Ind. Eng. Chem. Res. **1998,** *37,* 4715−4724

Multivariate Image Analysis for Real-Time Process Monitoring and Control

M. H. Bharati and J. F. MacGregor*,†

Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada

Information from on-line imaging sensors has great potential for the monitoring and control of spatially distributed systems. The major difficulty lies in the efficient extraction of information from the images in real-time, information such as the frequencies of occurrence of specific features and their locations in the process or product space. This paper uses multivariate image analysis (MIA) methods based on multiway principal component analysis to decompose the highly correlated data present in multispectral images. The frequencies of occurrence of certain features in the image, regardless of their spatial locations, can be easily monitored in the space of the principal components (PC). The spatial locations of these features in the original image space can then be obtained by transposing highlighted pixels from the PC space into the original image space. In this manner it is possible to easily detect and locate (even very subtle) features from real-time imaging sensors for the purpose of performing statistical process control or feedback control of spatial processes. Due to the current lack of availability of such multispectral sensors in industrial processes, the concepts and potential of this approach are illustrated using a sequence of multispectral images obtained from a LANDSAT satellite, as it passes over a certain geographical region of the earth's surface.

1. Introduction

Digital imaging sensors have recently become very popular in many off-line laboratory applications, particularly in the biological and medical areas. Digital microscopes are extensively used to gather images of cell cultures in a biological laboratory or of wood fiber slurries in a pulp and paper company quality control laboratory.[1,2] Magnetic resonance imaging (MRI), ultrasonic imaging, and positron emission tomography (PET) routinely provide three-dimensional image data in most hospitals. A large amount of digital image processing literature deals with approaches for image enhancement, restoration, analysis, compression, and synthesis. Methods such as edge detection filtering, histogram equalization, image segmentation, and morphological operations are developed to enhance digital images and quantitatively extract relevant information from them. Most of these methods are applied off-line and require considerable computer resources (hardware and software) to process the image data. Further information on these and other digital image processing approaches can be found in various texts.[3−6]

Imaging sensors have been applied to a much more limited extent for monitoring industrial processes. Examples include TV cameras for visually monitoring the state of combustion in engines,[7] robot-mounted digital cameras for monitoring the sizes and shapes of machine-made parts,[8] and laser scanners to monitor the surface properties and to detect faults in sheet forming processes.[9] These industrial imaging sensors are usually much less sophisticated than the laboratory and medical applications discussed earlier; they usually involve only grayscale or binary images. Due to the limited time available to analyze these images, the techniques used for analysis are usually quite simple (e.g. histogram thresholding and area counting).

The purpose of this paper is to investigate an on-line approach to extracting information from time-varying multivariate (3-dimensional) spectral images that will allow for the rapid monitoring, detection, and isolation of process faults and product quality in industrial processes. The approach uses multivariate image analysis (MIA) methods that are based on multiway principal component analysis (PCA). Section 2 provides an overview of these MIA methods including a brief literature review of their theory. The strength of the approach is illustrated through an application of MIA to a LANDSAT multispectral scanner (MSS) satellite image. These methods are then extended to the on-line situation in section 3 where one receives a sequence of images in time, and the purpose is to detect and isolate specific features from the images for statistical process control or feedback control of spatial processes. The approach is again illustrated using successive LANDSAT (MSS) satellite images as the satellite passes over the earth in a north−south trajectory.

2. Overview of Multivariate Image Analysis

Image data when collected in multiple spectral bands produce a multivariate image. It consists of a stack of congruent images, where each image in the stack is measured for a different wavelength, frequency, or energy.[10] Figure 1 illustrates a stack of 4 congruent 512 $\times$ 512 pixel images, with each image having a unique wavelength. Alternately, one could view a multivariate image as a two-way array of pixel intensity vectors with one vector at each pixel location in the (x, y) image plane (Figure 2).

Congruence in imaging is defined[10] as two or more stacked images such that for each pixel in one image there is a corresponding pixel in the other images that can be referred to the same position in the object or

† Tel.: (905) 525-9140 ext. 24951. Fax: (905) 521-1350. E-mail: macgreg@mcmaster.ca.

