On-site identification of *Ophiocordyceps sinensis* using multispectral imaging and chemometrics

Hongwei Duan, Xin Tong, Ruxiu Cui, Lujia Han, Guangqun Huang[*]

(*Laboratory of Biomass and Bioprocessing Engineering, College of Engineering, China Agricultural University, Beijing 100083, China*)

Abstract: For the reasonable and effective collection of *Ophiocordyceps sinensis*, a new method of on-site identification was attempted using a portable multispectral imaging (MSI) technique. Three dimensional (3D) data-cubes of representative *Ophiocordyceps sinensis* and weeds samples were acquired and pre-processed with standard normal variate transformation (SNV). Principal component analysis (PCA) and simulated annealing particle swarm optimisation (SAPSO) algorithms were used to extract characteristic images and develop the support vector classification (SVC) models. Results show that the fused feature model of SAPSO-SVC has the best performance, resulting in a recognition accuracy of the prediction set of 96.30%. Moreover, on-site distribution map of *Ophiocordyceps sinensis* and weeds was created using the spectral feature model of SAPSO-SVC, and the target could be easily identified from the distribution map. This work demonstrates the potential for on-site identification of *Ophiocordyceps sinensis* in the Qinghai–Tibet Plateau using a portable MSI technique combined with the SAPSO-SVC algorithm.

Keywords: *Ophiocordyceps sinensis*, MSI, SAPSO-SVC, On-site distribution map

DOI: 10.25165/j.ijabe.20201306.5425

1 Introduction

As a valuable wildlife resource with the dual-purpose of medicine and nutrition, *Ophiocordyceps sinensis* has strict requirements for its parasitism and growth environment, leading to its short supply in the high-end market. However, because of its small size, dark colour and wide distribution, the traditional manual search method has the disadvantages of high labour intensity and low efficiency in the process of collecting *Ophiocordyceps sinensis*. Therefore, developing a rapid and on-site identification technology is of major importance for the ecologically friendly and efficient excavation of *Ophiocordyceps sinensis* resources and to ensure sustainable use.

In recent, digital imaging technologies have been widely used in the classification of farm products, monitoring of crop diseases and insect pests, and other aspects of agricultural production based on the external physical features (e.g. colour, marbling and texture) without considering spectral fingerprints[1-4]. Similarly, many near-infrared spectroscopic studies have been reported to analyze the internal chemical or biological properties based on a small rounded region of the target sample[5-7]. As a fusion of imaging and spectroscopy[8-12], multispectral imaging (MSI) integrates both techniques in one configuration and can provide both spatial and spectral information for each pixel over the required wavelength range, thereby facilitating the fast and accurate identification of *Ophiocordyceps sinensis*.

Because of the heavy data-cube of the acquired original multispectral images, there may exist non-informative waveband images. Moreover, the unique external texture features of *Ophiocordyceps sinensis* are mainly influenced by its internal functional composition. Herein, principal component analysis (PCA)[13] and simulated annealing particle swarm optimisation (SAPSO)[14] were used to extract the characteristic images from two aspects of image and spectral analysis, and a more robust classification model may be developed to accurately identify *Ophiocordyceps sinensis* using the simplified three dimensional data-cube[15].

In this work, MSI was applied to fast identify and on-site visualize *Ophiocordyceps sinensis*. The specific objectives were to (1) extract and pre-process the spectra and images in regions of interest (ROIs) of *Ophiocordyceps sinensis* and representative weeds; (2) extract the characteristic images using PCA and SAPSO and obtain the texture features of characteristic images using Gray-level Co-occurrence Matrix (GLCM); (3) develop support vector classification (SVC) models and create an on-site classification distribution map.

2 Materials and methods

2.1 Sample collection and preparation

A total of 106 representative samples, including 24 *Ophiocordyceps sinensis* and 82 weeds from 11 categories, were collected from the main area of *Ophiocordyceps sinensis* production at 4200-4600 m above sea level in Nyingchi, Tibet Autonomous Region, China.

