

**Summary Sheet****➤ Paper citation:**

M. Eldessouki, M. Hassan, K. Qashqary, and E. Shady, "The Application of Principal Component Analysis to Boost The Performance of The Automated Fabric Fault Detector And Classifier," *FIBERS & TEXTILES in Eastern Europe*, vol. 22, no. 4(106), pp. 51–57, 2014.

**➤ Targeted problem:**

Fabric fault detection is performed *manually* through visual inspection by *human operators* with *fluctuating performance*, on the other hand, the feature datasets obtained from image analysis have *redundant data* that are *inter-correlated*

**➤ Objective(s):**

- Develop computer vision algorithms to analyze the fabrics' acquired images and extract different types of features
- Reduce the dimensionality of the obtained features dataset to remove the redundant data
- Create a smart system that is capable of detecting and classifying fabric's faults once appear in a fabric's image

**➤ Materials scope:**

- Fabrics produced with weaving technology with "plain" fabric structure
- Nine fabric fault categories were produced and tested

**➤ Computation method:**

- Spatial and spectral features were extracted from the acquired fabric images
- Principle component analysis was implemented to reduce the features dimensionality
- Artificial neural networks were implemented for the detecting and classification tasks

**➤ Paper significance:**

- The implemented principal component analysis (PCA) reduced the dimensionality of the input feature dataset without sacrificing the amount of information in the original dataset which enhanced the processing time and increased the classifier's performance

**➤ Software **

The software program developed with this paper utilizes the same GUI for the paper presented in the last chapter. Therefore, no need to present it here again.



## The Application of Principal Component Analysis to Boost The Performance of The Automated Fabric Fault Detector And Classifier

Mohamed Eldessouki<sup>2\*</sup>, Mounir Hassan<sup>1,2</sup>, Khadijah Qashqari<sup>3</sup> and Ebraheem Shady<sup>2</sup>

<sup>1</sup> Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia

<sup>2</sup> Department of Textile Engineering, Faculty of Engineering, Mansoura University, Egypt

<sup>3</sup> Department of Fashion Design, Faculty of Art & Design, King Abdulaziz University, Saudi Arabia

\* Corresponding author: eldesmo@auburn.edu

### Abstract

There is a growing need to replace the visual fabric inspection with automated systems that detect and classify fabric defects. The digital processing of the fabric images utilizes different methods that offer a large set of image features. The correlation between those features lead to problems during the fabric fault classification and reduces the performance of the classifiers. This work extracted a combination of statistical (spatial) and Fourier transform (spectral) features from the fabric images of the most frequent faults. The principal component analysis (PCA) was implemented to reduce the dimensionality of the input feature dataset which achieved a reduction to 36% of the original data size with preserving 99% of the information in the original dataset. The features processed using the PCA were fed to an artificial neural network (ANN) to classify the fault categories and compared to another ANN that worked with the whole feature dataset. The performance of the network that was implemented after the application of the PCA increased to 90% of correct classification rate as compared to 73.3% for the other network.

### 1. Introduction

The early intervention to fix, or remove, fabric faults is one of the mandatory tasks required by all fabric manufacturers. The undetected defects cause many problems downstream in the production line and result in end products with lower quality that result in cost burden and non-profitable products. Automatic fault detection systems offer good alternatives to replace the traditional human fabric inspection with computer vision systems that analyze fabrics in a systematic manner and aiming for consistent performance. The efficiency of these automated systems, however, depends on many parameters and varies according to the quality of the hardware and the analysis algorithm.

There are many research articles in the field of automatic fabric fault detection and classification that can be found in published reviews [1, 2]. Among those stand some common methods for feature extraction and fault classification as shown in Figure 1. The methods of feature extraction vary to include spatial (statistical) features [3], spectral features (fast Fourier

Transform) [4], combination of spatial and spectral features [5-10], as well as other methods that may utilize wavelet transformers [11, 12]. The extracted features are fed to a classification system that was implemented by researchers in different ways. Among these classifiers are the artificial neural networks (ANN) with its different types [13, 14], fuzzy inference systems [15], neuro-fuzzy systems [16, 17], as well as other classification systems [18].

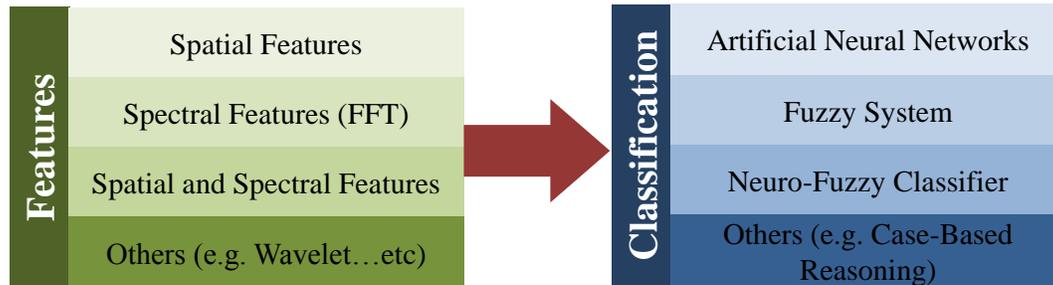


Figure 1. Different methods available in the literature for feature extraction and classification

In this work, some of the high frequently occurring defects that represent the main categories of faults (warp, weft, and areal directions) were studied. A system of image acquisition and enhancement was developed and a number of spatial and spectral features was extracted. The classification was done using pattern recognition artificial neural networks (ANNs) that were fed with the whole features and with the reduced dataset after the application of the principal component analysis (PCA) technique. The performance of the two ANN classifiers was evaluated.

## 2. Fabric faults

There is a large amount of fabric defects that may be caused by different sources and production technologies. Spinning faults, for example, should be mended before fabric production (either by weaving or knitting) otherwise it will lead to fabric faults that may not be fixed at all. Therefore, the scope of this study will be only the faults that occur during the weaving process and focus on plain woven fabrics only. The studied fabric faults might be considered as sever faults and had to be fixed or removed. The defect-free sample is shown in Figure 2-a and the defected samples can be categorized into three main categories; defects in warp direction, defects in weft direction, and areal defects. There are different names that can be found in the literature for the same defect; however the ASTM definition and description for these faults [19] will be considered in this work.

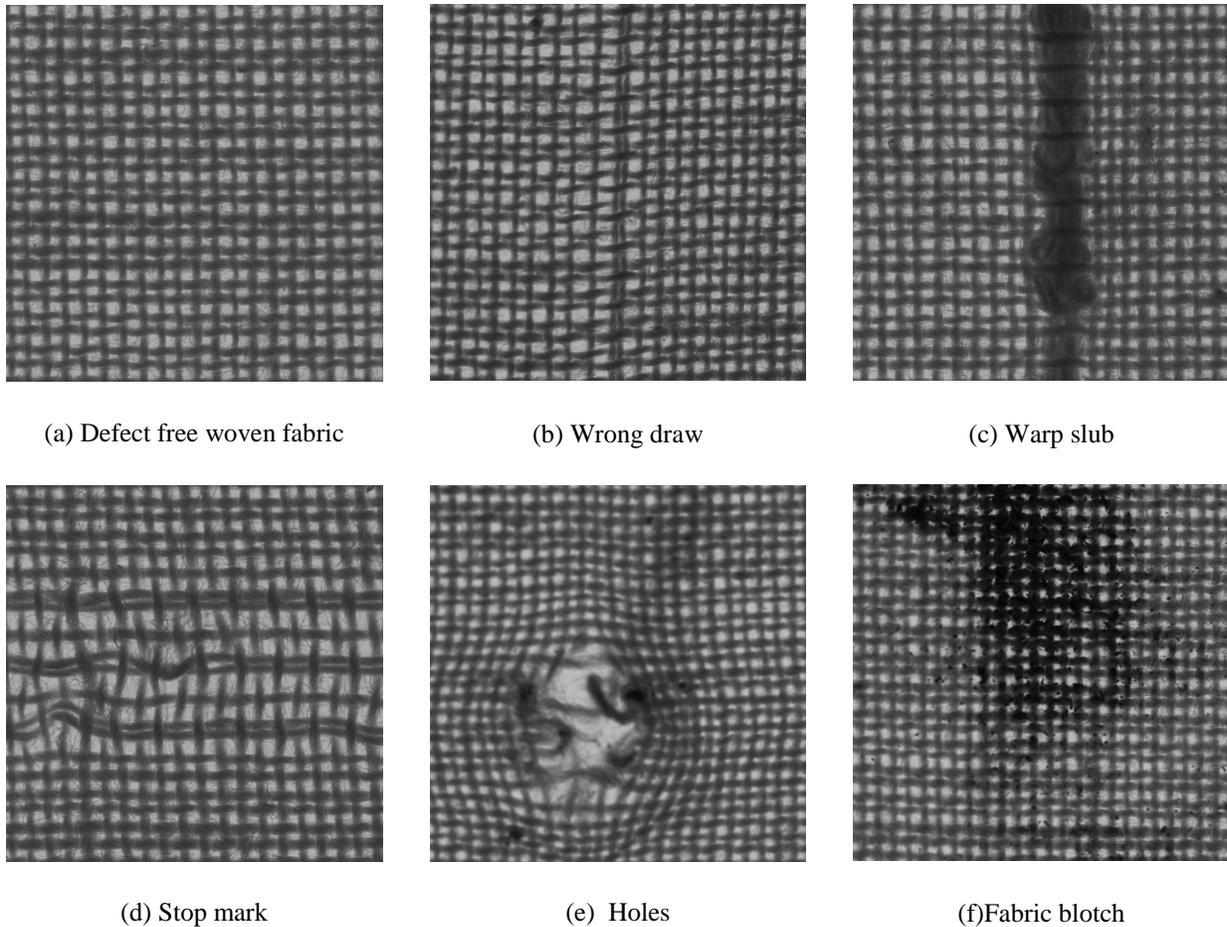


Figure 2. Images of defect-free (a) and some defected (b-f) fabrics

### Warp direction

*Wrong draw*: This fault results when one or more warp ends incorrectly drawn in the harness or the reed. The fault is shown in Figure 2-b and can also be called “wrong draft”, “misdraw”, or “double end”. This fault may be considered as a sever defect because it appears through the whole length of the fabric if not fixed.

