Shortwave infrared hyperspectral reflectance imaging for cotton foreign matter classification

Ruoyu Zhang, Changying Li, Mengyun Zhang, James Rodgers

Abstract

Cotton contaminants seriously reduce the commercial value of cotton lint and further degrade the quality of textile products. This research aims to investigate the potential of a non-contact technique, i.e., liquid crystal tunable filter (LCTF) hyperspectral imaging, to inspect foreign matter on the surface of cotton lint. The foreign matter samples used in this study included 11 types of botanical foreign matter and 5 types of non-botanical foreign matter. Hyperspectral images of the foreign matter were acquired using a LCTF hyperspectral imaging system with a spectral range from 900 to 1700 nm. The mean spectra of the foreign matter and lint samples were extracted manually from the images. Linear discriminant analysis was applied to classify different types of foreign matter and cotton lint according to their spectral features. Classification accuracies of 96.5% and 95.1% were achieved with leave-one-out and four fold cross-validation, respectively. For pixel-level image classification, a majority of the pixels for different types of foreign matter were classified correctly by a support vector machine, using the top features of the minimum noise fraction transformation. The results demonstrate that non-contact liquid crystal tunable filter hyperspectral imaging is a promising method to discriminate foreign matter materials from cotton lint.

1. Introduction

Cotton is an important living material produced globally, and cotton quality directly affects its potential profitability. China, India and the United States are the top three raw cotton producers, and combined they provide two-thirds of cotton produced globally (Zhang et al., 2014). Although the United States ranks third behind China and India in production, it is the largest exporter, with over 10 million bales of cotton exported each year (Cotton Incorporated, 2015). Among various cotton quality properties, contamination from foreign matter (FM) is one of the most important quality indices, as it directly affects ginning and spinning processes and the overall quality grade of cotton. Cotton contamination has become a more prominent issue with the introduction of new mechanical harvesting machines and processes and the application of new planting practices that are increasingly available around the world (Kan et al., 2011). For example, FM contaminants are being observed in cotton with the use of plastic mulch and drip irrigation belts as a result of new cotton planting practices in China. Similarly, fragmented plastic module cover and wrap have become sources of cotton FM contamination in the United States as more growers adopt the new mechanical harvesters with an onboard module builder (National Cotton Council, 2015a).

Traditional instruments like the High Volume Instrument (HVI), Advanced Fiber Information System (AFIS), and Shirley Analyzer (SA) have been used by the textile industry to evaluate cotton contamination. These systems, however, can only report an overall estimation of the total amount of FM content (Liu et al., 2012a), and they lack the ability to classify different types of FM. To classify FM, researchers have attempted to use spectroscopy and machine vision technologies in recent years. Spectroscopy is a fast and non-destructive technique which can measure multiple quality attributes simultaneously (Nicolai et al., 2006). The method detects FM according to the different spectral characteristics, such as reflectance, absorption, and fluorescence, of different materials. Liu et al. (2012b) examined a set of botanical FM (leaves, seed coats, hulls, stems) using Vis/NIR reflectance spectroscopy. Results showed that total trash, leaf trash, and non-leaf trash could be...
efficiently predicted with a ratio of performance to deviation (RPD) of 3.6, but the non-botanical FM were not investigated in this study. Fortier et al. (2011) explored using Fourier transform near infrared (FT-NIR) spectroscopy to identify lint, hull, leaf, seed coat, and stem. Although results showed a prediction accuracy of 97% for FM (hull, leaf, seed coat, and stem) identification, more reference spectra from new cultivars were needed to improve the robustness of the library. In addition to examining the absorption spectra of FM at the visible and near infrared range, some researchers studied fluorescence spectroscopy to distinguish FM from cotton. Gamble and Foulk (2007) utilized fluorescence spectroscopy and chemometrics method to quantitatively estimate the content of six botanical FM (leaf, stem, hull, bract, shale, and seed coat) according to their fluorescent characteristics. Promising results were reported ($R_{\text{hull}} = 0.94$, $\text{SECV}_{\text{hull}} = 6.2$%; $R_{\text{leaf}} = 0.93$, $\text{SECV}_{\text{leaf}} = 6.4$%) in detecting hull and leaf, but results for other FM were relatively poor. Despite the success that the spectroscopy technique achieved in detecting hull and leaf, but results for other FM were relatively poor. The main drawback of this approach is that it is a point-based measurement technique that lacks the spatial information (Ariana et al., 2006).

