Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 72 (2004) 57–71
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Image texture analysis: methods and comparisons

Manish H. Bharati[1], J. Jay Liu, John F. MacGregor*

Department of Chemical Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L7

Received 5 May 2003; received in revised form 5 February 2004; accepted 9 February 2004

Available online 6 May 2004

Abstract

Surface texture is an important quality characteristic of many products. This paper provides an overview of several different approaches to image texture analysis and demonstrates their use on the problem of classifying a set of rolled steel sheets into various quality grades. Methods covered include traditional statistical approaches such as gray level co-occurrence matrix (GLCM) methods, multivariate statistical approaches based on PCA and PLS, and wavelet texture analysis.

Traditional multivariate classification approaches, such as PLS-DA, applied *directly* to the images are shown to fail because of the loss of spatial identity of the variables (pixels) in those approaches, and the lack of congruency of the images. However, approaches that re-introduce spatial information, such as performing two-dimensional FFT on the images prior to applying multivariate methods can perform well. A new approach that re-introduces spatial information through image shifting and stacking, followed by multivariate image analysis (MIA) is presented and shown to work well. It can also be used to develop optimal spatial filters for extracting texture information. Wavelet texture analysis (WTA) methods are discussed and insight into their space/frequency decomposition behavior is used to show why they are generally considered to be state of the art in texture analysis.
© 2004 Published by Elsevier B.V.

Keywords: Texture analysis; Gray level co-occurrence matrix; Wavelet texture analysis; Principal component analysis; Partial least squares; Multivariate image analysis

1. Introduction

Although one can intuitively associate several image properties such as smoothness, coarseness, depth, regularity, etc. with texture [1], there is no formal or complete definition of texture. Many researchers have described texture using various definitions. Russ [2] loosely defined image texture as a descriptor of local brightness variation from pixel to pixel in a small neighborhood through an image. Alternatively, texture can be described as an attribute representing the spatial arrangement of the gray levels of the pixels in a region of a digital image [3]. Texture analysis has played an important role in many areas including medical imaging, remote sensing and industrial inspection, and its tasks are mainly classification, segmentation, and synthesis [4–6].

The approaches for analyzing texture are very diverse, and differ from each other mainly by the method used for *extracting textural features*. Four categories can be defined: (1) statistical methods, (2) structural methods, (3) model-based methods, and (4) transform-based methods.

Statistical texture analysis techniques primarily describe texture of regions in an image through higher-order moments of their grayscale histograms [7]. Probably, the most frequently cited method for texture analysis is based on extracting various textural features from a gray level co-occurrence matrix (GLCM) [8]. The GLCM approach is based on the use of second-order statistics of the grayscale image histograms. Alternatively, the run length matrix (RLM) encompasses higher-order statistics of the gray level histogram. The RLM texture analysis approach characterizes coarse textures as having many pixels in a constant gray level run and fine textures as having few pixels in such a run [9]. Besides traditional statistical texture analysis, multivariate statistical methods have also been proposed for textural feature extraction. Considering an image as a matrix, the Singular Value Decomposition (SVD) spectrum is a summary vector of image texture represented by its singular values. The SVD spectrum has been used as a textural feature vector for image classification [10,11].

* Corresponding author. Tel.: +1-905-525-9140x24951; fax: +1-905-521-1350.
E-mail address: macgreg@mcmaster.ca (J.F. MacGregor).
[1] Current address: Shell Global Solutions Inc., Westhollow Technology Center, P.O. Box 1380, Houston, TX, USA.

0169-7439/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.chemolab.2004.02.005

Structural texture analysis techniques describe a texture as the composition of well-defined *texture elements* such as regularly spaced parallel lines. The properties and placement rules of the texture elements define the image texture. Various structural texture analysis approaches have been proposed, ranging from using different shapes of structuring elements [12] to conceiving real textures as distorted versions of ideal textures [13]. However, these methods appear to be limited in practicality since they can only describe very regular textures [1].

Model-based texture analysis techniques generate an empirical model of each pixel in the image based on a weighted average of the pixel intensities in its neighborhood. The estimated parameters of the image models are used as textural feature descriptors. Examples of such model-based texture descriptors are autoregressive (AR) models [14], Markov random fields (MRF) [15], and fractal models [16].

Finally, *transform-based texture analysis* techniques convert the image into a new form using the spatial frequency properties of the pixel intensity variations. The success of these latter techniques lies in the type of transform used to extract textural characteristics from the image. Indhal and Næs [17] illustrated the use of spectra from 2-D Fast Fourier Transform (FFT) magnitude images for textural feature extraction. Image classification using Multi-way Principal Component Analysis (MPCA) on 2-D FFT magnitude images to extract features from various images was used by Geladi [18]. The Gabor or Wavelet transforms have been preferred recently in image texture analysis due to their space-frequency decomposition abilities. Features derived from a set of Gabor filters have been widely used in texture analysis for image segmentation [19]. Wavelet transform methods of feature extraction have been used to characterize texture and to treat the problems of texture segmentation and classification [4–6,20–22]. The Angle Measure Technique (AMT) has been used to extract textural features from unfolded image pixel values in order to characterize and predict externally measured reference textures using multivariate statistical techniques [11,23].

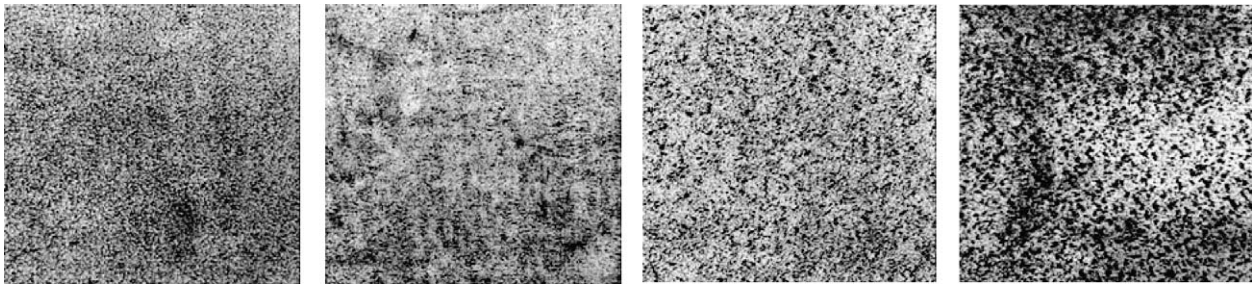
The purpose of this paper is to provide an overview and discussion of several of the above approaches and to contrast them by applying them to the classification of the texture of steel surface images. Differences among the approaches and the limitations of some of them are highlighted, and some new approaches are presented. The paper is organized as follows. In Section 2, a brief description of the image data set (used in this paper for illustration of the methods) is presented, and the classification objectives are defined. Sections 3–5 will outline various approaches for texture analysis and will apply them to the example image data set. Conclusions are given in Section 6.

2. Description of data set and classification objectives

Prior to shipping, steel quality is often monitored by performing random quality control checks on finished steel rolls. The quality of a steel sheet is reflected in the number and severity of pits on its surface. Good quality steel surfaces have few pits that are quite shallow and randomly distributed. When the pits become deeper, start to join, and result in deep craters throughout the steel, the surface quality is considered to be bad. Skilled graders visually determine the degree of steel surface pitting based on various criteria developed from previous experience and by comparison with standard samples. Classification using these criteria is time consuming and requires very experienced graders. Thus, an automated image-based grading system would be useful.

For this study, a total of 35 images of steel surfaces were obtained. Sheets with varying degrees of surface pits were cut from finished steel rolls. In order to highlight the surface pits prior to imaging, each slab is pre-treated by pouring black ink upon the surface. After the ink had filled into the pits, the steel slabs are lightly cleaned with a cloth. This results in the steel surface pits being represented by black areas. The stained steel slabs are then digitally imaged as grayscale images.

Fig. 1 shows examples of steel surface images with *excellent*, *good*, *medium* and *bad* surface qualities. An example of bad surface quality (see Fig. 1d) contains various 'snake'-like patterns representing deep pits that have joined to form craters. Fig. 1c illustrates an example of a medium



(a) (b) (c) (d)

Fig. 1. Examples of three types of steel surface grayscale images. (a) Excellent surface quality (E02), (b) good surface quality (G02), (c) medium surface quality (M06), and (d) bad surface quality (B12).

quality surface, which contains more pronounced pits as compared to the excellent and good quality samples. However, it does not contain the serpentine patterns exhibited by the bad quality steel. In all the data set images, ink smudge marks are also evident on the steel surfaces due to the manual cleaning of excessive ink with a cloth. The complete steel image data set from the four pre-labeled surface quality classes is available from the McMaster Advanced Control Consortium (MACC) FTP server [24]. Each image is an 8-bit grayscale image with pixel dimensions of 479 × 508. All images have been pre-processed to enhance their contrast via intensity histogram stretching [25]. Table 1 shows the division of 35 sample images into their respective pre-labeled classes, as determined by experienced graders. However, it is important to note that the selected classes are not clearly separated, but rather represent a progression from bad to excellent with the boundaries quite vague, particularly between the good and excellent classes.

The objective of using this data set is to illustrate classification based on the presented texture analysis techniques using the pre-labeled classes as a benchmark, and to comment on some of their strengths and weaknesses. No attempt is made to assess the performance of each method based on error rates of classification since the sample size (35 images) is inadequate for that purpose.