10.1021/ie980334l CCC: $15.00
Published on Web 11/14/1998

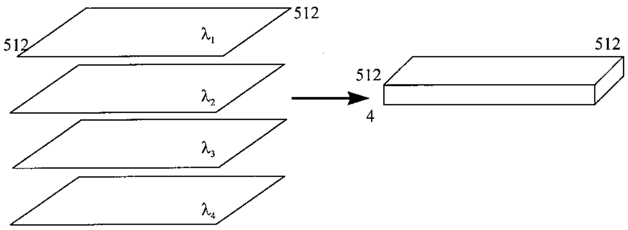

Figure 1. Stack of 4 congruent images (collected in 4 different wavelengths) to form a 512 × 512 × 4 multivariate image.

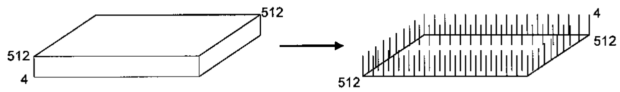

Figure 2. 512 × 512 × 4 multivariate image viewed as a two-way array of 4 × 1 variable vectors in the image plane. For graphical clarity not all variable vectors are shown.

scene depicted. Since data in multivariate images consists of several congruent images, each variable vector in such an image contains highly correlated pixel intensity values. Furthermore, a common multivariate image usually contains an enormous amount of data (e.g. 512 × 512 × 4 = 1 048 576 pixel intensities), making the analysis of such images computationally intensive. As a result, there is a definite need for data analysis techniques that can handle large volumes of highly correlated data.

Latent variable statistical methods like multiway principal component analysis (PCA) and multiway partial least squares or projection to latent structures (PLS) have been successfully used for MIA.[11−13] These methods efficiently compress highly correlated data and project it onto a reduced dimensional subspace through a few linear combinations of the original multivariate data. Multiway PCA of a three-dimensional ($n_x \times n_y \times n_z$) digital image array $\underline{\mathbf{X}}$ consists of decomposing it into a series of A ($<n_z$) principal components consisting of ($n_x \times n_y$) score matrices $\mathbf{T}_a$ and ($n_z \times 1$) loading vectors $\mathbf{p}_a$ plus a residual array $\underline{\mathbf{E}}$, i.e.

$$\underline{\mathbf{X}} = \sum_{a=1}^{A} \mathbf{T}_a \otimes \mathbf{p}_a + \underline{\mathbf{E}} \tag{1}$$

where ⊗ denotes the Kronecker product. The principal components are ordered in the sense that the first component explains the greatest amount of variance in $\underline{\mathbf{X}}$, the second component explains the next greatest variance, and so forth. The number of components (A) necessary to extract most of the meaningful information can be determined by various procedures.[14,15]

This method of multiway PCA is equivalent to unfolding the 3-way array $\underline{\mathbf{X}}$ into an extended 2-way matrix $\mathbf{X}$, as illustrated in Figure 3, and then performing ordinary PCA on it:

$$\underset{n_x \times n_y \times n_z}{\underline{\mathbf{X}}} \xrightarrow{\text{unfold}} \underset{(n_x\cdot n_y) \times n_z}{\mathbf{X}} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{\mathrm{T}} + \mathbf{E} \tag{2}$$

where $\mathbf{t}_a$ is a ($n_x \cdot n_y$) × 1 score vector, and $\mathbf{p}_a$ is a ($n_z \times 1$) loading vector. The score vectors $\mathbf{t}_a$ (a = 1, ..., A) are orthogonal, and the loading vectors $\mathbf{p}_a$ (a = 1, ..., A) are orthonormal. Although the 3-way digital image array $\underline{\mathbf{X}}$ could be unfolded in three different orientations to form 2-way arrays, only two of these orientations do so in such a way that the intensity variable vector z

