Comparative study on vision based rice seed varieties identification

Phan Thi Thu Hong
Dept. of Computer Science
Vietnam National Uni. of Agriculture
Hanoi, Vietnam
ptthong@vnua.edu.vn

Tran Thi Thanh Hai
MICA
Hanoi Uni. of Science and Technology
Hanoi, Vietnam
thanh-hai.tran@mica.edu.vn

Le Thi Lan
MICA
Hanoi Uni. of Science and Technology
Hanoi, Vietnam
Thi-Lan.Le@mica.edu.vn

Vo Ta Hoang
MICA
Hanoi Uni. of Science and Technology
Hanoi, Vietnam
tahoanght91@gmail.com

Vu Hai
MICA
Hanoi Uni. of Science and Technology
Hanoi, Vietnam
haicuhn@gmail.com

Thuy Thi Nguyen
Dept. of Computer Science
Vietnam National Uni. of Agriculture
Hanoi, Vietnam
ntthuy@vnua.edu.vn

***Abstract*— This paper presents a system for automated classification of rice variety for rice seed production using computer vision and image processing techniques. Rice seeds of different varieties are visually very similar in color, shape and texture that make the classification of rice seed varieties at high accuracy challenging. We investigated various feature extraction techniques for efficient rice seed image representation. We analyzed the performance of powerful classifiers on the extracted features for finding the robust one. Images of six different rice seed varieties in northern Vietnam were acquired and analyzed. Our experiments have demonstrated that the average accuracy of our classification system can reach 90.54% using Random Forest method with a simple feature extraction technique. This result can be used for developing a computer-aided machine vision system for automated assessment of rice seeds purity.**

Keywords—Computer vision; image processing; rice seed; morphological features; GIST feature; SIFT feature; KNN; SVM; Random Forest;

I. Introduction

Rice is the most important agricultural plant in Vietnam and many other countries. In general, to obtain high yield rice crops, it is necessary to well prepare all the stages for growing, one needs to have good rice seed quality, in which the purity of rice seed is one the most important factor. The rice seed of certain variety must not be mixed with seeds from other varieties. To ensure the purity of rice seeds of a certain rice variety, it is necessary to identify the unwanted seeds from other varieties that may be mixed in the interested rice seed samples. Currently in seed companies of Vietnam, the process is done manually by naked eyes of skillful experts/technicians based on visual features of rice seeds. It's laborious, time consuming, inefficient and may cause degrade in the quality of seeds and therefore losses in the productivity. Therefore, we need methods and techniques that can identify rice seed variety in mixed varieties. Hence, developing an automatic computer-aided machine vision system to assess rice seeds for determining seed's purity is a demanding task.

Computer vision and image processing have attracted more and more interest of researchers because of its wide applications in many fields ranging from industry product inspection, traffic surveillance, entertainment to medical operations [1]. In agricultural production, it has been successfully applied to automatic assessing, harvesting, grading of products such as food, fruit, vegetables or plant classification [2, 3]. Machine vision was also utilized for discriminating different varieties of wheat and for distinguishing wheat from non-wheat components [4, 5] or for identifying damaged kernels in wheat using a color machine vision system [6].

Regarding quality evaluation of rice grains, many computer-aided machine vision systems, that automatically inspect and quantitatively measure grains, have been widely developed [7, 8]. These systems apply computer vision technologies including several stages, which require advanced computer knowledge, especially in artificial intelligence. The most important steps are image data collection, feature extractions (such as shape, size, color, and orientation etc.) and their representation, model/algorithm selection and learning, and model testing. For example, Gerard van Dalen [8] extracted characteristics of rice using flatbed scanning and image analysis. Jose D Guzman et al. [9] investigated grain features extracted from each sample image. They then utilized multilayer artificial neural network models for automatic identification the sizes, shapes, and variety of samples of 52 rice grains in Philippine. Goodman et al. [10] measured physical dimensions such as grain contour, size, color variance and distribution, and damage; Lai et al. [11] applied interactive image analysis method for determining physical dimensions and classify the variety grains. Sakai et al. [12] demonstrated the use of two-dimensional image analysis for the determination of the shape of brown and polished rice grains of four varieties. Zhao-yan et al. [13] implemented a method of identification based on neural network to classify rice variety using color and shape features. Mousavi Rad et al. [14] used morphological features and back propagation neural network to identify five different varieties of rice. Kong et al. [19] proposed to use Near – Infrared hyperspectral imaging and multivariate data analysis for identifying rice seed cultivar.