Data classification using the latent variable spaces of multivariate statistical methods like PCA and PLS has been widely used in the chemometrics literature [26]. Unsupervised classification can be achieved through observing score clustering patterns in the latent space of a single PCA model. Supervised classification schemes based on building models for the known classes are Soft Independent Modeling of Class Analogy (SIMCA) approach [27] and Partial Least Squares Discriminant Analysis (PLS-DA) [28]. Because the purpose of this paper is to analyze different approaches for textural feature extraction, PCA and PLS-DA are used throughout this paper as unsupervised and supervised classification methods, respectively.

Table 1
Pre-labeled classes of the complete steel surface grayscale image data set and basic statistics of pixel intensities

Excellent surface	Good surface	Medium surface	Bad surface
Sample ID	Sample ID	Sample ID	Sample ID
E01	G01	M01	B01
E02	G02	M02	B02
E03	G03	M03	B03
E04	G04	M04	B04
E05	G05	M05	B05
E06	G06	M06	B06
E07	G07		B07
E08	G08		B08
	G09		B09
			B10
			B11
			B12

Underlined samples are used as test data in supervised classification.

3. Texture analysis using gray level co-occurrence matrix features

In this section, the GLCM is presented as representative of the statistical approaches to texture analysis. The GLCM of an image is an estimate of the second-order joint probability, $P_\delta(i,j)$ of the intensity values of two pixels (i and j), a distance δ apart along a given direction θ, i.e., the probability that i and j have the same intensity. This joint probability takes the form of a square array $\boldsymbol{P}_\delta$, with row and column dimensions equal to the number of discrete gray levels (intensities) in the image being examined. If an intensity image were entirely flat (i.e. contained no texture), the resulting GLCM would be completely diagonal. As the image texture increases (i.e. as the local pixel intensity variations increase), the off-diagonal values in the GLCM become larger.

The pixel intensity resolution of the steel surface grayscale images used in this paper is 8-bit, which result in GLCMs with dimensions of 256 rows × 256 columns for a given displacement vector. Finding GLCMs for all distances (δ) and angles (θ) would require a prohibitive amount of computation. Haralick et al. [8] suggested using GLCMs calculated from four displacement vectors with $\delta = 1$, or 2 pixels, and $\theta = 0°$, 45°, 90°, and 135°. In this example, only one GLCM was calculated for each of the 35 grayscale steel surface images using a single displacement vector with $\delta = 1$, and $\theta = 135°$ [$(x_{\text{lag}}, y_{\text{lag}})=(1,1)$]. The scale of the displacement vector was intentionally chosen to be 1 for sake of consistency and for comparison with the MIA-based texture analysis method described in Section 4.2. Because of the quite symmetric nature of the steel surfaces, a single angle was found to be adequate.

Haralick et al. [8] proposed a quantitative analysis of the GLCM through 14 textural descriptors calculated from $\boldsymbol{P}_\delta$, although typically only a few of these are widely used [7,29–31]. In this paper, four of the most commonly used descriptors (the angular second moment, contrast, correlation, and entropy) are used to extract textural features from the 35 GLCMs of the steel surface grayscale image data set.

$$\text{Angular Second Moment} = \sum_{i=1}^{n}\sum_{j=1}^{n}\{P_\delta(i,j)\}^2 \tag{1}$$

$$\text{Contrast} = \sum_{k=0}^{n-1} k^2 \sum_{i=1}^{n} \sum_{\substack{j=1 \\ |i-j|=k}}^{n} P_\delta\,(i,j) \tag{2}$$

$$\text{Correlation} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} i \cdot j P_\delta(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{3}$$

$$\text{Entropy} = -\sum_{i=1}^{n}\sum_{j=1}^{n} P_\delta(i,j)\log\{P_\delta(i,j)\} \tag{4}$$

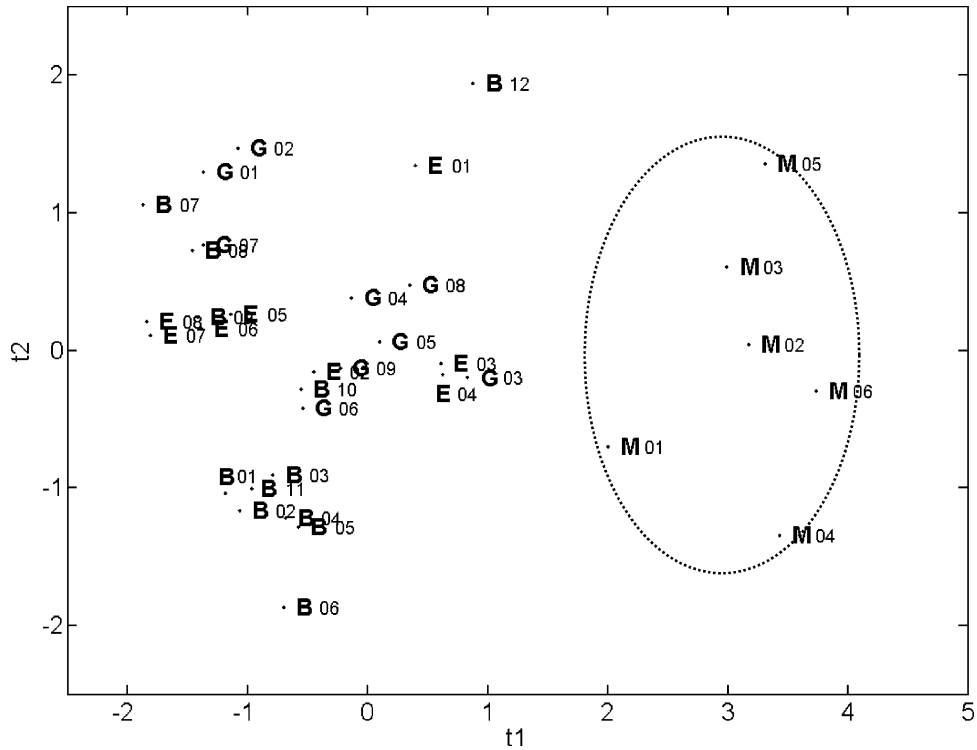


Fig. 2. Unsupervised classification of steel surface images in the latent space of GLCM features.

where the means and variances in the x and y direction are given by

$$\mu_x = \sum_{i=1}^{n} i \sum_{j=1}^{n} P_\delta(i,j), \mu_y = \sum_{j=1}^{n} j \sum_{i=1}^{n} P_\delta(i,j) \tag{5}$$

$$\sigma_x = \sum_{i=1}^{n} (i-\mu_x)^2 \sum_{j=1}^{n} P_\delta(i,j),$$
$$\sigma_y = \sum_{j=1}^{n} (j-\mu_y)^2 \sum_{i=1}^{n} P_\delta(i,j) \tag{6}$$

Fig. 2 illustrates the achieved unsupervised classification of the 35 steel surface images in the score space of PCA. It can be seen that excellent, good, and bad surfaces are not separable at all, although the t_1 axis separates the medium surfaces from the others quite well. (In Fig. 2, highlighting of the cluster boundary is done for visual purposes only.) Supervised classification using PLS-DA fails to show any improvement in separating the excellent, good, and bad surfaces.

4. Direct multivariate statistical approaches to texture analysis

In this section, we examine multivariate statistical approaches to extract textural information by applying PCA and PLS to image texture directly. The intention is to show the limitations of these approaches that arise because of their loss of spatial information, and to discuss and illustrate modifications of the approach that regain spatial information and thereby allow for efficient textural classification.

4.1. Supervised classification of steel surface images using a direct application of PLS-DA to image texture

A data matrix $\boldsymbol{X}$ is constructed, which contains n rows each corresponding to the unfolded pixel data from one steel surface image. Since the class belonging of each image is known a priori, this information is provided through a $\boldsymbol{Y}$ matrix of dummy (0,1) variables in order to train the PLS-DA regression model (Fig. 3). The model is built between $\boldsymbol{X}$ and $\boldsymbol{Y}$ for a training set of images comprising representative samples from each class. Once trained, the PLS-DA model can be used on a validation set of new images in order to predict their class belongings. This approach has been presented [32] using an equivalent but more parsimonious $\boldsymbol{X}$ matrix, with each row comprising a selected set of wavelet coefficients to represent each image.

Out of the 35 steel images, a training set of 25 images (ones without underline in Table 1) representing the four surface qualities was chosen to develop the PLS-DA regression model. The developed PLS-DA model was then tested on a test set of the remaining 10 steel sample images (underlined ones in Table 1).

Pixels from 25 steel surface grayscale images of the training set (each image has dimensions: 479 × 508 pixels) were unfolded into observations of the predictor array $\boldsymbol{X}$

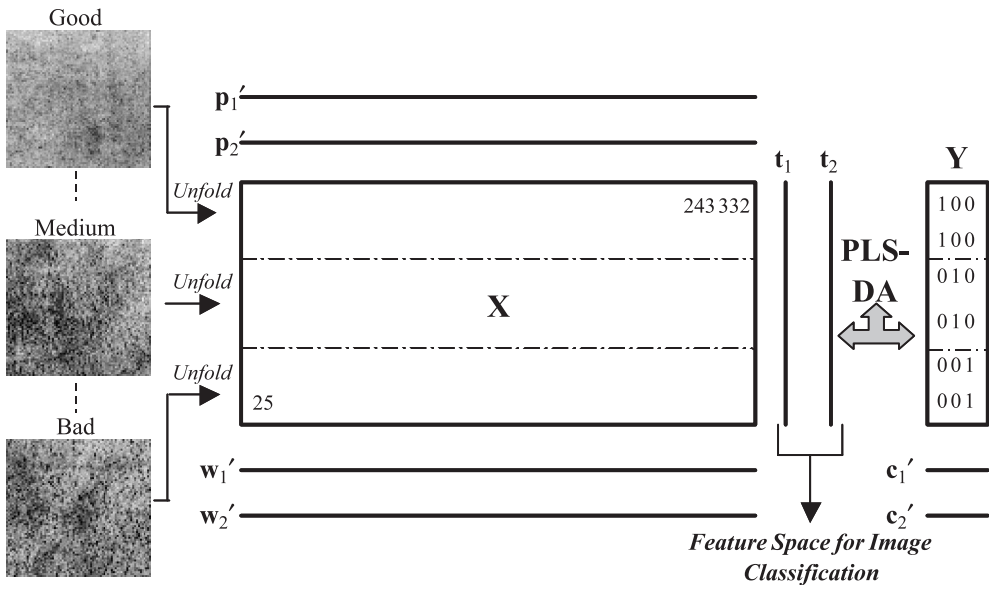


Fig. 3. Schematic of training a PLS-DA model to discriminate steel surface images.