Compared to the spectroscopy technique, the imaging approach can provide spatial information of foreign matter such as size, shape, and spatial distribution of the color (Mustafic et al., 2014). The color imaging systems are inexpensive, relatively easy to build, and therefore have been intensively investigated previously to detect cotton FM. Xu et al. (1999) measured color features of four types of FM (bark, leaf, seed coat inner, and seed coat with or without fiber) and fed these features into three clustering methods (sum of squares, fuzzy clustering, and neural networks) to classify FM. Results showed that the neural networks achieved the highest classification accuracy (95%). Most relevant studies conducted in the U.S. have focused on botanical FM detection, whereas studies in China have focused on non-botanical FM detection in cotton. Yang et al. (2009b, 2011b) built a line scan imaging system to detect and classify the botanical FM from cotton lint. Results showed that the mean classification accuracy was 92.34% for six types of FM with different colors (i.e., red cloth, red polypropylene twine, hemp rope, black plastic film, black hair, and black feather). Despite its success, this study only investigated non-botanical FM and another limitation of the system was that this color-imaging system can only detect FM with distinct colors. To overcome this limitation, Mustafic et al. (2014) developed a fluorescence imaging system to identify twelve types of cotton FM excited under blue and UV LED. Results showed that there were different levels of fluorescence among the twelve FM and cotton lint, and fluorescence imaging could be effective in detecting FM with similar color but different fluorescence emissions. This method, however, was less effective in detecting FM that did not contain fluorophores.

Hyperspectral imaging (HSI) is a relatively new and nondestructive imaging technology (Gown et al., 2007). This powerful technique combines the strengths of imaging and spectroscopy to acquire both spectral and spatial information simultaneously from samples. Due to the combined strengths, the HSI technique can greatly enhance the capability to detect and assess the quality of agricultural products. The HSI technique has also been explored by a few research groups in China and U.S. to detect cotton FM. Guo et al. (2012) developed a line scan HSI system in the spectral range from 500 to 900 nm to detect 8 types of non-botanical FM commonly found in China. Results indicated that gray polypropylene fiber and black hairs were detected effectively with an accuracy of 93%, but light color and transparent FM such as Polyethylene (PE) mulching film was detected with an accuracy of 53.8% in the test set. Jiang and Li (2015) developed a hyperspectral imaging system to detect 15 types of botanical FM commonly found in the United States. Results showed that 9 types of FM formed distinct clusters in principal component score plots, and all types of FM were different from each other at the significance level of 0.05 except between brown leaf and bract. One limitation of this study, however, is that it only classified FM based on the spectra extracted from hyperspectral images, but stopped short of performing pixel-level classification on the images, which is crucial to making this technique applicable to cotton contamination assessment in the classing office.