In Vietnam, Industrial Machinery and Instruments Holding Joint Stock Company (IMI) has developed a machine for sorting rice grains. Main functions are to classify grains utilizing simple boundary detection techniques and sensors for separating rice grains from artifacts (such as glass, brick rice) based on reflections of the IR light source. The system was developed for rice grain classification for colored and broken grains. It was not designed for rice seed purity assessment and rice variety recognition has not been used by seed production plants and farmers.

Da-Wen Sun showed that visual attributes of rice grains that affects the quality evaluation have been investigated using various computer vision techniques [7] and there are many computer vision systems for industrial applications as well as in agriculture as previously mentioned. However, up to our knowledge, there is no machine vision system for analyzing the visual features of rice seeds to determine the purity of variety in rice seeds processing for mass cultivation.

Therefore, in this paper we propose a machine vision system for rice seed variety identification. We focus on analyzing visual features (such as color, shape, and texture of the seeds) for efficient representation of rice seed images (each image is captured by our capture setup). We then implement different advanced machine learning models such as KNN, SVM, RF to evaluate rice seed images using these features. This allows one to select the best features for rice seed image description and a classifier with high accuracy to classify the rice seed images. The system can assist in recognizing the desired variety at high accuracy and can be deployed to aid technicians at the rice seeds producing plants in Vietnam. The remainder of this paper is organized as follows. Section 2 introduces materials and methods. Section 3 demonstrates our experimental results and discussion. Conclusion and future work are in Section 4.

II. MATERIALS AND METHODS

A. *Rice seed samples*

Six common cultivated rice seed varieties in Northern Vietnam, including *BC-15, Hương thơm 1, Nếp-87, Q-5, Thiên_ưu-8, Xi-23* were considered. The rice seeds are sampled from a rice seed production company where the rice varieties were grown and harvested following certain conditions for standard rice seeds production (Thaibinh and Hanoi regions in the north of Vietnam).

B. *Image Acquisition*

A CMOS image sensor color camera (NIKON D300S) with resolution of 640 x 480 pixels was used to acquire images. We set up a chamber with a white table as background for taking images. Rice seeds are manually spreaded inside an area of 10 * 16 cm2. Each image taken by this imaging system contains 30 to 60 seeds.

We have acquired totally 212 "big" images. Single rice seed image will be segmented from these images.

C. *Image segmentation*

In order to separate rice seed images from the acquired images into the individual rice seed images, we realized the image segmentation. Because the image background is unique in all experiments, we chose a threshold method for background subtraction. Moreover, we observed that the blue channel of images has an intensity that can distinguish the background and the rice seeds. That is why we used threshold method that is based on the similarity of intensity value of the image's blue channel. In the image's blue channel, the intensity of rice seed pixels is always less than or equal to 90 and the intensity of background is always higher than 90. In the image segmentation process, all the pixels with blue value greater than 90 were assigned the value 0, and all pixels with blue value less than 90 were assigned the value 255. After threshold image was created, we crop the rice seed images base on the object contours (Fig. 1.), each image now contains only one rice seed with a minimum bounding box. From now, when we say rice seed image, we refer to this set of images.

a. A sample of acquired image b. A thresholded image

Fig. 1. An example of acquired image and the segmentation

D. *Image description*

Once the image of a rice seed is segmented, image descriptor must be computed to represent the image, which will be input to a classifier. The image descriptor describes properties of an image, image regions or individual image location. These properties are typically called "features". Research in the field of image description or feature extraction started at the 60's. Until now, uncountable image descriptors have been proposed. They could be divided into categories following some criteria such as global vs. local, intensity vs. derivative or spectral based. In general, a good feature should be invariant to rotation, scaling, illumination, and viewpoint changes.

