Comparative Study of Different Pre-processing Methods for Discriminating Live Fish Based on Hyperspectral Imagery

Mohammadmehdi Saberioon, Petr Císař, Pavel Souček
University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, Nové Hrady, Czech Republic
e-mail: {msaberioon; cisar; psoucek}@frov.jcu.cz

Laurent Labbé, Pablo Pelissier
INRA, UE 0937 PEIMA (Pisciculture Expérimentale INRA des Monts d'Arrée), Sizun, France
e-mail: {laurent.labbe; pablo.pelissier}@inra.fr

Abstract—**The main aim of this study was to compare the performance of different pre-processing algorithms when coupled with Support vector machine as the classifier to differentiate live fish based on their diet received during cultivation using hyperspectral imagery system. Rainbow trout (*Oncorhynchus mykiss*) were fed either a fish meal-based diet or a 100 % plant-based diet. Hyperspectral images were made using a push-broom hyperspectral imaging system in the spectral region of 394-1009 nm. Six spectral pre-treatment methods were used, including Savitzky-Golay, First Derivative, Second Derivative, Standard Normal Variate and Multiplicative Scatter Correction were employed to improve the robustness and performance of the classifier. According to the criteria of correct classification rate and Kappa coefficient, the support vector machine with linear kernel when coupled with Savitzky-Golay pre-processing was determined as the best method for classifying live fish due to their diet.**

Keywords-Hyperspectral; Pre-processing algorithms; classification; Support vector machine;Fish diet.

I. INTRODUCTION

Researchers have used different optical sensors for measuring and determining light interaction with fish skin. Several studies used colorimeters to determine skin color, which usually provides readings in XYZ, RGB and CIE Lab color space [1]. These instruments allow accurate and reproducible measurements of the color with no influence by the observed or surroundings [2]. For instance, Yi et al. [3] used portable Minolta Chroma meter to show the effect of astaxanthin and xanthophylls as carotenoid sources on growth and skin color of large yellow croaker. Some also used visible and Near Infrared (Vis/NIR) spectroscopy to document the color changes on the fish skin. Lin et al. [4] showed a satisfactory application of Vis/NIR spectroscopy to detect bruises in Pacific pink salmon (*Oncorhynchus gorbuscha*) through the skin. Costa et al. [5] used Vis/NIR spectroscopy of skin to differentiate the sea bass (*Dicentrarchus labrax*) with 87% accuracy at 48hr post-mortem quality cultured in the tank from sea cage. Although fish skin color described with proximal sensors is accurate, their use has been criticized due to the small area measured by the machine, and that aspect of the overall colors are lost [6]. Also, for accurate measurement, the surface color should be quite uniform or homogenous and that may many points on the sample must be measured to obtain the representative color profile, which sometimes is destructive [7].

Researchers also used consumer grade cameras as a non-invasive tool for measuring skin color parameters. Digital images from consumer-grade cameras can overcome the deficiencies of visual and instrumental techniques and offer an objective measurement of color and other physical factors [8]. Wallat et al. [9] demonstrated how a compact true color camera could be employed for objective measurement of the skin color of live goldfish (*Carassius auratus*). Zatkova et al. [10] utilized a digital camera to estimate changes in skin color of wels catfish (*Silurus glanis*). They showed the feasibility of digital cameras for monitoring skin color changes due to diet alteration. Colihueque et al. [11] developed a method for estimating skin color, spottiness and darkness using consumer digital camera and digital image analysis for categorizing cultured rainbow trout (*Oncorhynchus mykiss*). Costa et al. [12] used a digital camera to analyze skin color to discriminate the effects of seabass fed organic or commercial diet. Segade et al. [13] also showed the effect of diet on seahorse (*Hippocampus hippocampus*) body color using images obtained from consumer-grade digital cameras. Consumer-grade cameras provide the capability to rapidly scan both larger areas, as well as smaller details but can only study color in Visible (Vis) bands. Furthermore, all above-mentioned studies have not described the interaction of light in Near Infrared (NIR) bands to show the chromatic changes on the fish skin.

Hyperspectral Imagery (HSI) is an emerging technology that integrates both spectroscopy and imaging in a single system; it has potential to capture the subtle spectral difference under different physiological and pathological conditions. HSI is enabling simultaneous acquisition of spatial and spectral information from an object. The system has the ability to image the same scene in hundreds of contiguous narrow wavebands, from the visible to the short-wave infrared region of electromagnetic spectrum (400-2500 nm). In other words, HSI has higher spatial resolution than the multispectral image, which is obtained by consumer grade digital cameras.