These prior studies of using hyperspectral imaging for cotton contamination detection were also limited by the following two aspects. First, the spectral range covered by the HSI systems was mainly in the visible and a small portion of the short wave near infrared range (i.e., 400–1000 nm). Although many types of FM can be discriminated in this spectral range, some contaminants such as plastic bags could be more effectively detected in the near infrared spectral range due to their unique chemical characteristics (Jiang and Li, 2015). Böhm et al. (2006) pointed out that the NIR spectral range (750–2500 nm) is most suitable to identify polypropylene and polyethylene in cotton. Yang et al. (2011a) also found that the near infrared spectrum from 780 to 1800 nm was more effective than the visible spectrum to detect plastic bags, hair, and leather. Most botanical FM such as stem, bract, hull, and seed are composed of lignin or protein, whereas cotton lint is mainly composed of cellulose (Himmelsbach et al., 2006). Lignin, protein, and cellulose are made of molecular bonds such as $\text{CH}_2$, $\text{OH}$, and $\text{NH}$—that have absorption bands in the NIRS spectral range (Wakelyn et al., 2006). In addition, previous studies mostly used line-scan or push-broom scanning which required the object to move in a direction perpendicular to scanning lines so that the entire object can be imaged (Lu and Chen, 1998). Liquid crystal tunable filter (LCTF) based hyperspectral imaging system is an area-scan technique that captures spatial 2-D images at selected wavelengths sequentially and then stacks these images from different wavelengths together to form a 3-D hyperspectral image cube (Wang et al., 2012b). Compared to the line scan technique, the LCTF system does not require a moving sample and is especially suitable for developing a multispectral imaging system that can randomly access user-selected wavelengths.

The overall goal of this study was to investigate the efficacy of a non-contact short wave near infrared hyperspectral imaging system to detect and classify various types of botanical and non-botanical cotton FM found in the United States and China. Specific objectives of this study were to: (1) evaluate the effect of the glass plate on the spectra of the FM and lint samples; (2) extract the spectra of FM and lint from the region of interest on the hyperspectral images and classify cotton FM using the extracted spectra; and (3) classify cotton FM at the pixel level on the image.

2. Materials and methods

2.1. Cotton lint and FM samples

To make the study relevant to the cotton industry in both the United States and China, a total of 13 types of FM collected in both countries were examined in this work (Fig. 1). The FM collected in the United States included eight types of botanical FM (stem, bark, bract, brown leaf, green leaf, hull, seed coat, and seed meat), and two types of non-botanical FM (twine and module cover). All types of the botanical FM were collected from the seed cotton and trash after ginning in December 2014. Twine (Lehigh Group 530 Jute Twine, Model No. 016033) was purchased from a local store and module cover was obtained from the Micro Gin at the University of Georgia Tifton Campus. Three types of non-botanical FM samples (plastic mulch, drip irrigation belt, and poly woven bag) com-
commonly found in China were collected from the cotton field in Xinjiang Province, China in September 2014. Three types of botanical FM (stem, bark, and seed coat) were examined for both the inner and outer surfaces due to the distinct features between the two surfaces, making a total of 16 categories of FM. Five commercial cotton cultivars (Delta Pine 1050, Delta Pine 1137, Delta Pine 1252, Phytogen 499, and Stoneville 6446) machine-harvested in 2014 were collected from the Micro Gin. The purpose of selecting five cultivars was to ensure the robustness and repeatability of our approach to detect cotton foreign matter regardless of cultivars, which was proved by our data. Therefore, the five cultivars were not differentiated in the following discussions. Six replicates of cotton lint from each cultivar were prepared for image collection, making a total of 30 pieces of cotton lint web. One sample from each type of FM was laid on top of each thin cotton lint web, making a total of 30 replicates for FM samples.

To make the size of the FM more similar to what was typically found in ginned cotton, eight types of FM samples (bract, brown leaf, green leaf, hull, seed coat inner, module cover, plastic mulch and drip irrigation belt) were cut into \( \frac{5}{2} \times \frac{5}{2} \) mm snippets with a scissor. The size of the above FM was only about a quarter of what had been investigated previously (Jiang and Li, 2015). Six types of FM (stem inner, stem outer, bark inner, bark outer, twine, and poly wave bag) were cut into rectangular shape with \( \sim 5 \) mm in length, while the width was maintained in their natural length that typically was less than 5 mm. The size of the above six types of FM was only about half of what had been investigated previously (Jiang and Li, 2015). The other types of FM (cotton seed coat and cotton seed meat) were kept their natural size.

