Scientific Research Publishing

Rapid, Non-Destructive, Textile Classification Using SIMCA on Diffuse Near-Infrared Reflectance Spectra

Christopher B. Davis[1], Kenneth W. Busch[1*], Dennis H. Rabbe[1], Marianna A. Busch[1], Judith R. Lusk[2]

[1]Department of Chemistry and Biochemistry, Baylor University, Waco, USA
[2]Department of Family and Consumer Sciences, Baylor University, Waco, USA
Email: [*]Kenneth_Busch@baylor.edu

Abstract

Soft independent modeling of class analogy (SIMCA) was successful in classifying a large library of 758 commercially available, non-blended samples of acetate, cotton, polyester, rayon, silk and wool 89% - 98% of the time at the 95% confidence level ($p = 0.05$ significance level). In the present study, cotton and silk had a 62% and 24% chance, respectively, of being classified with their own group and also with rayon. SIMCA correctly identified a counterfeit "silk" sample as polyester. When coupled with diffuse NIR reflectance spectroscopy and a large sample library, SIMCA shows considerable promise as a quick, non-destructive, multivariate method for fiber identification. A major advantage is simplicity. No sample pretreatment of any kind was required, and no adjustments were made for fiber origin, manufacturing process residues, topical finishes, weave pattern, or dye content. Increasing the sample library should make the models more robust and improve identification rates over those reported in this paper.

Keywords

1. Introduction

Textile screening has become increasingly important in the global economy to assure importers of fabric identity and to assess import fees and taxes. Four main methods of routine fabric characterization are currently used to

[*]Corresponding author.

determine fiber identity of a textile: visual inspection, microscopy, flame testing, and solubility testing [1] [2]. Visual inspection is subjective; the last two methods are destructive; and even microscopic examination is often not conclusive. As a result, it would be extremely beneficial to have another characterization method that is rapid, nondestructive and applicable to different fiber types regardless of differences in dyes, finishes, weaves, manufacturing process methods, or fiber origin.

To that end, the use of near-infrared (NIR) spectroscopy for textile screening has been studied by a number of workers. Because NIR spectra are generally broad and featureless, classification of samples into different categories by means of NIR spectra requires some form of statistical analysis. For example, Howell and coworkers [3] used Mahalanobis distances at four or six pre-selected wavelengths in diffuse NIR reflectance spectra from 1100 - 2500 nm to identify fibers of acrylic, rayon, nylon, and polyester, although dark fabric dyes were found to severely impact the results. Jasper and coworkers combined NIR absorbance spectra from 1000 - 2500 nm with neural networks to identify 17 categories of fibers from a 390-sample library, most with confidence levels of 95% or better [4]. Cotton, however, showed a 51% chance of being misidentified, with a 32% chance of being classified as linen [4]. Pattern recognition has also been used to classify fabric weave patterns [5] [6], identify dyes [7], and grade yarns [8]. Fourier transformation of diffuse NIR reflectance spectra was reported to aid pattern recognition in distinguishing a limited number of pure fabrics and cotton-polyester blends [9].

The purpose of this study was to investigate the feasibility of classifying commercial textile samples as to basic fiber type using diffuse reflectance NIR spectroscopy without any form of sample pretreatment and without any consideration of differences in textile weave, finish, dye, manufacturing process, or point of origin.

Statistical sample size is an important consideration in any statistical study, and one limitation of many of the previous studies cited above is that of limited population size. A key component of this study, therefore, is the development of an extensive textile library that more accurately reflects the wide assortment of six types of textiles that are among the most commonly encountered in the marketplace. This library is, to the best of our knowledge, one of the largest ever reported for a textile pattern recognition study. In the present study, we did not consider fabric blends nor were we interested in comparing different chemometric algorithms.

While a number of multivariate classification methods with different merits exist, including linear discriminant analysis (LDA) [10], linear discriminant analysis by regression (PLS-DA) [11], Mahalanobis distances [12] [13], *k*-nearest neighbors (*k*-NN) [14], and artificial neural networks [10] [15]-[17], soft-independent-modeling-of-class-analogy (SIMCA) was chosen for classification in this study [18].