In this work, we investigate four feature types that could be considered as representative of two main groups of features: global features (Morphological features, Color, Texture, GIST) and local feature (SIFT). Morphological features are the most classical features to describe shape of the object in image. Color and texture are very useful to distinguish objects when their shapes remain similar. GIST is a powerful global feature computed based Gabor filter bank applied on the whole image [15]. GIST has been shown to be very efficient for scene classification. SIFT is a local feature proposed by Lowe [18]. SIFT possesses all desired properties to be a good feature and now still keep its position in the field.

1) *Basic descriptor*

This is a combination of morphological features, color features and texture features to build a descriptor; we call it *basic descriptor* for reference.

a. Morphological descriptors

The morphological features were extracted from the images of individual rice seeds. A morphological feature descriptor with 8 dimensions is calculated as following:

- Area: It is the number of pixels inside, and including the seed boundary.
- Length: It is the length of the minimum bounding box of the rice seed.
- Width: It is the width of the minimum bounding box of the rice seed.
- Length/width: It is the ratio of Length to Width.
- Major axis length: It is the longest diameter of ellipse bounding rice.
- Minor axis length: It is the shorted diameter of ellipse bounding rice.
- Area of convex hull of a rice seed.
- Perimeter of convex hull of a rice seed.

b. Color

The RGB components of all images were analyzed. We got mean values of individual channels were computed. The color feature of rice seed for image analysis with 6 dimensions including:

- *R, G, B:* are the mean values of R, G, B channel.
- *RS, GS, BS* are square root of the value mean of channel R, G, B.

c. Texture

Texture feature are calculated as:

- Mean (m): $\sum_{i=1}^{L-1} z_i p(z_i)$
- Standard deviation (σ) : $\sqrt{(z_i - m)^2 . p(z_i)}$
- Uniformity : $\sum_{t=0}^{L-1} p^2(z_i)$
- Third moment : $\sum_{i=1}^{L-1} (z_i - m)^3 p(z_i)$

Where, z_i is the gray-scale intensity $p(z_i)$ is the ratio of number of pixels that have the intensity z_i and number of pixels in an image. The texture feature has 4 components.

Finally, we combine these component descriptors (morphological, color, texture) to obtain a descriptor of 18 dimensions.

2) GIST descriptor

Oliva and Torralba [15] proposed the GIST descriptor for scene classification. This descriptor represents the shape of scene itself, the relationship between the outlines of the surfaces and their properties while ignoring the local objects in the scene and their relationships. The main idea of this method is to develop a low dimensional representation of the scene, which does not require any form of segmentation. The representation of the structure of the scene is defined by a set of perceptual dimensions: naturalness, openness, roughness, expansion and ruggedness.

To compute GIST descriptor, firstly, an original image is converted and normalized to gray scale image I(x,y). We then apply a pre-filtering to I(x,y) in order to reduce illumination effects and to prevent some local image regions to dominate the energy spectrum. The filtered image I(x,y) then is decomposed by a set of Gabor filters. A 2-D Gabor filter is defined as follows:

$$h(x, y) = e^{-\frac{1}{2}\left(\frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2}\right)} e^{-j2\pi(u_0 x + v_0 y)}$$

Configuration of Gabor filters contains 4 spatial scales and 8 directions. At each scale (δ_x, δ_y), by passing the image I(x,y) through a Gabor filter *h(x,y),* we obtain all those components in the image that have their energies concentrated near the spatial frequency point (u_0, v_0). Therefore, the GIST vector is calculated using energy spectrum of 32 responses. To reduce dimensions of feature vector, we calculated averaging over grid of 4x4 on each response. Consequently, the GIST feature vector is reduced to 512 dimensions.