The weight of each cotton lint web was about 25 g with a thickness of 5 mm. To avoid potential interferences from other unknown contaminants, the cotton lint web was manually cleaned. A total of 16 types of FM samples were placed on top of the cotton lint web, and then a black cardboard with a 4 by 4 grid was laid on top of the FM samples to separate each type of FM. The cotton lint samples were fluffy and non-uniform under natural conditions, which could affect the integrity of the spectra. To reduce thickness variation of cotton lint and potential negative impact on the spectral data, a piece of floated borosilicate flat glass (BOROFLOAT® 33, thickness = 2.00 mm, Home Tech SCHOTT North America, Inc, Louisville, KY, USA) which has over 90% transmission in the near infrared spectral range was covered on the surface of lint samples.

**Fig. 1.** Sixteen types of foreign matter and cotton lint samples from five cultivars.
during image acquisition. To compare the spectral differences between the samples covered with and without the glass plate, the spectra of the samples were collected for both cases.

2.2. Short wave infrared hyperspectral imaging system

A short wave infrared (SWIR) HSI system developed by the Bio-Sensing and Instrumentation Lab was used to acquire hyperspectral images of FM and cotton lint samples. The system consists of a liquid crystal tunable filter (LCTF) (LNIR 20-HC-20, Cambridge Research & Instrumentation, Cambridge, MA, USA), an indium gallium arsenide (InGaAs) SWIR camera (SU320KTS-1.7RT, Goodrich, Sensors Unlimited, Inc., Princeton, NJ, USA) coupled with an near infrared lens (SOLO 50, Goodrich, Sensors Unlimited, Inc., Princeton, NJ, USA), and two 150 W halogen lamps for illumination (Fig. 2). A computer (Intel® Pentium® D Processor, 4 GB DDR3, Windows 7) was used to interface with the imaging system via a Camera Link frame grabber, and to acquire process the images. All the samples were scanned inside an enclosed light chamber to avoid interferences from ambient light. The nominal spectral response range of the imaging system was from 850 to 1850 nm. Due to the low signal to noise ratio at both ends of the spectrum, however, near infrared hyperspectral images were only acquired from 950 to 1650 nm with a 5 nm spectral interval. The distance between the lens and the samples was kept at 875 mm. After scanning a sample, a three-dimensional (x, y, λ) image cube with both spatial (320 × 256 pixels) and spectral data (141 wavelength bands) was constructed. The HSI system was controlled by an
in-house built LabVIEW computer program for image acquisition (Wang et al., 2012a).

The acquired hyperspectral images were corrected using the flat field correction algorithm (Eq. (1)) implemented in Interactive Dynamic Language (IDL4.7, Exelis Visual Information Solutions, Boulder, CO, USA) (Wang et al., 2012b). White reference images were acquired from a white Spectralon panel with 99% reflectance (SRT-99-050, Labsphere Inc., North Sutton, NH, USA) and dark images were acquired by covering the lens of the camera completely.

\[
I_{\text{Relative reflectance}} = \frac{4095(I_{\text{Sample}} - I_{\text{dark}})}{(I_{\text{White}} - I_{\text{dark}})}
\]

where \(I_{\text{Sample}}\) is the intensity of the raw image; \(I_{\text{dark}}\) is the intensity of the dark image; \(I_{\text{White}}\) is the intensity of the white reference image. The coefficient 4095 was used to display the image in its full dynamic range (12-bit).

2.3. Spectral data extraction, preprocessing, and classification

Image cropping was first implemented to remove noises around the border of the image due to misalignment of the raw image and reference images during flat field correction. Principal component analysis (PCA) is a projection method to extract systematic variations in a data set by forming principal components (PCs) (Johnson and Wichern, 1989). It is often used to eliminate the multicollinearity in the original data, and to replace the original variables (wavelengths) with fewer principal components (PCs) through an orthogonal transformation to maximize representation of the original data. In this study, PCA was applied to hyperspectral images in order to find the largest differences between the FM and lint. One of the PC images that showed the greatest contrast between the FM and lint was used to select the region of interest (ROI) for each FM sample. The ROIs of cotton lint and FM on this PC image were manually created in a zoom window using the rectangle shape provided by ENVI. Efforts were made to ensure that all the pixels in an ROI belonged to a specific sample type, and other undesired pixels were not included. Eventually, the mean spectra of cotton lint and FM were extracted from the ROIs on the hyperspectral images using ENVI.

