-5090-4134-3/16/$31.00 ©2016 IEEE

extraction approach. Section V analyses inter-and intra-class variations of the examined species. Section VI reports the classification results. Finally, Section VII concludes the work and suggests further research directions.

## II. Related work

There are many machine vision systems for food quality evaluation. A good survey can be found in Da-Wen Sun's textbook [4]. In Chapter 16 of this textbook, Y. Ogawa comprehensively surveys rice grain quality evaluation techniques using computer vision technology which analyses features such as physical property measurements, compound content and distribution. Lai et al. [5] applied an interactive image analysis method for determining the physical dimensions and classifying the variety of grains. In [6], the authors measured physical dimensions such as grain contour, size, colour variance and distribution and damage. Sakai et al. [7] demonstrated the use of two-dimensional image analysis for the determination of the shape of brown and polished rice grains of four varieties. Z. Lui et al. [1] implemented a method of identification based on neural networks to classify rice variety using colour and shape features. Guzman et al. [8] investigated grain features extracted from each sample image. They then utilized multilayer artificial neural network models for automatic identification of 52 rice grains. More recent works [1]–[3] focused on rice seed variety classification. Commonly, shape descriptors of the seed samples are extracted, then the classifiers such as Random Forests [3], Neural Networks [1] or Cubic B-Splines shape model [9] are trained. An automatic machine-vision system includes several stages, in which the most important steps are image data collection, feature extraction (such as shape, size, colour, and orientation etc.), and feature representations via models using pattern recognition algorithms or multivariate analysis techniques.

HSI specific systems for the food and agriculture engineering have been investigated also been addressed in the literature. The authors in [10] give a broad range of HSI applications for beef, pork, fruits, and plant products quality evaluations. For the rice grain quality inspection, [11] used a range of VIS/NIR spectral (400-1000 nm) information for discriminating three rice varieties. By using Principle Component Analysis (PCA) and Back Propagation Neural Network (BPNN), they achieved a classification accuracy of 89.18 and 89.91% for PCA and BPNN model, respectively. The authors in [12] suggest that a combination of the Least squares Support Vector Machine (LS-SVM) regression method and Vis/NIR spectroscopy at range 325-1075 nm provides a realisable technique to monitor the nitrogen status in rice. More recently, a HSI system has been used in [13] for identifying four rice seed cultivars. By utilizing the full spectral range 1,039-1,612 nm, they achieved very promising results, that is up to 100% accuracy with a Random Forest (RF) classifier. However, their evaluation is performed on four cultivars in [13] and therefore, it is unclear how the inter/intra class variations affect the performance. In this paper, we extract and combine both spatial and spectral features from the hyperspectral datacube acquired by a HSI system with range of NIR spectral. We argue that the combination of features increases classification performance.

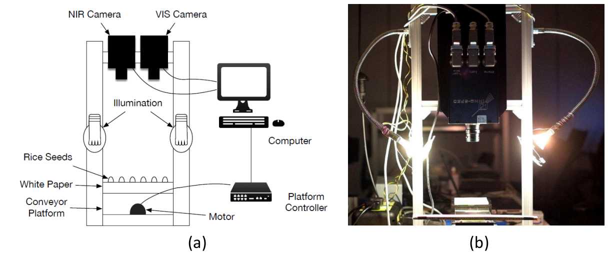


Fig. 2. Experimental setup of the data acquisition. (a) A schematic view . (b) A Photo of the real HSI system

## III. Rice Seed Sample Data Collections using a HyperSpectral Imaging System

### A. Hyperspectral Imaging System Setup

The experimental setup of the data acquisition system is shown in Figure 2(a). The NIR HSI system used to capture the data was the Inno-Spec$^{TM}$ Redeye 1.7 model (Inno-Spec GmbH, Germany) capturing 256 wavelengths from 950.73 - 1759.4 nm. The HSI device operates using a line scan where the spectral information from an entire row of pixels is captured at any given time. For this purpose, a conveyor platform (the stage) was positioned underneath the camera to allow scanning. Two halogen bulbs were used to illuminate the scene and the bulbs were positioned to create balanced illumination across the scan line. To ensure repeatable data acquisitions, the halogen bulbs were switched on and allowed to reach stable operating temperature before the data were acquired in a dark room, to minimise illumination variability between captures. To properly collect the data, the following parameters in the HSI system were adjusted:

- The exposure time of the camera (e.g., 500 ms) and the speed of movement of the stage (5 mm/s), was calibrated to avoid spatial distortions.
- The aperture (f=1/8) was set based on the exposure time to ensure a suitable light intensity and contrast.
- The height between the lens and the stage was set so that whole seed samples area are in FOV of the camera.