*Slub*: This fault shows in the fabric as an abruptly thickened place in a yarn. This fault can occur in warp or weft directions, but it was considered only in warp direction in this study. The other names of the same fault are lump, piecing, slough-off, and slug. All these names can be considered in ASTM standards. This fault may occur due to malfunctioning in warp sensors and it is shown in Figure 2-c.

### Weft direction

*Stop mark:* This appears as a visible change in the density of the weave across the width of the fabric caused by the tension on the warp not being adjusted properly after the loom has been stopped. This fault may be called as “set mark”, or “light beat-up” and is shown in Figure 2-d

### **Area faults**

*Hole:* It is an imperfection in the fabric where one or more yarns are sufficiently damaged to create an aperture. In case of a relatively large hole, it might be called “smash” that is characterized by broken warp ends and floating picks. The smash may be equivalently called a “break-out” and is shown in Figure 2-e.

*Stain:* It is an area of discoloration that penetrates the fabric surface. If this discoloration is caused by grease or oil and the off-colored area appears at any shape, it is called a “blotch” or “oil spot” and an example of this fault is shown in Figure 2-f.

## **3. Methodology**

### **3.1. Samples**

The fabric samples were manufactured on a Sulzer-Ruti weaving machine. The fabric structure is plain weave 1/1 with a yarn count of 20/1 Ne for warp and 14/1 Ne for weft. The densities of warp and weft yarns are 20 and 18 per cm, respectively. The chosen defects were intentionally introduced on the weaving machine based on the knowledge of defects sources.

### **3.2. Image acquisition**

The image acquisition setup is shown in Figure 3 and utilizes a Canon digital camera (model: EOS 450D) with CMOS sensor. The system is installed with a “Remote Live View Shooting” where online monitoring of the pictures and their adjustment can be done on the computer using the EOS Utility software. The camera uses 35mm EF-S lenses and captures images at resolution of 72 dot per inch (dpi). The fabric sample is placed on an inspection table that is equipped with a concentrated LED lights in a box that is placed directly under the shooting area. To remove the noise and interference of the surrounding lights, a suitable shield was installed between the camera and the shooting area.

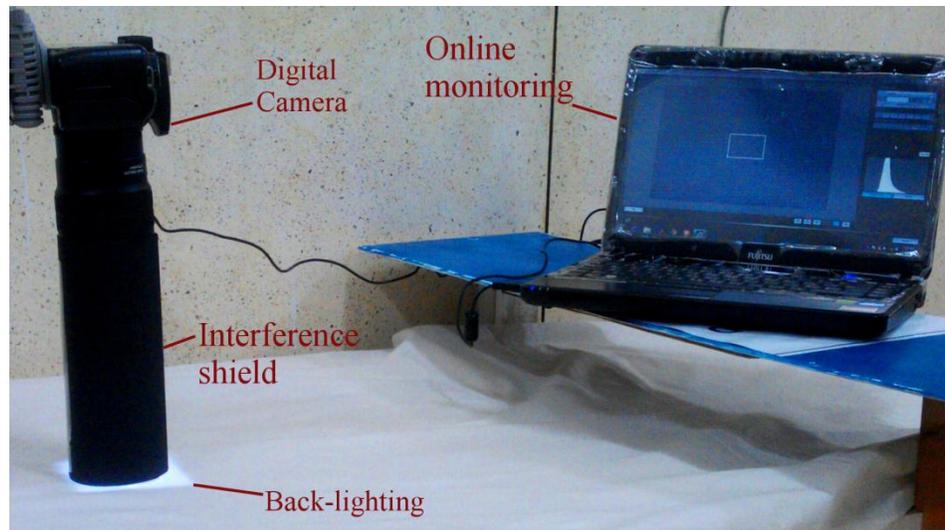


Figure 3. Fabric image acquisition setup

### 3.3. Image enhancement

The developed system applies initial enhancements on the original images to reduce the noise (e.g. hairiness) and improve their contrast. The system uses the contrast-limited adaptive histogram equalization (CLAHE) [20] algorithm to enhance the contrast of the grayscale image by transforming the values. The algorithm can be described briefly as it operates on small regions (windows) in the image. Each window's contrast is enhanced, so that the histogram of the output region approximately matches the specified histogram. The neighboring windows are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image.

### 3.4. Image analysis and feature extraction

If the fabric image can be defined in the spatial domain by the matrix  $P(x,y)$  where:  $x$  is the row number in the image ( $1 \leq x \leq N$ ),  $y$  is the column number ( $1 \leq y \leq M$ ), and  $N$  and  $M$  are the number of rows and columns, respectively. Therefore, the features that can be extracted from this image are summarized below.

#### 3.4.1. Statistical features

A total set of twenty spatial (statistical) features can be extracted from the fabric images. To obtain these features, the sum of individual gray-scale level values in the weft direction (rows) is calculated in the vector  $R(m_j)$  where:

$$R(m_j) = \sum_{i=1}^N P(x_i, y_j)$$

Similarly, the sum of individual gray-scale values in the warp direction (columns) is calculated as:

$$C(n_i) = \sum_{j=1}^M P(x_i, y_j)$$

The first selected feature is the sum of all the gray-scale values in the image which can be calculated as:

$$f_1 = \sum_{i=1}^N C(n_i)$$

The next two features  $f_2$  and  $f_3$  represent the mean of sum of rows ( $f_2$ ) and the mean of sum of columns ( $f_3$ ). The relation for the first feature in the weft direction is:

$$f_2 = \frac{1}{M} \sum_{j=1}^M R(m_j)$$

Similarly for the feature in the warp direction is:

$$f_3 = \frac{1}{N} \sum_{i=1}^N C(n_i)$$

For space constraints, the equations will be listed for the features in weft (rows) direction only and similar relations can be written for the warp (columns) direction by replacing  $R(m_i)$  with  $C(n_i)$ .

The standard deviation of the sum of rows ( $f_4$ ) and for the sum of columns ( $f_5$ ) can be calculated as:

$$f_4 = \sqrt{\frac{1}{M} \sum_{j=1}^M (R(m_j) - \bar{R})^2}, \bar{R} = f_2$$

The features  $f_6$  and  $f_7$  represent the median value of the sum of rows and columns:

$$f_6 = \text{median}(R(m_j))$$

The features  $f_8$  and  $f_{10}$  represent the minimum and maximum values, respectively, for the sum of rows and  $f_9$  and  $f_{11}$  represent the same values for the columns. These features can be written as:

$$f_8 = \min(R(m_j)), \quad f_{10} = \max(R(m_j))$$

The range of the sum of rows and columns was chosen to be, respectively, the features  $f_{12}$  and  $f_{13}$ :

$$f_{12} = \text{range}(R(m_j))$$

The entropy of the image represents the feature  $f_{14}$ :

$$f_{14} = - \sum_{i=1}^N \sum_{j=1}^M P(x_i, y_j) * \log(P(x_i, y_j))$$

The  $k^{\text{th}}$  order moment for the sum of rows can be calculated from the function

$$f_a = \frac{1}{M} \sum_{j=1}^M (R(m_j) - \bar{R})^k$$

The second, third, and fourth order moments can be calculated as features  $f_{15}$ ,  $f_{17}$ , and  $f_{19}$  ( $a=15, 17$  and  $19$ ) for the rows as well as the features  $f_{16}$ ,  $f_{18}$ , and  $f_{20}$  (i.e.  $a=16, 18$  and  $20$ ) for the columns.