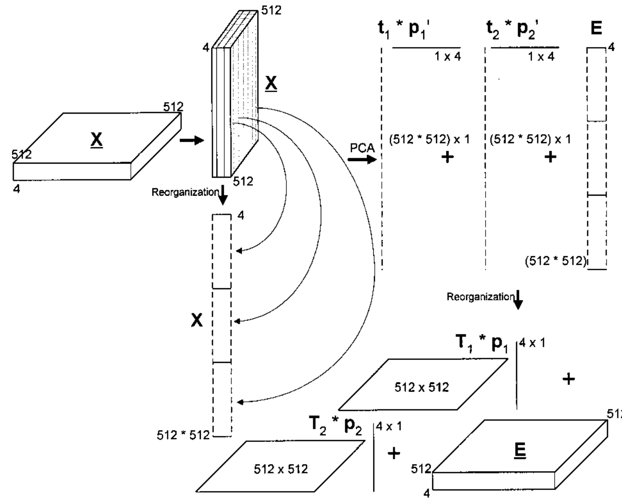

Figure 3. 512 × 512 × 4 multivariate image reorganized into a (512 × 512) × 4 array followed by PCA decomposition into reduced dimensional subspace.

corresponds to the columns in the 2-way array. The objects (rows) in either of these two unfolded forms then correspond to the pixel locations in the $n_x \times n_y$ image plane of the multivariate image.

The row dimension of the $\mathbf{X}$ matrix resulting from the unfolding operation is very large (equal to 262 144 for a 512 × 512 image space). Performing PCA on such a high-dimensional $\mathbf{X}$ matrix using the NIPALS algorithm[16] or using singular value decomposition (SVD)[17] would lead to excessive computational times. Therefore, with essentially all multivariate image data having long and thin unfolded matrices, a kernel algorithm[11] is used. In this algorithm the kernel matrix ($\mathbf{X}^{\mathrm{T}}\mathbf{X}$) is first formed, and then an SVD is performed on this very low dimensional ($n_z \times n_z$) matrix to obtain the loading vectors $\mathbf{p}_a$ (a = 1, ..., A). The corresponding score vectors $\mathbf{t}_a$ are then computed via eq 4. The only time consuming step in this kernel algorithm is the (one time) construction of the kernel matrix ($\mathbf{X}^{\mathrm{T}}\mathbf{X}$).

Upon completion of PCA on this long and thin two-dimensional matrix $\mathbf{X}$, the ($n_x \cdot n_y$) × 1 score vectors $\mathbf{t}_a$ (a = 1, ..., A) can then be reorganized back into ($n_x \times n_y$) score matrices $\mathbf{T}_a$ (a = 1, ..., A) giving a representation of the original $\underline{\mathbf{X}}$ array as expressed in eq 1 and illustrated in Figure 3. In this way one can see that the score matrices $\mathbf{T}_a$ (a = 1, ..., A) themselves represent images in the original ($n_x \times n_y$) scene space, $\mathbf{T}_1$ being the image with the largest variance, followed by $\mathbf{T}_2$ with the second largest variance, and so forth. A reconstructed multivariate image which eliminates much of the unstructured noise from the original image can be obtained by using only the dominant A principal components:

$$\hat{\underline{\mathbf{X}}} = \sum_{a=1}^{A} \mathbf{T}_a \otimes \mathbf{p}_a \tag{3}$$

where the residual component $\underline{\mathbf{E}}$ has been omitted. However, such multiway PCA (MPCA) methods are not very useful for image enhancement as they are not specifically designed to sharpen edges or enhance other specific features in the image. Other image analysis techniques such as various filtering methods, wavelets analysis, and morphological operations provide more powerful approaches to image enhancement or restoration.[4,6,18] The power of the MPCA approach lies in its ability to extract and isolate specific image features in a common region of the score space and, then, once the feature is detected, to reveal the locations where it occurs in the scene space. MPCA appears to be more powerful in this image analysis and interrogation aspect as compared to other image analysis methods.

In each dimension ($a = 1, ..., A$) MPCA extracts a principal component variable $\mathbf{t}_a$ which is defined as a linear combination of the intensities at each wavelength. The particular linear combination is given by the corresponding loading vector $\mathbf{p}_a$. The vector of values of the principal component at each pixel location in the scence space is defined by the score vector:

$$\mathbf{t}_a = \mathbf{X}\mathbf{p}_a \tag{4}$$

Hence each principal component variable extracts a particular spectral feature (i.e. a linear combination of the intensities over all the wavelengths) from the image. The reorganized score matrix $\mathbf{T}_a$ is a representation of the image in terms of that spectral feature.