A photo of the real HSI system is shown in Figure 2(b).

### B. Rice seed dataset acquisition

Six rice seed varieties (as shown in Figure 1(b)) were obtained from a seed production company in Vietnam: $BC15$, $BT07$, Khang Dan 18 (shortly named $KD18$), $N97$, Nep Lang Lieu ($LL$), and $Q5$. The selected varieties are the most frequently planted in North Vietnam. The producer screened the samples using experienced technical staff to ensure that each sample population only contained seeds of the corresponding species. The sample population of each variety consisted of 108 seeds with 648 seeds across all varieties. The 108 samples from each species was then divided to 3 batches with 36 samples each. The 36 seeds were positioned

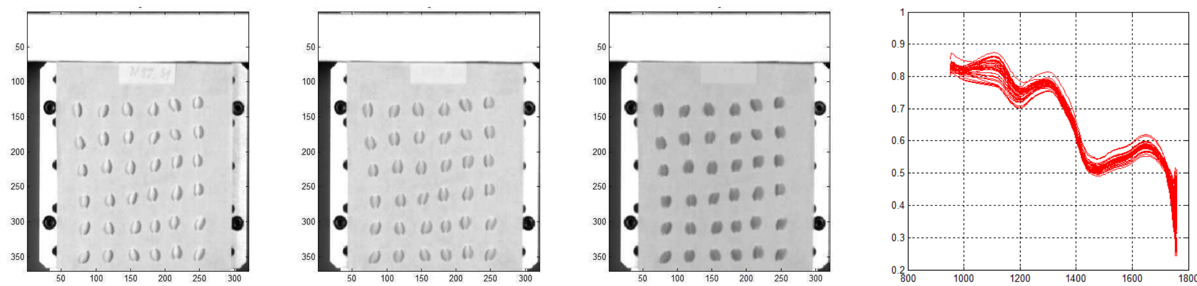


Fig. 3. Data correction results. (a) - (c) Normalized images at specific wavelengths. (d) Profiles of the normalised wavelength

on a white sheet of paper constructing a $6 \times 6$ matrix (e.g., as shown in Figures 3(a)-(c)), that was placed on a conveyor platform for imaging by the NIR camera. This resulted in 3 hyperspectral datacubes per variety resulting in 36 total number of the datacubes.

### C. Data correction

Let $y$ denote a datacube consisting of reflectance values $\lambda$ as a two-parameter set:

$$y_\lambda(x), x \in X, \lambda \in \Lambda \tag{1}$$

where $\lambda$ represents a wavelength belonging to $\Lambda$, that is a set of the wavelengths at NIR (Near-Infrared) range 950.73-1759.4 nm and $x$ represents a pixel in $X$ where $X$ is 2-D coordinate by row $m$ and column $n$. For each specific wavelength, the array of reflectance values can be regarded as an image where spatial relationships between the pixel reflectance values have meanings. For example, Figure 3 represents three images acquired by the device at specific wavelengths of 1109.3 nm, 1267.78 nm, and 1424.61 nm. It is noted that at each $x$, the raw reflectance value could vary due to different lighting conditions or manufacturing tolerance of the pixels in the imaging sensor. To reduce the variation in the acquired reflectance values among acquisitions the data are normalised relative to known maximum reflectance value as below:

$$y_\lambda(x) := \frac{y_{raw,\lambda}(x) - b(n, \lambda)}{w(n, \lambda) - b(\lambda, \lambda)}, \lambda \in \Lambda \tag{2}$$

where $b(n, \lambda)$ and $w(n, \lambda)$ are the reflectance values of reference dark and white objects. The dark object is setup by covering the lens with its cap and the white object is a white spectralon tile which is a highly reflective Lambertian scatter, commonly used to calibrate HSI systems. For each $\lambda$, $b(n)$ and $w(n)$ are averaged on reflectance values at column $n$ along the white tile height dimension. The images shown in Figures 3(a)-(c) have been normalized. Corresponding wavelength profiles of the rice seeds are shown in Figure 3(d).