### 3.4.2. Spectral features

A fabric image of size  $M \times N$  can be transformed from its spatial domain  $P(x,y)$  to the spectral domain  $\hat{P}(u,v)$  using the discrete Fourier transform (DFT) which can be expressed in mathematical form as:

$$\hat{P}(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P(x, y) e^{-2\pi i(\frac{ux}{M} + \frac{vy}{N})}$$

Where  $x$  and  $y$  are the image spatial variables that correspond to the coordinates inside the image, while  $u$  and  $v$  represent the transformed frequency variables. Once the image is transformed, the power spectrum of the transformed image can be calculated as:

$$PW(u, v) = |\hat{P}(u, v)|^2 = R^2(u, v) + I^2(u, v)$$

Where  $|\hat{P}(u, v)|$  is known as the Fourier spectrum,  $R^2(u, v)$  and  $I^2(u, v)$  are the real and imaginary parts of the transformed image  $\hat{P}(u, v)$ . Shifting the power spectrum is done by implementing the exponential properties of the transformer:

$$\mathfrak{F}[P(x, y)(-1)^{x+y}] = \hat{P}\left(u - \frac{M}{2}, v - \frac{N}{2}\right)$$

Where  $\mathcal{F}[\cdot]$  is the Fourier transform of an argument and this equation tells that the origin of the Fourier transform  $\hat{P}(0,0)$  of the image  $P(x,y)(-1)^{x+y}$  is located at  $u=M/2$  and  $v=N/2$  which causes the shifting of the spectrum to these coordinates. The first selected spectral feature is taken as the DC peak which represents the zero frequency in the image (hence originated the name DC referring to the direct current with zero frequency in electrical circuits). This peak is dominating because it represents the average gray level of the image as can be seen from the equation:

$$f_{21} = \hat{P}(0,0) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P(x,y)$$

The peaks at frequencies other than zero are important in summarizing information of the image and revealing its features. To visualize the other peaks and for the illustration purposes, the DC peak is suppressed to zero as illustrated in Figure 4. The peaks at the two basic orthogonal directions can be extracted as shown in Figure 5 and Figure 6 which represent the  $0^\circ$  and  $90^\circ$  directions, respectively. For each direction the first four peaks were considered for being features and for each peak both the magnitude (amplitude) and the frequency were extracted as individual features. Therefore, the features  $f_{22}$ ,  $f_{23}$ ,  $f_{24}$ , and  $f_{25}$  were extracted from the zero direction as the amplitudes of the peaks and features  $f_{26}$ ,  $f_{27}$ ,  $f_{28}$ , and  $f_{29}$  were extracted as the frequencies (locations) of the peaks. Similarly, the features  $f_{30}$  up to  $f_{37}$  were extracted from the  $90^\circ$  direction as the amplitudes and frequencies of the peaks in this direction.

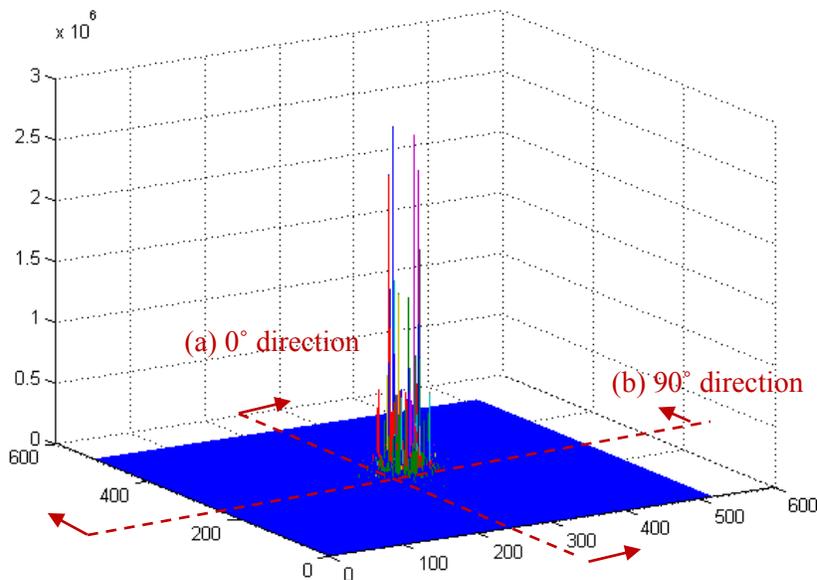


Figure 4. Fourier spectrum of the fabric image after suppressing the DC peak

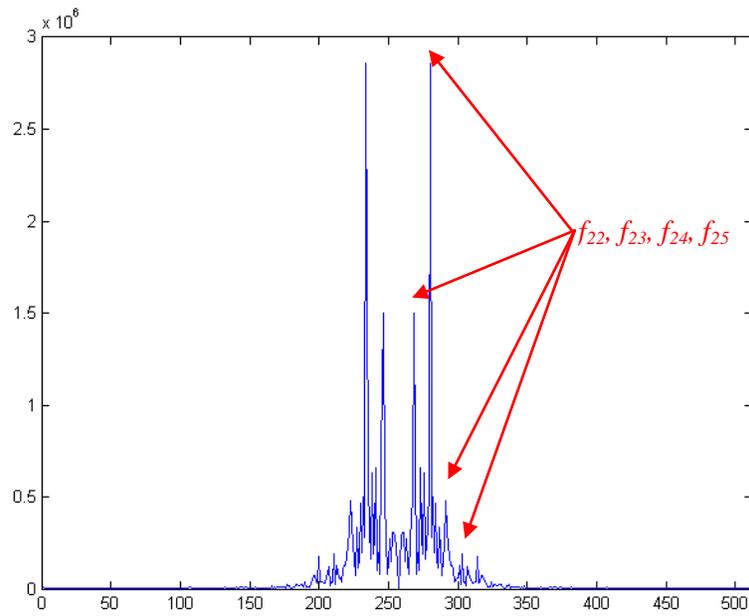


Figure 5. Peaks in the 0° direction (plan (a) in Figure 4)

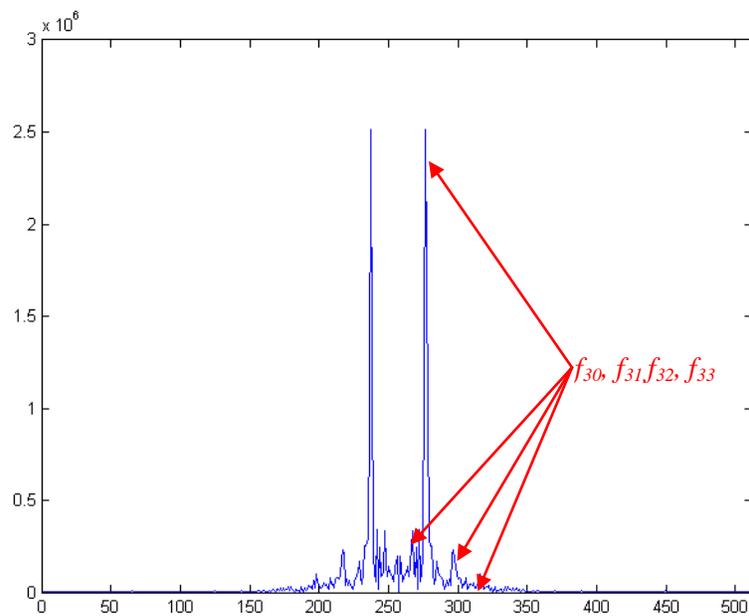


Figure 6. Peaks in the 90° direction (plan (b) in Figure 4)

### 3.5. Principal component analysis (PCA)

Although many features can be extracted from the images as shown before, some of them are highly correlated and some may not be affecting the model's predictability than others. The analysis for the most influential parameters can be performed using the principal component analysis (PCA) which is a method for linear transformation of a set of  $n$  dimensional data by projecting on an orthonormal set of  $r$  axes, where  $r \leq n$ . The new  $r$  axes are uncorrelated and called principal components because they are rotated in such a way that the axes are oriented along the direction of the highest variability of data. This in turn implies the highest amount of information represented by this data. In the situations where  $1 \leq r \ll n$ , a great reduction in the dimensionality can be achieved with the preservation of a high percentage of information in the original data. This high preservation is achieved because the first few principal components are usually chosen to represent the highest variability in the system. The dimensional reduction of the correlated data to uncorrelated components is very useful as it increases the robustness of the predictive models such as the artificial neural networks.

For a dataset  $A$  with  $n$  number of factors and  $m$  repeats or points for that factor and represented in the form:

$$A = \begin{bmatrix} a_{1,j} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,j} & \cdots & a_{n,m} \end{bmatrix}, i = 1, 2, \dots, n \text{ \& } j = 1, 2, \dots, m$$

The PCA procedure starts with normalizing the input dataset  $A$  to another set  $B$  that is translated to have a mean of zero and scaled to have a standard deviation of 1 for all the  $i^{\text{th}}$  factors. This normalization is important to neutralize the predictive models from any bias towards any of the input factors. The normalization can be achieved by constructing  $B=b_{i,j}$  where:

$$b_{i,j} = \frac{a_{i,j} - \bar{a}_i}{\sigma_i \sqrt{n}},$$

$$\bar{a}_i = \frac{1}{m} \sum_{j=1}^m a_{i,j},$$

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^m (a_{i,j} - \bar{a}_i)^2$$

In these relations,  $\bar{a}_i$  is the mean of events for the  $i^{th}$  factor, and  $\sigma_i$  is its standard deviation. After the data normalization, the correlation matrix  $C$  can be calculated from the normalized data  $B$  according to the relation [21]:

$$C_{ik} = \sum_{j=1}^m b_{j,i} b_{j,k} = \frac{1}{m} \sum_{j=1}^m \frac{(a_{i,j} - \bar{a}_i)(a_{i,k} - \bar{a}_k)}{\sigma_i \sigma_k}$$

The principle components are oriented toward the eigenvectors of the correlation matrix and have a variance equals to the associated eigenvalues. This can be represented mathematically in the form:

$$C \Psi = \Lambda \Psi$$

Where  $\Lambda$  and  $\Psi$  represent the eigenvalues and the eigenvectors, respectively. The first principal component  $PC_1$  is usually chosen to have the highest variance allocated with the highest eigenvalue  $\lambda_1$  and directed towards  $\psi_1$ . The second principal component  $PC_2$  is orthonormal to  $PC_1$  and is chosen to have the next highest variance associated with  $\lambda_2$ . In general, all principal components  $PC_k$  ( $k=1,2,\dots, r \leq n$ ) can be calculated in the same way. Each principal component contributes to the total variance by a percentage ( $v_k$ ) that can be calculated from the relation:

$$v_k = \frac{\lambda_k}{\sum_{i=1}^n \lambda_i} = \frac{\lambda_k}{n}$$

Since the first principal components are associated with the highest  $\lambda$  values, the dimensionality of the original data can be reduced to a limited number of components without losing the information (variability) embedded in the original data. Therefore, the PCA results in removing the redundancy in the original data (caused by the collinear variables) and reveals the effective dimensionality of the dataset [22].