3) SIFT descriptor

Lowe [18] introduced a scale invariant feature transform (SIFT) that is invariant to image scaling, translation, rotation, partially invariant to illumination changes. The computation of SIFT features consists of 4 steps: (i) scale-space extrema of Laplacian of Gaussian (LoG) extraction; (ii) keypoint localization; (iii) canonical orientation assignement; (iv) keypoint description. First, local extrem of Laplacian in scale space are extracted. This is efficiently done by constructing a Gaussian pyramid and detecting local extrema of difference of Gaussians (DoG). By this way, keypoints are invariant to scale change. These points detected will be next re-localized to improve precision in localization. Each point is then assigned a canonical orientation such that following which the description of the keypoint is invariant to rotation. The description of the keypoints is finally designed by building a array of histograms of gradient orientations. This description is more compact and signficantly discriminant than the signal image itself. Finally, to describe an image from SIFT features, state of art works are normally based upon Bag of Word (BoW) technique. The size of the descriptor depends of the preset size of vocabulary in BoW model *(200* in our experiments).

E. Classification

After feature extraction, a classifier is learned for identification of different rice varieties. In the following, we review some prominent classification models:

1) K-nearest neighbor

K-nearest neighbor (KNN) [20] is a method for classifying based on k nearest neighbors and then predicts the class of a new sample as the most frequent one occurring in the neighbors. This method has been used widely in classification problems because it is simple, effective and non- parametric [21].

2) Support vector machine

The basic idea of support vector machine (SVM) [16] is to find an optimal hyper-plane for linearly separable patterns in a high dimensional space where features are mapped onto. There is more than one hyper-plane satisfying this criterion. The task is to detect the one that maximizes the margin around the separating hyper-plane. This finding is based on the support vectors which are the data points that lie closest to the decision surface and have direct bearing on the optimum location of the decision surface.

SVMs are extended to classify patterns that are not linearly separable by transformations of original data into new space using kernel function into a higher dimensional space where classes become linearly separable. SVM is one of the most powerful and widely used in classifiers.

3) Random Forest

Breiman [17] proposed random forest (RF), a classification technique built by constructing an ensemble of decision trees. For each tree, RF uses a different bootstrap sample of the response variable and changes how the classification or regression trees are constructed: each node is splited using the best among a sub-set of predictors randomly chosen at that node, and then grows the tree to the maximum extent without pruning. For predicting new data, a RF aggregated the outputs of all trees. It is effective and fast to deal with a large amount of data and has shown that can perform very well compared to many other classifiers, including discriminant analysis, support vector machines and neural networks, and is robust against over-fitting [17].

III. EXPERIMENT AND DISCUSSION

1) Experiment data

We have conducted a set of experiments on the extracted feature types and classification models to evaluate their performance on image data of six common Vietnam rice seed varieties consisting BC-15, Hương thơm 1, Nếp-87, Q5, Thiên_ưu-8, Xi-23. Totally we have acquired six datasets; each dataset represents samples of rice seed of each variety. Some of examples of the rice seed images are shown in Fig2. Table 1. shows the number of rice seed images in each rice variety of each dataset.

2) Experiment set up

To conduct all experiments, we used a computer with 64bit Window 7, core i5, CPU 1.70 GHz (4 CPUs) and 4 GB main memory; matlab 2013a and R version 3.2.0. To build data set for each rice seed variety, we chose all of examples with positive labels and choose five other rice seed images for negative labels so that number of examples with positive labels approximate the number of examples with negative labels. To ensure fairly comparison of different classification methods, we fixed the test set and training set and used the Out-Of- Bag technique for estimating the generalization error [17]. So, about the 67% of the samples (for each rice variety data) were randomly selected as training set, while the rest of the samples were used as test set for classification.

