Poster Session | Others (including the category of JSAM and SASJ)
[6-1130-P] Other Categories (6th)
Fri. Sep 6, 2019 11:30 AM - 12:30 PM Poster Place (Entrance Hall)

[6-1130-P-13] Classification of Sugarcane Variety using Image Processing and Multivariate Analysis

*KITTIPON APARATANA[1], Hiroo Takaragawa[1,2], Yoshinari Izumikawa[1,2], Eizo Taira[1] (1. Faculty of agriculture, University of the Ryukyus, Okinawa 903-0213(Japan), 2. The United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima 890-0065(Japan))

Keywords: Sugarcane, Variety classification, PCA, PLS-DA, SVM-C

Sugarcane variety classification is essential for data collecting and learning for the breeder. It is difficult for a farmer to identify a sugarcane variety without specialist help. In this research, three Japanese sugarcane varieties (Ni15, Ni22, and Ni27) from six areas in the south of Japan were classified according to full pixel and color features of the sugarcane bud. The 54 images of sugarcane bud were acquired from the sugarcane field using a mobile phone’s digital camera, equipped with a fixed distance accessory. To develop classification models, two types of data; The full pixel data and color feature data from images were investigated for input to the model. The full pixel and color features were subjected to Principal component analysis (PCA) to describe the sugarcane bud samples. Then, the samples were classified into varieties by performing partial least squares discriminant analysis (PLS-DA) and support vector machine classification (SVM-C). The results of the full pixel show that the pooled classification rates (averaged three classification rate) by PLS-DA and SVM-C were 79.6% and 84.5%, respectively, while the pooled classification rates by PLS-DA and SVM-C of the color features were 75.9% and 74.1%, respectively. Therefore, these results show that the size and color spaces of sugarcane buds could be the keys to classifying sugarcane varieties and that the best way of classifying Japanese sugarcane (Ni15, Ni22, and Ni27) was the SVM-C method using full pixel of sugarcane bud.

Classification of Sugarcane Variety using Image Processing and Multivariate Analysis

Kittipon Aparatana[1*], Hiroo Takaragawa[1, 2], Yoshinari Izumikawa[1, 2] and Eizo Taira[1]

[1]Department of Agriculture, University of the Ryukyus, Japan
[2]United Graduate School of Agricultural Sciences, Kagoshima University, Japan

*Corresponding author: kittipon.aparatana@gmail.com

ABSTRACT

Sugarcane variety classification is essential for data collecting and learning for the breeder. It is difficult for a farmer to identify a sugarcane variety without specialist help. In this research, three Japanese sugarcane varieties (Ni15, Ni22, and Ni27) from six areas in the south of Japan were classified according to full pixel and color features of the sugarcane bud. The 54 images of sugarcane bud were acquired from the sugarcane field using a mobile phone's digital camera, equipped with a fixed distance accessory. To develop classification models, two types of data; The full pixel data and color feature data from images were investigated for input to the model. The full pixel and color features were subjected to Principal component analysis (PCA) to describe the sugarcane bud samples. Then, the samples were classified into varieties by performing partial least squares discriminant analysis (PLS-DA) and support vector machine classification (SVM-C). The results of the full pixel show that the pooled classification rates (averaged three classification rate) by PLS-DA and SVM-C were 79.6% and 84.5%, respectively, while the pooled classification rates by PLS-DA and SVM-C of the color features were 75.9% and 74.1%, respectively. Therefore, these results show that the size and color spaces of sugarcane buds could be the keys to classifying sugarcane varieties and that the best way of classifying Japanese sugarcane (Ni15, Ni22, and Ni27) was the SVM-C method using full pixel of sugarcane bud.

Keywords: Sugarcane, Variety classification, PCA, PLS-DA, SVM-C.

1. INTRODUCTION

Sugarcane is a critical economic crop in the world and Japan. Currently, in Japan, sugarcane is generally planted in the southern part during the summer or spring and then harvested in winter, and production has increased from 2015 to 2018 (Okinawa Prefectural, 2018). However, severe droughts and tropical storms (typhoons) frequently occur from July to September, which causes severe damage to sugarcane yield and sugar content through the loss of green leaves, lodging, and broken stalks (Takagi et al., 2005; Terauchi et al., 2012). Any delays in planting and ratooning due to planting and harvest conflicts will consistently affect the next season's harvest (Terauchi et al., 2012). Moreover, the loss of sugarcane continues for many reasons, including rotting, disease, parasitism, and harvesting (Kawanobe et al., 2017; Sharma and Tamta, 2015; Shinzato et al., 2015).

