IDENTIFICATION OF COTTON FOREIGN MATTER USING LINE-SCAN BASED HYPERSPETRAL IMAGING SYSTEM

Yu Jiang
Changying Li
College of Engineering, University of Georgia
Athens, GA

Abstract

Identification of cotton foreign matter is important to both the cotton and textile industry, because different foreign matter not only affects cotton's monetary value but also causes various damages to textile products. Hyperspectral imaging technique has shown the capability of classifying foreign matter, but it provided a large amount of redundant information, which limits the classification accuracy and processing speed. The goal of this study was to select the optimal wavelengths to be used for foreign matter classification. A total of 240 mean spectra were extracted from the hyperspectral images of cotton lint and 7 types foreign matter. Each type contained 30 replicates, and they were randomly separated into training and test set with the ratio of 15/15. The sequential forward selection method was applied on the dataset to select the wavelengths, which were used to reduce the dimensionality of original dataset as well as the optimal wavelengths for classification. Subsequently, the linear discriminant analysis with the wavelengths was performed to classify foreign matter categories with 2-fold cross validation. The experimental results showed that 22 out of 256 wavelengths were selected and they achieved classification accuracy of 93.33% or higher for all types. Therefore, the selected wavelengths could be used for online foreign matter classification systems.

Introduction

Cotton foreign matter (FM) is an important factor for the quality assessment, which determines the monetary value of cotton. There are two aspects for foreign matter measurement: the overall quantity and FM category. Currently, the overall quantity can be measured by instrument such as the High Volume Instrument (HVI), whereas the FM category has to be determined by classer because there is no instrument for classifying FM types (United States Department of Agriculture, 2005). Since the classer evaluation is time consuming and subjective, many research have been conducted to figure out new approaches for FM classification.

Since computer vision technologies have been widely used for quality assessment of agricultural products (Brosnan and Sun, 2002), traditional machine-vision based methods have been explored for FM classification, including color camera under white light, monochromatic camera under UV illumination (Li et al., 2010; Zhang et al., 2011), and monochromatic with multiple lens under LED arrays (Jia and Ding, 2005). Although these systems can be used for online FM classification, the FM with similar appearance was misclassified because the color information was insufficient for classifying them. Besides, the shape features are not reliable due to the FM change overtime (Xu et al., 1999). To improve the aforementioned issues, spectroscopic methods have been conducted, including Near-Infrared (NIR) (Fortier et al., 2011), Mid-Infrared (MIR) (Loudermilk et al., 2008), and Fourier-Transform Infrared (FT-IR) (Himmelsbach et al., 2006). These systems provided enough spectral information for FM classification, but they were impractical for industrial quality inspection due to the lacking of the FM spatial information and tedious sample preparation.

Hyperspectral imaging, which provides both spatial and spectral information, was recently used for the cotton FM classification (Jiang and Li, 2014). The preliminary results showed that 14 out of 15 types of FM usually found in U.S. cotton were statistically different from each other, indicating the capability of using the hyperspectral imaging to classify FM. However, the data analysis methods of the system have not been explored yet. One key aspect for the classification system based on hyperspectral imaging is the spectral dimension reduction. (Qin et al, 2013.).

The spectral dimension can be reduced by either feature extraction or feature selection. Feature extraction is to transform datasets from the original space to a lower dimensional space, and thus resulting in the loss of original meaning of features, whereas feature selection is to select a subset from the original feature space, which simultaneously reduces the data dimension and retains feature meaning (Isabelle and André, 2003). Thus, feature selection is more suitable for the spectral dimension reduction in hyperspectral imaging analysis.

The overall goal of this study was to select the optimal wavelengths to be used for FM classification. To this end, the specific objectives were to (i) apply the sequential forward selection algorithm to choose the wavelengths for FM classification, (ii) evaluate the performance of the selected wavelengths by the classification performance.