2.2 Multispectral image acquisition and processing

The developed MSI configuration consisted of an imaging

spectrometer (ImSpector, V10E-QE, Spectral Imaging Ltd., Finland) in the spectral range of 465-630 nm with a digital camera (CMV2000, Imec, Belgium), and computer installed with spectral data-cube acquisition software.

To maximize the information of *Ophiocordyceps sinensis* that could be obtained under suitable light conditions, the lens orientation was adjusted to approximately 30° downward in the horizontal direction. The distance between the lens and sample was adjusted to approximately 15 cm to ensure an appropriate image size. The system was then operated with a 3 ms exposure time to generate the multispectral image.

2.3 Regions of interest

The *Ophiocordyceps sinensis* stroma mainly consists of the pommel, fertile trunk and an infertile tip. Because of the more evident texture features and larger area weight of the fertile part, it was extracted as the ROIs of *Ophiocordyceps sinensis* with a size of 25×25 pixels. Similarly, the ROIs of different weed samples were determined by their characteristic areas. The averaged spectrum for each ROI was regarded as the representative spectrum of the *Ophiocordyceps sinensis* or weed sample.

2.4 Data process and analysis

Based on the acquired spectra of *Ophiocordyceps sinensis* and representative weed samples, the dataset was initially divided into the calibration set and prediction set using the concentration gradient method. Because of the influence of the uneven thickness of the tested samples, the spectral noise of light scattering needed to be removed using the preprocessing method of standard normal variate transformation (SNV). The characteristic images were then extracted from the denoised dataset using SAPSO and PCA, and its external texture features were analysed based on the GLCM method. SVC models were finally developed using the characteristic information, and it was also used to establish the On-site classification distribution map of *Ophiocordyceps sinensis*[16]. A high recognition accuracy of the prediction set signifies a robust prediction performance.

3 Results and discussion

3.1 Extraction of characteristic images

Prior to the extraction of image features, resampling was implemented to obtain the three dimensional data-cube for the ROIs. PCA was applied to this data-cube for image dimensional reduction, and PC_1 had a variance contribution rate of more than 80%. Additionally, it was observed that the larger weights in the projection coefficient matrix were mainly distributed in the wavelengths of 490.25 nm, 507.29 nm, 569.08 nm and 605.09 nm. These results were expected because two further wavelengths located at approximately 490.25 nm and 507.29 nm were the absorption peaks related to polysaccharide and α-tocopherol[16-22].

Similarly, SAPSO was used to extract four feature variables located at the wavelengths of 490.25 nm, 507.29 nm, 545.02 nm and 626.90 nm, as presented in Figure 1. The feature variables at wavelengths 490.25 nm and 507.29 nm were both obtained by PCA and SAPSO. Moreover, two wavelengths located at approximately 626.90 nm and 545.02 nm were obtained by SAPSO, which were the absorption peaks of polysaccharide and xanthine oxidase, respectively[19-21,23]. The results showed that the characteristic images extracted by PCA and SAPSO both contained the main feature information of *Ophiocordyceps sinensis*.
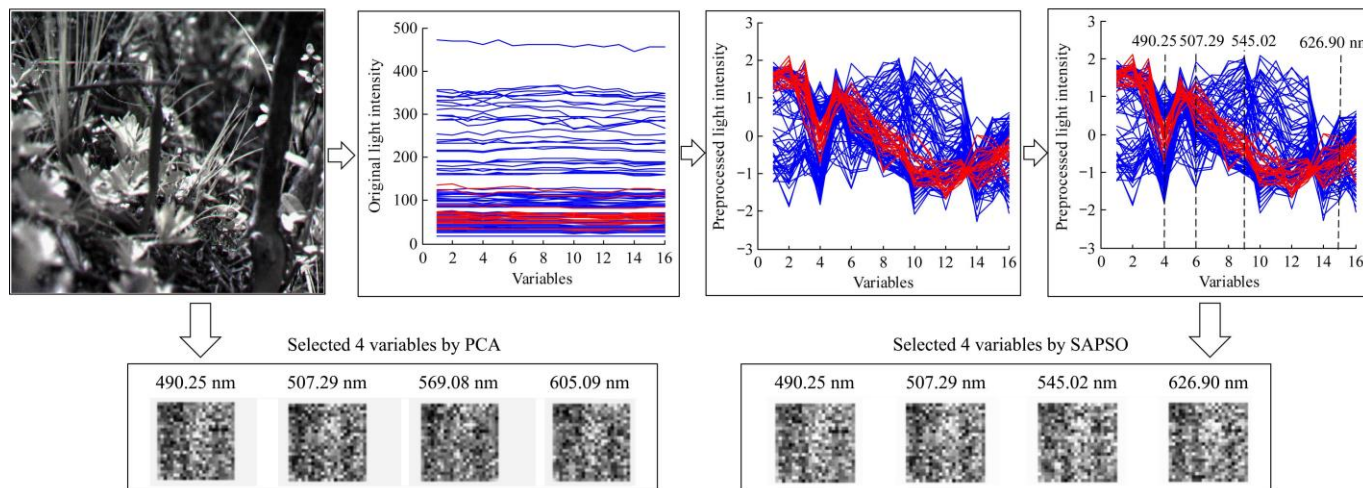