#### 4. Results and discussion:

The set of the combined spatial and spectral features was calculated for all fabric images. The extracted features were found to have different behaviors as some features were found to cluster and converge for a certain fabric category while diverge for other categories. To illustrate the features' behavior, the feature  $f_1$  is used as an example that is shown in Figure 7 where the feature values are represented on the y-axis and the fabric fault category numbers are represented on the x-axis. The category No. 1 represents the “defect free” fabrics, category No. 2 represents fabrics “wrong draw”, category No. 3 represents fabrics with “slubs”, category No. 4 represents

fabrics with “stop marks”, category No. 5 represents fabrics with “holes”, and category No. 6 represents “stained” fabrics.

It can be seen from the figure that  $f_I$  is concentrated with low dispersion for certain groups such as the defect free (category No. 1) where 65 readings are plotted on the graph with a relatively low variance. On the other hand, the same feature is scattered in representing other categories such as the case of stained fabric (category No. 7) where 20 values are plotted and having a high dispersion. It can be detected from the behavior of the features that some of them are able to distinguish a category or more from the other categories. The combination of the features allows the detection of the fault classes in the situations where no single feature can be used to distinguish the sample. The behavior of features also indicates the differences and similarities between the groups. For example, it can be seen that category No. 1 and category No. 4 are very close in their features values which may lead to a difficulty in differentiating these categories during the classification step.

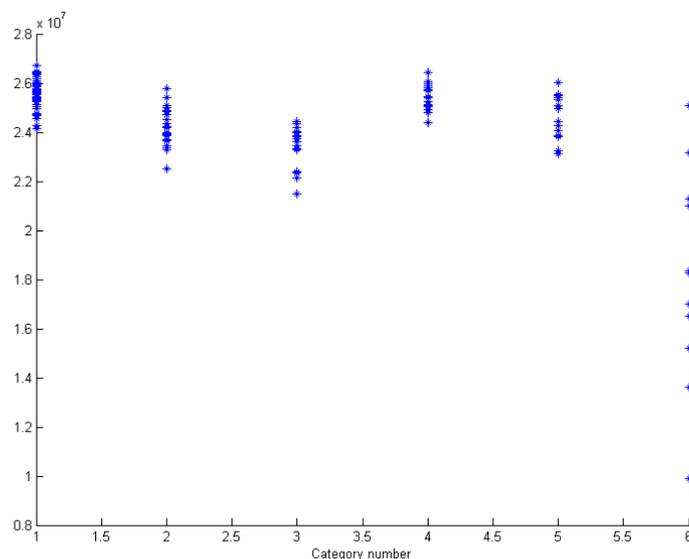


Figure 7. Feature  $f_I$  for different fabric fault categories

The original dataset of features has high correlations between the features and decreasing the dimensionality of the data will be useful during the classification. Applying the PCA to the feature dataset results in 36 principal components, which is the same number of the inputs. However, not all of these principal components are useful according to the information they preserve from the original dataset. It was found that the principal components that represent 99%

of the total variance are the first thirteen components, as shown in Table 1, which means a reduction of the data dimensionality by about 64%. The use of the remaining principal components will increase the dimensionality of the data without offering a lot of information regarding the original data. Since the principal components are uncorrelated, the presentation of the projected data on the new principal component axes demonstrates their usefulness. Figure 8 shows the scores plot obtained from the PCA with a representation of the PC2 versus PC1 for different fabric fault categories. Two main conclusions can be withdrawn from this figure; firstly, there is no distribution pattern for the plotted data which is scattered on the graph. This emphasizes the importance of the PCA procedure where no correlations between the principal components exist. Secondly, although the data are being scattered, they form clusters that might be useful in the classification of faults during the application of the ANN. Those clusters are not crisp for all groups but they have relatively distinguishable centers that are separated apart and even the category that is not shown in the figure (for stained fabrics) was excluded because it was far from the plotted categories which will shrink the scale of the figure if included.

Table 1. The the principal components' variance and their percentages

Component	Eigenvalue	Percentage (%)	Cumulative (%)
PC1	9.57	30.56	30.56
PC2	5.40	17.22	47.78
PC3	4.10	13.08	60.85
PC4	3.22	10.29	71.14
PC5	2.70	8.61	79.75
PC6	1.69	5.39	85.14
PC7	1.22	3.88	89.02
PC8	0.79	2.51	91.53
PC9	0.69	2.21	93.74
PC10	0.52	1.67	95.41
PC11	0.41	1.32	96.73
PC12	0.40	1.26	98.00
PC13	0.33	1.06	99.06

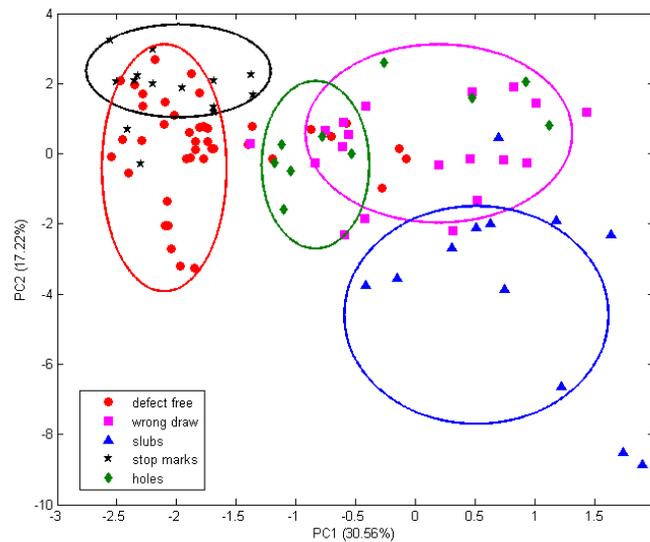


Figure 8. Scores plot for PC1 and PC2 for different fabric fault categories

The classification of the fabric faults was performed using pattern recognition artificial neural networks (ANN) which are special cases of the feed-forward neural networks that utilize certain transfer functions (tansig) and training algorithms. Two ANN were constructed during the study with the same architecture and only differ in the number of neurons in the input layer. The first network (called ANN1) was constructed using all features as inputs as demonstrated in Figure 9. ANN1 has two hidden layers that include 25 neuron in the hidden layer and 6 neurons in the output layer. The network was trained on a part of the dataset that represents all fabric fault categories and was selected randomly. The network also was tested on the remaining part of the dataset that was not used during the training. The second network (ANN2) was constructed with 13 neurons in the input layer as the PCA was applied on the inputs of this network. The network (ANN2) was trained and tested on the same training and testing datasets used with the first network (ANN1).

Table 2. Comparison of the performance of ANN1 and ANN2

	ANN1 (Full feature set)	ANN2 (Reduced feature set after PCA)
Correct Classification Rate (CCR) [%]	73.33	90
False Alarm Rate (FAR) [%]	15	5
False Negative Rate (FNR) [%]	20	7.5

Miss-Classification Rate (MCR) [%]

7.5

5

The performance of the ANNs can be quantitatively assessed from the results of the testing fabric images as summarized in Table 2 and shown in Figure 10 and Figure 11. Four measures were used to compare the performance of both ANNs. The first measure is the percentage of correct classification rate (CCR) which represents the percentage of all faults that were successfully classified on the networks. The CCR for the ANN1 was found to be 73.33% while it was increased to 90% for the ANN2. The false alarm rate (FAR) is the second comparison measure and refers to the number of defect free samples that were classified as faulty samples expressed as a percentage from the total number of the defect free samples. The FAR was found to be 15% for ANN1 and decreased to 5% for ANN2 which is considered as an improvement in the performance. On the other hand, the third measure is the false negative rate (FNR) which refers to the faulty samples that are classified as defect-free samples was found to be 20% for ANN1 while it was decreased to 7.5% with ANN2. The relatively high FNR of faulty fabrics that pass through the system (ANN1) without being detected should be taken serious because this will lead to a faulty end product if the system was industrially applied. The percentage of the faults that were detected but classified in the wrong category is considered as the fourth comparing measure and called the miss-classification rate (MCR). The MCR was found to be 7.5% for the ANN1 and improved to 5% by the application of the ANN2.

It is worthy to point out that, the similarity in the features of category No. 1 and category No. 4 (as noted earlier during the discussion of feature  $f_i$  Figure 7) might be the reason for a major part on the FNR in both networks that failed to detect the stop-mark defects. This result holds true because the first order statistical features applied in this study are usually not able to handle this type of fabric faults as long as the fabric picture has a similar number of threads even if they are irregularly distributed.