Each piece of spectrum extracted from the ROIs was normalized by dividing the raw intensity at each wavelength by the maximum intensity value found in the entire spectrum (i.e., 950–1650 nm). After normalization, the reflectance values of the FM and cotton lint were in the range from 0% to 100%.

Linear discriminant analysis (LDA) is a supervised classification method that performs a linear transformation by maximizing between-class variations and minimizing within-class variations. The discrimination performance was evaluated by the percentage of samples that were correctly classified using the leave-one-out and fourfold cross-validation.

2.4. Pixel level image classification

The minimum noise fraction (MNF) transformation is a common method used to reduce spectral dimensionality for hyperspectral images (Lu and Chen, 1998). It can maximize the signal to noise ratio through computing normalized linear combinations of the original wavelengths (Moham and Porwal, 2015). Support vector machine (SVM) is a supervised classification method. Like other kernel-based methods, it is robust when only a limited number of training samples are available (Cristianini and Shawe-Taylor, 2000). In this study, the radial basis function was selected as the kernel function, and important parameters (gamma in the kernel function and the penalty parameter) were automatically optimized by ENVI. Image classification was conducted in the following procedures. Effective information of the hyperspectral image was extracted by the MNF analysis. Meanwhile, the optimal number of MNF components was determined by the criterion that each MNF component should contribute at least 0.5% of the total variation. ROIs from each type of FM as well as cotton lint and the background were manually drawn on the MNF image to serve as the training set for the SVM classifier. To evaluate the accuracy of the classification, ground truth ROIs were manually drawn as well. Eventually, the SVM method was applied to classify FM and lint on the MNF images. The producer accuracy and user accuracy were used to assess the performance of the classification. The producer accuracy is the possibility of the pixels that are classified correctly in the ground truth ROIs. The user accuracy is the ratio between the number of pixels that are correctly classified into a class and all the pixels classified by the classifier into this class (Harris Geospatial Solutions and Envi, 2016).

Spectral normalization and canonical discrimination analysis were carried out using programs developed in MATLAB 2014 (The MathWorks Inc., Natick, MA, USA). All statistical tests were performed in SAS 9.4 (SAS Inc., Cary, NC, USA). Image preprocessing (i.e., flat field correction and image cropping) was done using programs developed in ENVI + IDL (ITT Visual Information Solutions, Boulder, CO, USA). Imaging processing using PCA, MNF and SVM was implemented in ENVI 4.6 (ITT Visual Information Solutions, Boulder, CO, USA).

3. Results and discussion

3.1. Spectra extraction and normalization

Raw hyperspectral images were cropped to remove the noise along the image border. It was difficult to identify and segment FM on top of cotton lint using a single-band spectral image. Therefore, Principal Component (PC) analysis was performed on the raw hyperspectral images and PC images were used for FM segmentation. Among the top six PC images, the fifth PC (PC5) image showed the best contrast between FM and lint (Fig. 3e), so PC5 image was applied to select ROIs.

The ROIs for each FM and lint were created using the rectangular tool in ENVI and marked on the PC5 image (Fig. 4a). The ROIs of each FM and lint were then mapped onto the original hyperspectral images to extract the full spectra (i.e., 950–1650 nm) and the spectra were averaged to represent the mean spectrum of each ROI (Fig. 4b). A total of 510 mean spectra (30 × 11 botanical FM samples, 30 × 5 non-botanical FM samples, and 30 cotton lint samples covered by a glass plate) were obtained.