(with dimension: 25 rows × 243,332 columns). As seen in Fig. 3, the columns of $\boldsymbol{X}$ represent unique pixel locations through the steel surface images. A kernel-based PLS algorithm [33] was used to develop the PLS-DA regression model between $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Two ($A=2$) latent variables were found to be significant and the variance explained on the fitted samples was $R_y^2 = 0.9$. The latent variables (t_1,t_2) of the resulting PLS-DA model provide a low-dimensional multivariate description of the images in the rows of $\boldsymbol{X}$ (which simultaneously accounts for their class memberships provided in $\boldsymbol{Y}$).

Fig. 4 illustrates a scatter plot of the two score vectors ($\boldsymbol{t}_1$ and $\boldsymbol{t}_2$) of the trained PLS-DA model using the steel surface training set data. Solid points represent the training set samples from four surface quality classes. According to the manually highlighted point clusters in Fig. 4 (done only for visual purposes), it can be seen that the discrimination of the training set steel sample images is very good.

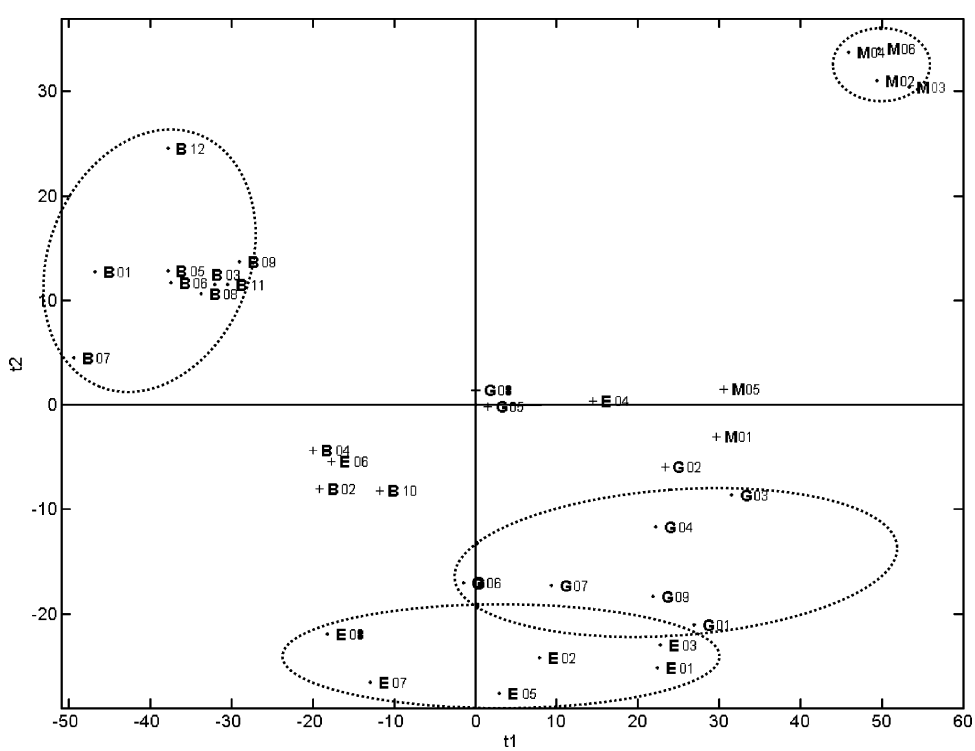


Fig. 4. Supervised classification of steel surface images using a direct application of PLS-DA to image texture.

A model developed from a training set is of little use if it cannot adequately classify new images, not included in the original training set. Hence, the developed PLS-DA model was used on the 10 steel surface images from the test set. The PLS-DA scores values (t_1,t_2) for these test images are indicated by cross ('+') points in the t_1 vs. t_2 score plot in Fig. 4. It can be seen from the figure that the PLS-DA model fails miserably in classifying the new steel surface images. Although the PLS-DA classification in the training stage produced tight and well separated score clusters (Fig. 4), none of the 10 validation set samples fell into their respective pre-labeled classes and all had scores clustering around (0,0).

The poor performance of the PLS-DA classification of texture is not surprising because spatial information is lost in these methods upon unfolding the grayscale steel surface images into row vectors of $\boldsymbol{X}$. Each column (variable) in $\boldsymbol{X}$ represents a particular pixel location, whereas each row (observation) represents a unique steel surface image. Upon PLS-DA decomposition of $\boldsymbol{X}$, the resulting weight vectors relate the importance of different pixel locations (variables) to the classification provided in $\boldsymbol{Y}$. However, unlike with chemical data or industrial data where a variable defined in $\boldsymbol{X}$ does have a consistent meaning from observation to observation, the pixel values at the same locations in the different images (columns of $\boldsymbol{X}$) do not have any meaningful relationship with one another. They represent pixel intensities from arbitrary locations on different steel surfaces. For all intensive purposes, the PLS-DA model would give similar results (in terms of data fit and predictive ability) if one were to construct $\boldsymbol{X}$ via unfolding each steel image by randomly selecting the pixel locations in each image and forming row vectors. This lack of congruency of the stacked images makes such an analysis by PCA or PLS meaningless. Furthermore, any permutations of the columns of $\boldsymbol{X}$ (e.g. exchanging columns 1 and 125 in $\boldsymbol{X}$) will result in exactly the same model. In other words, multivariate projection methods (PCA and PLS) contain no spatial information on the relative location of the pixels with respect to one another.

Since texture is a function of spatial variations in neighboring pixel intensities throughout an image, the absence of spatial information in PCA or PLS models means they cannot be used directly for texture analysis. However, there are several ways in which spatial information can be reintroduced into PCA and PLS image models. These are described in the following section.

4.2. Classification using MIA of steel surface images augmented with spatial information

The lost of spatial information (upon unfolding images) could be regained to a certain extent if each individual texture image were suitably augmented with different versions of itself using various techniques to form new variables of a multivariate image. The resulting data set may then be analyzed using MPCA and Multivariate Image Analysis (MIA) techniques [34]. In that case, the model would be forced to explain local variations of pixel intensities over a pre-defined neighborhood.

There are several ways of regaining spatial information through augmenting the steel image. One approach is to apply multivariate statistical techniques to a suitable spatial transform of the image data such as the two-dimensional Fast Fourier Transform (2-D FFT) spectrum. This approach is further discussed in the next section under transform-based methods.

Other approaches are to augment each image with several filtered versions of the same image using any set of the spatial filters available in most image processing software, and then use MIA based on multi-way PCA or PLS to build the classification models. Such an approach was used by Lied et al. [34] for texture-based discrimination of different classes within an image. Three different texture filters (Median filter, Laplace filter, and a compound filter with a combination of various filters) were used to filter the image at each pixel location and the filtered versions of the image were used as new channels of a multivariate image. A stack of congruent images was created consisting of the original image with all its filtered versions aligned and stacked above it. Hence at each pixel location, one had the original grayscale signal and directly above (or below) it the value of various spatial filters applied at that pixel location. Classification of new images is then performed by segmentation (masking) in the feature (score) space of PCA.

Difficulties in the above approach are to decide what 'potpourri' of spatial filters to apply in order to obtain the augmented image as well as the increase in the size of data. We therefore present an equivalent approach, but where the data itself is automatically used to determine the optimal spatial filters that should be applied in order to obtain the best classification. In this approach, the spatial pixel intensity distribution in the steel surface image is captured through spatially shifting the image in different directions, and then stacking the shifted images on top of each other to form a three-way pixel array, as shown in Fig. 5. The resulting three-dimensional image data is a multivariate image where the third dimension is the spatial shifting index. Each image in such a stack would contain the same feature information, but at each pixel location, the vertical direction would contain information at the pixel of interest, as well as at all spatial locations surrounding it. Multi-way PCA will then find scores that are linear combinations of these spatially shifted pixels defined by the loadings. Each score will therefore represent some form of spatial derivative filter that explain the greatest variability in the training image.