Figure 1   Extraction process of *Ophiocordyceps sinensis* characteristic images

3.2 Texture feature extraction of characteristic images

Based on the acquired characteristic images of *Ophiocordyceps sinensis*, texture feature parameters of angular second moment (ASM), contrast (CON), correlation (COR), inverse difference moment (IDM) and entropy (ENT) were extracted using GLCM from four directions of 0°, 45°, 90° and 135°, as shown in Figure 2.

In comparison with the other three directions, the dispersion and overall levels of ASM, COR and IDM in the 0° direction were high, while the values of CON and ENT were low. This indicates that the texture of the fertile trunk in the horizontal direction was more regular, fine and clearly visible. Furthermore, in all four directions, the CON values at wavelengths 545.02 nm and 626.90 nm acquired using SAPSO were both larger than those at wavelengths 569.08 nm and 605.09 nm obtained using PCA. This demonstrates that the characteristic images extracted using SAPSO had more distinct grooves.

3.3 Model prediction

The SVC models of *Ophiocordyceps sinensis* and weeds were developed based on the simplified characteristic images. The model results are shown in Table 1.

The characteristic images obtained by PCA were initially used to develop the PCA-SVC models, resulting in recognition accuracies of 88.89%, 85.19% and 92.60%, while the similar characteristic images obtained by SAPSO were then used to develop the SAPSO-SVC models, resulting in recognition accuracies of 92.59%, 92.59% and 96.30%. Results showed that

the SAPSO-SVC models had better performances than those of PCA-SVC models.   This was expected since the SAPSO algorithm could extract more texture features and dominant spectra related to polysaccharide, α-tocopherol and xanthine oxidase of *Ophiocordyceps sinensis*[24,25].   Furthermore, the fused feature model had the best performance in the three SAPSO-SVC models, and it can be applied for the on-site identification of *Ophiocordyceps sinensis*.