The main challenge limiting the application of hyperspectral for fish discrimination is finding suitable data pre-processing and classification strategy. Choosing the most robust technique can help to achieve a more reliable

 ISBN: 978-1-61208-638-5

classification model. In order to remove inappropriate information, which cannot be handled correctly by the classifier, pre-processing approaches are used. Usually, these approaches aim to decrease the noise and enhance possible spectral features. The most widely used spectral pre-processing methods can be commonly divided into three groups, namely smoothing, baseline removal, and scaling [14]. The first category is smoothing such as Savitzky-Golay (SG), which is used for noise reduction; the second category is baseline removal, such as the First Derivative (FD) and Second Derivative (SD), which is used for correcting background signals or baseline that is far away from zero level. Multiplicative Scatter Correction (MSC) is also another popular transformation method used to remove the scatter effects on spectral data. Another group is range scaling, this method is applicable when the total intensity in the spectra is sample-dependent, and samples need to be scaled in such a way that intensities can be compared. Standard Normal Variate (SNV) is one of the popular pre-processing methods, which centering and scaling each individual spectrum for correcting the multiplicative interferences of light scatter [15]. To the best of our knowledge, comparison of different pre-processing methods has not yet been commonly used as a forensic method to determine live fish diets based on their skin. The main aim of this study was to compare the performance of different pre-processing algorithms when coupled with Support vector machine as a classifier to differentiate live fish based on their diet received during cultivation using hyperspectral imagery system. In section 2, Materials and methods will be described. Afterward, results and discussion will be provided in section 3. Finally, the conclusion will be presented in section 4.

II. MATERIALS AND METHODS

A. Fish and Cultural Condition

The fish species were produced at INRA-PEIMA (Sizun, France). Experiments were designed in a split-block design with three replications for each diet; therefore, 80 fish were fed a commercial based diet (3 tanks) and 80 were fed a plant-based diet (3 tanks). After three weeks, all fish were used for hyperspectral image acquisition.

B. Diets and Feeding Controls

Diets were manufactured at the INRA NUMEA facility of Donzacq (France). The Fish meal-based diet contained fishmeal and fish oil as protein and lipid source respectively. Plant based diet (PBD) is contained a mixture of wheat gluten, extruded peas, corn gluten meal, soybean meal and white lupin as protein sources; and the combination of palm seed, rapeseed and linseed oil, rich in saturated, mono-unsaturated and n-3 poly-unsaturated fatty acids, respectively, as the lipid source. A mineral and a vitamin premix equally were added to both diets.

C. Image Acquisition

A push-broom, line-scanning reflectance hyperspectral imaging system was used to acquire the hyperspectral images of rainbow trout in a dark room to avoid the interference due to stray light and to get pure spectral reflectance. This system includes a high-performance CCD camera (Photonfocus 1312 CL) along with focus lens, a spectrograph (Specim V10E, Spectral Imaging Ltd., Oulu, Finland) attached to the camera acquire hyperspectral images in the wavelength range of 393-1009 nm, an illumination source (150 W halogen lamp attached to a fiber optic line light positioned at an angle of 45 degree to the moving table to reduce the shadowing effects), a moving table and a computer system equipped with an image acquisition software (SpectralScanner, DV optics).

To acquire the spectral and spatial information, each rainbow trout was placed on the sample loading device and they conveyed to the camera field of view (FoV) of the camera with adjusted speed (1.6 mm/s) and exposure time (10 s) to be scanned line by line. The procedure was controlled and implemented by image acquisition software (Specim Lumo software, Spectral Imaging Ltd., Oulu, Finland). The raw hyperspectral image for each sample consisted of several congruent images representing intensities at 784 wavelength bands from 393 to 1009 nm. Due to the low signal-to-noise rate at the two ends of the spectral ranges, only the wavelength ranging from 400 to1000 nm wavelength was used. Totally, 160 hyperspectral images were created, recorded and stored in raw format before being processed. Before measurement, each fish mildly anesthetized with Benzocaine to reduce the movement and minimize the stress. The surface of each rainbow trout also was wiped with piece of tissue paper to remove extra water from skin before data acquisition.