Table 1. Description of rice seed image dataset

Rice variety	**Number of individual rice seed images**
BC-15	3680
Hương thơm 1	4152
Nếp-87	2877
Q-5	3019
Thiên ưu-8	2011
Xi-23	4152

To use KNN, SVM and RF methods for classifying rice seeds, in the first step, we perform extract different types of features: global features (Morphological features, Color, Texture, GIST) and local feature. In the next step, after finishing the training process, the classification models were used to test with on the test datasets. The classification performance and classification accuracy of these methods were given in **Table 2., Table 3., and Table 4.**.

Fig. 2. Images of rice seed examples in 6 datasets

With the KNN method, one of the most important parameters is choice of suitable value of K. In our experiment, we test with different values of K (K = 1 to 55) and KNN model gave the best results when $K = 23$.

With support vector machine, we used linear function.

For random forest (RF), it is necessary to specify two parameters to train the model: *ntree* - number of trees to be constructed in the forest and *mtry* - number of input variables randomly sampled as candidates at each node. We used *ntree*=500, $mtry = \sqrt{number\ of\ features}$ for GIST and SIFT features. In particular, all of features (18 features) were chosen for simple features.

In this study, three measures were used to evaluate the performance of different classification methods on various feature types. These measures are defined as follows:

Precision (P) is the proportion of the predicted positive samples that were correctly classified:

$$Precision = \frac{tp}{tp + fp}$$

The recall (R) or true positive rate (TP) is the proportion of positive samples that were correctly identified

$$Recall = \frac{tp}{tp + fn}$$

Finally, Fmeasure (F) is a measure of a test's accuracy, as calculated using the equation:

$$Fmeasure = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

In which *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, respectively.

3) Results discussion

The reliability of classification models was based on classification performance and classification accuracy. The classification results of these methods using different types of features are shown in **Table 2., Table 3., and Table 4.**.

As can be seen from Table 2., the rice seed varieties were classified based on the basic feature. With this kind of feature, RF has proven a good capability for classifying all six rice seed cultivars, with classification accuracy above 85%. It obtained a highest classification accuracy of 95.71% for Nếp-87 and the average classification rate of 90.54%. The KNN model and SVM model indicate relatively low effectiveness

Table 2. Performance results of different classification models on basic feature

Rice seed name	KNN				SVM				RF			
	P	*R*	*F*	*Acc*	*P*	*R*	*F*	*Acc*	*P*	*R*	*F*	*Acc*
BC-15	80.03%	70.96%	75.22%	73.81%	60.76%	69.61%	64.88%	67.35%	85.65%	86.74%	86.19%	85.87%
Hương thơm 1	79.35%	84.37%	81.18%	81.85%	77.74%	82.12%	79.87%	79.87%	88.34%	89.09%	88.71%	88.63%
Nếp-87	81.07%	74.92%	77.87%	78.29%	91.55%	84.93%	88.12%	88.36%	95.24%	96.29%	95.76%	95.71%
Q-5	85.75%	72.86%	78.78%	76.55%	84.51%	75.93%	79.99%	78.52%	90.03%	89.90%	89.97%	90.40%
Thiên ưu-8	82.30%	83.45%	82.87%	72.41%	92.92%	90.39%	91.64%	63.42%	93.65%	94.02%	93.84%	93.95%
Xi-23	81.68%	72.52%	76.83%	81.27%	64.55%	68.37%	66.41%	90.66%	86.90%	87.42%	87.16%	88.67%
Average	81.70%	76.51%	78.79%	77.36%	78.67%	78.56%	78.49%	78.03%	89.97%	90.58%	90.27%	90.54%