SIMCA was chosen for this study for three reasons: 1) SIMCA is a well established classification method; 2) SIMCA is based on principal component analysis, which is well understood; and 3) the principal component models for different classes can be updated independently without having to redo the entire model. In addition, because SIMCA is a soft-modeling technique, samples can be classified into more than one category simultaneously (*i.e.*, it is possible, for example, for a hypothetical sample to be salty and sweet simultaneously). With these points in mind, it is worthwhile to discuss briefly how SIMCA goes about classification.

2. Theory

Consider a data matrix X that consists of the spectra of n different samples where each spectrum consists of intensity values taken at p wavelengths. Let us further suppose that all of the samples in X belong to a particular class or group. So X may be the spectra of 100 cotton samples, for example. The goal of principal component analysis is to break this data matrix down into a series of variance-scaled orthogonal vectors known as principal components so that

$$X = TL^{\mathrm{T}} + E \tag{1}$$

$$X = t_1 l_1^{\mathrm{T}} + t_2 l_2^{\mathrm{T}} + t_3 l_3^{\mathrm{T}} + \cdots + t_A l_A^{\mathrm{T}} + E. \tag{2}$$

In Equation (1), X is the original $n(\text{rows}) \times p(\text{columns})$ data matrix, T is an $n \times A$ matrix known as the scores matrix where A is the number of principal components in the model, L is a $p \times A$ matrix known as the loadings matrix, the superscript T indicates the transpose of L, and E is an $n \times p$ matrix that respresents the residual variance that is not explained by the model. In Equation (2), $t_i l_i^{\mathrm{T}}$ are the orthogonal principal components that make up the model, where $t_1 l_1^{\mathrm{T}}$, the first principal component, explains more of the variance in X than $t_2 l_2^{\mathrm{T}}$, and so on.

Now consider an unknown spectrum that we want to test to see if it belongs to the same class of samples that were used in X. The unknown spectrum x_u will be a $1 \times p$ vector. To transform this raw vector into principal component space, we need to post-multiply it by the $p \times A$ loadings matrix L from the PCA model made from X. This converts the original spectral data into an estimate of a $1 \times A$ scores vector for the unknown sample, $\hat{t}_u = x_u L$, where $\hat{t}$ is the estimate of the scores vector. Finally, we can post-multiply the estimate of the scores vector by the transpose of the loadings matrix. If we subtract the result from the original unknown spectrum x_u $(1 \times p)$, we can estimate the residual variance vector e_u $(1 \times p)$:

$$e_u = x_u - \hat{t}_u L^{\mathrm{T}}. \tag{3}$$

Equation (3) allows us to calculate the residual variance for the unknown sample that we are testing for class membership with the following equation:

$$s_u^2 = \sum_{j=1}^{p} \frac{e_{uj}^2}{p - A}. \tag{4}$$

The total residual variance for all the n samples in the PCA model will be

$$s_{\text{total}}^2 = \left(\frac{1}{n - A - 1} \right) \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \frac{e_{uj}^2}{p - A} \right)_i. \tag{5}$$

These two variances allow us to perform a Fischer's F-test, which forms the ultimate basis of the classification method,

$$F = \frac{s_u^2}{s_{\text{total}}^2}. \tag{6}$$

If, on the basis of the F-test, s_u^2 and s_{total}^2 are found to be the same statistically, the unknown sample is considered to belong to the group or category that made up the data matrix X.

If the samples of interest consist of q classes or categories, data matrices must be acquired for each class, X_q, and PCA models made. A given sample is then tested for membership in each class as described above.

3. Experimental

3.1. Textile Samples

Swatches of known commercially-available textiles were obtained from the Baylor University Department of Family and Consumer Science. Specimens, approximately one-inch square, were cut from these swatches and placed in acid-free polypropylene sleeves that were labeled with fiber type and a number indicating position in the data matrix. The identity of the textile was provided by the original supplier. NIR spectral analysis was conducted without any other form of sample pretreatment.

Because the textiles used in this study were commercially available and not pretreated prior to analysis, they could have a variety of fiber finishes and dyes. Therefore, the sample population was made as broad as possible so that models generated by SIMCA could account for as many variations within a fiber type as possible. A total of 758 textile samples was cataloged and their diffuse NIR reflectance spectra collected for training and testing. Table 1 shows the textile library used for this study with the number of samples in each fiber category.