a. GLCM-ASM

b. GLCM-CON

c. GLCM-COR

d. GLCM-IDM

e. GLCM-ENT
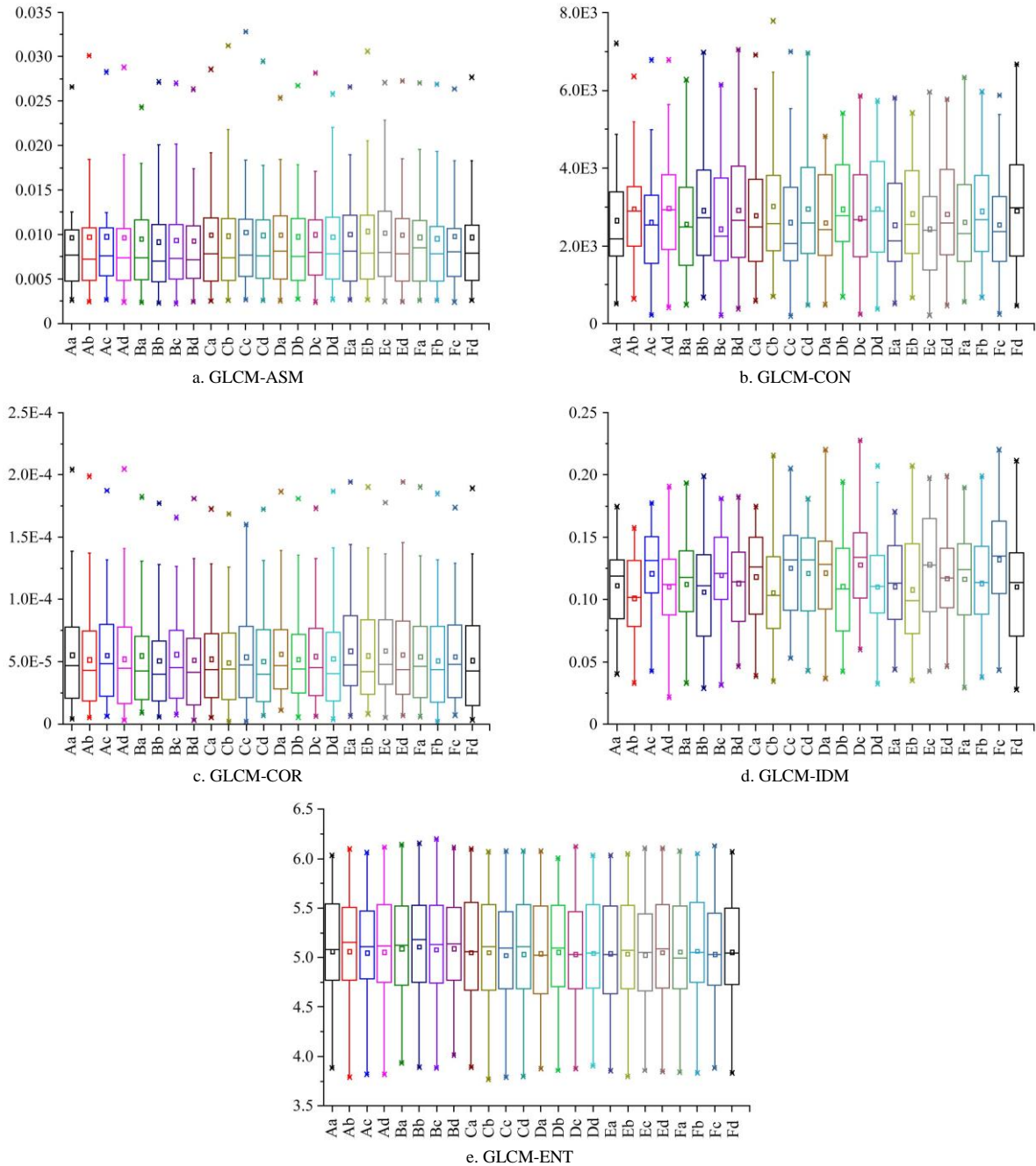
Note: A, B, C and D are the four feature spectra of 490.25 nm, 507.29 nm, 569.08 nm and 605.09 nm, respectively, of which, C, D are acquired using PCA; A, B are acquired using PCA and SAPSO; E and F are the feature spectra of 545.02 nm and 626.90 nm, respectively, acquired using SAPSO; and a, b, c and d are the four directions of 0°, 45°, 90° and 135°, respectively.