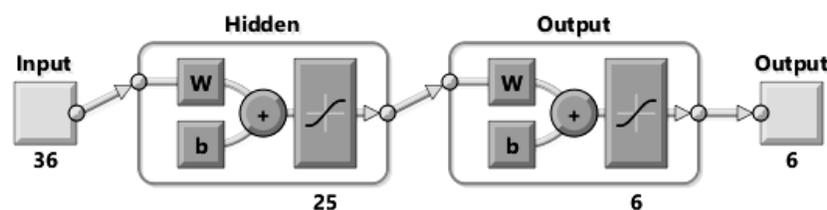


Figure 9. ANN with all features considered as inputs

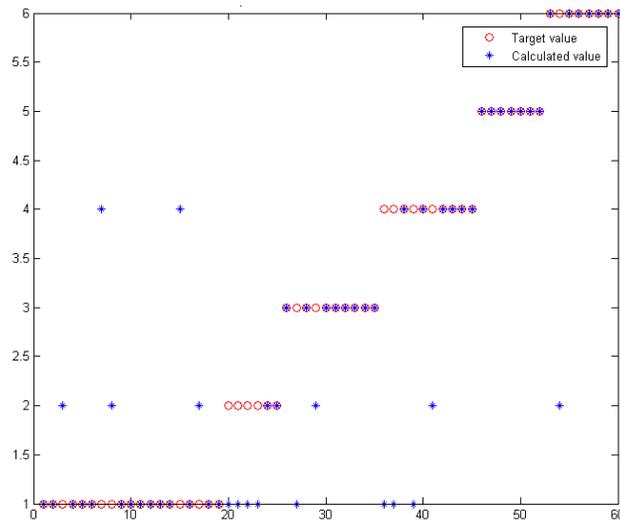


Figure 10. Performance of ANN1 (without PCA) in predicting different fabric fault categories

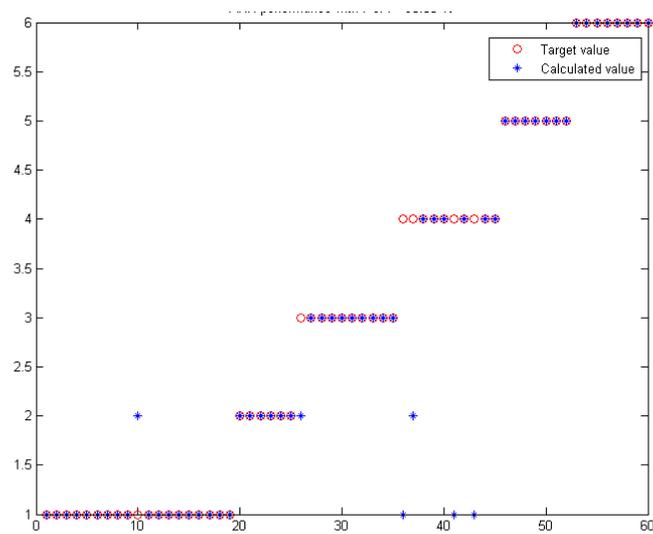


Figure 11. Performance of ANN2 (with PCA) in predicting different fabric fault categories

## 5. Conclusion

This work utilizes a digital camera to acquire and transmit fabric images to a computer which enhances and extracts the features of images. A large set of features composed of statistical and spectral features (using FFT) was used. The observation of the features shows different behaviors in the way they converge or diverge for a certain fabric fault. The PCA was used to reduce the number of features without losing the high variation in the data. A reduction to 36% of the original data size was achieved with preserving about 99% of the information in the original data. Two artificial neural networks were constructed with the same architecture and one of them was fed with the full feature dataset and the other was fed with the reduced dataset. The performance of the network that was implemented after the application of the PCA surpasses the performance of the other network in all aspect of characterization.

## References

- [1] H. Y. T. Ngan, G. K. H. Pang, and N. H. C. Yung, "Automated fabric defect detection—A review," *Image and Vision Computing*, vol. 29, pp. 442-458, 2011.
- [2] P. M. Mahajan, S. R. Kolhe, and P. M. Patil, "A review of automatic fabric defect detection techniques," *Advances in Computational Research*, vol. 1, pp. 18-29, 2009.
- [3] C.-F. J. Kuo and T.-L. Su, "Gray Relational Analysis for Recognizing Fabric Defects," *Textile Research Journal*, vol. 73, pp. 461-465, 2003.
- [4] B. Mallik and A. K. Datta, "Defect Detection in Fabrics with a Joint Transform Correlation Technique: Theoretical Basis and Simulation," *Textile Research Journal*, vol. 69, pp. 829-835, 1999.
- [5] B. Mallik-Goswami and A. K. Datta, "Detecting Defects in Fabric with Laser-Based Morphological Image Processing," *Textile Research Journal*, vol. 70, pp. 758-762, 2000.
- [6] A. Sakaguchi, G. H. Wen, Y.-i. Matsumoto, K. Toriumi, and H. Kim, "Image Analysis of Woven Fabric Surface Irregularity," *Textile Research Journal*, vol. 71, pp. 666-671, 2001.
- [7] E. Shady, Y. Gowayed, M. Abouiiiana, S. Youssef, and C. Pastore, "Detection and Classification of Defects in Knitted Fabric Structures," *Textile Research Journal*, vol. 76, pp. 295-300, 2006.
- [8] L. Xiuping, W. Zhijie, S. Zhixun, and K.-F. Choi, "Slub Extraction in Woven Fabric Images Using Gabor Filters," *Textile Research Journal*, vol. 78, pp. 320-325, 2008.
- [9] A. S. Malek, "Online fabric inspection by image processing technology," Université de Haute Alsace-Mulhouse, 2012.
- [10] H. G. Bu, X. B. Huang, J. Wang, and X. Chen, "Detection of Fabric Defects by Auto-Regressive Spectral Analysis and Support Vector Data Description," *Textile Research Journal*, vol. 80, pp. 579-589, 2010.
- [11] M. C. Hu and I. S. Tsai, "Fabric Inspection Based on Best Wavelet Packet Bases," *Textile Research Journal*, vol. 70, pp. 662-670, 2000.
- [12] C.-Y. Wen, S.-H. Chiu, W.-S. Hsu, and G.-H. Hsu, "Defect Segmentation of Texture Images with Wavelet Transform and a Co-occurrence Matrix," *Textile Research Journal*, vol. 71, pp. 743-749, 2001.
- [13] A. Tilocca, P. Borzone, S. Carosio, and A. Durante, "Detecting Fabric Defects with a Neural Network Using Two Kinds of Optical Patterns," *Textile Research Journal*, vol. 72, pp. 545-550, 2002.

- [14] C. W. M. Yuen, W. K. Wong, S. Q. Qian, D. D. Fan, L. K. Chan, and E. H. K. Fung, "Fabric Stitching Inspection Using Segmented Window Technique and BP Neural Network," *Textile Research Journal*, vol. 79, pp. 24-35, 2009.
- [15] H. T. Choi, S. H. Jeong, S. R. Kim, J. Y. Jaung, and S. H. Kim, "Detecting Fabric Defects with Computer Vision and Fuzzy Rule Generation. Part II: Defect Identification by a Fuzzy Expert System," *Textile Research Journal*, vol. 71, pp. 563-573, 2001.
- [16] C.-C. Huang and W.-H. Yu, "Fuzzy Neural Network Approach to Classifying Dyeing Defects," *Textile Research Journal*, vol. 71, pp. 100-104, 2001.
- [17] C.-C. Huang and I. C. Chen, "Neural-Fuzzy Classification for Fabric Defects," *Textile Research Journal*, vol. 71, pp. 220-224, 2001.
- [18] J.-J. Lin, "Pattern Recognition of Fabric Defects Using Case-Based Reasoning," *Textile Research Journal*, vol. 80, pp. 794-802, 2010.
- [19] D. Committee, "Terminology Relating to Fabric Defects," ASTM International 2012.
- [20] E. Pisano, S. Zong, B. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, *et al.*, "Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, pp. 193-200, 1998.
- [21] H. P. Singh, R. K. Gulati, and R. Gupta, "Stellar Spectral Classification using Principal Component Analysis and Artificial Neural Networks," *Monthly Notices of the Royal Astronomical Society*, vol. 295, pp. 312-318, 1998.
- [22] W. B. Faulkner, E. F. Hequet, J. Wanjura, and R. Boman, "Relationships of cotton fiber properties to ring-spun yarn quality on selected High Plains cottons," *Textile Research Journal*, vol. 82, pp. 400-414, 2012.

**Mohamed Eldessouki<sup>1</sup>,  
Mounir Hassan<sup>1,2</sup>,  
Khadiljah Qasbi<sup>3</sup>,  
Ebraheim Shady<sup>2</sup>**

<sup>1</sup> Computer Science Department,  
Faculty of Computing and Information Technology,  
King Abdulaziz University,  
Jeddah, Saudi Arabia  
<sup>2</sup> Telec Engineering Department,  
Faculty of Engineering,  
Mansoura University,  
Mansoura, Egypt  
<sup>3</sup> Fashion Design Department,  
Faculty of Art & Design,  
King Abdulaziz University,  
Jeddah, Saudi Arabia  
E-mail: mrehassan@kau.edu.sa

**Abstract**  
There is a growing need to replace visual fabric inspection with automated systems that detect and classify fabric defects. The digital processing of fabric images utilizes different methods that offer a large set of image features. The correlation between these features lead to problems during fabric fault classification and reduces the performance of the classifiers. This work extracted a combination of statistical (spatial) and Fourier transform (spectral) features from fabric images of the most frequent faults. Principal component analysis (PCA) was implemented to reduce the dimensionality of the input features dataset which achieved a reduction to 36% of the original data size while preserving 99% of information in the original dataset. The features processed using the PCA were fed to an artificial neural network (ANN) to classify the fault categories and then compared to another ANN that worked with the whole features dataset. The performance of the network that was implemented after application of the PCA increased to 90% of the correct classification rate as compared to 75.3% for the other network.