The only disadvantage in this scheme of image augmentation is that again one has a greatly expanded data set with many more columns in the unfolded $\boldsymbol{X}$ matrix. In this approach, a single training image must be used, since as discussed earlier, there is no congruency among pixels of different images. Hence, the training image must contain elements of all features that are important in the classification. Therefore, a composite image made up of digitally combining

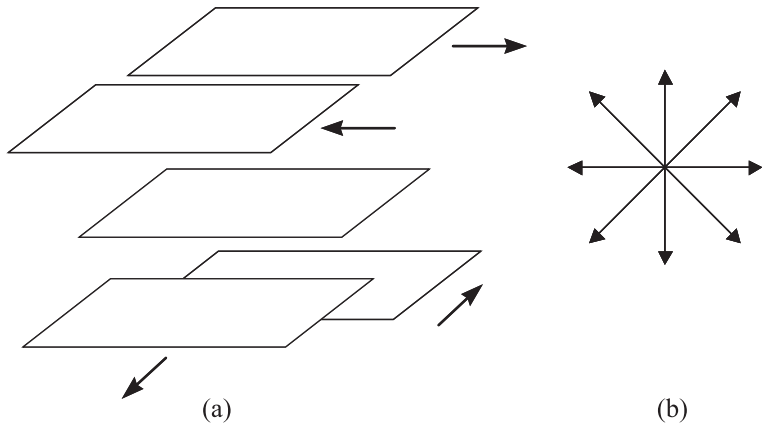


Fig. 5. (a) A multivariate image created via spatial shifting in four adjacent directions and stacking the shifted images. (b) Eight possible directions in which an image could be shifted.

several images (e.g., excellent, good, medium and bad steel images) is usually used. Features are then obtained by segmenting (masking) the PCA score plots (as commonly done in MIA) into regions that define different texture characteristics. New images are then classified by evaluating the number of pixels falling into these feature masks. This is illustrated below for the steel surface examples.

4.2.1. Texture analysis of steel images using MIA on shifted and stacked images

Performing MIA of a multivariate image resulting from spatially shifting and stacking an image (Fig. 5) allows the latent variables of multi-way PCA to extract almost any two-dimensional spatial filter structure allowable by the amount (and direction) of shifting. The steel surface image sample previously shown in Fig. 1d representing bad steel surface quality is used in this study as the training image. It contains almost all features present in the 35 steel samples, and hence was deemed to provide a reasonable training image. The original image was spatially shifted in eight adjacent directions (see Fig. 5b) by 1 pixel, and the shifted images were stacked above the original to form a nine-variable multivariate image. After shifting and stacking, the resulting three-way array was cropped at the edges to discard all the non-overlapping sections. MPCA decomposition was then performed on the resulting multivariate image array ($\underline{\boldsymbol{X}}_{\text{bad}}$) without any pre-scaling or mean centering of the data. The cumulative percent sum of squares explained by the first three PCs was 99.20%. Only the first three PCs have been used in subsequent analyses, with the rest of the PCs (four to nine) being attributed to explaining noise in the multivariate image. Table 2 shows the corresponding weights of the first three loading vectors ($\boldsymbol{p}_1$,$\boldsymbol{p}_2$,$\boldsymbol{p}_3$) with respect to the nine variable images of $\underline{\boldsymbol{X}}_{\text{bad}}$. The loading vector coefficients have been re-arranged as 3×3 arrays to correspond with the respective spatial locations of the shifted pixels in the original image.

(a)

$\boldsymbol{p}_1$	0.33	0.33	0.33
	0.34	0.34	0.34
	0.33	0.34	0.33

(b)

$\boldsymbol{p}_2$	-0.39	-0.45	-0.38
	0	0	0
	0.38	0.45	0.39

(c)

$\boldsymbol{p}_3$	0.38	0	-0.38
	0.45	0	-0.45
	0.39	0	-0.38

Since no pre-scaling of the image data was performed the first Principal Component (PC) explains mainly the average pixel intensity over their local area in the multivariate image. This is evident from the fact that all nine $\boldsymbol{p}_1$ loading coefficients are positive and almost equal (Table 2a). The resulting $\boldsymbol{T}_1$ score image of the first PC turns out to be a blurred version of the original image due to the averaging nature of the $\boldsymbol{p}_1$ loading coefficients. This is due to the fact that PC1 extracts only the pixel contrast information from the multivariate image via averaging over a 3×3-pixel neighborhood around each pixel of the steel surface image. Thus it can be said that PC1 serves as a smoothing filter on the original steel surface image.

Upon deflating the $\boldsymbol{X}$ matrix by removing this mean intensity prediction of the first PC, the second and third PCs of MIA extract the remaining feature information. Fig. 6a and b illustrates the second and third PC score images $\boldsymbol{T}_2$ and $\boldsymbol{T}_3$ of the original steel surface image, respectively. A close observation of the $\boldsymbol{T}_2$ image reveals that the 2nd PC predominantly extracts horizontal and diagonal edge information (i.e. 45° and 135°) with respect to the center of the image. On the other hand, it can be seen from the $\boldsymbol{T}_3$ score image that the main feature extracted by PC3 is the vertical and diagonal surface pit edge information.

The $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ loading vector coefficients are given in Table 2b and c, respectively. It can be seen that the second PC is simply a vertical first derivative filter, and hence will extract mainly horizontal edge information. Similarly, the loadings of the third PC shown in Table 2c shows that $\boldsymbol{p}_3$ is simply a horizontal derivative filter which highlights mainly the vertical edges of the surface pits as illustrated in the T3 image reconstruction in Fig. 6b. The sum of the loading coefficients for $\boldsymbol{p}_1$ is approximately 1, and the sum of the coefficients for both $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$ is approximately 0. This agrees with the convolution kernels of a low pass smoothing

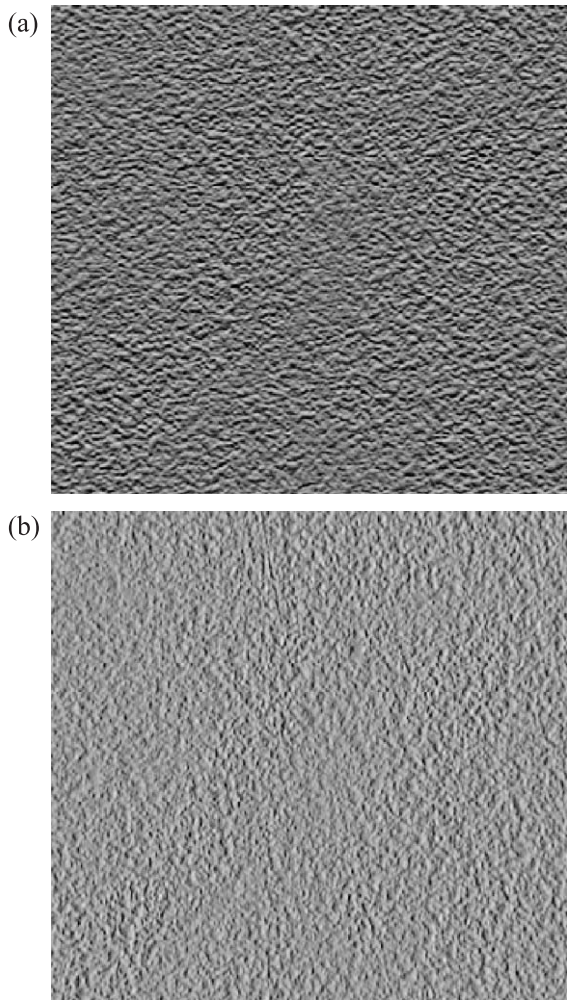


Fig. 6. (a) T_2 image of bad steel surface quality image; (b) T_3 image of bad steel surface quality image.

filter and first-derivative edge detection filters in the image processing literature [1,2,25,35,36].

It can be seen from this example that MIA on spatially shifted and stacked images automatically allows one to develop optimal filters as loading vectors based on pixel intensity variance over a pre-defined neighborhood. In general, these filters could be much more complex than the simple smoothing and first-derivative edge detection filters obtained above. Depending upon the number of pixel shifts and the chosen spatial direction(s) of shifting, the MPCA loading vectors could define much more complex filters, which one might never have anticipated in advance. These optimal spatial filters could then be used directly for all subsequent analysis. In this study, up to three shifts in each direction were tried with different shift intervals, with no real improvement. Only three significant PCs were still obtained and the loading vectors of these were essentially the same as obtained with a single shift, showing that only the first-derivative filters were needed.

Besides observing the MIA PCs as intensity images (image space), one could also use scatter plots of score vectors against each other and observe the pixels as point clusters (score space) in a color-coded two-dimensional histogram. Fig. 7 shows the PC12 (t_1 vs. t_2) and PC23 (t_2 vs. t_3) score plot of the bad steel surface image.

Further insight of the MIA score space can be gathered upon interrogating score point clusters using a masking strategy [37] in the score space to delineate pixels having similar features and then highlighting the pixels under the score plot mask in the $\boldsymbol{T}_1$ score image. Fig. 8a illustrates such a mask (shown as a gray rectangle) in the PC12 score plot of the training image that captures all pixels having low t_1 values. The corresponding masked pixels have been highlighted (as white) and overlaid on the $\boldsymbol{T}_1$ score image (shown in Fig. 8b). This low t_1 mask clearly captures all the deep pits in the image.