Table 3. Performance results of different classification models on GIST feature

Rice seed name	KNN				SVM				RF			
	P	*R*	*F*	*Acc*	*P*	*R*	*F*	*Acc*	*P*	*R*	*F*	*Acc*
BC-15	70.84%	64.76%	67.66%	66.39%	67.65%	66.39%	67.02%	66.94%	66.12%	69.24%	67.64%	67.26%
Hương thơm 1	90.21%	70.32%	79.04%	75.33%	79.78%	77.23%	78.48%	77.46%	75.24%	80.99%	78.01%	77.07%
Nếp-87	89.55%	87.35%	88.43%	89.20%	94.48%	93.51%	93.99%	94.43%	90.40%	91.27%	90.83%	90.83%
Q-5	74.51%	70.51%	72.46%	71.24%	71.45%	70.71%	71.08%	70.47%	73.01%	67.81%	70.32%	70.70%
Thiên ưu-8	88.26%	89.33%	88.79%	88.17%	92.42%	92.69%	92.55%	92.10%	92.84%	83.59%	87.98%	88.46%
Xi-23	89.21%	73.98%	80.89%	74.25%	77.30%	83.15%	80.12%	76.66%	79.13%	66.94%	72.53%	76.15%
Average	83.76%	76.04%	79.55%	77.43%	80.51%	80.61%	80.54%	79.68%	79.46%	76.64%	77.89%	78.41%

Table 4. Performance results of different classification models on SIFT feature

Rice seed name	KNN				SVM				RF			
	P	*R*	*F*	*Acc*	*P*	*R*	*F*	*Acc*	*P*	*R*	*F*	*Acc*
BC-15	95.99%	58.35%	72.58%	63.99%	81.45%	82.78%	82.10%	82.38%	83.14%	80.10%	81.59%	83.34%
Hương thơm 1	89.14%	81.49%	85.14%	83.95%	89.54%	91.29%	90.41%	90.20%	90.07%	87.91%	88.98%	89.12%
Nếp-87	53.48%	97.11%	68.98%	77.83%	87.33%	85.94%	86.63%	87.57%	87.16%	94.13%	90.51%	89.86%
Q-5	82.63%	62.05%	70.87%	65.51%	73.22%	74.40%	73.80%	73.59%	76.36%	73.10%	74.69%	75.45%
Thiên ưu-8	49.72%	90.32%	64.13%	70.46%	85.39%	85.68%	85.53%	84.68%	86.64%	85.43%	86.03%	86.41%
Xi-23	91.91%	69.46%	79.12%	70.38%	81.18%	88.94%	84.88%	82.34%	82.36%	85.19%	83.75%	85.13%
Average	77.15%	76.46%	73.47%	72.02%	83.02%	84.84%	83.89%	83.46%	84.29%	84.31%	84.26%	84.89%

with average classification performance of 78.79% and 78.49%. And BC-15 got poor prediction accuracy in all models. This result is similar to GIST feature based models (**Table 3.**), which implied that BC-15 was difficult to identify, and appropriate models could help to obtain more accurate classification.

In **Table 3.**, SVM model demonstrated the ability of classification better than RF and KNN method when using GIST feature. The SVM model obtained the highest classification accuracy of 94.43% and then 90.83%, 89.2%. for RF and KNN model with Nếp-87.

In **Table 4.**, considering the prediction performance, KNN was the worst classification model on the SIFT feature. In contrast, SVM and RF models give similar results (average rate 83.89%, and 84.26%).

Based on the results of classification of six rice seeds varieties (**Table 2.. 4.**), RF gave the best performance using basic feature (90.27%). In contrast, KNN indicated the least classification capability on all kinds of features. When using GIST features, SVM model demonstrated the ability of classification better than the two remaining methods.

From the results, we see that basic feature (morphological features, color and texture) with RF method has demonstrated its strengths to identify rice seed (average accuracy achieves 90.54%) in comparison with the two remaining features. GIST is a global feature and has been shown to be very efficient for scene classification but it is not been strong enough for describing in detail to distinguish the rice seed varieties.