3.2. Data Acquisition and Analysis

NIR analysis was carried out using a dual-channel spectrometer assembled and tested in our laboratory [19]. To acquire diffuse NIR reflectance spectra, the sample beam was directed into an eight-inch diameter barium sulfate ($BaSO_4$) coated integrating sphere (Model 70451, Oriel) that had a lead sulfide (PbS) detector (Model 7202, Oriel) located in the detector port and mounted 45° from the sample window. The integrating sphere was equipped with a baffle and a light trap to reduce specular reflection.

Textile samples were placed in the sample holder at the sample window of the integrating sphere. Because the sensitivity of PbS detectors drops off at wavelength limits of approximately 1000 and 2500 nm, the wavelength

Table 1. Textile library.

Fabric type	No. of samples
Acetate	61
Cotton	274
Polyester	109
Rayon	77
Silk	45
Wool	192

region selected for these studies ranged from 1334 to 1906 nm. Wavelengths outside this region produced signals with significantly reduced signal-to-noise ratios.

Diffuse NIR reflectance scans were recorded with a sampling interval of 2.0 nm, and measurements were reported as $\log_{10}(1/R)$, where R is equal to the diffuse reflectance signal. Data were written into ASCII tables by the LabView™ software (version 5.1, National Instruments, Austin, TX) that controls the instrument and converted to Microsoft® Excel spreadsheets as required for chemometric analysis. The spectral data, contained in the Excel spreadsheets, were imported into the Unscrambler© software package (versions 9.1, 9.6, and 9.7, Camo, Inc., Woodbridge, NJ). The data were then transposed to list the wavelengths as the x-variables and the reflectance as the y-variables.

Except where stated otherwise, raw spectral data was subjected to Savitzky-Golay smoothing [20], using a zero-degree polynomial and 11 averaging points, 5 on each side of a given wavelength. The multivariate principal component analysis for each textile category was performed on mean-centered data at the 95% confidence level (p = 0.05 significance level). A mean-centered plot, calculated by the Unscrambler© statistical package, was obtained by first averaging the individual spectra on a wavelength-by-wavelength basis (*i.e.*, adding the $\log_{10}(1/R)$ values of each spectrum on a wavelength-by-wavelength basis and dividing each sum by the number of samples). This average spectrum was then subtracted from each individual spectrum on a wavelength-by-wavelength basis to give mean-centered spectra.

4. Results and Discussion

Diffuse reflectance NIR spectra were obtained on 758 textile samples representing acetate, cotton, polyester, rayon, silk, and wool, using a Savitzky-Golay algorithm for smoothing. As an example, **Figure 1** shows the smoothed diffuse NIR reflectance spectra for 274 cotton textiles, plotted as $\log_{10}(1/R)$ over the wavelength range 1334 - 1906 nm.

4.1. PCA Modeling

Individual principal component analyses were conducted for each fabric type using the smoothed reflectance spectra collected for each group.