A more important analysis of the image comes from the compressed representation of the intensity information at each pixel location in terms of the score values (t_1, t_2, ...) of the dominant principal components. These score values summarize the dominant spectral features of the image at each pixel location. If, at different pixel locations in an image, the same feature is present (e.g. surface dirt particles in a sheet forming process), the score value combination (t_1, t_2) would be almost identical for these pixels. Regardless of the spatial locations of the various occurrences of this feature in the image space, MPCA would represent it by the same combination of score values (t_1, t_2). Therefore, by the plotting of the score values of the dominant principal components (t_1, t_2) for each object (i.e. each pixel location) against each other in a scatter plot, the score combinations for all pixel locations in the scene space having the same spectral characteristics would plot on top of one another or at least in the same neighborhood in this score plot. These score plots will therefore be invaluable in the analysis and monitoring of on-line multispectral images. The above MPCA decomposition of images and subsequent analysis of the results are best presented by way of an example.

2.1. Example: LANDSAT (MSS) Satellite Image. Although the focus of this paper is on the on-line monitoring of industrial processes using real-time multispectral imaging sensors, there are as yet almost no industrial examples of this. Therefore, we have chosen to illustrate these methods with multivariate image data from a satellite as it scans a section of the earth's surface. The image used in this example is a LANDSAT (MSS) satellite image of size 512 × 512 consisting of 4 wavelength bands ranging 500−1100 nm. The image has been geometrically corrected with each pixel representing a surface area of 80 × 80 m^2. It has also been used previously by Geladi et al.[10,19] to demonstrate the use of MIA. Technical details regarding the raw multivariate image data are provided in these references. The satellite image depicts a scene of the city of Mobile in Alabama, U.S. (in the center of the image), along with the Alabama river delta (toward the top-right corner of the image) and the Gulf of Mexico (toward the bottom-right corner of the image). Figure 4 represents a false color composite provided by the first score matrix $\mathbf{T}_1$ from the MPCA analysis.

Figure 4. False color composite of the $\mathbf{T}_1$ score image representing Mobile, AL.

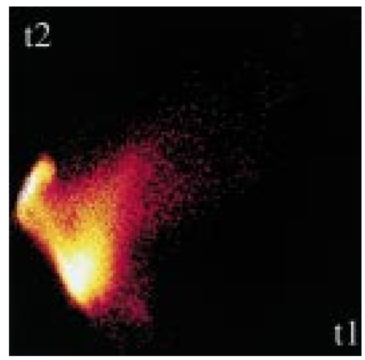

Figure 5. Color-coded $\mathbf{t}_1-\mathbf{t}_2$ score plot of the satellite multivariate image.

The cumulative percent sum of squares explained by the first two principal components is 99.7% (95.0% and 4.7%, respectively). Therefore, only $A = 2$ components will be used in the subsequent analyses. The loading vectors for these two dimensions are $\mathbf{p}_1^{\mathrm{T}} = [0.356\ 0.402\ 0.607\ 0.586]$, $\mathbf{p}_2^{\mathrm{T}} = [-0.550\ -0.583\ 0.149\ 0.579]$. From these loading values it can be seen that the first principal component (PC) represents some type of an average of the pixel intensities at each wavelength, while the second PC represents a contrast or difference among the pixel intensities at various wavelengths. Further details regarding the MPCA decomposition and interpretation of the multivariate image can be found in Bharati.[20]

A scatter plot of the first two score vectors ($\mathbf{t}_1$ vs $\mathbf{t}_2$) is illustrated in Figure 5. In this plot there are 262 144 score combinations plotted, one for each of the 512 × 512 pixel locations in the original image. Since similar features in the original image will yield almost identical (t_1, t_2) score combinations, many points overlap in this scatter plot. If the score plots are drawn as simple black and white (binary) scatter plots, there would be no way of visualizing the number of overlapping pixels at a particular point. The number of pixels represented by a single point in a score plot is called the pixel density. Following Geladi et al.[10] an intensity matrix can be constructed which represents the number of pixels at each point in the score space. The score plot ($\mathbf{t}_1$ vs $\mathbf{t}_2$)

is constructed as a 3D histogram with a grid of 256 × 256 bins. Each bin is filled with a bar giving the count of pixels belonging to that bin. This matrix is then color coded depending upon the number of pixels in each bin (i.e. pixel density) using a color scheme ranging from cold colors (e.g. black) representing bins with a low number of overlapping pixels to hot colors (e.g. white) representing bins having the highest pixel density.