## IV. Spatial and Spectral Feature Extractions

### A. Separating seed samples from background

In the proposed system, rice seed samples need to be separated from background regions in order to allow the extraction of the physical properties of the grain as well as spectral features. However, seed segmentation is not a straightforward procedure because of artefacts such as shadow and lighting conditions. Particularly, to correctly measure physical properties, the seed segmentation procedure suffered due low spatial resolution of the hypercubes data. Thus, we deploy a series of image processing techniques to overcome these difficulties. Firstly, we consider the difference of low-band and high-band images, as shown in Figure 3(a) and Figure 3(c), respectively, to boost contrast. This difference image $I_{diff}$ differentiates between background regions and foreground ones (e.g., rice seed regions). A morphological opening operator is then applied on the difference image, named background image $I_{bg}$. As inherited from a Top-hat transform, the foreground image is the subtraction of $I_{diff}$ and $I_{bg}$; i.e. $I_{fg} = |I_{diff} - I_{bg}|$. An example of $I_{fg}$ is shown in Figure 4(a). Comparing with original image at a specific wavelength (e.g., Figure 3(a)-(c), the shadow and artefacts in $I_{fg}$ are eliminated. Subsequently, a thresholding operator using Otsu's threshold [14] is applied on $I_{fg}$. The extracted seeds are marked by the red boundary in Figure 4(b).

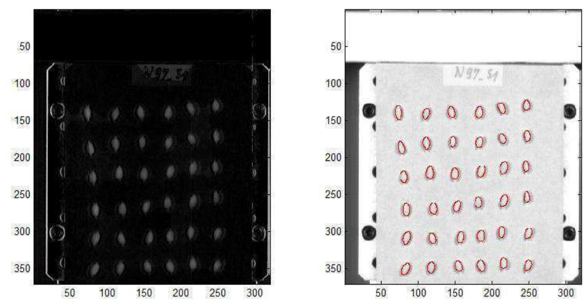


Fig. 4. The procedures separating rice seeds from background

### B. Spatial Feature extractions

Given an individual rice seed from batch samples, we measures spatial/morphological features. The selected features are similar to those presented in recent works such as [3] and [1] because of their effectiveness for discriminating among species. A morphological feature descriptor $f$ with 6 dimensions is calculated as follows:

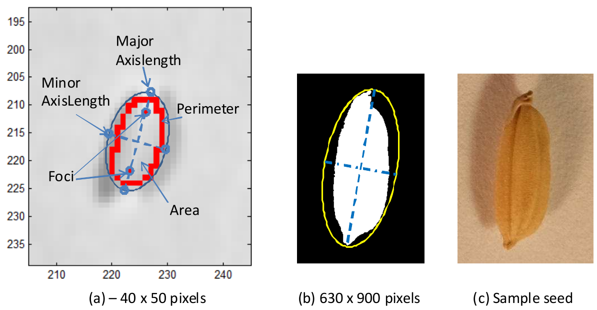


Fig. 5. Spatial feature extraction (a) on hyperspectral datacube; (b) on CCD camera. (c) Photo of a seed sample for reference

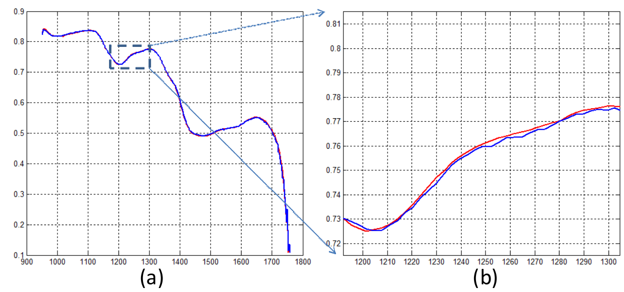


Fig. 6. PCA reconstruction using the first ten components. (a) Overlapping the original and the reconstructed profiles of a seed sample. (b) A close-up region of graph. Blue line is original data; red line is reconstructed data

- $f_1$: is the number of pixels inside a seed sample.
- $f_2, f_3$: are the MajorAxisLength and MinorAxisLength respectively, that specify the length (in pixels) of the major/minor axis of the ellipse that covers the boundary of the sample seeds.
- $f_4$: is the aspect ratio $\frac{f_3}{f_2}$ and
- $f_5 = \frac{Perimeter}{Area}$, where $Perimeter$ is number of pixels along the seed boundary; and $Area$ is $f_1$ feature.
- $f_6 = \frac{FociDistance}{MajorAxisLength}$ is eccentricity specified by $FociDistance$ which is the distance between two foci of the ellipse, and the major axis length.