Sugarcane variety databases are key indices to learning and improving sugarcane variety for high yielding, high sucrose content, high biomass (Matsuoka, 2006), and high durability of a natural disaster. Generally, several sugarcane factories in Japan obtain data of sugarcane variety through inquiries with farmers and with the experience of a specialist. The sugarcane variety classification method nowadays is inconvenient, and it is difficult for a farmer to identify a sugarcane variety without specialist help, which can affect the quality of the database. Sugarcane variety classification mostly uses pictorial identification techniques based on bud shape, dewlap shape, leaf shape, etc. (Gravois et al., 2018; Takaragawa et al., 2019). Nevertheless, these techniques need a long time to correct the data and need the experience to be identified. Thus, there is a need for new tools or methods that could work faster, be more accurate, and be more convenient to use to identify sugarcane variety.

With recent advancements in computer technology, the image can extract much information from image data, such as many types of color space and intensity of color, with the image processing technique, which was widely used for detection or identification in the agriculture and food industry because it is fast, accurate, and cost-efficient (Chen et al., 2010; Khan and Yadav, 2017; Moshashai et al., 2008). However, sugarcane variety classification using image processing has not been researched yet.

Therefore, the current research focuses on classifying Japanese sugarcane varieties using image processing. This research aims to use full pixels and color features of sugarcane bud images to describe and separate sugarcane varieties using multivariate analysis methods.

2. MATERIALS AND METHODS

Matlab R2018a (version: 9.4.0.813654, The Math Works, Natick, MA, USA, 2017) with the PLS toolbox (Eigenvector Research, Inc., Manson, WA, USA, 2017) was used for data processing and analysis.

2.1 Sugarcane samples

In this research, as shown in Figure 1, three Japanese sugarcane varieties (Ni15, Ni22, and Ni27) from six areas in Southern Japan (Minami island, Ishigaki island, Miyako island, Okinawa Nanbu, and two difference crops in University of the Ryukyus) were selected as sugarcane variety samples for classification according to full pixel image and color features. The image of 54 sugarcane bud samples (18 samples per variety) was acquired from the sugarcane field using a mobile phone's digital camera (iPhone SE, Apple Inc, USA) equipped with a fixed distance accessory. The first dimensions of the images were 3024 x 4032 pixels in JPG-format.

(a) (b) (c)

Figure 1. The sample image of sugarcane variety samples; (a) Ni15, (b) Ni22, and (c) Ni27.

2.2 Image processing

As shown in Figure 2, the images were then cropped on the bud area and their sizes reduced to 100 x 100 pixels in order to diminish the time and load of the analysis process. The acquired image is generally displayed in three-dimensional RGB color space. However, RGB color space is not perceptually uniform, and the proximity of colors does not indicate a color similarity. Color space transformations make for an effective means of distinguishing color images. The classification performance could be improved by weighting each color component differently. For this research, The RGB color space was evaluated as normalized RGB, YCbCr, and HSV color spaces.

The normalized RGB can be obtained from Eq. (1); in order to remove the brightness from the RGB color space, one can normalize the values of red, green, and blue.

$$
\begin{aligned}
r &= R/(R+G+B)\\
g &= G/(R+G+B)\\
b &= B/(R+G+B)
\end{aligned}
\tag{1}
$$

The YCbCr can be obtained from Eq. (2) (Umbaugh, 2005).

$$
\begin{aligned}
&Y= 0.299R-0.587G+0.114B \\
&Cb=-0.1687-0.3313G+0.500B+128 \\
&Cr=0.500R-0.4187G+0.0813B+128
\end{aligned} \quad (2)
$$

As such, the Y element represents the luminance component, and the Cb and Cr elements represent two chrominance components.
The 12 color spaces were then extracted to 24 color features by computing the mean and standard deviation of color spaces. Subsequently, two types of data, the full pixel data and color feature data from images, were investigated for input to description analysis and classification analysis.