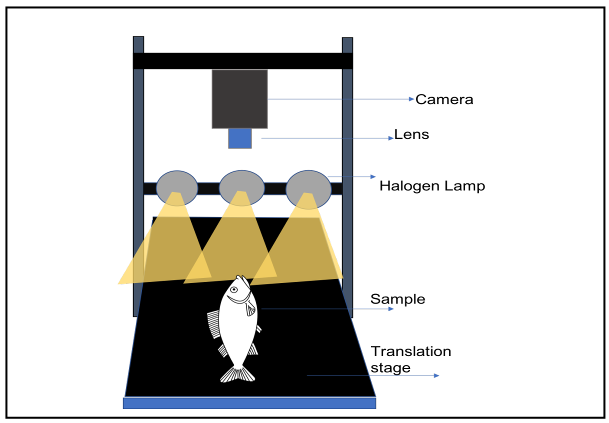

Figure 1. Schematic demonstration of the main components of the used hyperspectral imaging system

D. Hyperspectral Image Calibration

The acquired hyperspectral images were corrected using (1).

$$I_i = (R_i - D_i) / (W_i - D_i) \tag{1}$$

where *I* is the corrected hyperspectral image, R is the raw hyperspectral image, W is the white reference image and D is the dark reference image, as well as *i* is the pixel index. Afterward, the reflectance spectrum from the region of interest (ROI) was computed by averaging the spectral value of all pixels in the ROI for each sample using the Environment for Visualizing Images software (ENVI 5.3, Harris geospatial solutions, FL, USA).

E. Spectral Pre-processing

Six forms of spectra pre-processing were used in this study to remove the non-constituent-related effects in spectra data and to develop optimal classifier model. The six forms were SG smoothing with second order polynomial fit and 11 smoothing points, the FD and SD transformation, MSC and SNV.

F. Classifier

After pre-treatment, Support vector machine (SVM) with the linear kernel as a classifier was employed to develop the classification models for discriminating two different diets. The dataset from 160 hyperspectral for rainbow trout was divided into training set (80% of total samples) used to develop the classifier models and a validation set (20% of total samples) used to assess the prediction accuracy of each model. Further details can be found in Vapnik [16]. R package Caret [17] used for SVM classification model.

G. Evaluation of Classification

The classifier was evaluated through the analysis of correct classification rate (CCR, %) and Cohen's Kappa coefficient in the validation set. CCR can be calculated using (2).

$$CCR= (N_1 / N_0) \times 100 \quad (2)$$

where N1 is number of correct estimation of samples and N0 is the total number of samples. Kappa coefficient (K) also calculated using (3).

$$K= (Pr(a) – Pr (e)) / (1 – Pr(e)) \quad (3)$$

where Pr(a) is observed agreement and Pr(e) is probability of random agreement.

III. RESULTS AND DISCUSSION

The raw spectra of the two diets were shown in Figure 1. Also, different pre-processing algorithms for all samples were displayed in Figure 2.

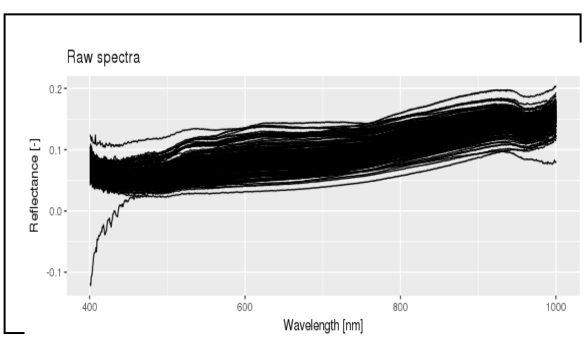

Figure 2. representative Vis/NIR Spectra of fish skin raw.

Table 1 shows the average CCRs for testing set with six spectral pre-processing techniques based on the whole spectral range. When the raw and pre-processed spectra were used to build the classification models, CCRs ranged between 0.740 and 0.871. After application of SD, classification accuracy decreased compare to classification using spectra without pre-processing, however, application of SG, FD, SNV and MSC improved the classification compared to raw spectra. Thus, spectral pre-processing of SD was not helpful and to some extend reduced the accuracy of classification, but other pre-processing treatments improved the performance of classifier for the full range of wavelengths.

TABLE I. MODEL PERFORMANCE FOR IDENTIFICATION OF DIFFERENT DIET ON VALIDATION SET

Method	CCR	Kappa
Raw- SVM	0.741	0.485
SG-SVM	0.871	0.741
FD-SVM	0.806	0.612
SD-SVM	0.730	0.483
SNV-SVM	0.774	0.548
MSC-SVM	0.838	0.676

The best performance for classification with the highest CCR of 0.871 and Kappa coefficient of 0.740 achieved when smoothing solely used as a pre-processing method for full wavelength.

