

## Identification of Pure Cotton and Mercerized Cotton Fabrics Based on 2-D Correlation Spectra Analysis

CAO Kai<sup>1</sup>, ZHAO Zhong<sup>1\*</sup>, YUAN Hong-fu<sup>2</sup>, LI Bin<sup>2</sup>

1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
2. College of Materials Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China

**Abstract** Pure cotton and mercerized cotton products are widely used in daily life. It is difficult to classify the pure cotton and mercerized cotton products with simple methods because they are similar in chemical and physical structures. In this work, a new method of rapid identification of pure cotton and mercerized cotton products with two-dimension correlation spectra analysis was proposed. In this work, 200 textile samples including 100 pure cotton fiber products and 100 mercerized cotton fiber products were collected. For each sample, the water content was changed 4 times and one-dimension spectra was collected, among them, the water content of 4 times was 20.20%, 14.52%, 7.77% and 0% respectively. Then their simultaneous two-dimension correlation spectra were calculated based on correlational analysis. Three kinds of classification features were extracted from the synchronous two-dimension correlation spectra. Support Vector Machine (SVM) was combined with different kind of the classification features to construct different classifiers. In this work, an information fusion method was proposed to make the multi-classifier decision. To verify the feasibility and effectiveness of the proposed method, the comparative experiments have been done. The accuracy of identification with the classifier based on extracted one-dimensional spectra features with PCA was only 76%. The accuracy of identification with the three classifiers based on extracted features from two-dimensional correlation spectra were 88%, 90% and 88% respectively. The accuracy of identification with the proposed method was 92%. Compared with one-dimension spectra based feature extraction, the two-dimension correlation spectra based feature extraction achieved feature enhancement and the multi-classifier fusion decision could improve the accuracy of classification obviously. Two-dimensional correlation spectroscopy extended spectral information to higher dimensions, unfolded hidden fold peaks in one-dimensional spectra, and had higher classification accuracy. The proposed method provided a new way for rapid identification of pure cotton and mercerized cotton products.

**Keywords** Two-dimension correlation spectroscopy; Cotton and mercerized cotton; Support vector machine; Information entropy; Information fusion

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2019)05-1653-08

### Introduction

Pure cotton and mercerized cotton products are fiber products commonly used in daily life. It is usually difficult to

identify pure cotton and mercerized cotton products with traditional methods<sup>[1]</sup>. For these cases, Some methods based on pattern recognition have been proposed to solve these problems. Classification features on the internal chemical and physical structures are extracted from the molecular spec-

**Received:** 2018-04-07; **accepted:** 2018-09-12

**Foundation item:** Beijing Natural Science Foundation (4172044), Chaoyang District Collaborative Innovation Project (CYXC1707), National Key R&D plan(2016YFF0102505)

**Biography:** CAO Kai, (1992—), the postgraduate of College of Information Science and Technology, Beijing University of Chemical Technology e-mail: caok\_5062@163.com \* Corresponding author e-mail: zhaozhong@mail.buct.edu.cn

tra<sup>[2-3]</sup> which can be classified in pattern recognition methods. But till now, identification of the materials with similar chemical and physical structures, such as the pure cotton and mercerized cotton products, is still an ad hoc research. Classification feature extraction is the most important part in the spectra analysis based on pattern recognition method. A good feature extraction method for a classifier needs to reduce the data dimension for lower computing cost, and more importantly, to transform the original data to the final feature space where the samples can be classified<sup>[7]</sup>. Two-dimensional correlation spectroscopy was first proposed in 1986, and was used in the field of nuclear magnetic resonance<sup>[8]</sup>. Two-dimensional correlation spectroscopy can enhance the apparent spectral resolution. Spectra change caused by the perturbations can be used to study the dynamic correlations of different spectra. In this work, 200 textile samples were collected including 100 pure cotton fiber products and 100 mercerized cotton fiber products. For each sample, the water content was changed several times and its corresponding one-dimensional spectra was collected. Then the simultaneous two-dimensional correlation spectra were calculated based on correlational analysis. Three kinds of classification features were extracted from the synchronous two-dimension correlation spectra. Support Vector Machine (SVM) was combined with different kind of the classification features to construct different classifiers. In this work, an information fusion method was proposed to make the multi-classifier decision. To verify the feasibility and effectiveness of the proposed method, the comparative experiments have been done. The remainder of the paper is organized as follows: Section 2 introduces the methods. In Section 3, the experiment is presented. In Section 4, the results and discussion are presented. Section 5 is the summary.