Since the second and third PCs have been shown to represent first-derivative filters in the vertical and horizontal directions, respectively, the large positive or negative values of t_2 and t_3 will correspond to pixel locations where rapid intensity changes are occurring. Fig. 9a illustrates a mask

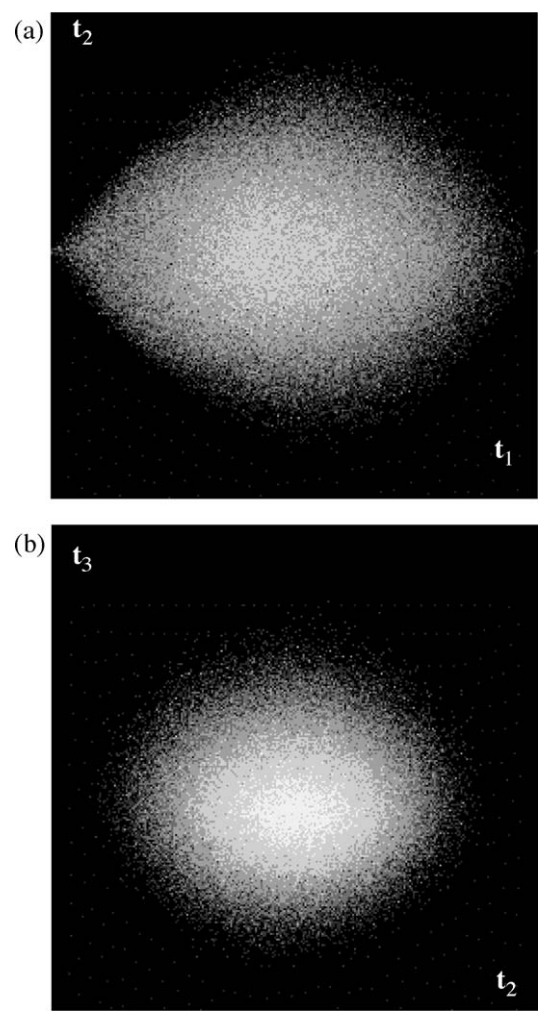


Fig. 7. (a) Score space of PC12 for bad steel surface quality image; (b) score space of PC23 for bad steel surface quality image.

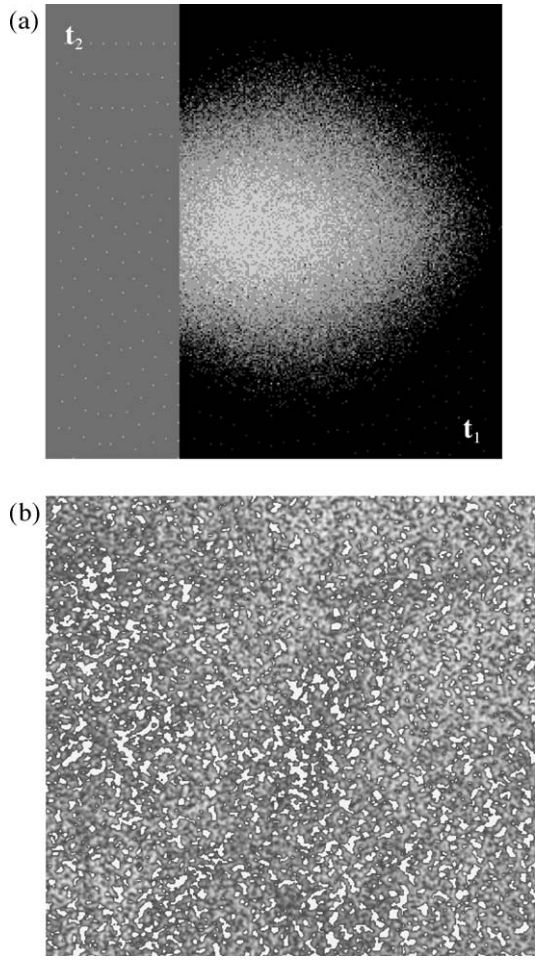


Fig. 8. (a) Manually applied mask on PC12 score space of bad steel surface image; (b) corresponding feature pixels under PC12 mask highlighted (in white) and overlaid on T_1 score image.

(shown in gray around the central cluster) that highlights the extreme score combinations in the PC23 score plot of the training image. The corresponding pixels covered by this mask have been highlighted (as white) and overlaid on the $\boldsymbol{T}_1$ image as shown in Fig. 9b. Clearly, this mask highlights those pixels belonging to the pit edges in the image.

4.2.2. Classification of steel images using MIA model features

The MIA model and the score space masks can now be used to extract similar texture properties from each steel surface image in the data set and to classify the steel surfaces. Each new steel surface image is shifted and stacked and then unfolded in the same way as the training image, and the PCA model used to obtained score plots (t_1,t_2,t_3) for that image. Features for classification are then taken as the count of the total number of pixels falling under each of the score space masks for each steel surface image. These features provide an objective measure of image texture through a count of pixels belonging to pit cores and pit edges in each steel surface image [38].

Classification can then be achieved by plotting the respective pixel counts of pit cores and edges against each other for every image [39] as shown in Fig. 10. Such a plot produces an appropriate feature space for image classification. Steel surfaces depicting similar overall texture characteristics should have (on average) similar feature pixel counts of pit cores and edges. The abscissa separates samples representing bad steel surface quality from the other samples. This trend is expected since the bad surface quality samples contain mainly deeper pit cores, which occupy a larger surface area as they are joined together in ‘snake’-like patterns. The ordinate represents derivative (edge) information, and as seen in Fig. 10, this direction mainly separates samples in the excellent and good surface quality from those belonging to medium surface quality. The above implies that medium quality steel surfaces have a similar number of deep pit cores as the excellent and good surface quality surfaces, but have more pit edges than excellent or good surfaces.

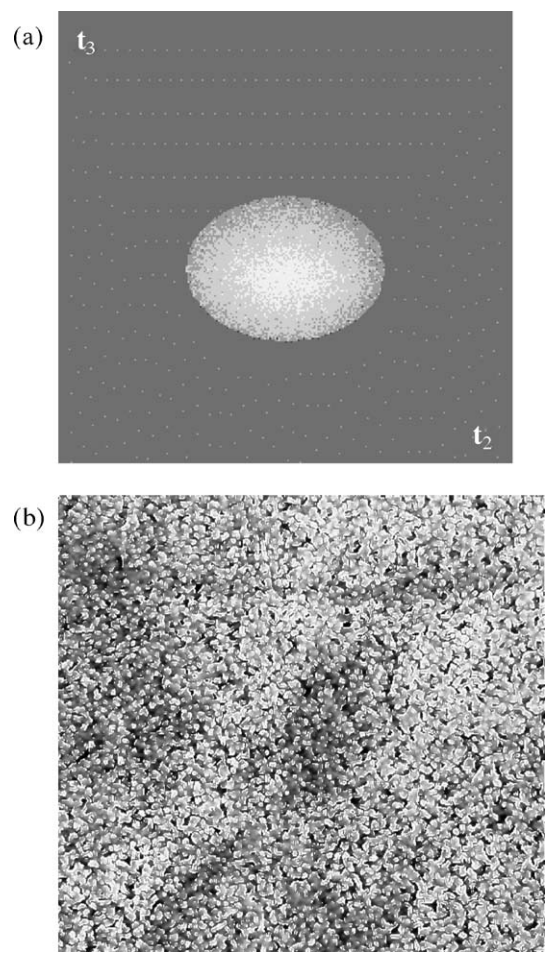


Fig. 9. (a) Manually applied mask on PC23 score space of bad steel surface image; (b) corresponding feature pixels under PC23 mask highlighted (in white) and overlaid on T_1 score image.

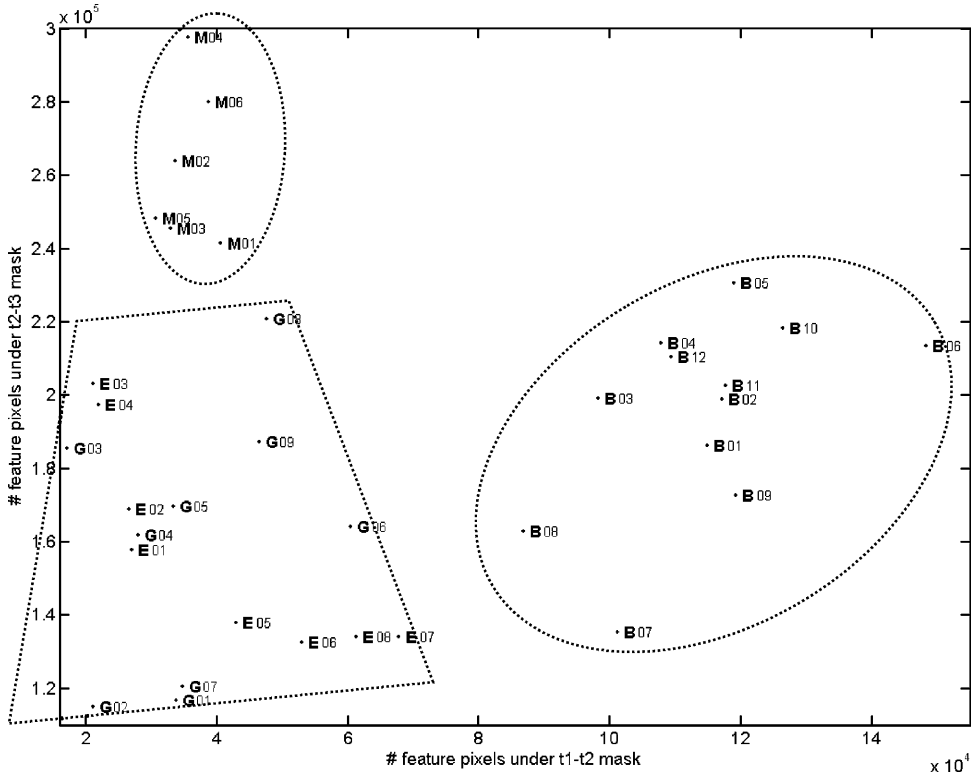


Fig. 10. Steel sample image classification based on surface pits and edges detected by MIA texture analysis strategy.