As an example, **Figure 2** shows the results of the principal component analysis obtained using the spectra in **Figure 1** for cotton. The three-dimensional scores plot in **Figure 2(a)** shows that the samples are segregated into a relatively distinct group with approximately 100% of the relationship being explained by the first and second principal components. **Figure 2(b)** is a plot of the residual x-variance as a function of leverage. This graph provides a measure of how much the residual x-variance of any individual sample varies compared to other samples, and how distant that sample point is from the model center (leverage). As can be seen in **Figure 2(b)**, with a few exceptions, most samples fit relatively well into the cotton category. **Figure 2(c)** is a plot of the regression coefficients versus wavelength for the second PC. This plot shows that some wavelengths contribute positively to the model while others contribute negatively. The curve falls gradually, but aside from a small, broad feature between 1600 and 1700 nm, there are few distinguishing characteristics. **Figure 2(d)** is a plot of residual variance as a function of the number of principal components. Two principal components appear to explain approximately 100% of the spectral.
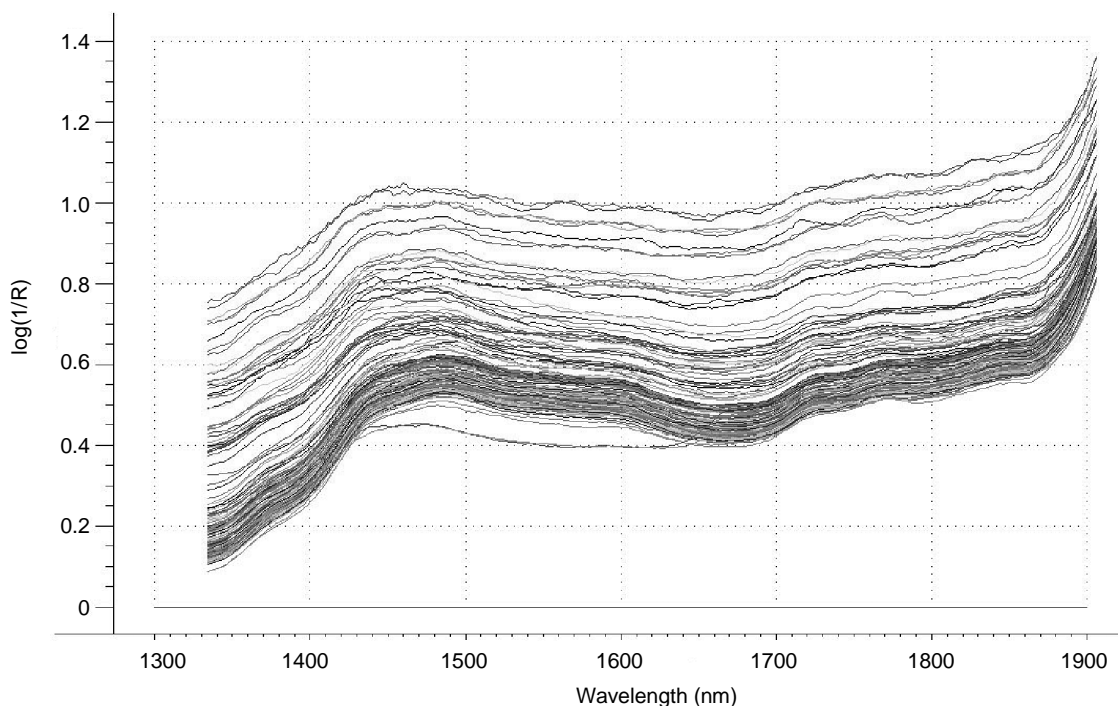
Figure 1. Savitzky-Golay smoothed diffuse near-infrared reflectance spectra of 274 cotton textiles plotted as $\log_{10}(1/R)$ over the wavelength range 1334 - 1906 nm.
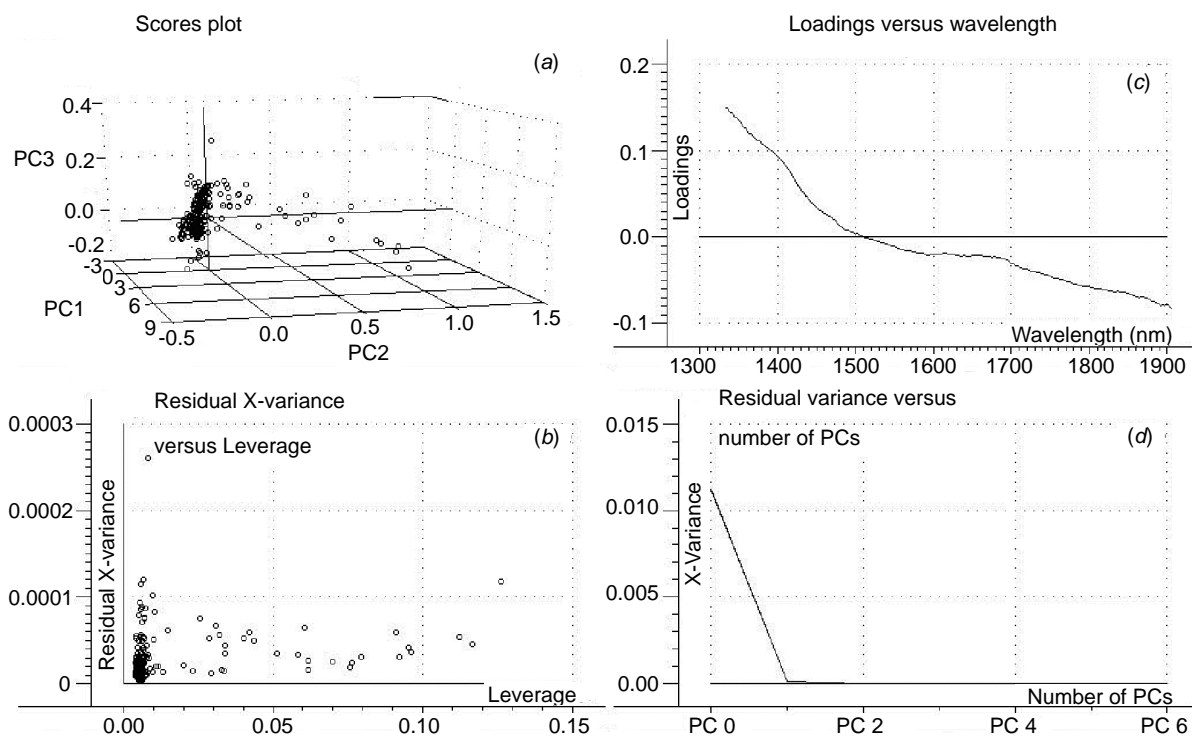