## 1 Method

### 1.1 Two-dimension correlation spectra

#### 1.1.1 Acquisition of two-dimensional correlation spectra

The principle of two-dimensional correlation spectroscopy<sup>[9]</sup> is illustrated in Fig. 1. In the process of measuring the spectra of substances to be classified, a specific external perturbation is applied to them, and this external perturbation can be any physical or chemical change that can change the internal sample environment. The spectra which change with external perturbations are called dynamic spectra. Two-dimensional correlation spectra can be acquired by two-dimensional correlation calculation of dynamic spectra<sup>[10]</sup>.

##### (1) Calculation of dynamic spectra

First select an external disturbance, which can be any proper physical or chemical perturbation such as temperature and humidity<sup>[11]</sup>. In this work, the water content of samples

is selected as the external disturbance and the minimum/maximum value of external disturbance is defined as  $t_{\min}$ ,  $t_{\max}$ , and spectral intensity of external disturbances is defined as  $y(v, t)$ . Where  $v$  denotes the spectral changes,  $t$  denotes the external disturbance, and the dynamic spectra  $\bar{y}(v, t)$  can be defined as

$$\bar{y}(v, t) = \begin{cases} y(v, t) - \bar{y}(v) & t_{\min} \leq t \leq t_{\max} \\ 0 & \text{Others} \end{cases} \quad (1)$$

where  $\bar{y}(v)$  denotes the reference spectra. In this work, the average spectra is chosen as the reference spectra and can be calculated as

$$\bar{y}(v) = \frac{\sum_{j=1}^m y(v, t_j)}{m} \quad (2)$$

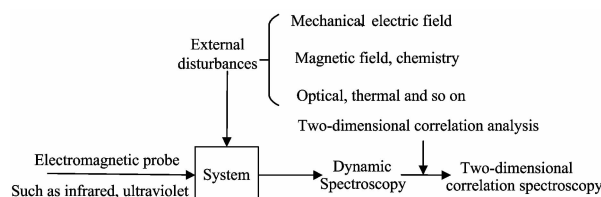


Fig. 1 Principle of two-dimensional correlation spectra acquisition

(2) Two-dimensional synchronous and asynchronous spectra calculation

The generalized two-dimensional correlation spectra is divided into synchronous two-dimensional correlation spectra and asynchronous two-dimensional correlation spectra. For two different spectral wavelength points  $v_1$  and  $v_2$ , their synchronization correlation strength  $\Phi(v_1, v_2)$  is defined as

$$\Phi(v_1, v_2) = \frac{1}{m-1} \sum_{j=1}^m y_j(v_1) y_j(v_2) \quad (3)$$

Asynchronous correlation intensity  $\Psi(v_1, v_2)$  is defined as

$$\Psi(v_1, v_2) = \frac{1}{m-1} \sum_{j=1}^m y_j(v_1) \sum_{k=1}^m (N_{kj} y_j(v_2)) \quad (4)$$

where  $m$  denotes the number of spectra measured during perturbation,  $N \in R^{m \times m}$  denotes the Hilbert-Noda matrix,  $y_j(v)$  denotes the spectral intensity of the  $j$ th spectra at the wavelength  $v$  in the dynamic spectra and

$$N_{kj} = \begin{cases} 0 & j = k \\ 1/\pi(k-j) & j \neq k \end{cases} \quad (5)$$

The dynamic spectra can be described in vector form, then the synchronization spectra and the intensity of asynchronous spectra can be defined as

$$\Phi(v_1, v_2) = \frac{1}{m-1} \tilde{y}(v_1)^T \tilde{y}(v_2) \quad (6)$$

$$\Psi(v_1, v_2) = \frac{1}{m-1} \tilde{y}(v_1)^T N \tilde{y}(v_2) \quad (7)$$

where  $\tilde{y}(v)$  denotes the dynamic spectra of the sample at wavelength  $v$ .