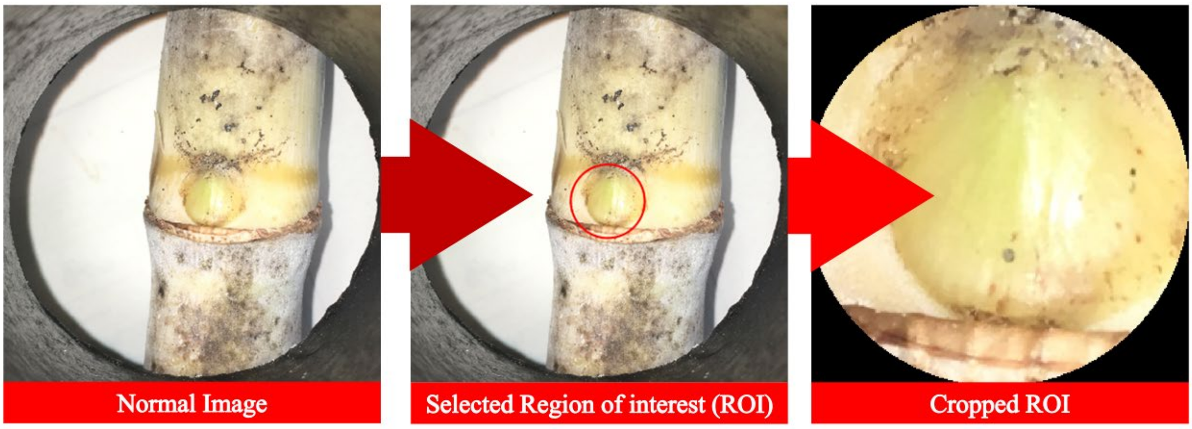
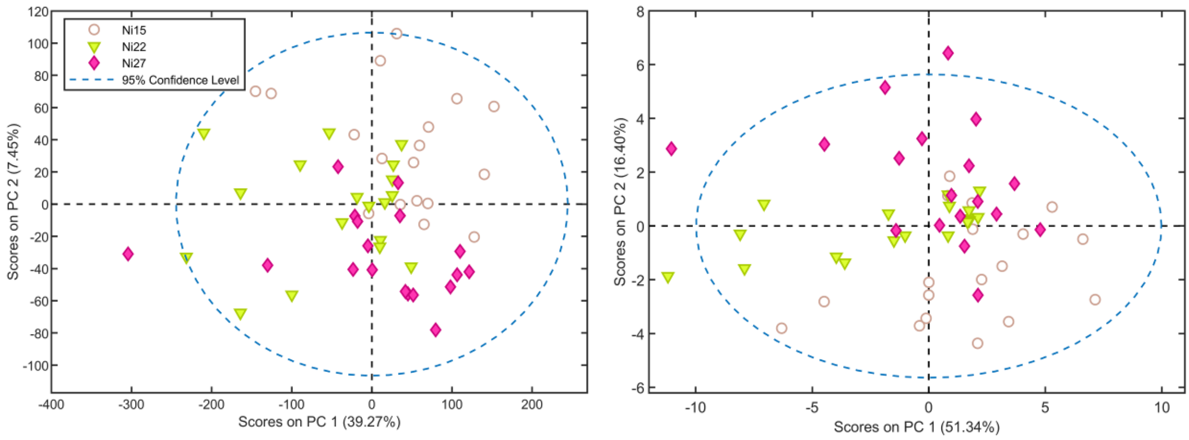


Figure 2. The example of sugarcane bud sample processing

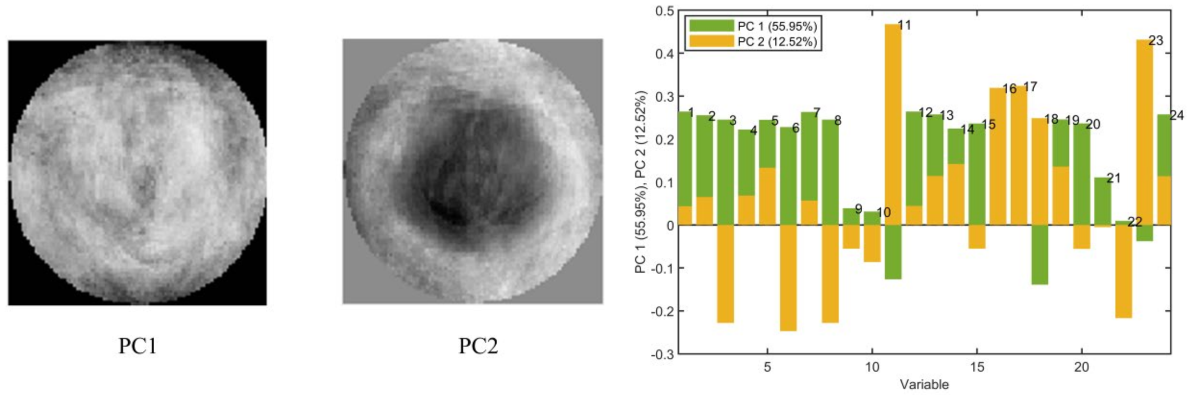
2.3 Multivariate analysis
The multivariate analysis techniques were objectives for description, classification, and prediction analysis. There are many types of multivariate data analysis techniques to choose from nowadays. The principal component analysis (PCA) is one of the frequently used methods for data description and explorative data structure modeling (Esbensen, 2000) and it's also one of the most critical and influential ways to decompose complex data (Bro and Smilde, 2014). Moreover, PCA could be used on a digital image for the benefit of learning and reducing size, as it enables locating the highest variance in data (Ng, 2017). The same goes for partial least squares discriminant analysis (PLS-DA) (Amigo et al., 2009) and support vector machine (SVM-C) (Zhang, 2012), which are the dominant methods for classifying data. Thus, this research chose PCA to describe the sugarcane bud samples and both PLS-DA and SVM-C for classifying sugarcane varieties.