Figure 2. Summary of the principal component analysis of Savitzky-Golay smoothed diffuse near-infrared spectra plotted as $\log_{10}(1/R)$ of 274 cotton samples over the wavelength range 1334 - 1906 nm: (a) Three-dimensional scores plot resulting from three principal components; (b) Residual *x*-variance as a function of leverage; (c) *x*-loadings for the second principal component as a function of wavelength; (d) Residual variance as a function of the number of principal components.

Using cross-validation, the Unscrambler© program automatically defaults to a suggested number of principal components that the program deems necessary to explain the variation in the spectral data, which in this case was two (**Figure 2(d)**). But, close examination of **Figure 2(d)** shows that the residual variance continues to drop slowly beyond the second principal component. Several different numbers of PCs were tried, and models were computed using both the default number and values up to six [21]-[24]. For cotton as well as the other fabric types studied, six PCs did not appear to over-determine the data and were used in the SIMCA classification studies.

4.2. SIMCA Classification

SIMCA classification was performed on the fiber classes of acetate, cotton, polyester, rayon, silk, and wool using the principal component models for each class of fabric. The methodology involved removing a randomly selected number of samples (validation samples in **Table 2**) in the full set for each textile category (**Table 1**) and classifying those with a model built from all remaining samples in the same population (calibration samples in **Table 2**). Validation was performed ten times with a new set of validation samples for each of the six categories, and the percentage of correct classifications (positive result) was calculated at the 95% confidence level. The results are shown in **Table 2** and indicate that SIMCA correctly identified the textiles in 89% - 98% of the cases.

SIMCA was also used to check for misclassifications. Each one of the 758 textile samples was individually tested for membership in each of the six textile categories. Tests were performed using PCA models constructed from all members of each textile group (total samples in **Table 3**). Four results were possible (**Figure 3**): (A) correctly classified in only one group; (B) classified in more than one group, with one group correct; (C) classified in one or more groups, with no group correct; and (D) not classified in any group. The best outcome is A. Outcome B is still a positive result, but the sample would require further testing. The results (A, B, C or D) are indicated in **Table 3** for each textile. The percentage is relative to the total number of samples in the textile class.

As shown in **Table 3**, acetate, polyester, rayon and wool models performed well, with few misclassifications as other fiber types (columns B and C). Of these four, wool had the most misclassifications with a 5% chance of

Table 2. SIMCA validation results.

	Acetate	Cotton	Polyester	Rayon	Silk	Wool
No. of calibration samples	51	254	89	67	40	172
No. of validation samples	10	20	20	10	5	20
% Correct[a]	91	89	95	92	96	98

[a]Average of ten, separate validation sets randomly selected and analyzed in each category using the indicated number of samples and 6 principal components.

Table 3. SIMCA classification outcomes.

Outcome[a] (see **Figure 3**)	A	B	C	D	Total samples
Acetate samples[b]	57 93%	1 2%	0	3 5%	61
Cotton samples[c]	79 29%	171 62%	4 1%	20 7%	274
Polyester samples	101 93%	0	0	8 7%	109
Rayon samples[d]	70 91%	1 1%	0	6 8%	77
Silk samples[e]	31 69%	11 24%	1 2%	2 4%	45
Wool samples[f]	182 95%	5 3%	3 2%	2 1%	192