**Key words:** fabric fault detector, image processing, artificial neural networks, principal component analysis.

tration and fault classification can be found as shown in Figure 1. The methods of feature extraction vary, including spatial (statistical) features [3], spectral features (fast Fourier Transform) [4], a combination of spatial and spectral features [5 - 10], as well as other methods that may utilize wavelet transforms [11, 12]. The features extracted are fed to a classification system that has been implemented by researchers in different ways. Among these classifiers are artificial neural networks (ANN) of different types [13, 14], fuzzy inference systems [15], neuro-fuzzy systems [16, 17], as well as other classification systems [18].

**Introduction**

Early intervention to fix, or remove, fabric faults is one of the mandatory tasks required by all fabric manufacturers. Undetected defects cause many problems downstream in the production line and result in end products with lower quality, creating a cost burden and non-profitable products. Automatic fault detection systems offer good alternatives to replace traditional human fabric inspection with computer vision systems that analyse fabrics in a systematic manner and aim at consistent performance. The efficiency of these automated systems, however, depends on many parameters and varies according to the quality of the hardware and the analysis algorithm.

There are many research articles in the field of automatic fabric fault detection and classification that can be found in published reviews [1, 2]. Among those some common methods for feature ex-

traction and fault classification can be found as shown in Figure 1. The methods of feature extraction vary, including spatial (statistical) features [3], spectral features (fast Fourier Transform) [4], a combination of spatial and spectral features [5 - 10], as well as other methods that may utilize wavelet transforms [11, 12]. The features extracted are fed to a classification system that has been implemented by researchers in different ways. Among these classifiers are artificial neural networks (ANN) of different types [13, 14], fuzzy inference systems [15], neuro-fuzzy systems [16, 17], as well as other classification systems [18].



Figure 1. Different methods available in the literature for feature extraction and classification.

Eldessouki M, Hassan M, Qasbi K, Shady E. Application of Principal Component Analysis to Boost the Performance of an Automated Fabric Fault Detector and Classifier. *FIBRES & TEXTILES in Eastern Europe* 2014; 22(410): 51-67.

51

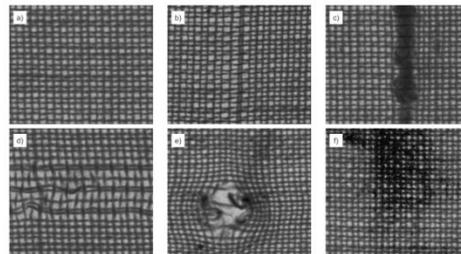


Figure 2. Images of: (a) defect-free, (b) wrong draw, (c) warp slub, (d) slub mark, (e) holes, (f) fabric, (g) fabric-Match woven fabrics.

**Wrong draft**, "misdraw", or "double end". This fault may be considered as a severe defect because it appears throughout the whole length of the fabric if not fixed.

**Slub:** This fault shows in the fabric as an abnormally thickened place in a yarn. It can occur in the warp or weft direction, but it was considered only in the warp direction in this study. Other names of the same fault are hump, piling, slough-off, and slug. All these names are considered in ASTM standards. This fault may occur due to malfunctioning in warp sensors, shown in Figure 2e.

**Weft direction**  
**Slub mark:** This appears as a visible change in the density of the weave across the width of the fabric caused by the tension on the warp not being adjusted properly after the loom has been stopped. This fault may be called a "wet mark" or "light beat-up", shown in Figure 2d.

**Area faults**  
**Hinte:** It is an imperfection in the fabric where one or more yarns are sufficiently damaged to create an aperture. In case of a relatively large hole, it might be called a "small", which is characterized by broken warp ends and floating picks. The small may be equivalently called a "weik-out", shown in Figure 2e.

**Stain:** It is an area of discoloration that penetrates the fabric surface. If this discoloration is caused by grease or oil and the oil-colored area appears in any shape, it is called a "blotch" or "oil spot", as an example of which is shown in Figure 2f.

**Methodology**

**Samples**

Fabric samples were manufactured on a Sulzer-Runi weaving machine. The fabric structure was plain weave 1/1 with a yarn count of 29.5 tex for the warp and 42 tex for the weft. The densities of warp and weft yarns are 20 and 18 per cm, respectively. The defects chosen were intentionally introduced on the weaving machine based on knowledge of the defects sources.

**Image acquisition**  
The image acquisition setup is shown in Figure 3, utilizing a Canon digital camera (model: EOS 450D) with CMOS sensor. The system is installed with "Remote Live View Shooting", where online monitoring of the pictures and their adjustment can be done on a computer using EOS Utility software. The camera uses 35 mm EF-S lenses and captures images at a resolution of 72 dots per inch (dpi). The fabric sample is placed on an inspection table that is equipped with concentrated LED lights in a box placed

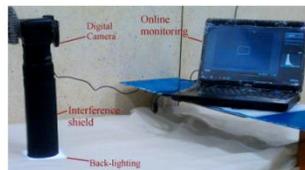


Figure 3. Fabric image acquisition setup.

FIBRES & TEXTILES in Eastern Europe 2014; 22(410): 51/52

52

directly under the shooting area. To remove the noise and interference of the surrounding lights, a suitable shield was installed between the camera and shooting area.

**Image enhancement**

The system developed applies initial enhancements to the original images to reduce noise (e.g. hairiness) and improve their contrast. The system uses the contrast-limited adaptive histogram equalisation (CLAHE) [20] algorithm to enhance the contrast of the grayscale image by transforming the values. The algorithm can be described briefly as it operates in small regions (windows) in the image. Each window's contrast is enhanced so that the histogram of the output region approximately matches the histogram specified. The neighboring windows are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image.

**Image analysis and feature extraction**

If the fabric image can be defined in the spatial domain by the matrix  $P(x,y)$ , where  $x$  is the row number in the image ( $1 \leq x \leq N$ ), the column number ( $1 \leq y \leq M$ ), and  $N$  &  $M$  are the number of rows and columns, respectively, the features that can be extracted from this image are summarized as below.

**Statistical features**

A total set of twenty spatial (statistical) features can be extracted from the fabric images. To obtain these features, the sum of individual gray-scale level values in the weft direction (rows) is calculated in the vector  $R(m)$ , where:

$$R(m) = \sum_{x=1}^N P(x,y) \quad (1)$$

Similarly, the sum of individual gray-scale values in the warp direction (columns) is calculated as:

$$C(n) = \sum_{y=1}^M P(x,y) \quad (2)$$

The first feature selected is the sum of all the gray-scale values in the image, which can be calculated as:

$$f_1 = \sum_{x=1}^N C(n) \quad (3)$$

The next two features,  $f_2$  and  $f_3$ , represent the mean of the sum of rows ( $f_2$ ) and that of the sum of columns ( $f_3$ ). The relation for the first feature in the weft direction is:

$$f_2 = \frac{1}{M} \sum_{m=1}^M R(m) \quad (4)$$

Similarly for the feature in the warp direction:

$$f_3 = \frac{1}{N} \sum_{n=1}^N C(n) \quad (5)$$

For space constraints, the equations will be listed for the features in the weft (rows) direction only and similar relations can be written for the warp (columns) direction by replacing  $R(m)$  with  $C(n)$ .

The standard deviation of the sum of rows ( $f_4$ ) and for the sum of columns ( $f_5$ ) can be calculated as:

$$f_4 = \frac{1}{M} \sqrt{\frac{1}{M} \sum_{m=1}^M (R(m) - R)^2} \quad (6)$$

$$R = f_2$$

Features  $f_6$  and  $f_7$  represent the median value of the sum of rows and columns:

$$f_6 = \text{median}_m(R(m))$$

Features  $f_8$  and  $f_9$  represent the minimum and maximum values, respectively, for the sum of rows and  $f_{10}$  and  $f_{11}$  represent the same values for the columns. These features can be written as:

$$f_8 = \min_m(R(m)), f_9 = \max_m(R(m))$$

The range of the sum of rows and columns was chosen to be, respectively, features  $f_{12}$  and  $f_{13}$ :

$$f_{12} = \text{range}(R(m))$$

The entropy of the image represents feature  $f_{14}$ :

$$f_{14} = - \sum_{x=1}^N \sum_{y=1}^M P(x,y) \times \log_2(P(x,y)) \quad (7)$$

The  $k^{\text{th}}$  order moment for the sum of rows can be calculated from the function

$$f_{15} = \frac{1}{M} \sum_{m=1}^M (R(m)) \cdot R^k \quad (8)$$

The second, third and fourth order moments can be calculated as features  $f_{15}$ ,  $f_{16}$ , and  $f_{17}$  ( $k = 2, 3, 4$ ) for the rows as well as features  $f_{18}$ ,  $f_{19}$ , and  $f_{20}$  ( $k = 2, 3, 4$ ) for the columns.