### 1.2 Feature classification method

#### 1.2.1 Support Vector Machine (SVM)

SVM is a feature classification method that is to find a classification hyperplane that could guarantee the high classification accuracy and make the classification interval maximization<sup>[12]</sup>. In this work, the dichotomous classification problem is studied with SVM. Assuming the given training sample set is  $(x_i, y_i)$ ,  $i=1, 2, \dots, l$ ,  $x \in R^n$ ,  $y \in \{\pm 1\}$  and the classification hyperplane of  $w x + b = 0$  exists, to maximize the classification accuracy and classification interval, then (8) must be satisfied

$$y_i [w x_i + b] \geq 1, \quad i = 1, 2, \dots, l \quad (8)$$

where  $x_i$  is the classification feature of the  $i$ th training sample and  $y_i$  is the classification category label of the  $i$ th training sample,  $w = (w_1; w_2; \dots; w_d)$  is the normal vector of the classification hyperplane, which determines the orientation of the hyperplane,  $b$  is the displacement term, which determines the distance between the hyperplane and the origin.

The distance from any point  $x$  in the sample space to hyperplane  $(w, b)$  can be written as

$$r = \frac{|x^T w + b|}{\|w\|} \quad (9)$$

The sum of distances between two heterogeneous support vectors and the hyperplane is calculated as  $r = 2 / \|w\|$ , which maximizes the interval. Then, the above problem of finding optimal hyperplane can be transformed into the minimization problem as

$$\min \Phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w'w) \quad (10)$$

Introduce Lagrange function as

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l a_i y_i (w x_i + b) + \sum_{i=1}^l a_i \quad (11)$$

where  $a_i > 0$  is the Lagrange multiplier. The QP (quadratic programming) problem as (11) can be transformed into the following duality problem ( $a = (a_1, a_2, \dots, a_l)$ ) as

$$\begin{aligned} \max Q(a) &= \sum_{j=1}^l a_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (x_i x_j) \\ \text{s. t. } &\sum_{j=1}^l a_j y_j = 0 \quad a_j > 0, \quad j = 1, 2, \dots, l \end{aligned} \quad (12)$$

The optimal weight vector  $w^*$  and optimal bias  $b^*$  can be calculated respectively as

$$w^* = \sum_{j=1}^l a_j^* y_j x_j \quad (13)$$

$$b^* = y_i - \sum_{j=1}^l y_j a_j^* (x_j x_i) \quad (14)$$

All  $a_j$  greater than 0 is set as  $a_j^*$ ,  $j \in \{j | a_j^* > 0\}$ , so the optimal classification surface is  $w^* x + b^* = 0$ , then the optimal classification function can be calculated as

$$f(x) = \text{sgn}\{w^* x + b\} = \text{sgn}\left\{\sum_{j=1}^l a_j^* y_j (x_j x_i) + b^*\right\} \quad (15)$$

where  $x \in R^n$ .

### 1. 2. 2 Multi-classifier fusion with voting mechanism

Random forest is a classifier that uses multiple decision trees to train and predict samples<sup>[13]</sup>. It can be used as a multi-classifier fusion method. Decision tree is a weak classifier. According to the voting mechanism, random forest fuses the multiple (odd) different weak classifiers to a strong classifier. According to probability principle, the accuracy of classification results of the fused strong classifier would be higher than that of each individual weak classifier. However, the idea of random forest with voting mechanism may amplify the effect of poor classifier and cause the lower accuracy of classification.

### 1. 2. 3 Multi-classifier fusion based on information entropy decision

Assuming a sample  $X$  may exist in several different states as  $x_1, x_2, \dots, x_n$ ,  $p(x_i)$  denotes the probability of state  $x_i$  ( $i=1, 2, \dots, n$ ) appearing, then the sample information entropy  $H(X)$  is defined as

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (16)$$

where  $0 \leq p(x_i) \leq 1$  and  $\sum_{i=1}^n p(x_i) = 1$ , when  $p(x_i) = 0$ , define  $0 \log 0 = 0$ . Information entropy is a measure of the degree of information disorder. Assuming a sample set has a total of  $K$  samples and the probability of each sample in different states under the classifier  $S$  is known, the information entropy  $H(S)$  of the sample set under the classifier  $S$  can be defined as<sup>[14-15]</sup>

$$H(S) = - \sum_{j=1}^K \sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (17)$$