[a]Six principal components. [b]Also classified as polyester, column B. [c]Also classified as rayon, column B; misclassified as rayon, column C. [d]Also classified as cotton, column B. [e]Also classified as rayon, column B; misclassified as rayon, column C. [f]Also classified as silk, column B; misclassified as silk, column C.
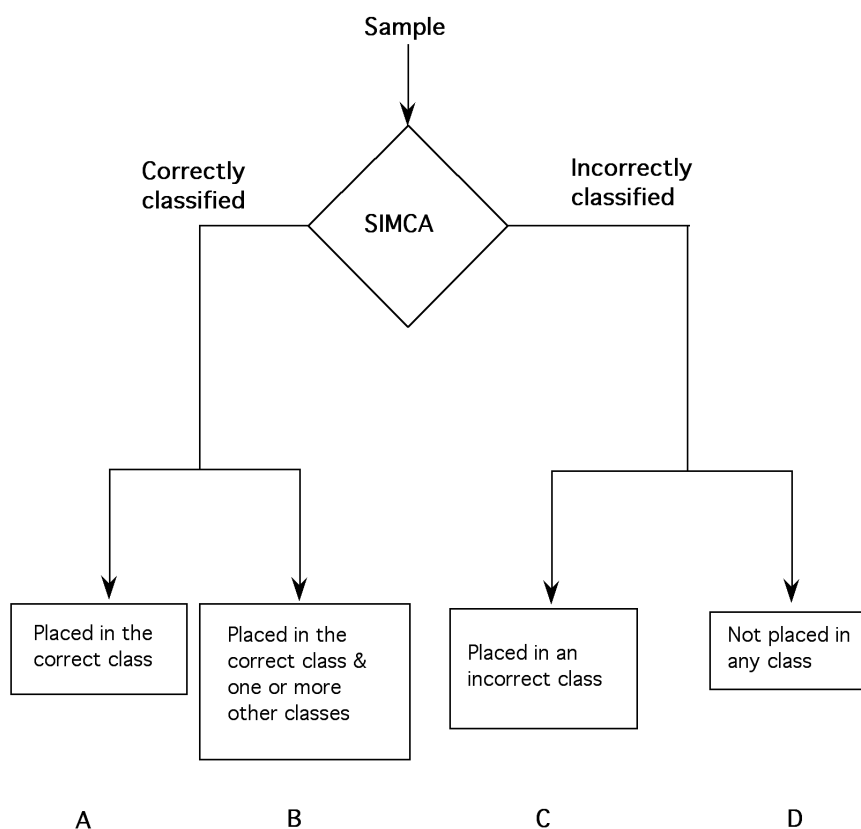
Figure 3. Possible SIMCA outcomes.

being identified as both silk and wool (3%) or just silk (2%). Both wool and silk are protein based and differ primarily in the relative amounts of amino acids present and the type of hydrogen-bonding holding the protein chains together, making classification difficult.

Cotton samples were correctly classified as cotton in the large majority of cases, but were frequently marked as both rayon and cotton (62%). This result can be anticipated from the similarity in the chemical structures of the two fibers (cellulosic). Silk samples were also correctly classified as silk in the large majority of cases, but were frequently marked as both silk and rayon (24%). The diffuse NIR reflectance spectra of both fibers are broad with few differentiating features. In cases where chemical structures and spectra are very similar, classification should be improved by adding more samples to the PCA models.

4.3. Identification of a Counterfeit Silk

Silk is an expensive luxury fabric, and cheaper look-alikes are sometimes substituted for the genuine article. The initial textile library contained one "silk" sample, labeled as Silk-120, that SIMCA flagged as belonging to the polyester category, but not to silk. A visual inspection of the $\log_{10}(1/R)$ spectrum for Silk-120, compared to representative spectra of polyesters also suggested that the sample was counterfeit. A burn test confirmed that the fabric was polyester.

5. Conclusions

SIMCA, coupled with diffuse NIR reflectance spectroscopy, shows considerable promise as a quick, simple, non-destructive method for pure fiber identification. No sample pretreatment of any kind was required, and 89% - 98% of the samples of non-blended acetate, cotton, polyester, rayon, silk and wool textiles could be successfully classified during SIMCA validation without regard for origin, manufacturing process residues, topical finishes, weave pattern, or dye content. The method was used to correctly identify a counterfeit silk sample as polyester.