3. RESULTS AND DISCUSSION
3.1 Principal component analysis results
Fifty-four of the sugarcane variety samples with two types of data (full pixel and 24 color features) were divided into three variety classes, resulting in 18 samples per variety. Then, a PCA analysis using the scores was undertaken to create a scattering plot of principal components 1 and 2, as shown in Figure 3. The sugarcane variety Ni15 and Ni27 were distinguished, but Ni22 overlapped a little with the other two varieties per the implementation of the first and second principal components in the two types of data. After recreating an image from full pixel loadings of principal component analysis, figure 4 (a) shows the first principal component was related to the lightness of the image; the second was related to the size of the sugarcane bud in the case of a full pixel by recreating an image from the loading of principal. Figure 4 (b) show in the color space of the sugarcane buds, the first principal component loading was mainly related to the mean of RGB, normalized RGB, Y, Cb, and V; the standard deviation of RGB, Y, Cb, and V. The second principal component loading mainly related to the mean of B, normalized b, Cb and S; the standard deviation of normalized RGB, H, and S.



(a) (b)

Figure 3. Score plots of principal component analysis for modeling samples (18 samples of sugarcane variety Ni15 in circle mark, Ni22 in triangle mark, and Ni27 in theta mark, respectively); (a) full pixel (b) color features.



(a) (b)

Figure 4. Loadings of principal component analysis; (a) full pixel (b) color features (where 1, 2, and 3 means mean of RGB. 4, 5, and 6 means mean of normalized RGB. 7, 8, and 9 means mean of YCbCr. 10, 11, and 12 means mean of HSV. 13, 14, and 15 means standard deviation of RGB. 16, 17, and 18 means standard deviation of normalized RGB. 19, 20, and 21 means standard deviation of YCbCr. 22, 23, and 24 means standard deviation of HSV.)

3.2 Classification of PLS-DA and SVM-V

Two types of data (full-pixel and 24-color features) and a two-class (model variety and other varieties) PLS-DA and SVM-C model were developed for variety classification. The 54 samples (18 model variety and 36 other variety) were used to create the PLS-DA and SVM-C model with Venetian blind cross-validation to determine the number of factors and evaluate the classification rate. The results presented in Table 1 reveal that the variety of Ni15 classification rates results of the full pixel by PLS-DA and SVM-C were 83.3% and 87.0%, respectively; the Ni22 classification rates by PLS-DA and SVM-C of color spaces were 74.1% and 83.3%, respectively; the Ni27 classification rates by PLS-DA and SVM-C of color spaces were 81.5% and 83.3%, respectively. Moreover, the results of the color features show that the sugarcane variety Ni15 classification rates by PLS-DA and SVM-C were 88.9% and 85.2%, respectively; the Ni22 classification rates by PLS-DA and SVM-C of color spaces were 66.7% and 72.2%, respectively; the Ni27 classification rates by PLS-DA and SVM-C of color spaces were 72.2% and 64.8%, respectively.

Table 1. Classification results of PLS-DA and SVM-C using cross-validation.

Classification methods		Classification rates	
		PLS-DA	SVM-C
Data types	Variety	%	%
Full pixel	Ni15	83.3	87.0
	Ni22	74.1	83.3
	Ni27	81.5	83.3
Color features	Ni15	88.9	85.2
	Ni22	66.7	72.2
	Ni27	72.2	64.8

4. CONCLUSION

The three Japanese sugarcane varieties (Ni15, Ni27, and Ni22) were mostly distinguished by the implementation of first and second principal components in the full pixel and color features. The samples were classified into varieties by performing a partial least squares discriminant analysis (PLS-DA) and a support vector machine classification (SVM-C). The results of the full pixel set show that the pooled classification rates by PLS-DA and SVM-C were 79.6% and 84.5%, respectively. Meanwhile, the pooled classification rates by PLS-DA and SVM-C of the color features set were 75.9% and 74.1%, respectively. However, this research could not correctly classify the sugarcane variety because the input images had various factors that might have affected the results, such as sunlight, a damaged sugarcane bud, the age of the cane, and fertility in the field. These results therefore show that the size and color spaces of sugarcane buds could be the keys to classifying sugarcane varieties. Moreover, the best way to classify Japanese sugarcane (Ni15, Ni22, and Ni27) was the SVM-C method using a full pixel of sugarcane bud.