Figure 2   Texture feature statistics of simplified characteristic images

Table 1   Comparison of SVC model results

Data	Models	Parameters		Accuracy		
		c	g	Calibration set	Cross-validation set	Prediction set
Spectra	PCA-SVC	63.6753	93.2611	93.67%	77.22%	88.89%
	SAPSO-SVC	67.7858	100	98.73%	84.81%	92.59%
Texture	PCA-SVC	85.4048	65.0283	86.08%	77.22%	85.19%
	SAPSO-SVC	12.2453	0.49125	100%	86.08%	92.59%
Fusion	PCA-SVC	7.5589	0.0647	98.73%	78.48%	92.59%
	SAPSO-SVC	22.943	0.29409	100%	92.41%	96.30%

Note: c signifies penalty parameter; g signifies kernel parameter.   79 samples were used as calibration set, and 27 samples were used as prediction set.

3.4   On-site visualization of Ophiocordyceps sinensis and weeds

Because of the incomputability of texture features on each single pixel, the on-site classification distribution map of *Ophiocordyceps sinensis* was therefore created based on the spectral feature model of SAPSO-SVC. As shown in Figure 3, pixels marked by blue and yellow represent the portions of *Ophiocordyceps sinensis* and weeds, respectively. The densely distributed blue pixels could automatically generate the rod-shaped images of *Ophiocordyceps sinensis* stroma, and its precise position could be intuitively identified from the complex background. This indicates that the distribution map could be used for on-site identification of *Ophiocordyceps sinensis*. However, there remained a large amount of dispersed blue 'noise points' on the map. This may be because some representative weed samples in the main producing areas of *Ophiocordyceps sinensis* were neglected because of the complexity of the growth environment.
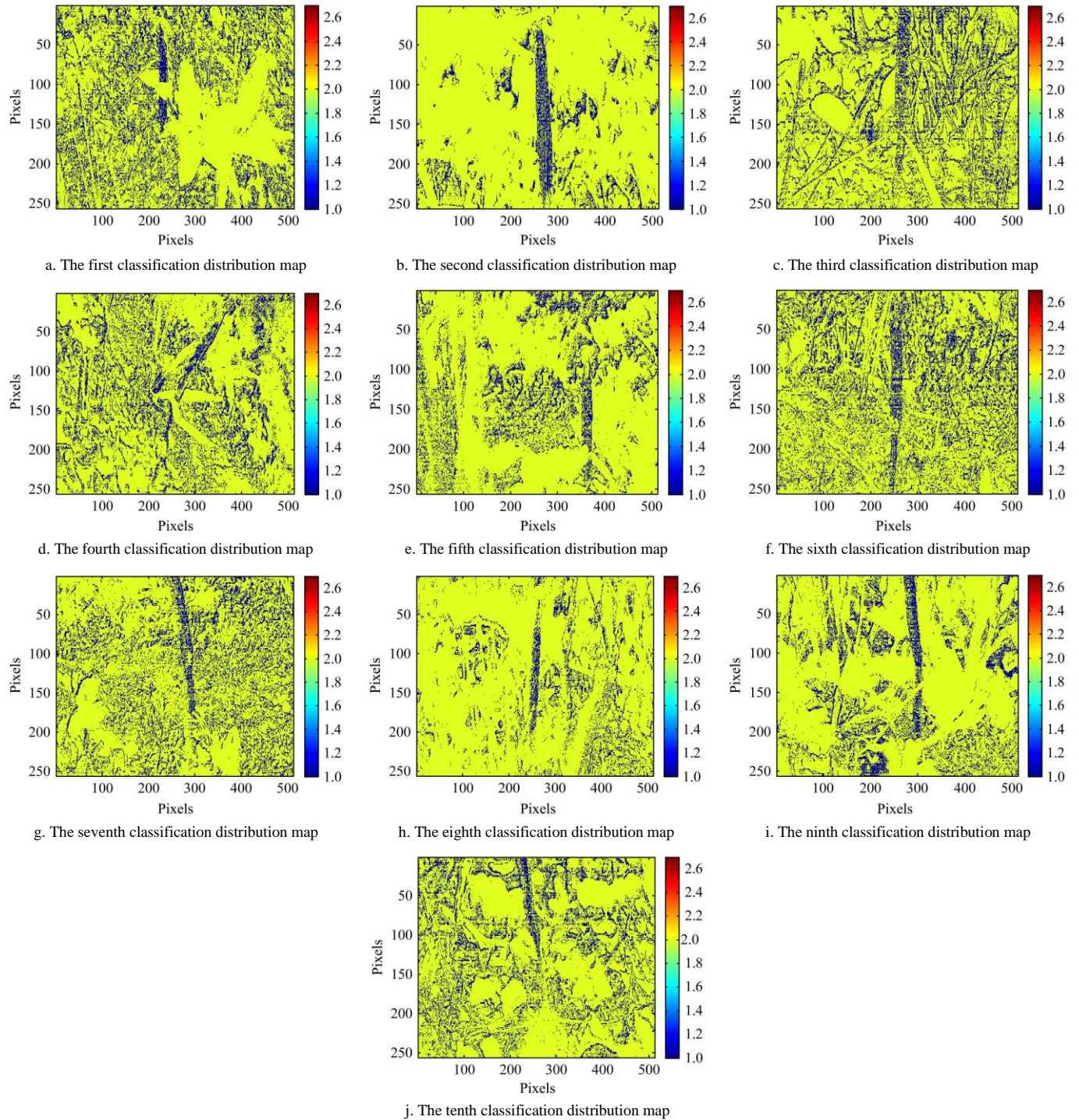
a. The first classification distribution map