Materials and Methods

Mean Spectra of Lint and Foreign Matter Samples

A spectral dataset was collected from hyperspectral reflectance images (Jiang and Li, 2014) of cotton lint and 7 types FM (Figure 1). The FM included bark inner and outer, brown leaf, bract, twine, green leaf, plastic bag, and plastic bale packaging. There were 30 replicates for each type of samples, and each replicate was represented by its mean spectra extracted using a region-of-interest (ROI) based approach. Thus, the spectral dataset consisted of 240 mean spectra of the samples.

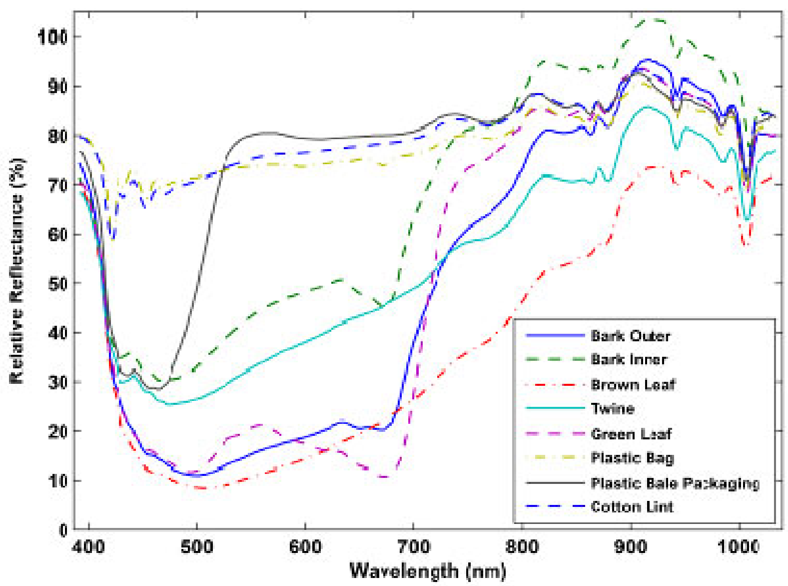

Figure 1. Mean spectra of lint and 7 types FM samples

Pre-processing on the Spectral Dataset

Prior to processing the dataset, the number of wavelengths was reduced from 256 to 223 based on the criterion of signal-to-noise ratio (SNR). The wavelength with SNR less than 10 dB was considered as noise and removed from the dataset. Subsequently, the dataset were equally divided into training and test sets by a partition method provided in MATLAB (MATLAB R2014a, The MathWorks, Inc., Natick, MA). Since the partition method randomly separated the dataset using class information, both training and test sets had 15 samples for each class.

Sequential Forward Selection

The sequential feature selection (SFS) is a widely used approach to search the most informative features from a feature set by sequentially adding (forward) or removing (backward) individual feature until certain conditions occur. A forward sequential selection was performed on the dataset to reduce the data dimensionality and extract the key wavelengths. The method started from an empty set and iteratively created candidate feature subsets by adding each of the features not yet selected into the best subset in the last iteration. Each candidate subset was evaluated using the mean misclassification rate (MCE) of a 10-fold cross validation calculated by the linear discriminant analysis (LDA), and the candidate provided the minimum mean MCE was considered as the best in this iteration. This process continued until all 223 bands were selected, and generated 223 subsets with their performance (MCE value). According to the computational efficiency and performance, the subset with both fewer features and lower MCE was considered as the final SFS result. In order to remove the trivial performance fluctuation, the MCEs less than 0.6 were treated as providing the same performance of classifying cotton FM.

Classification of Foreign Matter

The LDA is to find a linear combination of features which characterizes or separates classes of observations. The LDA used in the present study was the Fisher's linear discriminant which provided by MATLAB (MATLAB R2014a, the MathWorks, Inc., Natick, MA). The method was to use the mean spectra of each type of samples to train a linear model which maximized the variance between classes and minimized the variance within individual classes. The training and test sets were used to conduct the 2-fold cross validation, which trained on training set and test on test set, followed by training on test set and testing on training set.