It is relatively easy to detect outlier pixels which are remote from the pixel clusters in the score space. It is also easy to detect high density clusters and the various pixel density gradients that exist both within and between clusters. Figure 5 reveals two major dense classes of pixels which are separated by a prominent area of between-class pixels. As mentioned earlier, pixels having similar spectral features in the multivariate image will have comparable combinations of score values and result in point clusters in the score plot. This fact can be put to use in segmentation of features from the multivariate image through delineating pixel classes in the score plots. In effect, one can delineate a tentative data class corresponding to pixels having similar spectral fingerprints.[12]

Pixel class delineation may be carried out in more than one way. An area in the score space may be selected and the corresponding pixels belonging to this area highlighted in the image space. The selected area in the score space is in fact a local model which is chosen to delineate a tentative class of pixel data from the rest. The procedure of selecting an area in the score space is called "masking" in the MIA literature. Since score plots can themselves be represented as 256 × 256 images, various sizes and shapes of masks can be selected be carrying out simple graphical operations on these score plot images. Using the "*roipoly*" (region of interest selection by polygon) command in the Image Processing Toolbox (v1.0b) of MATLAB (v4.2c1), it is relatively easy to choose any arbitrary size and shape of polygon to highlight particular pixels in the score space. The pixels contained within the polygon can be isolated (into a local model) and projected back to the image space.

This procedure of masking point clusters and outlier pixels in the score space and highlighting the chosen pixels in the image space forms the backbone of the MIA off-line feature extraction strategy. To successfully delineate a class of pixels it becomes imperative to study both the score and image spaces simultaneously. The ability to toggle between the image space and score space is quite fast and easy since both spaces are represented as images in MIA. As a result, it is relatively simple to interrogate the score space by masking several clusters in the score space and projecting the masked pixels back to the image space. Figure 6a replots the $\mathbf{t}_1-\mathbf{t}_2$ score plot of Figure 5 but with a maroon polygon mask covering a portion of the lower dense point cluster. The pixel class which has been masked in the score plot of Figure 6a is outlined in Figure 6b where each pixel with a $\mathbf{t}_1-\mathbf{t}_2$ score combination lying under the mask has been replotted as an overlaid white pixel on the false color image of the first latent variable $\mathbf{T}_1$. From this figure it is evident that the class of pixels masked in the score space belongs to major roadways, paved areas, and building tops in and around the city. Since all the highlighted pixels have similar spectral combinations, they map into the region masked by the polygon in the score space.

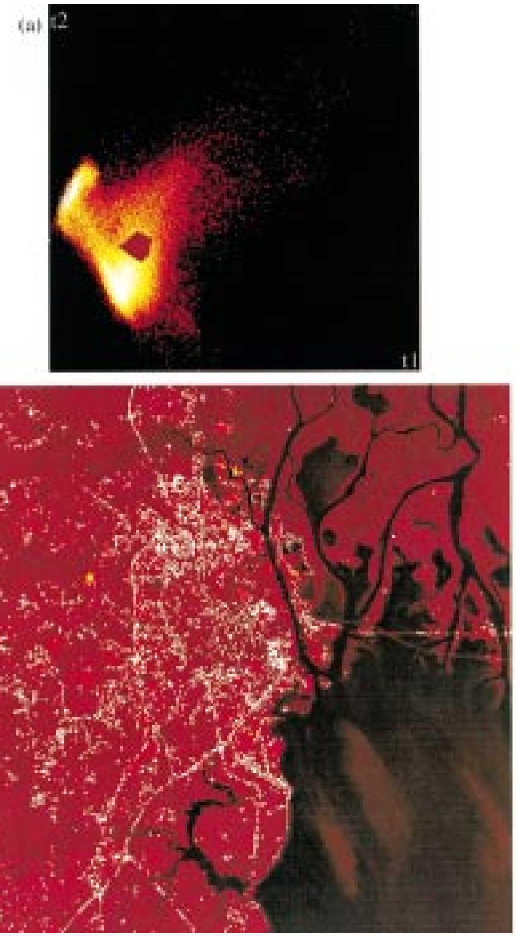

Figure 6. (a) $\mathbf{t}_1-\mathbf{t}_2$ score plot of the satellite multivariate image with a class-masking of the upper part of the lower dense pixel cluster (masked in maroon). (b) Overlay of $\mathbf{T}_1$ image with highlighted pixels from class outlined in (a).