In this work, identification of pure cotton and mercerized cotton fabrics is a dichotomous problem. The information entropy  $H(S)$  of the sample set in classifier  $S$  can be redefined as

$$H(S) = - \sum_{j=1}^K (p(x_1) \log(p(x_1)) + p(x_2) \log(p(x_2))) \quad (18)$$

Define the reciprocal of  $H(S)$  for classifier  $S$  as the weight of information fusion decision to avoid the unbalanced decision-making problem caused by the same weight of different classifiers. In this work, the identification problem is a two-class identification problem. Define the class labels are 1 and -1. In this work, there are three classifiers ( $S_1, S_2, S_3$ ) for information fusion decision. The three classifier's information entropy are defined as  $H(S_1)$ ,  $H(S_2)$ ,  $H(S_3)$  and the corresponding weights are  $1/H(S_1)$ ,  $1/H(S_2)$ ,  $1/H(S_3)$  respectively. For the same sample  $A$ , the discriminant labels of the three classifiers are  $P_1, P_2, P_3$  (the value of  $P_i$  is -1 or 1) respectively, then the final multi-classifier fusion result of the sample can be defined as

$$P_A = \text{sgn}(1/H(S_1)P_1 + 1/H(S_2)P_2 + 1/H(S_3)P_3) \quad (19)$$

## 2 Experiment

### 2.1 Samples

Totally 200 samples including 100 Cotton and 100 mercerized cottonfabrics shown in Fig.2 were provided by CSEPAT (Beijing, China) Technology Co. Ltd. Samples were taken from sweaters and shirts of different colors, both in size 15 cm×15 cm. The sample type number, calibration set, validation set, and sample size are shown in Table 1.



Fig. 2 Samples of pure cotton and mercerized cotton fabrics

Table 1 The statistics of the samples

| Fabrics type      | Number of samples | Number of samples of training set | Number of samples of validationset | Category number |
|-------------------|-------------------|-----------------------------------|------------------------------------|-----------------|
| Cotton            | 100               | 50                                | 50                                 | 1               |
| Mercerized cotton | 100               | 50                                | 50                                 | -1              |

### 2.2 Collection of near infrared spectra

The instrument used in this experiment was an Antaris II FT-NIR spectrometer with a spectral range of 4 000~10 000  $\text{cm}^{-1}$  and a resolution of 3.8  $\text{cm}^{-1}$  as shown in Fig. 3. Water content was used as an external disturbance to acquire the dynamic spectra. First, 200 original samples were evenly laid in an oven, then dried at 105  $^{\circ}\text{C}$  and weighed. Each time a sample was taken out, folded into 4 layers and tiled to collected the spectra on the instrument with scanning 32 times. Then the dried samples were tiled on a clean bench, sprayed pure water on both sides of the sample uniformly to make the water content be 20% and collected each sample spectra in the second time. After all the samples were exposed to the air for a short period of time to make the water content drop to about 14%~15%, then collect each sample spectra in the third time. Similarly, spectra of samples with water content declining to 7%~8% were collected in the fourth time. Finally, each sample had 4 NIR spectra with different water contents as shown in Table 2 to make a dynamic spectral matrix.

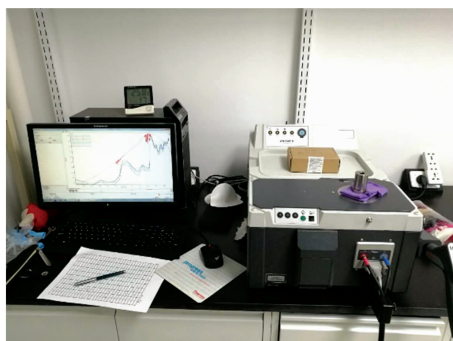


Fig. 3 NIR spectrometer

### 2.3 Data processing

The MATLAB R2014a was used for data processing.

First, the multivariate scatter correction and first-order S-G derivation were used to preprocess spectral data. Two-dimensional correlation spectra for each sample were calculated and classified using three SVM classifiers as shown in Table 3. The correct and incorrect probabilities of each sample were calculated with the SVM Matlab toolkit for each classifier. The entropy and weight of each classifier were calculated with (16)—(18). The final classification result of each samples was calculated with (19). The classification accuracy was calculated with (20)<sup>[16]</sup>.

$$\text{Accuracy} = \frac{\text{Correct classification number}}{\text{Total number of samples}} \times 100\% \quad (20)$$

Table 2 The water contents of samples for the dynamic spectra

| The number of acquisitions | The total sample weight/g | Sample water content/% |
|----------------------------|---------------------------|------------------------|
| 1                          | 6 432.49                  | 20.20                  |
| 2                          | 7 021.13                  | 14.52                  |
| 3                          | 5 678.51                  | 7.77                   |
| 4                          | 5 272.53                  | 0                      |