**Spectral features**

A fabric image of size  $M \times N$  can be transformed from its spatial domain  $P(x,y)$  to the spectral domain  $\hat{P}(u,v)$  using the discrete Fourier transform (DFT), which can be expressed in a mathematical form as:

$$\hat{P}(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P(x,y) e^{j2\pi(ux+vy)} \quad (9)$$

Where  $u$  and  $v$  are the image spatial variables that correspond to the coordinates inside the image, while  $x$  and  $y$  represent the frequency variables transformed. Once the image is transformed, its power spectrum can be calculated as:

$$P(u,v) = |\hat{P}(u,v)|^2 = R^2(u,v) \quad (10)$$

Where  $\hat{P}(u,v)$  is known as the Fourier spectrum, and  $R^2(u,v)$  and  $P(u,v)$  are the real and imaginary parts of the image transformed  $\hat{P}(u,v)$ . Shifting the power spectrum is done by implementing the exponential properties of the transform:

$$R[\hat{P}(x)(-1)^{u+v}] = \hat{P}(u-M/2, v-N/2) \quad (11)$$

Where  $[-1]$  is the Fourier transform of an argument, stating the origin of the Fourier transform  $\hat{P}(u,v)$  of image  $P(x,y)(-1)^{u+v}$  is located at  $u = M/2$  and  $v = N/2$ , which causes the shifting of the spectrum to these coordinates. The first spectral feature selected is taken as the zero frequency in electrical circuits. This peak dominates because it represents the average gray level of the image, as can be seen from the equation:

$$f_{21} = P(0,0) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P(x,y) \quad (12)$$

The peaks at frequencies other than zero are important in summarizing information of the image and revealing its features. To visualise the other peaks and for illustration purposes, the DC peak is suppressed to zero, as illustrated in Figure 4. The peaks in the two basic orthogonal directions can be extracted as shown in Figure 4 and 6, which represent the 90° and 0° directions, respectively. For each direction the first four peaks were considered as features and for each peak both the magnitude (amplitude) and frequency

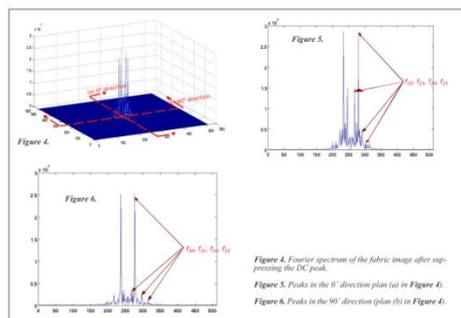


Figure 4. Fourier spectrum of the fabric image after suppressing the DC peak.

Figure 5. Peaks in the 0° direction (phi) in Figure 4.

Figure 6. Peaks in the 90° direction (theta) in Figure 4.

were extracted as individual features. Therefore features  $f_{22}$ ,  $f_{23}$ ,  $f_{24}$ , and  $f_{25}$  were extracted from the zero direction as the amplitudes of the peaks, and features  $f_{26}$ ,  $f_{27}$ ,  $f_{28}$ , and  $f_{29}$  were extracted in the 90° direction as the amplitudes and frequencies of the peaks in this direction.

**Principal component analysis (PCA)**

Although many features can be extracted from the images as shown before, some of them are highly correlated and some may not affect the model's predictability as others. Analysis to establish the most influential parameters can be performed using principal component analysis (PCA), which is a method for the linear transformation of a set of dimensional data by projecting on an orthonormal set of axes, where  $r \leq n$ . The new axes are uncorrelated and called principal components because they are rotated in such a way that the axes are oriented along the direction of the highest variability of data. This, in turn, implies the highest amount of information represented by this data. In situations where  $r < n$ , a great reduction in dimensionality can

be achieved with the preservation of a high percentage of information in the original data. This high preservation is achieved because the first few principal components are usually chosen to represent the highest variability in the system. The dimensional reduction of the correlated data to uncorrelated components is very useful as it increases the robustness of predictive models such as artificial neural networks.

Dataset  $A$  with  $n$  number of factors and  $m$  inputs or points for that factor can be represented in the form:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

$$i = 1, 2, \dots, m; j = 1, 2, \dots, n$$

$$C_k = \frac{1}{m} \sum_{j=1}^m b_{kj} \quad (16)$$

$$= \frac{1}{m} \left( \frac{m}{m} (a_{1j} - \bar{a}_j) (a_{1j} - \bar{a}_j) + \dots \right)$$

The PCA procedure starts with normalizing the input dataset  $A$  to another set  $B$  that is translated to have a mean of zero and scaled to have a standard deviation of 1 for all the  $P$  factors. This normalisation is important to neutralise the predictive models from any bias towards any of the input factors. The normalisation can be achieved by constructing  $B = A - \bar{A}$  where:

$$C = P \cdot A \cdot P^T$$

$$\bar{a}_j = \frac{\sum_{i=1}^m a_{ij}}{m} \quad (13)$$

$$\bar{a}_j = \frac{1}{m} \sum_{i=1}^m a_{ij} \quad (14)$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^m (a_{ij} - \bar{a}_j)^2 \quad (15)$$

In these relations,  $\bar{a}_j$  is the mean of events for the  $j^{\text{th}}$  factor, and  $\sigma_j$  is its standard deviation. After data normalization, the correlation matrix  $C$  can be calculated from the normalized data  $B$  according to the relation [21]:

$$C_{ij} = \frac{1}{m} \sum_{k=1}^m b_{ki} \cdot b_{kj} = \frac{1}{m} \left( \frac{m}{m} (a_{1j} - \bar{a}_j) (a_{1i} - \bar{a}_i) + \dots \right)$$

$$= \frac{1}{m} \sum_{k=1}^m (a_{1j} - \bar{a}_j) (a_{1i} - \bar{a}_i) + \dots$$

The principle components are oriented toward the eigenvectors of the correlation matrix and have a variance equal to the associated eigenvalues. This can be represented mathematically in the form:

$$C \cdot P = \Lambda \cdot P^T$$

FIBRES & TEXTILES in Eastern Europe 2014; 22(410): 51/52

54

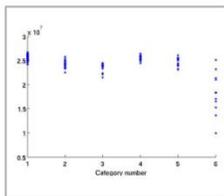


Figure 7. Feature  $f_j$  for different fabric fault categories.

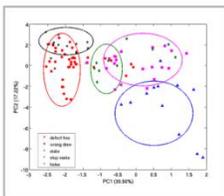


Figure 8. Scores plot for PC1 and PC2 for different fabric fault categories.

Where  $A$  and  $F$  represent the eigenvalues and eigenvectors, respectively. The first principal component  $PC_1$  is usually chosen to have the highest variance allocated with the highest eigenvalue  $\lambda_1$  and directed towards  $w_1$ . The second principal component  $PC_2$  is orthogonal to  $PC_1$  and is chosen to have the next highest variance associated with  $\lambda_2$ . In general, all principal components  $PC_i$  ( $i = 1, 2, \dots, r \leq n$ ) can be calculated in the same way. Each principal component contributes to the total variance by a percentage ( $\alpha_i$ ) that can be calculated from the relation:

$$\alpha_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \times 100 \quad (17)$$

Since the first principal components are associated with the highest  $\alpha$  values, the dimensionality of the original data can be reduced to a limited number of components without losing the information (variability) embedded in the original data. Therefore the PCA results in removing the redundancy in the original data (formed by the collinear variables) and reveals the effective dimensionality of the dataset [22].

III Results and discussion

The set of combined spatial and spectral features was calculated for all fabric images. The features extracted were found to have different behaviours as some were found to cluster and converge for a certain fabric category while diverge

for other categories. To illustrate the features' behaviour, feature  $f_j$  is used as an example, shown in Figure 7, where the feature values are represented on the y-axis and fabric fault category numbers on the x-axis. Category No. 1 represents 'defect free' fabrics, category No. 2 'wrong draw' fabrics, category No. 3 fabrics with 'holes', category No. 4 fabrics with 'stop marks', category No. 5 fabrics with 'holes', and category No. 6 represents 'stained' fabrics.

It can be seen from the figure that  $f_j$  is concentrated with low dispersion for certain groups such as the defect free (category No. 1), where 65 readings are plotted on the graph with relatively low variance. On the other hand, the same feature is scattered in representing other categories such as the case of stained fabric (category No. 6), where 20 values are plotted and have a high dispersion. It can be detected from the behaviour of

the features that from some of them we are able to distinguish a category or more from the other categories. The combination of features allows the detection of fault classes in situations where no single feature can be used to distinguish the sample. The behaviour of features also indicates the differences and similarities between the groups. For example, it can be seen that categories No. 1 and No. 4 are very close in their feature values, which may lead to difficulty in differentiating these categories during the classification step.

The original dataset of features has high correlation between the features and decreasing the dimensionality of the data will be useful during the classification. Applying PCA to the feature dataset results in 16 principal components, which is the same number of the inputs. However, not all of these principal components are useful according to the information

Table 1. Variance of principal components and their percentages.