REFERENCES

Amigo, J. M., Ravn, C., Gallagher, N. B., and Bro, R. 2009. A Comparison of a Common Approach to Partial Least Squares-Discriminant Analysis and Classical Least Squares in Hyperspectral Imaging. *International Journal of Pharmaceutics*, 373(1–2):179–82.

Bro, R. and Smilde, A. K. 2014. Principal Component Analysis. *Anal. Methods*, 6(9):2812–31.

Chen, X., Xun, Y., Li, W., and Zhang, J. 2010. Combining Discriminant Analysis and Neural Networks for Corn Variety Identification. *Computers and Electronics in Agriculture*, 71:S48–53.

Esbensen, K. H. (HiT/TF). 2000. *Multivariate Data Analysis In Pracetice*. 4th ed. CAMO ASA.

Gravois, K. et al. 2018. *Louisiana Sugarcane Variety Identification Guide*.

Kawanobe, M., Miyamaru, N., Yoshida, K., Kawanaka, T., and Toyota, K. 2017. A Field Experiment with Nematicide Treatment Revealed Potential Sugarcane Yield Loss Caused by Plant-Parasitic Nematodes in Okinawa, Japan. *Nematological Research (Japanese Journal of Nematology)*, 46(1):9–16.

Khan, A. and Yadav, M. S. 2017. Image Processing Based Disease Detection for Sugarcane Leaves. *International Journal Of Advance Research, Ideas And Inovations In Technology*, 3(4):497–502.

Matsuoka, M. 2006. Sugarcane Cultivation and Sugar Industry in Japan. *Sugar Tech*, 8(1):3–9.

Moshashai, K., Almasi, M., Minaei, S., and Borghaee, A. M. 2008. Identification of Sugarcane Nodes Using Image Processing.Pdf. *International Journal of Agricultural Research*, 3(5):357–64.

Ng, C. 2017. Principal Component Component Analysis Analysis to to Reduce Reduce Dimension Dimension on on Digital Digital Image Image Principal. Pp. 113–19 in *Procedia Computer Science*. Procedia Computer Science.

Okinawa Prefectural. 2018. *The Annual Report of Sugarcane and Sugar Production Results in Okinawa Prefecture 2017/18 (In Japanese)*.

Sharma, R. and Tamta, S. 2015. A Review on Red Rot : The “ Cancer ” of Sugarcane. *Journal of Plant Pathology & Microbiology*, 1–8.

Shinzato, Y., Uehara, K., and Ueno, M. 2015. Adaptability of Small-Sized Sugarcane Harvesters in Okinawa. *Engineering in Agriculture, Environment and Food*, 8(4):207–11.

Takagi, H., Sato, M., and Matsuoka, M. 2005. *A Guidebook for Sugarcane in Japan*. edited by M. Sato. Organizing Committee of the ISSCT Joint 12th Germplasm & Breeding and 9th Molecular Biology Workshops and Japanese Society of Sugar Cane Technologists (JSSCT).

Takaragawa, H., Watanabe, K., Kobashikawa, R., Hoang, D. T., and Kawamitsu, Y. 2019. Development of Sugarcane Leaf Erectness Index Using Leaf Morphological. *Trop. Agr. Develop.*, 63(2):55–60.

Terauchi, T. et al. 2012. Sugarcane Breeding of Early Maturing Clone with High Sucrose Content for Earlier Harvest in Japan. *Japan Agricultural Research Quarterly*, 46(3):227–35.

Umbaugh, S. . E. 2005. *Computer Imaging: Digital Image Analysis and Processing*. 1st ed. New York: TAYLOR & FRANCIS GROUP.

Zhang, Y. 2012. Support Vector Machine Classification Algorithm and Its Application. Pp. 179–86 in *Communications in Computer and Information Science. Information Computing and Applications. ICICA 2012.*, edited by Y. A. Liu C., Wang L. Berlin, Heidelberg: Springer.