Table 3 The classification features for each classifier

| Classifier number | Classification features   |
|-------------------|---|
| 1                 | The normalized mean and variance of the numerical values of the wavenumbers region [4 471; 5 537, 5 932; 6 696] in the two-dimensional correlation matrix |
| 2                 | The minimum value array of three sub-matrices with wavenumbers region [4 385; 5 156, 4 000; 4 771]  |
| 3                 | The projection vector of wavenumbers region [6 313; 7 081, 4 385; 5 156]  |

### 3 Results and discussion

#### 3.1 Comparison of near infrared one-dimension and two-dimension correlation spectra

One-dimension spectra of the pure cotton and mercerized cotton samples with water content of 0 are shown in Fig. 4. From Fig. 4, it can be seen that there are serious overlap of one-dimension spectra of pure cotton and mercerized cotton samples. It is usually difficult to classify the pure cotton and mercerized cotton samples with one-dimension spectra as shown in Fig. 4. The average spectra of two-dimension correlation spectra of 100 pure cotton textiles and 100 mercerized cotton textiles are shown in Fig. 5. From Fig. 5, it can be seen that there is obvious difference between the two-dimension correlation spectra of pure cotton textiles and mercerized cotton textiles in the red region. Compared with the one-dimensional spectra, the two-dimensional correlation spectra can enhance the apparent spectral resolution.

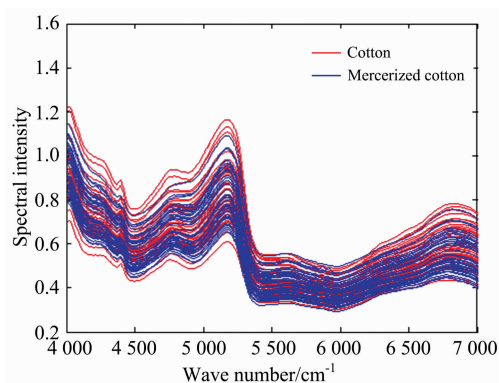


Fig. 4 First-order S-G derivative spectra of pure cotton and mercerized cotton samples

#### 3.2 Classification based on PCA and SVM with one-dimension spectra

One-dimension spectra with water content of 0 for all samples were taken. PCA was used to compress the one-dimension spectra and extract the classification features of each sample. Since the sum of the contribution of the first two principal components reached 99%, all the samples were compressed to three dimensions and then classified with SVM. The sample space distribution after classification feature extraction with PCA was shown in Fig. 6. 76 samples of total 100 test samples was correctly classified and the final classification accuracy was 76%. Because there are serious overlap of one-dimension spectra of pure cotton and mercerized cotton samples as shown in Fig. 4, it is not easy to classify the pure cotton and mercerized cotton samples using PCA to extract classification features from a one-dimension spectra.

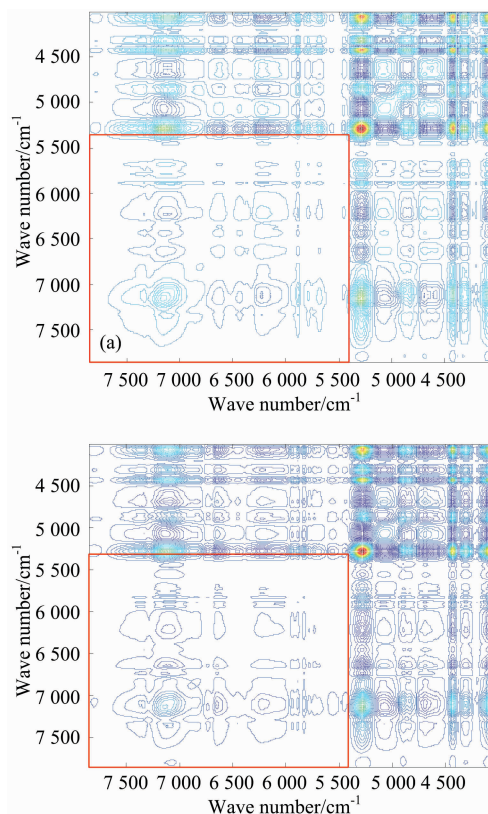


Fig. 5 The averaged two-dimension correlation Synchronous spectra of pure cotton and mercerized cotton samples  
(a): Two-dimension correlation spectra of cotton; (b): Two-dimension correlation spectra of mercerized cotton