Results and Discussion

Criterion of Wavelength Selection

The best feature subsets were determined for the SFS, GA, and LOB, respectively. The MCE of the feature subsets selected by the SFS firstly achieved less than 0.06 when adding 22 wavelengths, and the curve staid at a relatively flat at the feature size range from 22 to 173. After adding more than 173 features, the MCE started to increase again, which meant over-fitting occurred (Figure 2) The number of the selected wavelengths significantly affects the computational efficiency of the model training and prediction process, it is preferable to select fewer wavelengths when the MCEs of the feature subsets are same or similar. Two MCEs were considered as providing the same performance when their difference was less than 0.01. Consequently, although the subset with 88 bands provided the best performance (MCE was 0.05), the subset with 22 bands was selected as the final result due to its similar performance (MCE was 0.06) and fewer features.

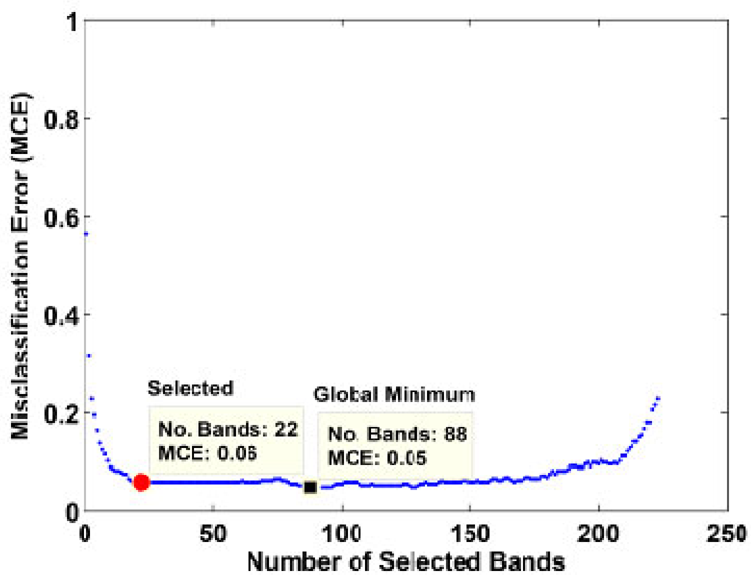

Figure 2. The MCE curve for different number of selected bands from SFS

Analysis of the Selected Wavelengths

The spectral responses of the samples were shown at the selected wavelengths (Figure 3). The wavelengths in the visible range were selected due to the natural or synthetic pigments in the FM. All the botanical FM (bark, leaf, etc.) contained at least three natural pigments: chlorophyll a and b, and carotenoid. The chlorophyll a and b have absorption peaks at ~440 nm and ~660 nm in the diethyl ether, which correlates to the 437 to 488 nm and 646 to 696 nm in the spectral images, whereas the carotenoid has the peak at ~470 nm in the diethyl ether, which correlates to the 503 to 515 nm in the spectral images. Since the botanical FM contained different amount of the chlorophyll and carotenoid, they could be differentiated from each other. For example, the green leaf contained a larger quantity of chlorophyll than the brown leaf, so it showed less reflectance at 646 nm to 696 nm. The non-botanical FM also contained pigments, for instance, the plastic bale packaging was made with a synthetic pigment which absorbed light at ~500 nm. The near-infrared (NIR) range 795 nm, 831 nm, and 982 nm) was selected because of the special chemical compositions. Since the NIR range is not affected by the color factor, the FM with similar appearance could be separated. For example, plastic bag was different from cotton lint.