b. The second classification distribution map

c. The third classification distribution map

d. The fourth classification distribution map

e. The fifth classification distribution map

f. The sixth classification distribution map

g. The seventh classification distribution map

h. The eighth classification distribution map

i. The ninth classification distribution map

j. The tenth classification distribution map

Figure 3   On-site classification distribution map of ten *Ophiocordyceps sinensis*

4   Conclusions

To avoid the disadvantages of high labour intensity and low efficiency in the process of collecting *Ophiocordyceps sinensis*, MSI was used to fast identify *Ophiocordyceps sinensis* from complex background. A comparison between the models developed by characteristic images indicated that the fused feature model of SAPSO-SVC had the best performance, resulting in a recognition accuracy of 96.30%. Furthermore, the on-site classification distribution map of *Ophiocordyceps sinensis* was

created based on the spectral feature model of SAPSO-SVC. Results showed that the MSI technique combined with the SAPSO-SVC algorithm could quickly find out *Ophiocordyceps sinensis*.

Acknowledgements

[References]

[1] Brosnan T, Sun D W. Inspection and grading of agricultural and food products by computer vision systems—A review. Computers and Electronics in Agriculture, 2002; 36: 193–213.

[2] Ma J, Sun D W, Qu J H, Liu D, Pu H, Gao W H, et al. Applications of computer vision for assessing quality of agri-food products: a Review of recent research advances. Crit Rev Food Sci Nutr, 2016; 56(1): 113–127.

[3] Rong D, Rao X, Ying Y. Computer vision detection of surface defect on oranges by means of a sliding comparison window local segmentation algorithm. Computers and Electronics in Agriculture, 2017; 137: 59–68.

[4] Tellaeche A, Pajares G, Burgos-Artizzu X P, Ribeiro A. A computer vision approach for weeds identification through Support Vector Machines. Applied Soft Computing, 2011; 11(1): 908–915.

[5] Khulal U, Zhao J, Hu W, Chen Q. Nondestructive quantifying total volatile basic nitrogen (TVB-N) content in chicken using hyperspectral imaging (HSI) technique combined with different data dimension reduction algorithms. Food Chem, 2016; 197: 1191–1199.

[6] Liu L, Cozzolino D, Cynkar W U, Gishen M, Colby C B. Geographic classification of Spanish and Australian tempranillo red wines by visible and near-infrared spectroscopy combined with multivariate analysis. J. Agric. Food Chem., 2006; 54: 6754–6759.

[7] Wang D, Dowell F E, Ram M S, Schapaugh W T. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. International Journal of Food Properties, 2004; 7(1): 75–82.