To test the hypothesis that the glass plate placed on top of lint does not affect the spectra of the samples, ten hyperspectral images of a piece of white paper with and without glass plate coverage were acquired, respectively. Similarly, normalized mean spectra were obtained from a black cardboard with and without glass plate coverage. Normalized mean spectra were extracted from each image. Although the magnitude of reflectance for the samples covered with the glass plate was slightly lower than that without glass coverage from 1350 nm to 1650 nm, the trends of the spectra were the same for both white paper and the black cardboard (Fig. 5a). The phenomenon was supported by the optical property of borosilicate flat glass (Schott, 2015). Further, specular reflectance from the glass surface may have slightly impacted the total reflectance from the white paper and black cardboard. Overall, the glass plate did not affect the spectral properties of the samples, and therefore it was used to cover the FM and lint samples in order to keep the density of the cotton lint uniform.

Normalized mean spectra (ranging from 0% to 100%) of FM and cotton lint (Fig. 5b) showed different reflectance characteristics.
between different types of FM and cotton lint covered by the glass plate. Each mean spectrum of the FM and cotton lint sample was an average of 30 samples within each category. The spectra of botanical FM in the spectral range from 1100 to 1650 nm agreed overall with previous NIR spectral results (Fortier et al., 2012). The mean spectra of seed coat inner, seed coat outer, and seed meat were largely different from other FM categories as well as lint because the seed coat and meat have different chemical components, displaying strong absorption peaks at 1200 nm and 1500 nm (Fig. 5c). Small differences in the mean spectra were observed in two pairs of botanical FM (i.e., bark inner and hull; stem outer and bract), primarily because they have similar chemical components (i.e., lignin). The mean spectra of most non-botanical FM (except twine) were clearly different from those of cotton lint. Overall, the spectral results for the non-botanical FM were in agreement with NIR region spectral differences for non-botanical FM observed previously (Fortier et al., 2012). The differences between the twine and lint were quite small primarily because the jute twine contains similar chemical components (e.g. cellulose, waxes) as cotton lint (National Cotton Council, 2015b). The drip irrigation belt samples are made of thick polyethylene and black color concentrate (carbon black) which resulted in strong absorption of light (Povacz et al., 2014). In addition, the use of solid concentrates can significantly impact the diffuse reflectance of the fiber (Rodgers, 2002). As a result, the drip irrigation belt showed the greatest difference from the lint and other FM categories throughout the entire spectrum. Compared to the difference between plastic mulch and lint, the difference between the poly woven bag and lint was greater. Since the plastic mulch was thin and transparent, most of its spectral characteristics in reflectance were similar to those of cotton lint. Poly woven bag was made of polypropylene (PP) and fluorescent brightening agents with special molecular structures (Chemnet, 2015). Generally, module cover is made of polyethylene (PE) or polyethylene terephthalate (PET) and stabilizer for polymers (Klein, 2011), resulting in significant spectral differences from lint and other FM categories.

Fig. 3. The top 6 principal component (PC) images.