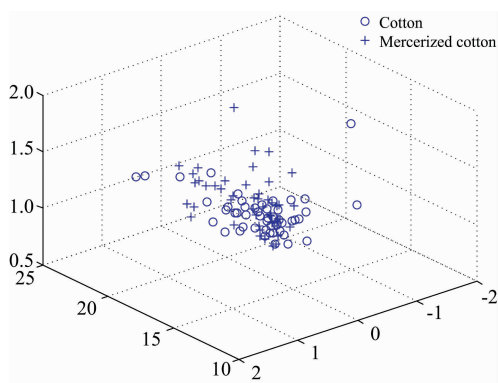


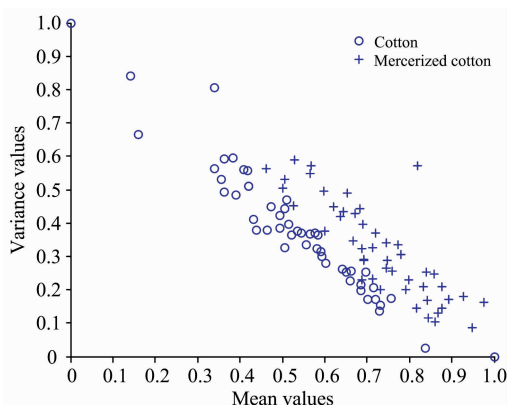
Fig. 6 The sample space distribution after classification feature extraction with PCA

#### 3.3 Classification based on two-dimension correlation spectra

Compared with one-dimensional spectra, two-dimensional correlation spectra have higher resolution as shown in Fig. 5. In this work, three kinds of classification features were extracted from the synchronous two-dimension correlation spectra. Support Vector Machine (SVM) was combined with different kind of the classification features to construct different classifiers. In this work, an information fusion method was proposed to make the multi-classifier decision.

### 3.3.1 Classification with classifier 1

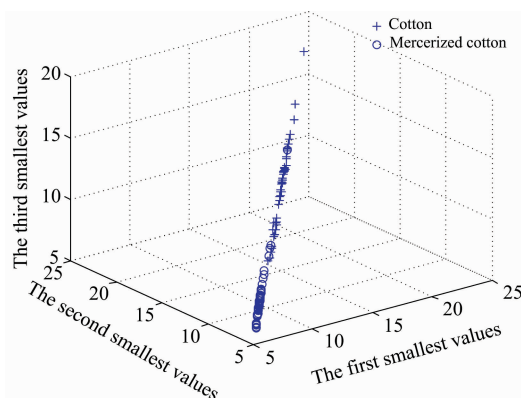
From Fig. 5, it can be seen that there are significant differences between the two types of samples in the red region indicating that the mean and variance of the two are also different in this region. In this work, the mean and variance of the two-dimension correlation spectra were chosen as the classification features in the red region. The whole two-dimension spectra were traversed using the moving window technique. Each time the data point number was  $400 \times 400$ , and the mean and variance of the region were extracted as the classification features. It was found that the wavenumber coordinates  $[4\ 471; 5\ 537, 5\ 932; 6\ 696]$  had the most significant differences. The SVM using the RBF kernel function with a penalty coefficient 0.8 and a gamma value 1 was chosen as the classifier 1 to classify the spectra of wavenumber coordinates  $[4\ 471; 5\ 537, 5\ 932; 6\ 696]$ . The mean and variance for all samples in the wavenumber  $[4\ 471; 5\ 537, 5\ 932; 6\ 696]$  region were plotted in Fig. 7, where the  $x$ -axis is the mean of all samples and the  $y$ -axis is the variance of all samples. 100 validation samples, including 50 cottons and 50 mercerized cottons fabrics, were classified correctly by the classifier 1 and the classification accuracy was 88%.



**Fig. 7** The plot of mean values vs variances of all samples in the region with wavenumber coordinates  $[4\ 471; 5\ 537, 5\ 932; 6\ 696]$

### 3.3.2 Classification with classifier 2

Similar to classifier 1, using the moving window analysis technique, the whole two-dimension correlation spectra were traversed. A  $400 \times 400$  size submatrix of the two-dimension correlation matrix of each sample were taken and the minimum three numbers were extracted from the sub-matrix region as the classification features for classifier 2. It was found that the wavenumber coordinates  $[4\ 385; 5\ 156, 4\ 000; 4\ 771]$  had the most significant differences. All three samples in the region  $[4\ 385; 5\ 156, 4\ 000; 4\ 771]$  are shown in Fig. 8, where  $x$ -axis,  $y$ -axis and  $z$ -axis are the smallest three digits respectively.