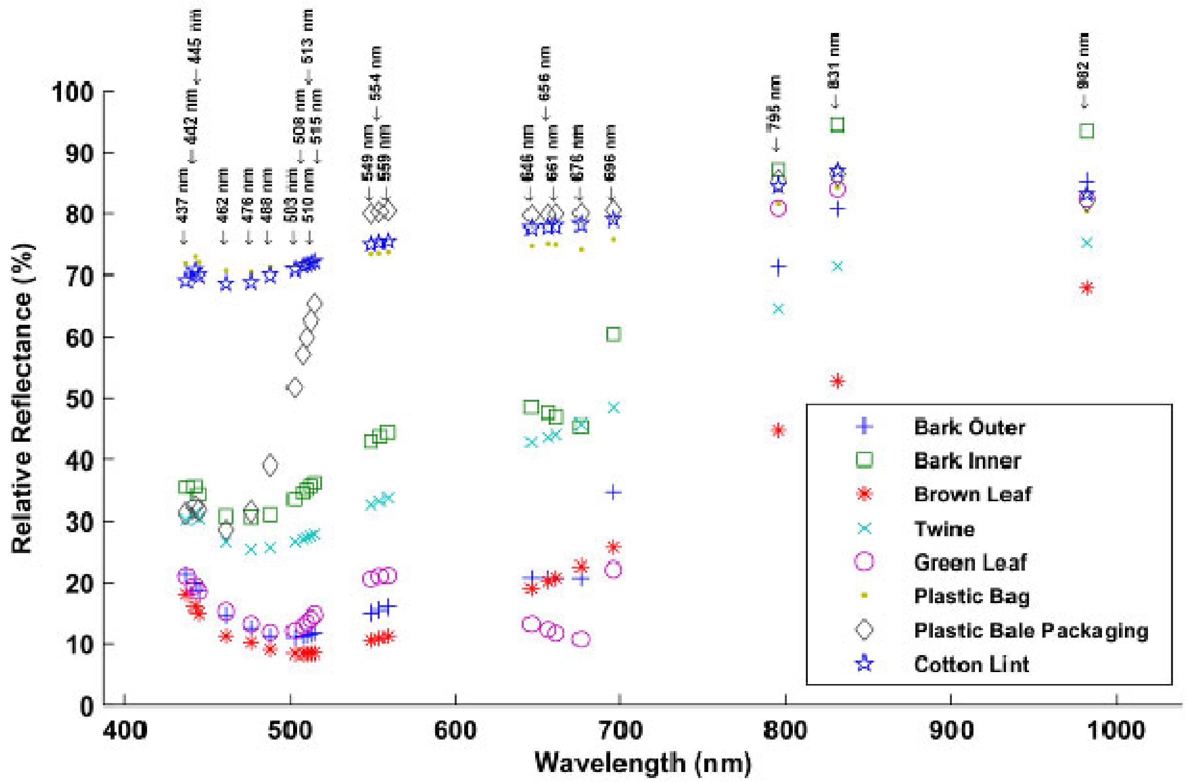

Figure 3. The spectral responses of the samples at the selected 22 wavelengths

Classification using the selected Wavelengths

The selected wavelengths achieved at least 93.33% for all 8 types of the samples (Figure 4). Only one bark inner was misclassified as bark outer, but the quality measurement was not affected by this misclassification. Because both inner and outer bark were considered as bark in the quality measurement.

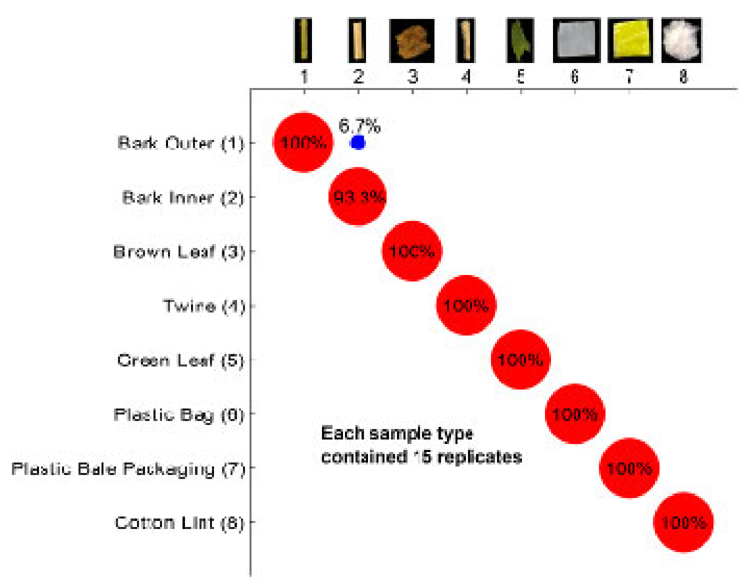

Figure 4. The classification result calculated by the LDA using the selected wavelengths