Fig. 4. Regions of interest (ROI) defined on a principal component image (a) and on a spectral image at one wavelength (b). The ROIs were used to extract the mean spectra from foreign matter and cotton lint.
The differences between various FM and cotton lint largely attribute to bond vibration regions which are related to chemical components of FM and lint. In the short wave infrared range, these bond vibration regions are C–H third overtone (1300–1420 nm), O–H first overtone (1420–1600 nm), and N–H first overtone (1420–1600 nm) (Aenugu et al., 2011). There are several wavelengths in the standard absorption region including HC=CH second overtone (1170 nm), CH\_3 second overtone (1195 nm), CH\_2 combination (1395 nm), ROH first overtone (1410 nm), protein first overtone (1500 nm), starch first overtone (1540 nm). Other than the above bands, 1145 nm is also close to the aromatic second overtone (~1143 nm). Cotton and other cellulosic products exhibit a spectral band at approximately 1215–1250 nm due to the C–H second overtone, a slight band at approximately 980 nm due to the overtone of the 1930–1960 nm moisture band (O–H), and a band between 1000 and 1050 nm due to the overtone of the 2100 nm cotton and moisture (water) band (O–H and C–O combination) (Fortier et al., 2012; Rodgers et al., 2010). The broad band from approximately 1350 to 1650 nm has been identified as a combination of cotton (cellulosic) and moisture band, due primarily to the C–H combination first overtone vibrations for cotton and other cellulosic materials and the O–H first overtone vibrations for both cotton/cellulosic and water (Fortier et al., 2012; Rodgers et al., 2010). Some FM categories (i.e., stem and bract) are composed of lignin which mainly contains aromatic elements. It is known that cotton lint is composed of cellulose (93–95%), in addition to ~5–7% of pectin and ~0.7–1.1% of wax on its surface (Himmelsbach et al., 2006). Lignin consists of structures such as OH=–, –CH=–OH, CH\_2=O=, C\_6H\_4O=, and C–O (Wikipedia, 2015a), and the C\_6H\_4O= band appears at 1145 nm. Jute twine is mainly made of cellulose which contains special structures such as C–OH, C–O–C, and OH (Wikipedia, 2015b).

In contrast, most synthetic FM categories are composed of polymers, such as Polyethylene (PE), Polypropylene (PP), Polyethylene terephthalate (PET), some color concentrate and stabilizer. The monomer of Polyethylene is ethylene, a hydrocarbon with the formula C\_2H\_4, which contains a pair of methylene groups (\(=\text{CH}_2\)) (Whiteley et al., 2000). Polypropylene contains CH\_3\_2, \(=\text{CH}=-\), and \(=\text{CH}\_2\) (Beswick and Dunn, 2002), and the absorption of \text{CH}\_3\_2 takes place at 1195 nm. Polyethylene terephthalate (PET) consists of polymerized units of monomer ethylene terephthalate, with repeating \(\text{C}_6\text{H}_4\text{O}_2\) units which mainly contain \(-\text{CH}=\_\text{CH}+\text{CH}=\_\text{OH}\) (at 1395 nm) and COO= (Lepoittevin and Roger, 2011). Various non-botanical bands observed from approximately 1350 to 1650 nm are in good agreement with bands assigned to C–H combinations, third overtone of CH and CH\_2 bands that occur at approximately 2300–2350 nm, and the O–H bands (e.g., PET) at approximately 1530–1570 nm (Fortier et al., 2012; Rodgers, 2002).

3.2. Spectral classification of FM and lint

The full wavelengths were used to discriminate 16 types of FM and cotton lint. The 17 classes were generally separable using the top three canonical variables (CV) (Fig. 6). In particular, drip irrigation belt, module cover, and seed meat were far from other FM categories and lint. This pattern was in agreement with the spectral characteristics of these FM.

Linear discriminant analysis was applied to classify 16 types of FM and cotton lint using leave-one-out and fourfold cross validation methods and the two methods achieved similar results (Fig. 7). Cotton lint, seed coat inner, seed coat outer, seed meat, green leaf, brown leaf, bract and 4 synthetic FM categories (module cover, plastic mulch, poly woven bag, and drip irrigation belt) were all correctly identified with a classification accuracy of 100%. These results suggest that near infrared spectra is suitable for detecting cotton lint, most botanical FM, and these 4 synthetic FM categories. The classification results related to synthetic FM were in agreement with other studies (Yang et al., 2009a). Seed coat outer, seed coat inner, and seed meat also can be easily classified with an accuracy.
of 100.0%, suggesting that the composition of the seed and seed coat is substantially different from that of other FM and lint.