[8] Dissing B S, Papadopoulou O S, Tassou C, Ersbøll B K, Carstensen J M, Panagou E Z, et al. Using multispectral imaging for spoilage detection of pork meat. Food and Bioprocess Technology, 2012; 6(9): 2268–2279.

[9] Pu H, Kamruzzaman M, Sun D W. Selection of feature wavelengths for developing multispectral imaging systems for quality, safety and authenticity of muscle foods-a review. Trends in Food Science & Technology, 2015; 45(1): 86–104.

[10] Ropodi A I, Pavlidis D E, Mohareb F, Panagou E Z, Nychas G J E. Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. Food Research International, 2015; 67: 12–18.

[11] Sendin K, Manley M, Williams P J. Classification of white maize defects with multispectral imaging. Food Chem, 2018; 243: 311–318.

[12] Tang C, He H, Li E, Li H. Multispectral imaging for predicting sugar content of 'Fuji' apples. Optics & Laser Technology, 2018; 106: 280–285.

[13] Li H, Chen Q, Zhao J, Wu M. Nondestructive detection of total volatile basic nitrogen (TVB-N) content in pork meat by integrating hyperspectral imaging and colorimetric sensor combined with a nonlinear data fusion. LWT - Food Science and Technology, 2015; 63(1): 268–274.

[14] Qi C M, Zhou Z B, Sun Y C, Song H B, Hu L S, Wang Q. Feature selection and multiple kernel boosting framework based on PSO with mutation mechanism for hyperspectral classification. Neurocomputing, 2017; 220: 181–190.

[15] Yang Y C, Sun D W, Pu H, Wang N N, Zhu Z. Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. Postharvest Biology and Technology, 2015; 103: 55–65.

[16] Duan H W, Zhu R G, Yao X D, Lewis E. Sensitive variables extraction, non-destructive detection and visualization of total viable count (TVC) and pH in vacuum packaged lamb using hyperspectral imaging. Analytical Methods, 2017; 9(21): 3172–3183.

[17] Eberhart R, Kennedy J. A new optimizer using particle swarm theory. Sixth International Symposium on Micro Machine and Human Science. IEEE, 1995; pp.39–43. DOI: 10.1109/MHS.1995.494215

[18] Haralick R M, Shanmugam K. Textural features for image classification. IEEE Trans. Syst. Man Cybern, 1973; 6: 610–621.

[19] Kuo Y C, Tsai W J, Shiao M S, Chen C F, Lin C Y. Cordyceps sinensis as an immunomodulatory agent. Am. J. Chinese Med, 1996; 24: 111–125.

[20] Li S P, Li P, Dong T T, Tsim K W. Anti-oxidation activity of different types of natural Cordyceps sinensis and cultured Cordyceps mycelia. Phytomedicine, 2001; 8(3): 207–212.

[21] Li S P, Zhao K J, Ji Z N, Song Z H, Dong T T X, Lo C K, et al. A polysaccharide isolated from Cordyceps sinensis, a traditional Chinese medicine, protects PC12 cells against hydrogen peroxide-induced injury. Life Sciences, 2003; 73(19): 2503–2513.

[22] Wang Z M, Peng X, Lee K L D, Tang J C-O, Cheung P C-K, Wu J Y. Structural characterisation and immunomodulatory property of an acidic polysaccharide from mycelial culture of *Cordyceps sinensis* fungus Cs-HK1. Food Chemistry, 2011; 125(2): 637–643.

[23] Dong C H, Yao Y J. In vitro evaluation of antioxidant activities of aqueous extracts from natural and cultured mycelia of Cordyceps sinensis. LWT - Food Science and Technology, 2008; 41(4): 669–677.

[24] Lin H. Development and application of feature selection techniques in protein data Analysis and Prediction. Lett. Org. Chem, 2017; 14: 619–620.

[25] Saini S, Zakaria N, Rambli D R, Sulaiman S. Markerless human motion tracking using hierarchical multi-swarm cooperative particle swarm optimization. PLoS One, 2015; 10(5): e0127833.