The features are illustrated in Figure 5(a). Note that these features are similar but different from those in [3]; the HSI camera gives a low spatial resolution for each seed (e.g., $40 \times 50$ pixels) versus the corresponding images collected from high resolution CCD camera [3] (e.g., $630 \times 900$ pixels; shown in Figures 5(b) and (c)). This difference hinders the discrimination ability between the rice species when only spatial features are used.

### C. Spectral feature extractions

A hyperspectral datacube contains spectral information for every pixel of the seed regions. For each wavelength, the mean normalised intensity across all pixels in the seed region can be computed resulting in 256 spectral features. The per pixel spectral profiles for one rice seed species are shown in Figure 3(d). As denoted in Equation (2), the raw spectral feature vector of a rice seed sample is a set of $y_\lambda$ in which $\lambda$ is one of 256 bands belonging a range $\Lambda = 950.73 - 1759.4nm$. Figures 7(a)-(b) show the spectral features of two seed samples from two variates (Q5 and N97). The mean spectrum over each seed regions are shown in the corresponding right panels.

The 256 spectral features of the mean normalised intensity of pixels in the seed lead to "Curse of Dimensionality" and dimensionally reduction techniques are commonly applied in spectral data analysis to avoid overfitting, reduce redundancy and co-linearity of spectral data. This also facilitates the construction of simple, stable and practical classification models. In particular, we use Principal Component Analysis (PCA) to transform the original data into a small number of uncorrelated variables. The PCA transformation is applied on the spectrum profile of the all the collected data:

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{648} \end{bmatrix} = \begin{bmatrix} f_{\lambda_1} & f_{\lambda_2} & \dots & f_{\lambda_{256}} \\ f_{\lambda_1} & f_{\lambda_2} & \dots & f_{\lambda_{256}} \\ \dots & & & \\ f_{\lambda_1} & f_{\lambda_2} & \dots & f_{\lambda_{256}} \end{bmatrix}$$

Results of PCA provide a mean $\mu_\Theta$ and eigenvectors $\nu_\Theta$. Given the normalized datacube $y_\Lambda$, the projected data in a PCA space is defined by:

$$y_\Lambda \rightarrow f_{pc1,pc2,..,pc10} : \nu_\Theta^T (c_\lambda - \mu_\Theta) \quad (3)$$

The first few principal components (PCs) can be used to capture most of the variability in the data reducing dimensionality. In our dataset, the first ten Principal Components (PCs) ($f_{pc1}, .., f_{pc10}$) can reconstruct 99% original data. The consistency of the reconstructed data using first ten PCs against all of the original data is illustrated in Figure 6. In this paper, we therefore use the the first ten PCA-based features for each rice seed rather than entire spectral data.

## V. Discriminant Analysis and Classification

### A. Species discrimination using spectral features

Many relevant works [1]–[3] have evaluated shape-based properties for identifying/separating rice seed species. However, these measurements strongly depend on how the grains are spread on the captured surface. For example, Figure 7(a) - right panel shows an illustration of two seeds from the same species and although it was expected that the shape-based properties would be similar this is not the case. The positioning of the seed on the surface is such that a different perspective (side) is captured. Conversely, the shape appearance of two seeds from different species may be very similar, e.g. two seeds shown in Figure 7(b)- right panel. The benefit of the HSI system is that it can measure hidden information inside the seeds. As expected, the wavelength profiles of two seeds in Figure 7(a) are very similar. In the same way, the wavelength profiles of the seed samples from two different species in Figure 7(b) are separable. Statistically speaking, the wavelength profiles of each species is averaged based on a hyperspectral datacube collected from 108 seed samples. Pair comparisons of the spectral profiles between one species with others are shown in Figures 8(a)-(e).