Summary

There are three limitations on the present study: FM size, ROI-based spectra extraction, and internal detection. In order to obtain the accurate spectra of the samples, the sample was cut into ~1×1 cm^2piece, which was too large compared with the size in the industrial situation. Additionally, the spectra extracted by ROI method were smoother than that of individual pixel, so the variation of the pixel spectra was larger than the ROI one. The larger variation could significantly lower the classification performance. Besides, the ROIs were manually selected, which cannot be applied on any automatic online detection system. Currently, the system can only detect and classify the FM on the surface of the cotton lint because it is based on the reflectance mode, which cannot acquire information of the FM inside the lint. So, the classification result can be used only for the surficial FM measurement.

In summary, 22 wavelengths were selected as the optimal set for the FM classification, and they were highly correlated to the chemical properties of the lint and FM samples. Thus, they can be used for building multispectral online detection system. The future study will be focused on exploring the size issue and implementing the online pixel-level image classification algorithm using the selected wavelengths.

Acknowledgements

This study was funded by the Cotton Incorporated and Georgia Cotton Commission. The authors gratefully acknowledge Andy Knowlton (Research Engineer/Gin Manager) at the UGA Cotton Micro Gin in Tifton, GA, USA for kindly providing us with cotton lint and cotton foreign matter samples. The authors also would like to thank Dr. Adnan Mustafic for sample preparation.

References

Brosnan, T., Sun, D.W., 2002. Inspection and grading of agricultural and food products by computer vision systems - a review. Computers and Electronics in Agriculture 36, 193-213.

Fortier, C.A., Rodgers, J.E., Cintron, M.S., Cui, X.L., Foulk, J.A., 2011. Identification of cotton and cotton trash components by Fourier transform near-infrared spectroscopy. Text Res J 81, 230-238.

Himmelsbach, D.S., Hellgeth, J.W., McAlister, D.D., 2006. Development and use of an attenuated total reflectance/Fourier transform infrared (ATR/FT-IR) spectral database to identify foreign matter in cotton. J Agr Food Chem 54, 7405-7412.

Isabelle, G., André, E., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157-1182.

Jia, D.Y., Ding, T.H., 2005. Detection of foreign materials in cotton using a multi-wavelength imaging method. Meas Sci Technol 16, 1355-1362.

Jiang, Y., Li, C., 2014. A Push-broom based Hyperspectral Imaging System for Cotton Trash Identification, 2014 Montreal, Quebec Canada July 13 – July 16, 2014, p. 1.

Li, D., Yang, W., Wang, S., 2010. Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. Computers and Electronics in Agriculture 74, 274–279.

Loudermilk, J.B., Himmelsbach, D.S., Barton, F.E., de Haseth, J.A., 2008. Novel search algorithms for a mid-infrared spectral library of cotton contaminants. Appl Spectrosc 62, 661-670.

Qin, J.W., Chao, K.L., Kim, M.S., Lu, R.F., Burks, T.F., 2013. Hyperspectral and multispectral imaging for evaluating food safety and quality. J Food Eng 118, 157-171.

United States Department of Agriculture, A.M.S., 2005. Cotton classification.

Xu, B., Fang, C., Watson, M.D., 1999. Clustering analysis for cotton trash classification. Text Res J 69, 656-662.

Zhang, X., Li, D., Yang, W., Wang, J., Liu, S., 2011. A fast segmentation method for high-resolution color images of foreign fibers in cotton. Computers and Electronics in Agriculture 78, 71-79.