In summary, as seen from the manually highlighted clusters in Fig. 10 (done for visual purposes only), one can easily separate medium and bad surfaces from the excellent and good surfaces. However, the excellent and good surfaces cannot be distinguished from one another. The proposed MIA-based image texture analysis method produces better classification results than the earlier discussed PLS-DA and GLCM approaches. Furthermore, being inherently multivariate in nature the MIA texture analysis technique has the added advantage that it can be applied not just to grayscale images, but also to true color (RGB) and other multi-spectral images.

5. Transform-based methods for texture classification

5.1. Unsupervised classification using MPCA on 2-D FFT magnitude images

The two-dimensional Fast Fourier Transform (2-D FFT) of an image can be thought of as a two-dimensional representation of the spatial power spectrum of the image. Theoretical and mathematical details regarding the 2-D FFT can be obtained from the image-processing literature [1,2,25,35,36]. Various researchers have proposed the use of FFT spectra as texture feature descriptors [7] that can be used for characterizing images based on overall texture [40], or for multivariate prediction of externally measured textural data [17].

The approach "ASUNIM" (analysis of a set of univariate images) suggested by Geladi [40] is used in this section to classify the steel surface images according to their surface texture. A multivariate image consists of several *congruent* variable images. In order to transform several incongruent grayscale images into a common base, Geladi suggested converting them into their respective 2-D FFT magnitude images and stacking them as variables of a multivariate image. In doing so, the resulting multivariate image conserves textural frequency information in its variables. Upon decomposing the multivariate image using MPCA, classification can be performed in the scatter plots of the MPCA loading vectors.

The 35 steel surface images in the data set were transformed into their respective 2-D FFT magnitude images, followed by stacking them into a (35 variable) multivariate image. However, prior to stacking, each 2-D FFT magnitude image was passed through a "Gaussian" filter (9×9 pixel convolution kernel, with $\sigma = 0.5$) [35]. This filter serves as a low-pass smoothing function, which is mainly used to remove traces of high-frequency noise from the texture images. Further details on the Gaussian filter and other similar windowing functions may be found in Jenkins and Watts [41]. Finally, each of the 35 2-D FFT magnitude images was cropped. Only the right half was used since the left half was a symmetric mirror of the right half.

99.78% of the total variation of multivariate image $\underline{X}$ was explained by the first three PCs (PC1 = 99.76%;

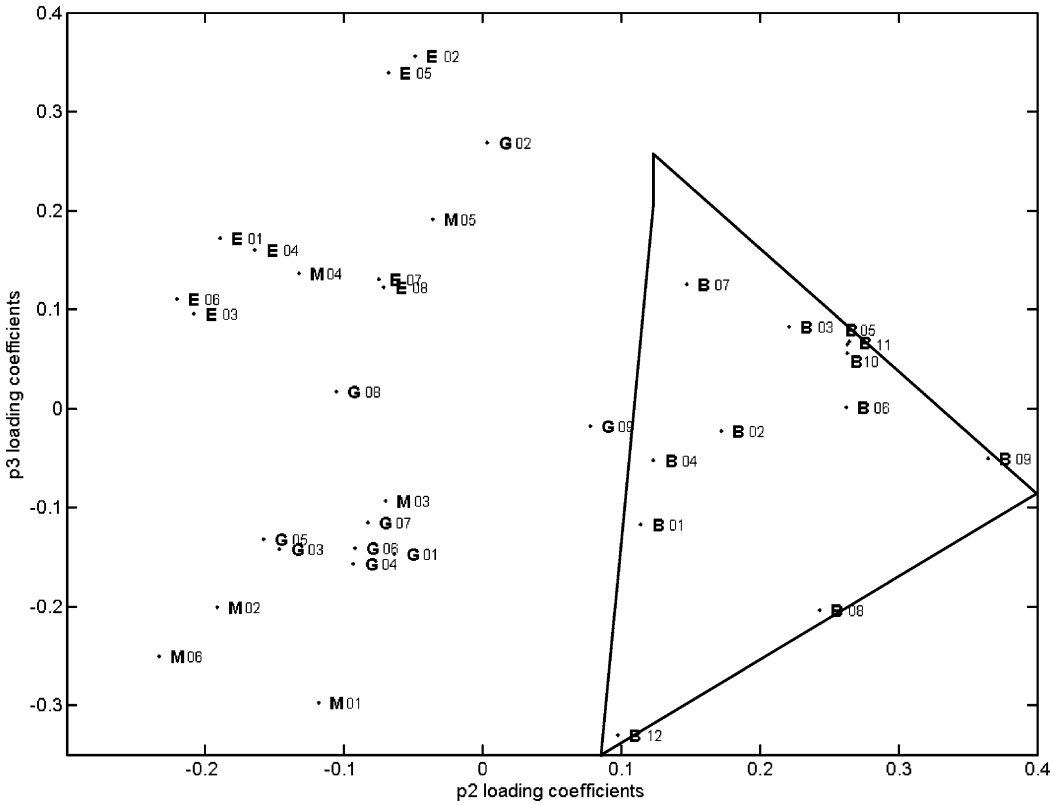


Fig. 11. Steel surface image classification using MPCA on 2-D FFT magnitude of steel images.

PC2 = 0.014%; PC3 = 0.011%). The first PC explains the average magnitude information in the variable images of $\underline{\boldsymbol{X}}$, whereas contrast differences between the variables (i.e., images) is emphasized by subsequent PCs. Keeping this in mind, the loading space of PC2 and PC3 is used to discriminate the 35 variable images of $\underline{\boldsymbol{X}}$. Upon scatter plotting, the $\boldsymbol{p}_2$ loading coefficients of the 35 variable images against those of $\boldsymbol{p}_3$, each image is represented as a single point in the feature space. The resulting point clusters in the feature space may then be used to test the achieved classification of the steel surface images. Fig. 11 illustrates the results of this unsupervised classi-

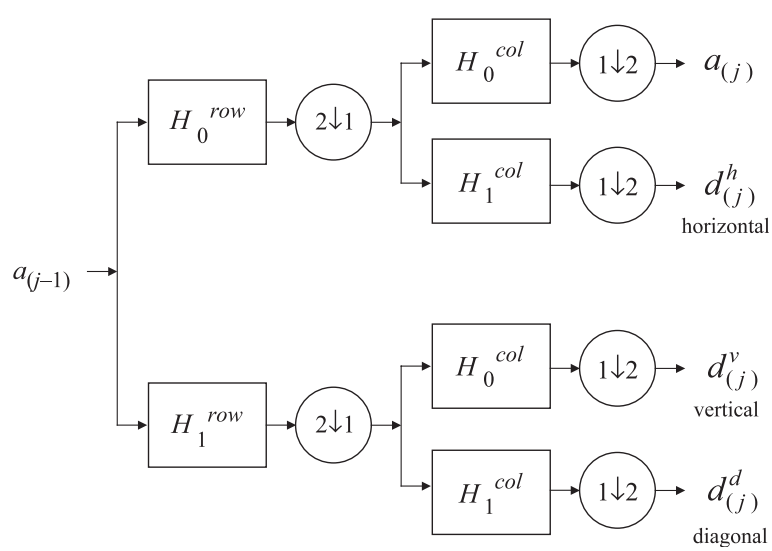


Fig. 12. Two-dimensional discrete wavelet transform implemented by a separable two-dimensional filter bank (only the *j*th decomposition stage shown here). It consists of a cascade of horizontal (indicated as *row*) vertical (indicated as *col*) filtering. Since this results in strong directionality, three detail subimages are called horizontal (*h*) vertical (*v*) and diagonal (*d*), respectively.

fication scheme. Images exhibiting similar spatial frequency patterns of steel surface roughness are grouped together to form a class.

According to the manually highlighted classes in Fig. 11 (done for visual purposes), it can be seen that the achieved classification using the above strategy is better than those of GLCM and direct PLS-DA, but not as good as that based on MIA of the shifted and stacked images. The bad surfaces are well separated, but there is no clear separation of the others (although the excellent ones are reasonably well clustered).

5.2. Texture classification using wavelet texture analysis (WTA)

Whereas the 2-D FFT performs a frequency decomposition of an image, the Gabor transform and 2-D wavelet transform perform a space-frequency decomposition, which is more suitable for texture analysis. The wavelet transform is preferred to the Gabor transform in the respect that the wavelet transform can maintain good space and frequency localization when discretized [42,43]. For these reasons, a wavelet-based method, which is often called wavelet