Component	Eigenvalue	Percentage, %	Cumulative, %
PC1	9.57	30.56	30.56
PC2	5.40	17.22	47.78
PC3	4.10	13.08	60.85
PC4	3.22	10.29	71.14
PC5	2.70	8.61	79.75
PC6	1.69	5.39	85.14
PC7	1.22	3.88	89.02
PC8	0.79	2.51	91.53
PC9	0.69	2.21	93.74
PC10	0.52	1.67	95.41
PC11	0.41	1.32	96.73
PC12	0.40	1.26	98.00
PC13	0.33	1.06	99.06

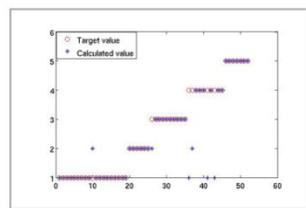


Figure 11. Performance of ANN2 (with PCA) in predicting different fabric fault categories.

considered as the fourth comparing network, called the miss-classification rate (MCR). The MCR was found to be 7.5% for ANN1, which improved to 3% with the application of ANN2.

It is worthy to note that the similarity in the features of categories No. 1 and No. 4 (as noted earlier during the discussion of feature  $f_j$ ) (Figure 7) might be the reason for a major part of the FNR in both networks failing to detect stop-mark defects. This result holds true because the first order statistical features applied in this study are usually not able to handle these types of fabric faults as long as the fabric picture has a similar number of threads, even if they are irregularly distributed.

III Conclusion

This work utilized a digital camera to acquire and transmit fabric images to a computer which enhances and extracts the features thereof. A large set of features composed of statistical and spectral features (using FFT) was used. Observation of the features shows different behaviours in the way they converge or diverge for a certain fabric fault. PCA was used to reduce the number of features without losing the high variation in data. A reduction to 36% of the original data size was achieved while preserving about 99% of the information in the original data. Two artificial neural networks were constructed with the same architecture and one of them was fed with the full feature dataset and the other with the reduced dataset. The performance of the

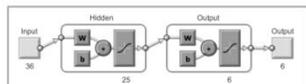


Figure 9. ANN with all features considered as inputs, 36, 25, 6, and 6 number of neurons.

Table 2. Comparison of the performance of ANN1 and ANN2.

	ANN1 (Full feature set)	ANN2 (Reduced feature set after PCA)
Correct classification rate (CCR), %	73.33	90
False alarm rate (FAR), %	15	5
False negative rate (FNR), %	20	7.5
Miss-classification rate (MCR), %	7.5	5

they preserve from the original dataset. It was found that the principal components, which represent 99% of the total variance, are the first thirteen components, as shown in Table 1, which means a reduction in data dimensionality of about 64%. The use of the remaining principal components will increase the dimensionality of the data without offering a lot of information regarding the original data. Since the principal components are uncorrelated, presentation of the data projected on the new principal component axes demonstrates their usefulness.

Figure 4 shows the scores plot obtained from the PCA with a representation of PC2 versus PC1 for different fabric fault categories. Two main conclusions can be drawn from this figure: Firstly there is no

distribution pattern for the data plotted, which is scattered on the graph. This emphasises the importance of the PCA procedure, where no correlations between the principal components exist. Secondly, although the data are scattered, they form clusters that might be useful in the classification of faults during application of the ANN. These clusters are not crisp for all groups, but they have relatively distinguishable centres that are separated apart, and even the category that is not shown in the figure (for stained fabric) was excluded because it was far from the categories plotted, which will skew the scale of the figure if included.

Classification of the fabric faults was performed using pattern recognition arti-

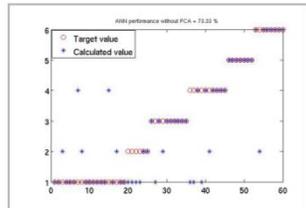


Figure 10. Performance of ANN1 (without PCA) in predicting different fabric fault categories.

56

ficial neural networks (ANN), which are special cases of feedforward neural networks that utilise certain transfer functions (tuning) and training algorithms. There is a weight for each connection link ( $W$ ) and a bias term ( $b$ ) which are adjusted during the training process. Two ANNs were constructed during the study with the same architecture, only differing in the number of neurons in the input layer. The first network (called ANN1) was constructed using all features as inputs, as demonstrated in Figure 9. ANN1 has two hidden layers that include 25 neurons in the hidden layer and 6 neurons in the output layer. The network was trained for a part of the dataset that represents all fabric fault categories and was selected randomly. The network also was tested for the remaining part of the dataset that was not used during the training. A second network (ANN2) was constructed with 13 neurons in the input layer as the PCA was applied to the inputs of this network. The network (ANN2) was trained and tested for the same training and testing datasets used with the first network (ANN1).

The performance of the ANNs can be quantitatively assessed from the results of the fabric images tested, as summarised in Table 2 and shown in Figures 10 & 11. Four measures were used to compare the performance of both ANNs. The first measure is the percentage of the correct classification rate (CCR), which represents the percentage of all faults that were successfully classified in the networks. The CCR for ANN1 was found to be 73.33%, while it increased to 90% for ANN2. The false alarm rate (FAR) is the second comparison measure and refers to the number of defect free samples that were classified as faulty samples, expressed as a percentage of the total number of defect free samples. The FAR was found to be 15% for ANN1, while it decreased to 5% for ANN2, which is considered as an improvement in performance. On the other hand, the third measure, i.e. the false negative rate (FNR), which refers to the faulty samples that are classified as defect-free, was found to be 20% for ANN1, while it decreased to 7.5% with ANN2. The relatively high FNR of faulty fabrics that pass through the system (ANN1) without being detected should be taken seriously because this will lead to a faulty end product if the system is industrially applied. The percentage of faults that were detected but classified in the wrong category is con-

sidered as the fourth comparing network, called the miss-classification rate (MCR). The MCR was found to be 7.5% for ANN1, which improved to 3% with the application of ANN2.

57

Acknowledgment

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia, under grant No. (1503/964/422). The authors, therefore, acknowledge with thanks DSR technical and financial support.

References

- Ngan H Y T, Pang G H, Yung N H C. Automated fabric defect detection - A review. *Image and Vision Computing* 2011; 29: 442-458.
- Mahajan PM, Kolhe SR, Patel PM A. A review of automatic fabric defect detection techniques. *Advances in Computational Research* 2009; 1: 18-29.
- Kim C-F, Su T-L. Gray Relational Analysis for Recognizing Fabric Defects. *Textile Research Journal* 2003; 73: 481-482.
- Malik B, Datta AK. Defect Detection in Fabrics with a Joint Transform Correlation Technique: Theoretical Basis and Simulation. *Textile Research Journal* 1990; 60: 828-835.
- Malik-Gowami B, Datta AK. Detecting Defects in Fabric with Laser-Scanned Morphological Image Processing. *Textile Research Journal* 2002; 70: 756-762.
- Sakaguchi A, Wen G H, Matsumoto Y, Torumi K, Kim H. Image Analysis of Woven Fabric Surface Irregularity. *Textile Research Journal* 2001; 71: 666-671.
- Shady E, Goward Y, Aboulana M, Yousef S, Pantare C. Detection and Classification of Defects in Knitted Fabric Structures. *Textile Research Journal* 2006; 76: 295-300.

- Xing L, Zhijie W, Zhiqun S, Chao K-F. Silk Extraction in Woven Fabric Images Using Color Filters. *Textile Research Journal* 2008; 78: 320-325.
- Malik AS. Online fabric inspection by image processing technology. *University of Huddersfield, Huddersfield*, 2012.
- Bu H-G, Huang X-B, Wang J, Chen X. Detection of Fabric Defects by Adaptive Spectral Analysis and Support Vector Data Description. *Textile Research Journal* 2010; 80: 579-589.
- Hu M-C, Tai S-S. Fabric Inspection Based on Best Wavelet Packet Bases. *Textile Research Journal* 2002; 70: 662-670.
- Wang C, Chen S-H, Hu W-S, Hu G-H. Defect Segmentation of Texture Images with Wavelet Transform and a Co-occurrence Matrix. *Textile Research Journal* 2001; 71: 743-748.
- Tizoca A, Boccone P, Carosio S, Durante A. Detecting Fabric Defects with a Neural Network Using Two Kinds of Optical Patterns. *Textile Research Journal* 2002; 72: 545-550.
- Yuan CHM, Wong WK, S-Q, Qian D, Fan L-K, Chan, and E H K. Fung. Fabric Defect Inspection Using Segmented Window Technique and BP Neural Network. *Textile Research Journal* 2000; 70: 24-35.
- Choi H-T, Jeong S-H, Kim S-R, Jaung J-Y, Kim S-H. Detecting Fabric Defects with Computer Vision and Fuzzy Rule Generation. Part II: Defect Identification by a Fuzzy Expert System. *Textile Research Journal* 2001; 71: 563-573.
- Huang C-C, Yu W-H. Fuzzy Neural Network Approach to Classifying Dyeing Defects. *Textile Research Journal* 2001; 71: 100-104.
- Huang C-C, Chen J-C. Neural-Fuzzy Classification for Fabric Defects. *Textile Research Journal* 2001; 71: 220-224.
- Phan S-Z, Zong S, Homminger B, Deluca M, Johnson RE, Muller K, et al. Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated speculations in dense mammograms. *Journal of Digital Imaging* 1998; 11: 193-200.
- Singh H P, Gullik R K, Gupta R. Statist. Spectral Classification using Principal Component Analysis and Artificial Neural Networks. *Statist. Advances of the Royal Astronomical Society* 1998; 295: 312-313.
- Faulkner WB, Haquet EF, Warjuna J, Boman H. Relationships of cotton fiber properties to ring-spun yarn quality on selected High Plains cottons. *Textile Research Journal* 2012; 82: 40-414.

Received 04.11.2013 Revised 09.01.2014