By the repeated use of the masking/highlighting procedure with different polygon masks a *signature* of every feature existing in the image space (regardless of its subtlety or spatial location) can be isolated in the score space. Due to the ability to switch easily between score and image spaces, MIA can also be employed as a reverse mode image analysis tool. Specific pixels belonging to known features of interest in the image space can be highlighted in the score space to determine the region which represents their corresponding score combinations. The area surrounding the highlighted score points can then be masked using a reasonably sized polygon. As a result, subtle features that are subjectively difficult to identify in the image space may easily be identified using the reverse mode application of MIA. The polygon mask used in Figure 6a was developed using this methodology. Geladi et al.[10] list several other modes of MIA that employ the use of the image space/score space relationship.

3. Real-Time Monitoring Using MIA

In this section we proceed to the main thrust of this paper—the extension of MIA methods to the real-time monitoring of time-varying processes (i.e. real-time image analysis). MIA methods have largely been applied only to off-line analysis of fixed images. However, the purpose in monitoring an industrial process with imaging equipment is to detect and isolate faults or quality defects in the process or product being monitored. For example, in industrial sheet forming processes on-line multivariate imaging could be extremely useful in detecting and isolating the appearance of various faults in the sheet (e.g. dirt particles, streaks, variations in layer thickness, etc.). The major impediments to the use of on-line imaging at present are the difficulties in handling the large volumes of data collected in real-time and speed limitations in being able to process all these data with current image analysis methods. It is shown in this section that an on-line multiway PCA approach to MIA can provide a powerful method for monitoring and analyzing many time-varying industrial processes.

The main concepts of this approach are described as follows. A multiway PCA model is built off-line on a training or calibration image which contains all typical features that one might be interested in detecting using the on-line monitoring scheme. This training image may be a single image which contains all such features of interest, or it may be a composite image put together from sections selected out of many different images. The training image should be of the same dimensions as the subsequent images that are obtained sequentially by the on-line imaging system.

From the off-line analysis of this image masks are developed in the score space which correspond to each feature in the image space that one desires to monitor. Upon application of the fixed PCA model to the new images as they become available, values of the scores are computed for the dominant principal components $\mathbf{t}_a$ (eq 4) using the loading vectors $\mathbf{p}_a$. The pixel densities in the score space (Figure 5) can then be updated for the new image. By monitoring the changing score point cluster intensities (pixel densities) under each mask area in the score plot, one can then track the appearance and disappearance of each feature in the current image. Upper tolerance limits can be set on the pixel densities in each mask area. These limits might be chosen simply on a subjective basis or, as in statistical process control (SPC) charts, on the underlying statistical distribution of pixel densities when the process is subject only to common cause variation. Violation of these limits in future monitoring could then be handled as in any SPC monitoring scheme. Upon discovering violation of the SPC tolerance limits for any feature being monitored, one can investigate further by switching to the image space to reveal those pixel locations where the fault feature was present. The locations of this feature in the image (e.g. all on side of a moving sheet) might indicate possible assignable causes which can be corrected. This on-line monitoring approach is now illustrated by way of an example.

3.1. On-Line Monitoring of Surface Features from LANDSAT (MSS) Images. A modified version of the LANDSAT (MSS) multispectral satellite image example (depicting Mobile, AL) introduced in the earlier example is used here to illustrate the main ideas of feature monitoring. The original 512 × 512 × 4 multivariate image has been segmented into nine sections of 256 × 256 × 4 multivariate images as the LANDSAT satellite moves in a north to south trajectory over this region.

As discussed above, it is extremely important to select a training image that contains representative samples of all the features of interest to build a good PCA training model. In the satellite image example there are four main features of interest that are chosen from various landscape features in and around Mobile, AL. Some of these features of interest are subjectively easy to detect, whereas others are not so obvious. The four main features of interest that are chosen to be monitored throughout the sequence of multivariate images, as the satellite moves over the region, are (a) deep water areas, (b) shallow water areas, (c) major roads and paved areas, and (d) golf courses. As a result, a good representative training multivariate image must contain ample pixels belonging to all of the above four features of interest. The center 256 × 256 × 4 pixel subsection of the full scene (Figure 4) in the original multivariate image adequately meets this requirement.