A large sample library, such as the one used in this study, is critical in establishing a representative population with as many variations as possible within a given textile type. Application of the method to areas where large numbers of textiles must be tested, such as the import/export business, should provide an opportunity to increase the sample library significantly. Existing PCA models can easily be expanded to include additional fabric samples in any category, making the model more robust, while entirely new PCA models can easily be added, as necessary, to train the procedure to recognize new textiles that will almost certainly be introduced into the marketplace. Thus, actual application of the method to commercially available fibers should expand and improve identification rates over those reported in this study.

Acknowledgements

References

[1]   Kadolph, S.J. (2007) Textiles. 10th Edition, Prentice Hall, Upper Saddle River, 34-36.

[2]   King, R.R. (1985) Textile Identification, Conservation and Preservation. Noyes Publications, Parkridge, 5-121.

[3]   Howell, H.E. and Davis, J.R. (1991) *Textile Chemist and Colorist*, **23**, 69.

[4]   Jasper, W.J. and Kovacs, E.T. (1994) *Textile Research Journal*, **64**, 444.
       http://dx.doi.org/10.1177/004051759406400803

[5]   Kuo, C.-F.J., Shih, C.-Y. and Lee, J.-Y. (2004) *Textile Research Journal*, **74**, 107.
       http://dx.doi.org/10.1177/004051750407400204

[6]   Jeon, B.S., Bae, J.H. and Suh, M.W. (2003) *Textile Research Journal*, **73**, 645.
       http://dx.doi.org/10.1177/004051750307300714

[7]   Chen, C.-S., Brown, C.W. and Bide, M.J. (1997) *Journal of the Society of Dyers and Colourists*, **113**, 51.
       http://dx.doi.org/10.1111/j.1478-4408.1997.tb01867.x

[8]   Lien, H.-C. and Lee, S. (2002) *Textile Research Journal*, **72**, 320. http://dx.doi.org/10.1177/004051750207200408

[9]   Nitta, M., Yoshimura, N. and Takayanagi, M. (2004) *Bunko Kenkyu*, **4**, 249. http://dx.doi.org/10.5111/bunkou.53.249

[10]   Varmuza, K. and Werther, W. (1996) *Journal of Chemical Information and Computer Science*, **36**, 323.
       http://dx.doi.org/10.1021/ci9501406

[11]   Wold, S. (1989) Computer Applications in Chemical Research and Education. In: Brandt, J. and Ugi, I.K., Eds., Huthig, Heidelberg, 101-128.

[12]   Mark, H.L. and Tunnel, D. (1985) *Analytical Chemistry*, **57**, 1449. http://dx.doi.org/10.1021/ac00284a061

[13]   Howell, H.E. and Davis, J.R. (1991) *Textile Chemist and Colorist*, **23**, 69.

[14]   Kowalski, B.R. and Bender, C.F. (1972) *Journal of the American Chemical Society*, **94**, 5632.
       http://dx.doi.org/10.1021/ja00771a016

[15]   Jasper, W.J. and Kovacs, E.T. (1994) *Textile Research Journal*, **64**, 444.
       http://dx.doi.org/10.1177/004051759406400803

[16]   Shanmugam, N. and Doke, S.S. (2006) *Indian Journal of Fibre and Textile Research*, **31**, 583.

[17]   Vangheluwe, L. (1997) *Textile Research Journal*, **67**, 809.

[18]   Otto, M. (1999) Chemometrics. Wiley-VCH, Weinheim, 169.

[19]   Soyemi, O.O., Rabbe, D., Busch, M.A. and Busch, K.W. (2000) *Spectroscopy*, **54**, 1092.

[20]   Savitzky, A. and Golay, M.J.E. (1964) *Analytical Chemistry*, **36**, 1627. http://dx.doi.org/10.1021/ac60214a047

[21]   Ghosh, S. and Rodgers, J. (2008) NIR Analysis of Textiles. In: *Handbook of Near-Infrared Analysis*, 3*rd Edition*, *Practical Spectroscopy Series*, Vol. 35, CRC Press, Boca Raton, 485-527.

[22]   Esbensen, K.H. (2004) Multivariate Data Analysis: An Introduction to Multivariate Data Analysis and Experimental Design. 5th Edition, Camo Process AS, Woodbridge, 163-168.

[23]   (1984) Encyclopedia of Textiles, Fibers, and Nonwoven Fabrics. In: Grayson, M., Ed., Vol. 1, John Wiley & Sons, New York.

[24]   Davis, C.B. (2007) Ph.D. Dissertation, Baylor University, 157.