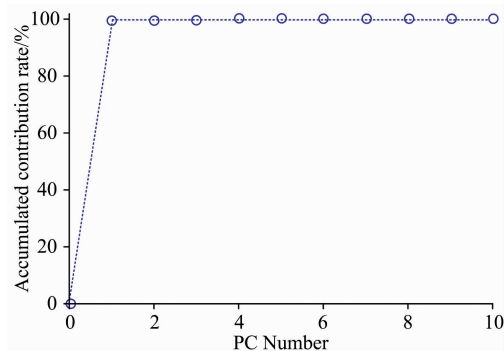


**Fig. 8** The plot of the three smallest values of all the samples in the region of coordinates  $[4\ 385; 5\ 156, 4\ 000; 4\ 771]$

The SVM using the RBF kernel function with a penalty coefficient 1 and a gamma value 1 was chosen as the classifier 2 to classify the three smallest values of all the samples in the region of coordinates  $[4\ 385; 5\ 156, 4\ 000; 4\ 771]$ . 100 validation samples, including 50 cottons and 50 mercerized cottons fabrics were classified with the classifier 2 and the classification accuracy was 90%.

### 3.3.3 Classification with classifier 3

Similar to the classifier 1, using the moving window analysis technique, the whole two-dimension correlation spectra were traversed. A  $400 \times 400$  size submatrix of the two-dimension correlation matrix of each sample were taken. PCA was used to extract the classification features for classifier 3 from the wavenumber coordinates  $[6\ 313; 7\ 081, 4\ 385; 5\ 156]$ . The number of principal components and the corresponding cumulative contribution rate are shown in Fig. 9. The RBF kernel function with a penalty coefficient 1 and a gamma value 1.5 was chosen as classifier 3. 100 validation samples, including 50 cottons and 50 mercerized cottons fabrics, were classified with the classifier 3 and the classification accuracy was 88%.



**Fig. 9** PC Number and accumulated contribution rate in the area of coordinates  $[6\ 313; 7\ 081, 4\ 385; 5\ 156]$

### 3.3.4 Multi-classifier fusion based on information entropy decision

The entropy and the weight of each classifier were calculated with (17) and its reciprocal respectively. According to the classification result of each sample in different classifiers and the proportion of the classifier in the multi-classifier

fusionbased on information entropy decision, the final classification result of each sample was obtained according to (19). The classification results for all samples in the test set with each classifier and with the multi-classifier fusion based on information entropy decision are shown in Table 4.

**Table 4 The classification result of each classifier and the fusion**

|                         | The number of test set | Entropy of classifiers | The classifier weights of Information fusion | Correct classification numbers | Accuracy /% |
|-------------------------|------------------------|------------------------|--|--------------------------------|-------------|
| Classifier 1            | 100                    | 0.55                   | 1.81   | 88                             | 88          |
| Classifier 2            | 100                    | 0.63                   | 1.58   | 90                             | 90          |
| Classifier 3            | 100                    | 0.91                   | 1.11   | 88                             | 88          |
| Multi-classifier fusion | 100                    |                        |  | 92                             | 92          |

## 4 Conclusions

In this work, a new method is proposed for rapid identification of pure cotton and mercerized cotton products with two-dimensional correlation spectra analysis. Two-dimension correlation spectra were calculated in which humidity is the external disturbance. Three kinds of classification features were extracted from the synchronous two-dimension correlation spectra. Then Support Vector Machine (SVM) was combined with different kind of classification features to construct

diverse classifiers respectively. Finally, to achieve the fusion of multiple classifier results, an information entropy-based information fusion method was used. The comparative experiment results have verified that, using information entropy fusion method, two-dimension correlation spectra have higher classification resolution than one-dimension spectra. What's more, multiple weak classifiers can form a strong classifier and can reduce the classification error caused by the unilaterality of single classifier. The proposed method in this work provided a new way for rapid identification of pure cotton and mercerized cotton products.