IV. CONCLUSION

The present study provided an alternative tool for the classification of live fish based on their information acquired from skin using hyperspectral images. Overall, classification models established showed good performance when SG used as pre-treatment of spectra. Further studies should be carried out to not only improve the classification accuracy using different machine learning algorithms but also increasing classification power for online evaluation of fish skin at industrial scale.

ACKNOWLEDGMENT

This work was funded by projects CENAKVA [CZ.1.05/2.1.00/01.0024] and CENAKVA II (the results of the project LO1205 were obtained with a financial support from the MEYS of the CR under the NPU I program); The CENAKVA Centre Development [No. CZ.1.05/2.1.00/19.0380]; and the European Union's Horizon 2020 research and innovation program under grant agreement No. 652831" (Aquauexcel2020).

REFERENCES

[1] M. Saberioon, A. Gholizadeh, P. Cisar, A. Pautsina, and J. Urban, "Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues," Reviews in Aquaculture, vol. 9, no 4, pp. 369-387, 2016.

[2] F. Clydesdale and E. Ahmed, “Colorimetry — methodology and applications∗ ,” C R C Critical Reviews in Food Science and Nutrition, vol. 10, no. 3, pp. 243–301, 1978.

[3] X. Yi et al., “Effects of dietary astaxanthin and xanthophylls on the growth and skin pigmentation of large yellow croaker Larimichthys croceus,” Aquaculture, vol. 433, pp. 377, 2014.

[4] M. Lin et al., “Bruise Detection in Pacific Pink Salmon (Oncorhynchus gorbuscha) by Visible and Short-Wavelength Near-Infrared (SW-NIR) Spectroscopy (600−1100 nm),” Journal of Agricultural and Food Chemistry, vol. 51, no. 22, pp. 6404–6408, 2003.

[5] C. Costa, S. D’Andrea, R. Russo, F. Antonucci, F. Pallottino, and P. Menesatti, “Application of non-invasive techniques to differentiate sea bass (Dicentrarchus labrax, L. 1758) quality cultured under different conditions,” Aquaculture International, vol. 19, no. 4, pp. 765-778, 2011.

[6] F. Mendoza and J. Aguilera, “Application of image analysis for classification of ripening bananas,” Journal of Food Science, vol. 69, no. 9, pp. E471–E477, 2004.

[7] K. Yam and S. Papadakis, “A simple digital imaging method for measuring and analyzing color of food surfaces,” Journal of Food Engineering, vol. 61, no. 1, pp. 137–142, 2004.

[8] Y.-R. Chen, K. Chao, and M. Kim, “Machine vision technology for agricultural applications,” Computers and Electronics in Agriculture, vol. 36, no. 2–3, pp. 173-191, 2002.

[9] G. Wallat, A. Lazur, and F. Chapman, “Carotenoids of Different Types and Concentrations in Commercial Formulated Fish Diets Affect Color and Its Development in the Skin of the Red Oranda Variety of Goldfish,” North American Journal of Aquaculture, vol. 67, no. 1, pp. 42-51, 2017.

[10] I. Zaťková, M. Sergejevová, J. Urban, R. Vachta, D. Štys, and J. Masojídek, “Carotenoid-enriched microalgal biomass as feed supplement for freshwater ornamentals: albinic form of wels catfish (Silurus glanis),” Aquaculture Nutrition, vol. 17, no. 3, pp. 278-286, 2009.

[11] N. Colihueque, M. Parraguez, F. Estay, and N. Diaz, “Skin Color Characterization in Rainbow Trout by Use of Computer-Based Image Analysis,” North American Journal of Aquaculture, vol. 73, no. 3, pp. 249-258, 2011.

[12] C. Costa, P. Menesatti, E. Rambaldi, L. Argenti, and M. Bianchini, “Preliminary evidences of color differences in European sea bass reared under organic protocols,” Aquacultural Engineering, vol. 57, pp. 82-88, 2013.

[13] Á. Segade, L. Robaina, O. Ferrer, G. Romero, and M. Domínguez, “Effects of the diet on seahorse (Hippocampus hippocampus) growth, body color and biochemical composition,” Aquaculture Nutrition, vol. 21, no. 6, pp. 1-7, 2014.