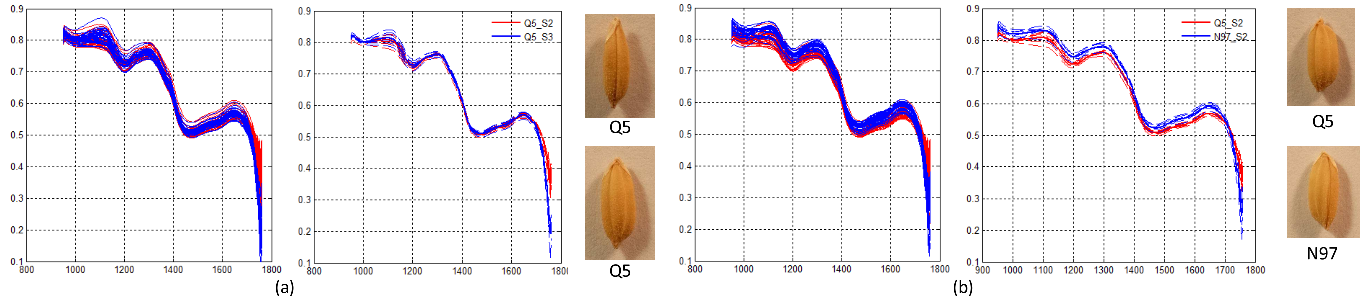


Fig. 7. Discriminant analysis examples. (a) Two seeds of same species. (b). Two seeds from different species. On (a)-(b) Left panel: detailed wavelength profiles; Right panel: Averaged wavelength profiles. Photos of the rice seed samples for references

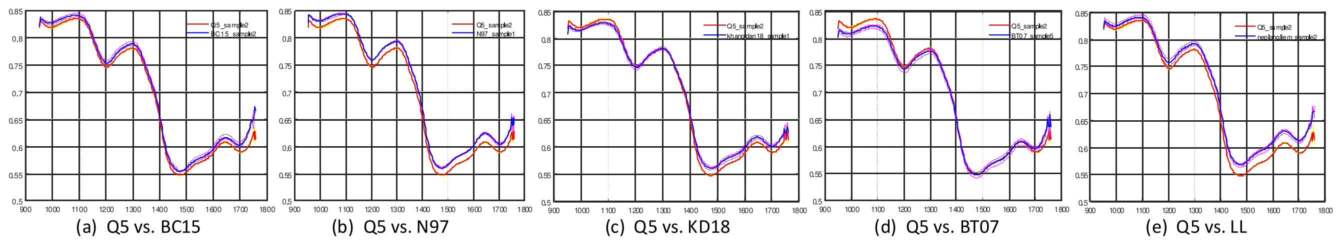


Fig. 8. Comparison wavelength profiles of a species (Q5) with others. The analysis utilized the hyperspectral datacubes of six examined rice seed species

### B. Building Classifiers

Utilizing the spatial and spectral features, as well as their combination, we build classifiers that come from two different classification methods.

**A Random Forest** classifier contains many decision trees, and each tree is grown from a bootstrap sample of the response variables. The best split is selected from a random subset of variables at each node of the tree and then the tree grows to the maximum extent without pruning. Prediction can be made from new data by aggregating the outputs of all trees. RF is a fast and effective algorithm for dealing with a large amount of data. RF has shown the advantages that it reduces variance and achieves comparable classification accuracy. In this study, the number of the decision trees is set to 500.

**Support Vector Machine** is a widely used supervised statistical learning algorithm. SVM shows advantages in dealing with small sample, linear and high dimensional data. SVM is based on the structural risk minimum (SRM) and SVM has high generalization capacity and could provide a flexible and easy-to-compute solution. Selection of the kernel function in SVM models has a significant influence in model performance, and in this study, the commonly used Radial Bias Function (RBF) is employed as kernel function.

## VI. Experimental results

We evaluate the performance of each set of features (spatial, spectral individually, and their combination) on the collected dataset, as described in Section III-B. The feature extraction procedures were implemented using Matlab on a PC Core i5 3.10GHz CPU, 4GB RAM. The LibSVM library [15] and a Random Forest Library [16] were employed to build the classifier models. To validate the proposed method, leave-$p$-out-cross-validation was utilized. For each classifier, 50 seed samples were collected randomly as positive samples, the negative samples were collected in a balanced fashion from all other species so that total negative samples are equal 50 (in other words, 10 from each other species). To evaluate the performance, two criteria measures are defined:

$$Precision(P) = \frac{tp}{tp+fp}, \text{ and } Recall(R) = \frac{tp}{tp+fn} \quad (4)$$

where $tp$ is the number of true positive, $fp$ is the number of the false positive, $tn$ is the number of true negative and $fn$ is the number of false negative. The performance evaluation results are obtained by averaging over 10 runs.