Some bark outer samples were misclassified into bract, because the spectral features of bark outer are similar to those of bract. Therefore, the bark outer samples were classified with the lowest accuracies of 83% in leave-one-out and 77% in fourfold cross-validation. Similarly, bark inner and stem outer were misclassified into other botanical FM (i.e., hull, stem inner, and bract). About 6.7% of twine samples were misclassified into lint, which was likely because the jute twine contains similar chemical components with lint, supported by their similar spectral profiles. Overall, leave-one-out and fourfold cross validation methods achieved 96.5% and 95.1% classification accuracy, respectively.
3.3. Pixel-level image classification

The first twelve MNF components were selected based on the selection criterion described earlier. The training set ROIs and ground truth ROIs for each FM category as well as cotton lint and background were selected and marked with different colors (Fig. 8a and b). On the initial classification map (Fig. 8c), the majority of the pixels within each class were correctly classified, although there were many sporadic misclassified pixels on the image. After applying morphological image processing by an average filter (kernel = 3), most of these noises were removed and every class was marked with one specific color on the image (Fig. 8d). It appeared that most misclassifications occurred between plastic mulch and lint: many pixels of lint (white pixels) were misclassified as plastic mulch (green1 pixels). This was mainly plastic mulch was very thin and almost transparent, and as a result, the difference between plastic mulch and lint on the spectral image was quite small (Jiang and Li, 2015). The lack of distinction between the two also led to the difficulty in selecting the training set ROI for plastic mulch. Certain erroneous selection of training pixels could lead to misclassifications of the two classes. There were also some minor misclassifications around the edges between seed coat inner and outer, and between drip irrigation belt and black cardboard background. The overall accuracy of the image classification on this particular image was 96% (correctly classified pixels divided by total pixel numbers from the ground truth ROIs) with a Kappa coefficient of 0.94 (a value between 0 and 1; 1 indicates perfect agreement between classification and ground truth pixels).

To test the repeatability of the image classification method, three randomly selected hyperspectral images were used for accuracy evaluation. Overall, image classification results were similar among the three images. The mean user accuracy and producer accuracy were 94.63% and 82.56%, respectively (Fig. 9). High user accuracy suggested that SVM was an effective classifier. For each class, the classifier achieved producer accuracies greater than 65%. Compared to spectral discrimination results, image classification accuracies for each individual class was lower. This result was reasonable because the spectral classification was performed in cross validation (leave-one-out and fourfold), while the image classification was evaluated on many unseen pixels that were not in the training ROIs. Nevertheless, image classification results

Fig. 8. Pixel-level image classification of foreign matter and cotton lint using SVM.

1 For interpretation of color in Fig. 8, the reader is referred to the web version of this article.
demonstrated the feasibility of pixel-level image discrimination for different types of FM and lint, which could be adopted for online inspection in the future.

4. Conclusions
This work investigated a non-contact near infrared imaging method to classify cotton lint and typical cotton foreign matter materials that are relevant to the cotton industry in both the United States and China. The data showed that the borosilicate plate did not have a significant effect on the spectral properties of FM and lint samples. By utilizing the spectra extracted from hyperspectral images, the LDA classification model achieved over 95% accuracy in classifying 16 types of foreign matter materials and cotton lint. At the pixel level on the hyperspectral images, the SVM model also achieved satisfactory classification accuracies for the foreign matter materials on top of lint. The results established the efficacy of the short wave infrared LCTF hyperspectral imaging technique for cotton foreign matter detection and classification at both the spectral and imaging domain. This technology could complement the current HVI in the U.S. classing offices for classifying and quantitatively assessing various botanical and non-botanical foreign matter simultaneously on the lint. In future studies, FM samples at different depths of cotton lint will be investigated by hyperspectral transmittance imaging.

Acknowledgements
The authors gratefully acknowledge the financial support provided by Cotton Incorporated and Georgia Cotton Commission of the United States, as well as China Scholarship Council Program (201205215041) and National Natural Science Foundation of China (No. 51265047). The authors also are in debt to 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 thank Delmaries Gonzalez and Wesley Passos for sample preparations.

References

Rodgers, J.E., 2002. Influences of carpet and instrumental parameters on the identification of carpet face fiber by nir. AATCC Rev. 2.