## References

- [1] Liu Fang, Zhang Baokun. *Journal of Donghua University, Natural Sciences*, 2008, 10(5): 171.
- [2] Jia Shiqiang, Lao Haochuan, An Dong, et al. *Spectroscopy and Spectral Analysis*, 2017, 37(5): 1634.
- [3] Zhu Xiangrong, Li Gaoyang. *Journal of Chinese Institute of Food Science and Technology*, 2016, 16(12): 206.
- [4] Wei Yujuan, Li Lin. *China Dairy Industry*, 2016, 44(10): 48.
- [5] D'Aspremont A, El Ghaoui L, Jordan M I, et al. *International Conference on Neural Information Processing Systems*. MIT Press, 2004. 41.
- [6] Wang Yasheng, Yang Meng, Luo Zhiyuan, et al. *Spectroscopy and Spectral Analysis*, 2016, 36(6): 1685.
- [7] Lecun Y, Bengio Y. *Deep Learning Nature*, 2015, 521(7553): 436.
- [8] Noda I. *Applied Spectroscopy*, 1990, 44(4): 550.
- [9] Chen Bin, Zhang Yuying, Lu Daoli, et al. *Journal of the Chinese Cereals and Oils Association*, 2016, (6): 153.
- [10] Noda I. *Applied Spectroscopy*, 1988, 42(2): 203.
- [11] Sonoyam M, Shoda K. *Applied Spectroscopy*, 1996, 50(3): 377.
- [12] Feng Changliao, Shi Zhong. *Journal of Computer Research and Development*, 2016, 50(9): 1971.
- [13] Ding Meijun, Liu Guiquan, Li Hui. *Pattern Recognition and Artificial Intelligence*, 2015, (11): 1003.
- [14] Shannon C. *Bell System Technical Journal*, 1948, 27(1-2): 623.
- [15] Howard Barnum, Jonathan Barrett, Lisa Orloff Clark, et al. *New Journal of Physics*, 2010, (3): 1.
- [16] Cheng Shuxi, Kong Wenwen, Zhang Chu, et al. *Spectroscopy and Spectral Analysis*, 2014, 34(9): 2519.

## 基于二维相关光谱分析的纯棉与丝光棉制品鉴别分析

曹 凯<sup>1</sup>, 赵 众<sup>1\*</sup>, 袁洪福<sup>2</sup>, 李 彬<sup>2</sup>

1. 北京化工大学信息科学与技术学院, 北京 100029

2. 北京化工大学材料科学与技术学院, 北京 100029

**摘 要** 纯棉与丝光棉制品是日常生活中常用的两种纤维制品,但是由于二者在物理结构和化学结构上非常相似,以至于使用一些简单的方法难以准确识别一部分纯棉与丝光棉制品。提出一种使用含水量作为扰动的二维相关光谱结合机器学习方法来对二者进行鉴别的新方法。共使用从专业机构获得的 200 个标准样本设计实验对新方法进行验证,其中包括 100 个纯棉样本与 100 个丝光棉样本。对每一个样本,使用含水量作为扰动,分 4 次改变样本含水量并采集该含水量下样本的一维光谱,其中 4 次的水含量分别为 20.20%, 14.52%, 7.77% 与 0%。根据四条不同的一维构造每一个样本的动态光谱,再通过二维相关算法来计算其同步二维相关光谱,从该同步二维相关光谱中使用移动窗口技术提取三组不同的分类特征,每组特征分别对应一个设计好的支持向量机(SVM)分类器。之后本文提出一种基于信息熵的多分类器融合方法,根据权值不同,将三个分类器融合为一个具有更优效果的强分类器。为了验证方法的准确性与有效性,设计了严谨的实验对方法进行验证。实验首先按照传统的从一维光谱中提取特征的方法对纯棉与丝光棉样本进行鉴别,使用两种样本各 50 个来进行分类模型建立,剩余的进行模型验证,分类效果最高只有 76%。但是基于从二维相关光谱中提取的三组特征设计的三个支持向量机(SVM)分类器的准确率分别可以达到 88%, 90%, 88%, 最后根据提出的基于信息熵的多分类器信息融合方法将三个分类器进行融合同一可以得到 92% 的分类准确率,比三个基础分类器准确率都有提升。与从一维光谱中提取特征并设计分类器进行分别鉴别相比,从二维相关光谱中提取特征设计多个分类器并使用基于信息熵的多分类器信息融合方法进行鉴别具有更高的分类准确率。二维相关光谱将光谱信息扩展到更高的维度,将一维光谱中隐藏的折叠峰进行展开,因此具有更高的分类准确率。提出的方法是一种快速准确鉴别纯棉与丝光棉制品的新方法。

**关键词** 二维相关光谱; 纯棉与丝光棉; 支持向量机; 信息熵; 信息融合

(收稿日期: 2018-04-07, 修订日期: 2018-09-12)

\* 通讯联系人