The performance of the classifiers using only the spatial features ($f = f_1, .., f_6$) is given in Table I. As shown, the RF classifier is slightly better than the SVM and the best performance is achieved with the $LL$ species. Utilizing all the 256 spectral features, Table II shows better performances compared to spatial features especially for $BT07, KD18, LL, N97$. Combining the spatial and all the spectral features together, a feature vector consists of $256 + 6$ dimensions. The results for this combination are shown in Table III. The performance increases; from 77-78% precision for spatial only and spectral only features to 81% when combined. Finally, using the 6 spatial features along with the 10 principal components to reduce the dimensionality of the spectral data and over-fitting issues higher performance is achieved with precision at 84%. These evaluations confirm the benefits of the features extracted from a HSI system.

TABLE I
SPATIAL FEATURE PERFORMANCES

| Speices | SVM | | | RF | | |
|---|---|---|---|---|---|---|
| | P | Recall | F-Measure | P | Recall | F-Measure |
| BC15 | 0.77 | 0.71 | 0.74 | 0.8 | 0.8 | 0.8 |
| BT07 | 0.73 | 0.73 | 0.73 | 0.78 | 0.74 | 0.76 |
| KD18 | 0.7 | 0.75 | 0.72 | 0.71 | 0.7 | 0.7 |
| LL | 0.81 | 0.71 | 0.76 | **0.89** | 0.81 | 0.85 |
| N97 | 0.68 | 0.55 | 0.61 | 0.72 | 0.62 | 0.67 |
| Q5 | 0.61 | 0.56 | 0.59 | 0.7 | 0.71 | 0.7 |
| Average | 0.72 | 0.67 | 0.69 | **0.77** | **0.73** | 0.75 |

TABLE II
SPECTRAL FEATURE PERFORMANCES

| Speices | SVM | | | RF | | |
|---|---|---|---|---|---|---|
| | P | Recall | F-Measure | P | Recall | F-Measure |
| BC15 | 0.75 | 0.27 | 0.39 | 0.67 | 0.66 | 0.66 |
| BT07 | 0.98 | 0.53 | 0.69 | **0.86** | 0.83 | 0.84 |
| KD18 | 0.91 | 0.6 | 0.72 | 0.82 | 0.85 | 0.83 |
| LL | 0.69 | 0.91 | 0.78 | 0.82 | 0.82 | 0.82 |
| N97 | 0.78 | 0.58 | 0.67 | 0.8 | 0.79 | 0.79 |
| Q5 | 0.65 | 0.43 | 0.52 | 0.73 | 0.74 | 0.74 |
| Average | 0.79 | 0.55 | 0.63 | **0.78** | **0.78** | 0.78 |

TABLE III
SPATIAL AND FULL BANK SPECTRAL FEATURE COMBINATIONS

| Speices | SVM | | | RF | | |
|---|---|---|---|---|---|---|
| | P | Recall | F-Measure | P | Recall | F-Measure |
| BC15 | 0.75 | 0.47 | 0.58 | 0.71 | 0.72 | 0.71 |
| BT07 | 0.81 | 0.66 | 0.73 | 0.87 | 0.85 | 0.86 |
| KD18 | 0.76 | 0.73 | 0.74 | 0.86 | 0.87 | 0.86 |
| LL | 0.71 | 0.56 | 0.63 | **0.88** | 0.84 | 0.86 |
| N97 | 0.81 | 0.48 | 0.6 | 0.8 | 0.81 | 0.8 |
| Q5 | 0.65 | 0.51 | 0.57 | 0.76 | 0.79 | 0.77 |
| Average | 0.75 | 0.57 | 0.64 | **0.81** | **0.81** | 0.81 |

TABLE IV
SPATIAL AND 10 PCA-BASED FEATURES COMBINATIONS

| Speices | SVM | | | RF | | |
|---|---|---|---|---|---|---|
| | P | Recall | F-Measure | P | Recall | F-Measure |
| BC15 | 0.73 | 0.69 | 0.71 | 0.78 | 0.83 | 0.81 |
| BT07 | 0.8 | 0.69 | 0.74 | 0.87 | 0.89 | 0.88 |
| KD18 | 0.77 | 0.72 | 0.74 | **0.92** | 0.89 | 0.9 |
| LL | 0.8 | 0.73 | 0.76 | 0.89 | 0.87 | 0.88 |
| N97 | 0.73 | 0.53 | 0.61 | 0.81 | 0.82 | 0.82 |
| Q5 | 0.62 | 0.55 | 0.58 | 0.74 | 0.75 | 0.75 |
| Average | 0.74 | 0.65 | 0.69 | **0.84** | **0.84** | 0.84 |

## VII. CONCLUSION

This paper describes a HSI system supporting rice seed varietal purity inspection. The proposed system combines a hardware camera setup and a tool for extracting features from the collected hyperspectral datacubes. We have confirmed that by taking advantage of a HSI system on both spatial and spectral features, we achieve very promising results on eliminating varietal impurity of species from large seed samples.