Unlike Gist feature, Sift is a local feature and has all properties to be a good feature. However, with rice seed varieties identification, Sift does not prove its advantage in describing the shape of rice seeds, particularly their shapes remain similar, one crucial factor for classification.

From the above analysis, we propose to use basic features combined with RF for identifying of rice seeds cultivars.

IV. CONCLUSION AND FUTURE WORKS

In this study, we focused on analysing visual features of rice seed images such as colour, shape, texture, GIST and SIFT. We then applied different classification models using these types of features. This research indicated that image processing techniques can combine with classification techniques such as KNN, SVM, RF to identify rice seeds in mixed samples. RF method using simple features proved the best capability and accuracy of classification, average achieved 90, 27%, 90.54% respectively.

The present work can be deployed at the rice seeds production plants in Vietnam and extended for other varieties also other types of features can be extracted to increase the performance of classification models.

ACKNOWLEDGMENT

The authors thank the CUD Programme of Belgium with Vietnam National University of Agriculture (VNUA) 2014 - 2019 for supporting this work.

REFERENCES

[1] Szeliski Richard , "Computer Vision: Algorithms and Applications," *Springer*, 2010.

[2] Tadhg Brosnan and Da-Wen Sun, "Inspection and grading of agricultural and food products by computer vision systems—a review," *Computers and Electronics in Agriculture*, vol. 36, no. 2-3, pp. 193-213, November 2002.

[3] Cheng-Jin Du and Da-Wen Sun, "Learning techniques used in computer vision for food quality evaluation: a review," *Food Engineering*, vol. 72, no. 1, pp. 39-55, January 2006.

[4] I Zayas , F S Lai , and L Y Pomeranz, "Discrimination between wheat classes and varieties by image analysis," *Cereal Chemistry*, vol. 63, pp. 52-56, 1986.

[5] I Zayas, Y Pomeranz, and F S Lai, "Discrimination of wheat and nonwheat components in grain samples by image analysis," *Cereal Chemistry*, vol. 66, pp. 233-237, 1989.

[6] X Luo, D S Jayas, and S J Symons, "Identification of damaged kernels in wheat using a color machine vision," *Cereal Science*, vol. 30, pp. 45-59, 1999.

[7] Da-Wen Sun, "Computer Vision Technology for Food Quality Evaluation," , 2008.

[8] Gerard van Dalen, *Characterisation of rice using flatbed scanning and image analysis*, Arthur P. Riley, Ed, 2006

[9] Guzman D Jose and Peralta K. Engelbert, "Classification of Philippine Rice Grains Using Machine Vision and Artificial Neural Networks," in *World conference on Agricultural information and IT*, 2008, pp. 41-48.

[10] D.E. Goodman and R.M. 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, Mar. 1984.

[11] F.S. 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.

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

[13] L. Zhao-yan, C. Fang, Y. Yi-bin, and R-Xiu-qin, "Identification of rice seed varieties using neural networks," *Journal of Zhejiang University Sscience*, vol. 11, pp. 1095-1100, 2005.

[14] S.J. Mousavirad, F. A. Tab, and K. Mollazade, "Design of an Expert System for Rice Kernel Identification using Optimal Morphological Features and Back Propagation Neural Network," *International Journal of Applied information systems*, vol. 3, pp. 33-37, 2012.

[15] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vision*, vol. 42, pp. 145-175, May 2001.

[16] V. Vapnik, "The Nature of Statistical Learning Theory," *Springer-Verlag*, 1995.

[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[18] D. Lowe, "Distinctive Image Features from Scale-Invariant Key-points," *Int'l J. Computer Vision*, vol. 2, no. 60, pp. 91-110, 2004.

[19] 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, pp. 8916-8927, 2013.

[20] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175-185, Aug. 1992.

[21] W. A. David, K. Dennis, and K. A. Marc , "Instance–based learning algorithms," *Machine learning*, vol. 6, no. 1, pp. 37-66, 1999.