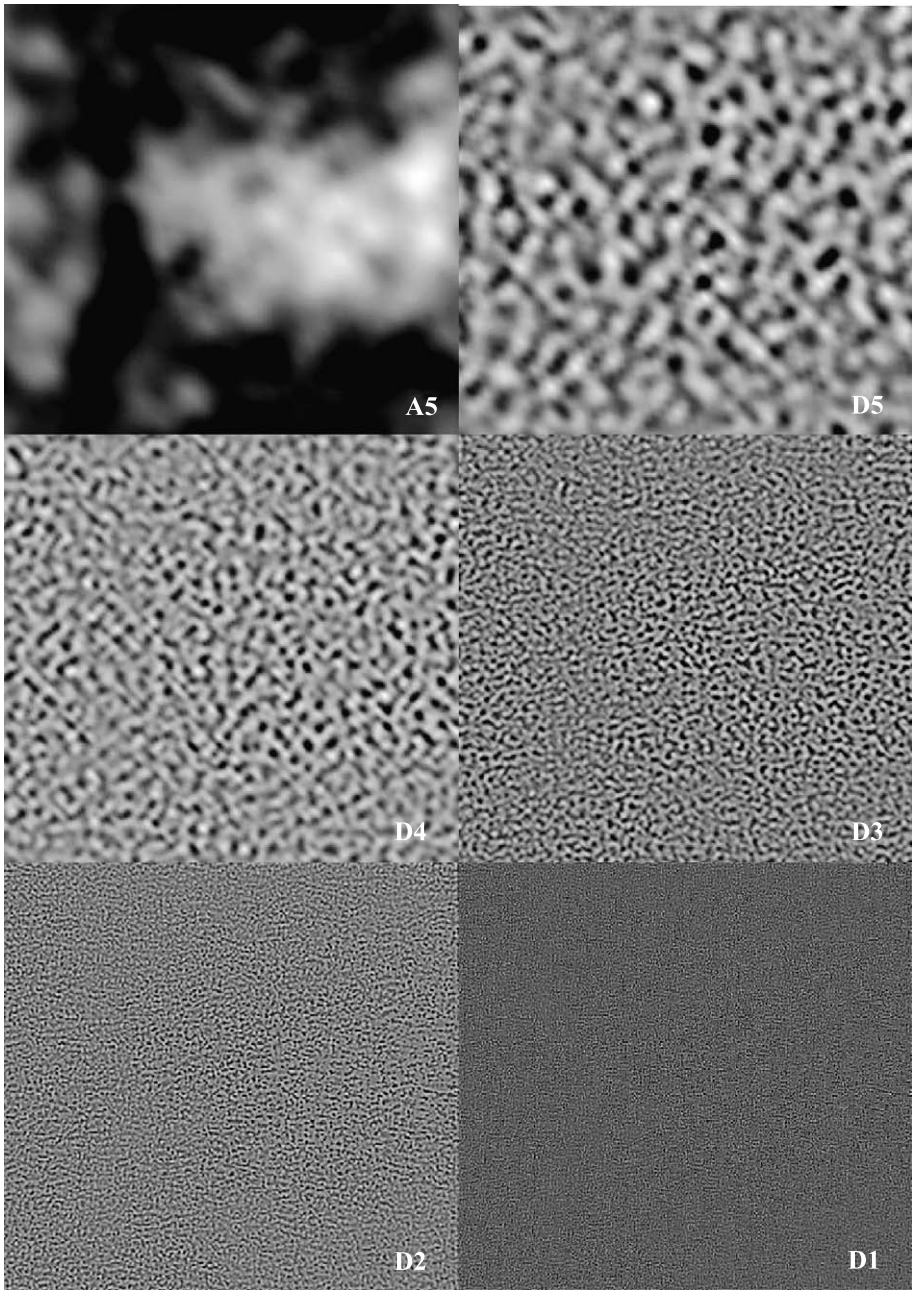


Fig. 13. The reconstructed sub-images of steel sample B12 for the 5 details (D1–D5) and the approximation (A5).

texture analysis (WTA), is considered to be the current state of the art among other texture analysis methods and also has shown better performance than other methods in many cases [5].

A basic idea of WTA is to generate a textural feature from wavelet coefficients at each resolution. Then it is assumed that each texture has its unique distribution of features at all resolutions. Therefore, if the frequency spectrum is decomposed appropriately, different texture will have different features. Denote by $d_{(j)}^{k}$ ($j=1, 2, \ldots, J$; $k=h,v,d$) the k-component of the detail sub-images of a grayscale image at the jth decomposition stage, where h, v, and d denote horizontal, vertical, and diagonal, respectively, as shown in Fig. 12. When each detail sub-image is treated as a matrix, then the energy of the detail sub-image is defined as

$$E_{jk} = \| d_{(j)}^{k} \|_F^2 \tag{7}$$

where $\| \cdot \|_F$ denotes the Frobenius norm. Often this is divided by the number of pixels, yielding normalized energy. When these energies are employed as elements of the textural feature vector, it is called the *wavelet energy signature* [6], which is the most popular feature used in WTA. Other popular textural features are entropy or averaged l_1-norm. The size of the energy signature is then $3J$ (J is the number of decomposition stages) for a grayscale image. The variations induced by lighting or illumination are usually captured in the approximation sub-image and thus it is generally not included. Based on this idea, many publications on WTA have appeared [4–6,20–22]. Since mean values of detail coefficients or sub-images are equal to zero [21], the wavelet energy signatures are equal to channel variances when divided by the number of pixels. In addition, the entropy signatures are equivalent to high-order moments of pixel values.

For WTA of the steel surfaces, each image was decomposed to the fifth stage using orthogonal Coiflet wavelet with a 6-tap filter, which is orthogonal and nearly symmetric [44]. For the choice of the wavelets, we followed the general guidelines given in Ref. [45]. Based on experience, we selected the number of decomposition stages such that the size of the smallest sub-image was greater than 10×10. This criterion is similar to that of Ref. [20].

To illustrate why WTA is a very effective tool for texture analysis, the reconstructed images for each of the five detail stages and the approximation are shown in Fig. 13 for the bad steel sheet sample B12 in Fig. 1d. To accomplish this, the three reconstructed detail sub-images (horizontal, vertical, and diagonal) at each decomposition stage are added to give one reconstructed detail sub-image.

The very informative nature of WTA is clearly apparent in the sub-images in Fig. 13. The image D1, which corresponds to the detail sub-image with highest spatial frequency, has very fine bright and dark dots and they are evenly distributed. The size of features (dots) in the sub-images gets progressively bigger as one moves from D1 to D5. The approximation sub-image A5 captures the average lighting intensity variation across the image including all the ink smudge marks (compare with the original image in Fig. 1d). Clearly, these features in A5 are unrelated to the important textural differences among the samples and should be omitted from the analysis. Omitting approximation coefficients (A5) is equivalent to performing powerful high-pass filtering of the image data. When comparing images D1–D5 with the original image Fig. 1d, one can easily note that dark spots in the reconstructed detail sub-images correspond to the pits in the original image; D1 and D2 have very small pits and D5 and D4 have big and deep pits. Then, it can be expected from the explanation of surface qualities in Section 2 that excellent surfaces will have larger energy signature in the details D1 and D2 whereas bad surfaces will have larger energy signatures in the details D4 and D5.

To illustrate this, frequency distributions of the energy signatures for the five details of three images in Fig. 1 are shown in Fig. 14 by plotting the energy vs. frequency (D5, …, D1). Again, energies of three detail sub-images (h,v,d) at each decomposition stage are summed for illustration. The shift in the energy distribution from D1 towards D5 as one moves from the excellent sheet to the bad one is clearly evident. Also evident are the lower energy values at all scales for the excellent sample.

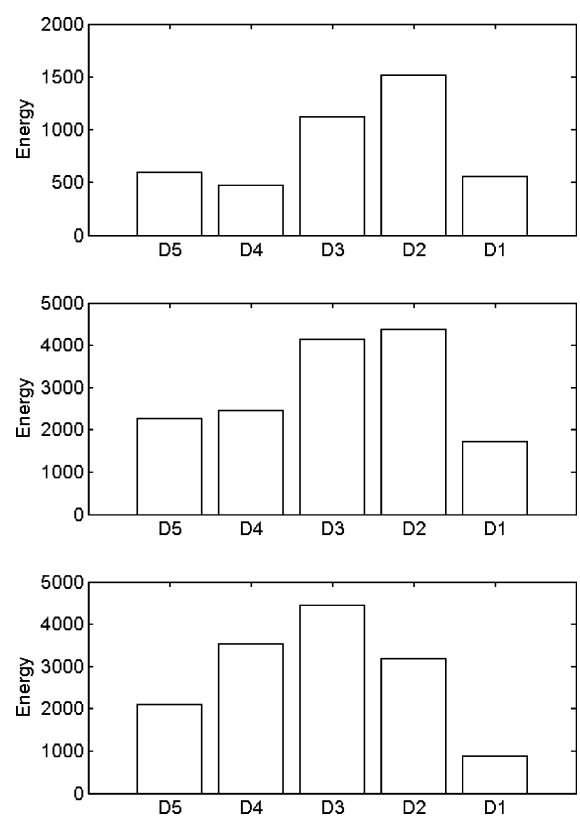


Fig. 14. The WTA energy distribution of steel images. Top-E02, middle-M06, bxottom-B12.