Although the precision obtained from spatial features is lower compared to that reported in [3] it must be noted that the imaging acquisition in [3] is a high resolution camera to extract shape properties of the segmented seeds and consequently shape features were described more precisely compared to the HSI system. Despite this, HSI system provides additional discrimination capability to the shape descriptors as shown from the results in this study and this suggests that a new imaging modality can be used to improve performance further. Going forward, we propose to combine data from registered high resolution images from a high resolution CCD camera with spectral images from HSI system. Moreover, we believe that utilizing the spectral data at each pixel rather than mean spectrum on all of the pixels of the seed regions can be useful to investigate chemical features of a seed and therefore, discrimination of species would improve.

## ACKNOWLEDGEMENTS

The authors would like to thank the Newton Fund for providing funding (NRCP1516/1/65) to support this internationally collaborative program of work. Thanks also to Adam Polack for his support and guidance in all aspects of optimizing the acquisition of hyperspectral data.

## REFERENCES

[1] Z.-y. Liu, F. Cheng, Y.-b. Ying, and X.-q. Rao, "Identification of rice seed varieties using neural network," *Journal of Zhejiang University. Science. B*, vol. 6, no. 11, pp. 1095–1100, 11 2005.

[2] A. G. OuYang and R. j. Gao et al, "An automatic method for identifying different variety of rice seeds using machine vision technology," in *2010 Sixth Int. Conf. on Natural Computation*, vol. 1, Aug 2010, pp. 84–88.

[3] P. T. T. Hong and T. T. T. H. et al., "Comparative study on vision based rice seed varieties identification," in *Knowledge and Systems Engineering (KSE), 2015 Seventh Int. Conf. on*, Oct 2015, pp. 377–382.

[4] D.-W. Sun, *Computer Vision Technology for Food Quality Evaluation*. Elsevier, 2008.

[5] F. Lai, I.Zayas, and Y.Pomeranz, "Application of pattern recognition techniques in the analysis of cereal grains," *Cereal Chemistry*, vol. 63, no. 2, pp. 168–172, 1982.

[6] D. Goodman and R. Rao, "A new, rapid, interactive image analysis method for determining physical dimensions of milled rice kernels," *Journal of Food Science*, vol. 49, no. 2, pp. 648–649, 1984.

[7] N.Sakai, S.Yonekawa, A.Matsuzaki, and H.Morishima, "Twodimensional image analysis of the shape of rice and its application to separating varieties," *Journal of Food Engineering*, vol. 27, pp. 397–407, 1996.

[8] D. J. Guzman and P. Engelbert, "Classification of philippine rice grains using machine vision and artificial neural networks," in *World conference on Agricultural information and IT*, 2008.

[9] C. Peralta, "Modeling shapes using uniform cubic b-splines for rice seed image analysis," in *2016 Sixth Int. Conf. on Comm. and Elec.*, July 2016.

[10] G. Barbosa-Canovas, *Hyperspectral Imaging Technology in Food and Agriculture*. Springer, 2015.

[11] L.Wang, D.Liu, H.Pu, and D.W.Sun, "Use of hyperspectral imaging to discriminate the variety and quality of rice," *Food Anal. Methods*, vol. 8, pp. 515–523, 2015.

[12] Y. Shao, C.Zhao, Y.Bao, and Y.He, "Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy," *Food Bioprocess Technol*, vol. 5, pp. 100–107, 2012.

[13] W. Kong, C. Zhang, F. Liu, P. Nie, and Y. He, "Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis," *Sensors*, vol. 13, no. 7, p. 8916, 2013.

[14] N.Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Sys., Man., Cyber*, vol. 9, pp. 62–66, 1979.

[15] C. Chang and C. Lin. (2016, May) Libsvm – a library for support vector machines. [Online]. Available: https://www.csie.ntu.edu.tw/ cjlin/libsvm/

[16] A. Liaw. (2016, May) Classification random forest. [Online]. Available: http://cran.r-project.org/web/packages/randomForest