[14] A. Gholizadeh, L. Borůvka, M. Saberioon, J. Kozák, R. Vašát, and K. Němeček, “Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features,” Soil and Water Research, vol. 10, no. 4, pp. 218-227, 2015.

[15] J. Duckworth, C. Roberts, J. Workman Jr. , and J. Reeves III, “Mathematical Data Preprocessing,” vol. agronomymonogra, in Near-infrared Spectroscopy in Agriculture, ASA, CSSA, SSSA, 2004, pp. 115–132.

[16] V. Vapnik, “Statistical learning theory,” John Wiley & Sons: Hoboken, NJ, USA, 1998.

[17] M. Kuhn, “Building Predictive Models in R Using the caret Package,” Journal of Statistical Software, vol. 28, no. 5, pp. 1-62, 2008.

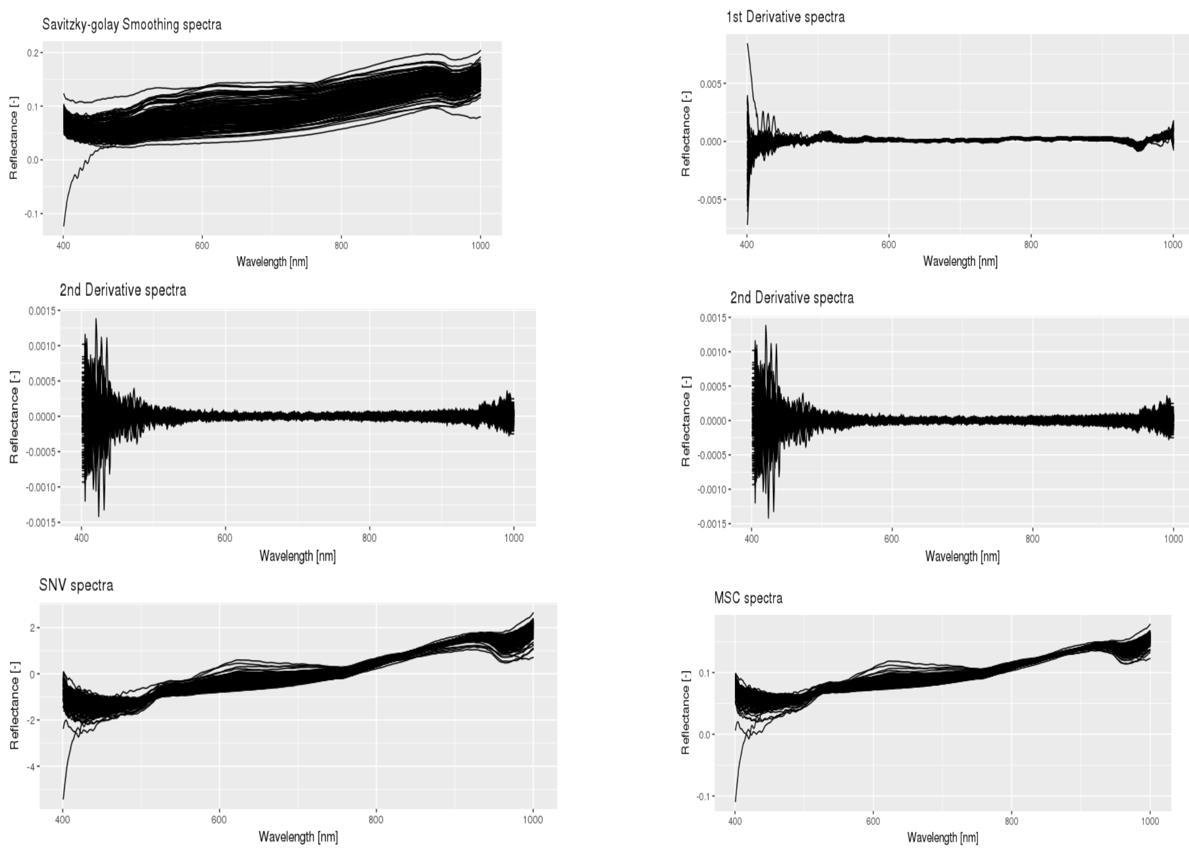

Figure 3. representative Vis/NIR Spectra of fish skin after different pre-processing methods; SG: Savitzky-golay, FD: 1st Derivative, SD: 2nd Derivative, SNV: Standard normal variate, and MSC: multiplicative scatter correction.