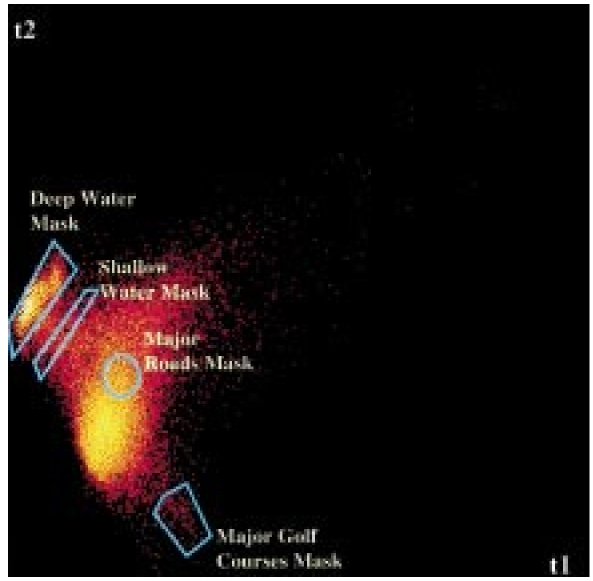

Figure 7. Detailed $\mathbf{t}_1-\mathbf{t}_2$ score plot of training image with highlighted masks that adequately represent the four features of interest.

The PCA training model is calibrated by masking various point clusters with customized local area score masks. The sizes and shapes of these masks are determined by using the previously discussed iterative procedure of feature extraction from multivariate images. In this case, since the features of interest are known beforehand, the reverse mode of MIA is used to determine the spatial locations of selected feature pixels in the score space. Once the score points are highlighted, manually shaped masks are applied such that the areas surrounding these scores are covered. All pixels that are covered by these masks are highlighted in the image space to determine if the particular masks adequately segment the features of interest. This procedure is repeated until an adequate number of pixels belonging to the above four features of interest are highlighted in the image space. More optimal procedures for determining the mask boundaries could be developed but are not of direct concern in this paper. Score space masks that adequately represent the four features of interest in the image space are highlighted in Figure 7. Using the four customized local area masks in Figure 7, along with the reduced loading matrix $\mathbf{P}_{\mathrm{tr}}$, the calibration of the PCA training model is complete.

To emulate a moving satellite gathering scans of the earth surface only the left half of the scene is used (i.e. columns 1−256). This half of the full scene is further divided into 9 overlapping equal sized scans of 256 × 256 pixels in all four spectral bands. Since the LANDSAT satellite moves in a north to south trajectory while

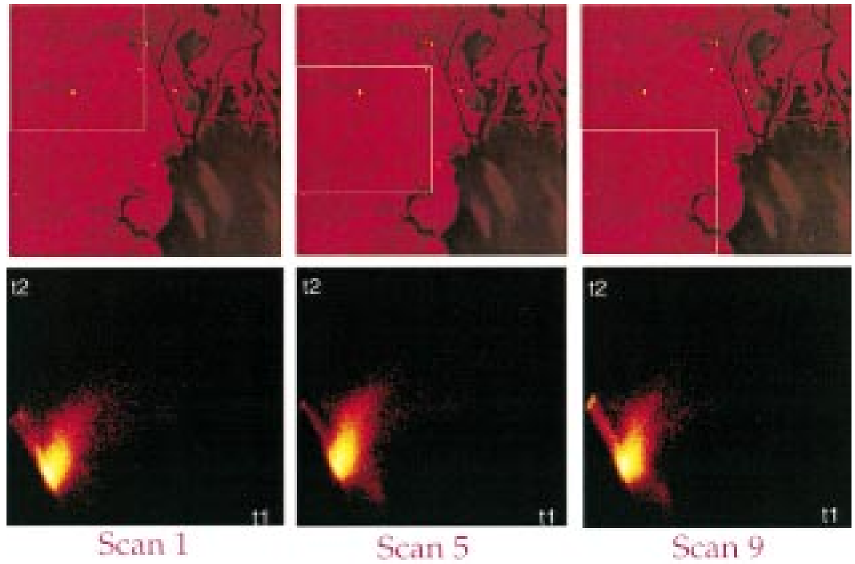

Figure 8. Three of nine (scan 1, 5, 9) input images outlined as white boxes on false color $\mathbf{T}_1$ full scene score image with corresponding color-coded $\mathbf{t}_1-\mathbf{t}_2$ score plot images.