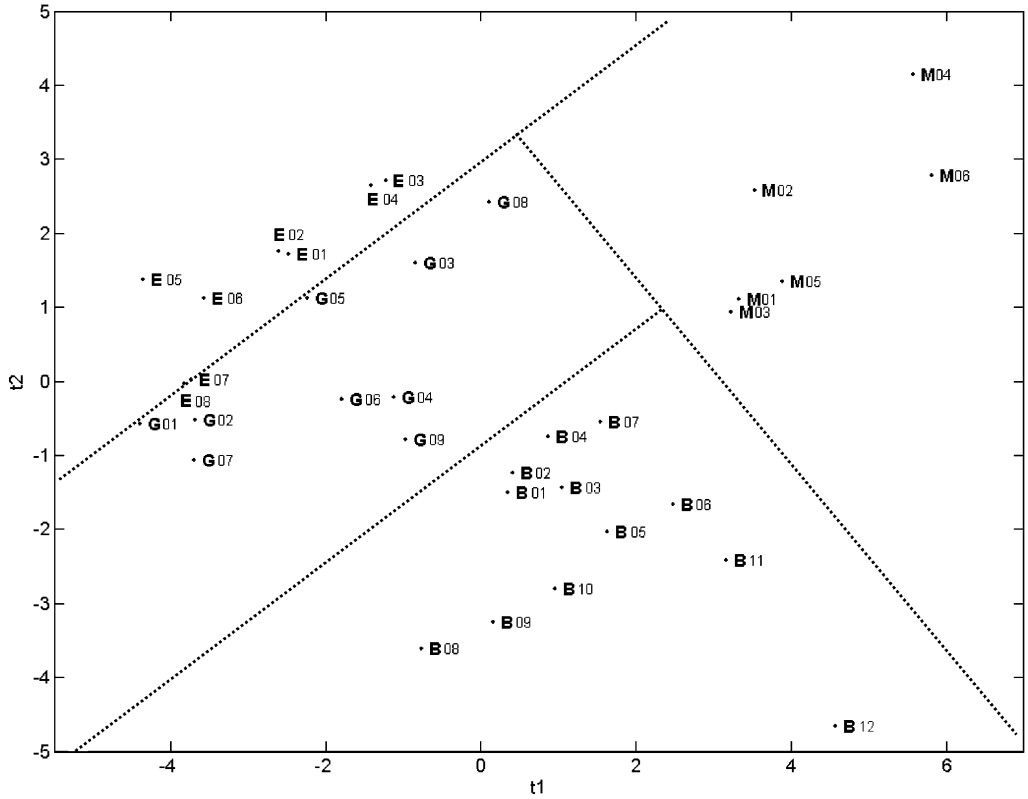


Fig. 15. Unsupervised classification of steel surface images using PCA score plot (t_1 vs. t_2) of wavelet energy signature.

Classification is performed using the energies of all the detail sub-images as textural features. Thus, the dimension of a feature vector is 15 ($=3*J$). PCA and PLS-DA are then used as a clustering and a classification method. The number of significant components determined by cross-validation was five, which is equal to the number of decomposition stages. However, the class representations are easily seen in the score plot of the first two components (t_1 vs. t_2) shown in Fig. 15. This wavelet texture analysis approach is the only one among the approaches discussed in this paper that provided clear separation lines to be drawn among the classes. It also shows the expected progressive behavior between the good and excellent classes while other approaches failed to show that. The PLS-DA t_1 vs. t_2 score plot is nearly as same as the PCA score plot.

6. Conclusions

Several approaches to image texture analysis are overviewed and are used to classify a set of industrial rolled steel sheet samples into various quality grades. Some of the strengths and weaknesses of the different methods become apparent in the discussion of the methodology and in the application to the steel sheet classification. The older, more traditional statistical approaches, represented by the GLCM method in this paper, appear to have been supplanted by more recent approaches based on multivariate image analysis and, in particular, by wavelet texture analysis.

The use of multivariate classification approaches such as PLS discriminant analysis applied to *directly* to the image data are shown to fail because of the loss of spatial identity of the pixels (variables) in those approaches, and because of the lack of congruency of the images. The congruency problem can be overcome by working with the 2-D FFT of the images rather than the raw image themselves. The 2-D FFT also captures some of the spatial frequency variations in the image, and multivariate image analysis (MIA) methods based on multi-way PCA can then be used to classify the textural information in the images. An alternative approach to regain spatial information is to create several versions of each image using different spatial filters, and to use MIA to analyze the expanded set of multivariate images. A new alternative, proposed in this paper, is to use spatially shifted versions of the image and analyze the stacked images by MIA. The latter approach is also shown to provide a way of optimally selecting spatial filters for texture analysis.

Wavelet texture analysis (WTA) methods currently appear to be the most powerful approach to image texture analysis. Two dimensional wavelet transforms perform a space-frequency decomposition which is more suitable than the frequency only decomposition provided by the 2-D FFT. The ability to selectively filter out features such as lighting

intensity variations and smudges from the analysis by removing the low-frequency approximation sub-image or other relevant detail sub-images also makes this approach more flexible and powerful.

The steel sheet image set, used in this paper to illustrate the various methods, gives insight into the methods, and provides a reasonable comparison of their strengths and weaknesses.

References

[1] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison-Wesley, Reading, MA, 1992.
[2] J.C. Russ, The Image Processing Handbook, 3rd edition, CRC Press, Florida, 1999.
[3] IEEE Standard 610.4-1990.
[4] S. Livens, *Image Analysis for Material Characterization*, PhD thesis, University of Antwerp, Antwerp, Belgium, 1998.
[5] T. Randen, *Filter and Filter Bank Design for Image Texture Recognition*, PhD thesis, NTNU, Stavanger, Norway, 1997.
[6] G. Van de Wouwer, *Wavelets for Multiscale Texture Analysis*, PhD thesis, University of Antwerp, Antwerp, Belgium, 1998.
[7] F. Tomita, S. Tsuji, Computer Analysis of Visual Textures, Kluwer Academic Publishing, Massachusetts, 1990.
[8] R.M. Haralick, K. Shanmugam, I. Dinstein, IEEE Trans. Syst. Man Cybern. 3 (1973) 610–621.
[9] M.M. Galloway, Comput. Vis. Graph. Image Process. 4 (1975) 172–179.
[10] B. Ashjari, Singular Value Decomposition Texture Measurement for Image Classification, PhD thesis, University of Southern California, Los Angeles, CA, 1982.
[11] K. Kvaal, J.P. Wold, U.G. Indhal, P. Baardseth, T. Næs, Chemometr. Intell. Lab. Syst. 42 (1998) 141–158.
[12] L. Carlucci, Pattern Recogn. 4 (1972) 53–72.
[13] S.W. Zucker, Comput. Vis. Graph. Image Process. 5 (1976) 190–202.
[14] A. Sarkar, K.M.S. Sharma, R.V. Sonak, IEEE Trans. Image Process. 6 (1997) 407–413.
[15] G. Cross, A. Jain, IEEE Trans. Pattern Anal. Mach. Intell. 5 (1983) 25–39.
[16] J.M. Keller, S. Chen, R.M. Crownover, Comput. Vis. Graph. Image Process. 45 (1989) 150–166.
[17] U. Indhal, T. Næs, J. Chemometr. 12 (1998) 261–278.
[18] P. Geladi, Chemometr. Intell. Lab. Syst. 14 (1992) 375–390.
[19] A. Bovik, M. Clark, W. Geisler, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1990) 55–73.
[20] T. Chang, C.C.J. Kuo, IEEE Trans. Image Process. 2 (1993) 429–441.
[21] M. Unser, IEEE Trans. Image Process. 4 (1995) 1549–1560.
[22] A. Laine, J. Fan, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1995) 1186–1191.
[23] J. Huang, K.H. Esbensen, Chemometr. Intell. Lab. Syst. 54 (2000) 1–19.
[24] ftp://130.113.90.10/misc/Steel_Images.zip.
[25] C.M. Thompson, L. Shure, Image Processing Toolbox for use with MATLAB®: User's Guide, The MathWorks, Natick, MA, 1995.
[26] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA and PLS), Umetrics, Umeå, Sweden, 1999.
[27] S. Wold, C. Albano, W.J. Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg, M. Sjöström, in: B.R. Kowalski (Ed.), Chemometrics: Mathematics and Statistics in Chemistry, D. Reidel Publishing, Dordrecht, Holland, 1984, 17–95.
[28] M. Sjöström, S. Wold, B. Söderström, in: E.S. Gelsema, L.N. Kanal (Eds.), Pattern Recognition in Practice II, Elsevier, Amsterdam, 1986.
[29] J.S. Weszka, C.R. Dyer, A. Rosenfeld, IEEE Trans. Syst. Man Cybern. 5 (1976) 269–285.
[30] J.D. Wood, *The Geomorphological Characterization of Digital Elevation Models*, PhD thesis, University of Leicester, UK, 1996.
[31] A. Al-Janobi, Pattern Recogn. 34 (2001) 171–180.
[32] Umetrics, SIMCA-P 9.0: User's Guide, Umetrics, Umeå, Sweden, 2001.
[33] S. Rännar, F. Lindgren, P. Geladi, S. Wold, J. Chemom. 8 (1994) 111–125.
[34] T. Lied, P. Geladi, K. Esbensen, J. Chemom. 14 (2000) 585–598.
[35] G.A. Baxes, Digital Image Processing: Principles and Applications, Wiley, New York, 1994.
[36] W.K. Pratt, Digital Image Processing, Wiley, New York, 1978.
[37] K. Esbensen, P. Geladi, Chemometr. Intell. Lab. Syst. 7 (1989) 67–86.
[38] M.H. Bharati, Multivariate Image Analysis and Regression for Industrial Process Monitoring and Product Quality Control, PhD thesis, McMaster University, Hamilton, Canada, 2002.
[39] M.H. Bharati, J.F. MacGregor, Proceedings of SPIE-Process Imaging for Automatic Control, Boston, MA, November 5–6.
[40] P. Geladi, Chemometr. Intell. Lab. Syst. 14 (1992) 375–390.
[41] G.M. Jenkins, D.G. Watts, Spectral Analysis and its Applications, Holden-Day, San Francisco, CA, 1969.
[42] M. Vetterli, J. Kovačević, Wavelets and Subband Coding, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[43] S.G. Mallat, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 674–693.
[44] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, Wavelet toolbox for use with MATLAB, The MathWorks, Natick, MA, 1996.
[45] U.E. Ruttimann, M. Unser, R.R. Rawlings, D. Rio, N.F. Ramsey, V.S. Mattay, D.W. Hommer, J.A. Frank, D.R. Weinberger, IEEE Trans. Med. Imag. 17 (1998) 142–154.