gathering images of the earth surface, the first 256 × 256 × 4 scan is located in the top left hand corner of the image (Figure 8). Each successive scan is gathered as the satellite moves south in a certain time which is represented as 32 rows. As a result, the second 256 × 256 × 4 scan spans rows 33–288 whereas the ninth scan spans rows 257–512 in the left half of the full scene. Using these nine scans a north to south moving LANDSAT satellite is emulated that gathers intermittent multispectral images as it passes over the region of Mobile, AL.

The 9 input multivariate images are decomposed into their score spaces with the help of the PCA training model. Each 256 × 256 × 4 pixel multivariate scan image is rearranged into a 65 536 × 4 two-way array $\mathbf{X}_{\text{test}}$ and multiplied with the training loading vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ to produce the corresponding score vectors $\mathbf{t}_1$ and $\mathbf{t}_2$. These score vectors are rearranged into 256 × 256 intensity arrays and viewed as score images, as well as plotted against each other as color-coded scatter plot images (256 × 256 pixels). As a result of the decomposition, 9 $\mathbf{t}_1-\mathbf{t}_2$ score plot images can be plotted. Figure 8 illustrates three (scans 1, 5, 9) of the 9 scan images (outlined by white boxes) along with their corresponding $\mathbf{t}_1-\mathbf{t}_2$ color-coded score plot images. Upon examination of the 3 scan images in Figure 8, it can be seen that scan no. 1 captures regions that are mainly north of the city. The majority of the city areas are captured by scan no. 5; finally, scan no. 9 captures areas that are south of the city. Some subjective impressions regarding the four features of interest can be obtained upon observing these scan images. For example, it is obvious that scan no. 9 contains more pixels that belong to deep and shallow water areas as compared to scan no. 5. However, subtle features like golf courses and roadways are more difficult to identify with a visual observation of these image space scans.

A more objective feature monitoring scheme is obtained by monitoring the pixel densities of the corresponding point clusters under the masks in the score space of Figure 7. The pixel densities under the four masks in the nine score space images covering the north to south movement of the satellite are monitored by enumerating the exact number of pixels belonging to these four features of interest for each satellite scan. The total number of pixels belonging to each of the four features are recorded and plotted as control charts for all nine scan images. The exact spatial locations of these feature pixels in the image space are then obtained by highlighting them in the score space and transposing them into the image space. The resulting control charts for the feature pixel counts and the highlighted spatial locations of these pixels in the image space are illustrated in Figures 9–12. The score images used to highlight the spatial locations of pixels belonging to the four features of interest are false color images of the first score vectors $\mathbf{T}_1$ in each of these figures. The bar charts through Figures 9–12 show the exact number of pixels belonging to each of the four features of interest for all nine scan images. These provide on-line monitoring charts for the time variation of all four features throughout the sequence of the 9 multivariate images. The deep and shallow water pixels increase as the satellite moves to the south of the city (scans 8 and 9) whereas the paved area pixels are more pronounced in scans over the city (scans 4 and 5). A total of three major golf courses were detected by the satellite as it moved over the region covered by the nine scans. The monitoring chart in Figure 12 is seen to detect one major golf course in scan 1, two in scans 2 and 3, and three in scans 4–9. The locations of the courses are highlighted in the corresponding image spaces in Figure 12. The above results show that all four prechosen features of interest were adequately monitored through the sequence of 9 multivariate satellite images.

In order to apply this type of strategy on a time-varying industrial process it is important to be able to process and display the results in a reasonable period of time. Table 1 provides average times (averaged over all 9 scans per feature of interest) required to perform three major tasks in the above feature monitoring example. The total time required to obtain results for each feature is dependent on several factors including size of the local area mask in the score space and the number of pixels belonging to the feature of interest per individual scan. The simulations for the above example were performed on an IBM compatible Pentium 166 MHz personal computer with 64 MB of RAM running in a Windows '95 environment and using MATLAB

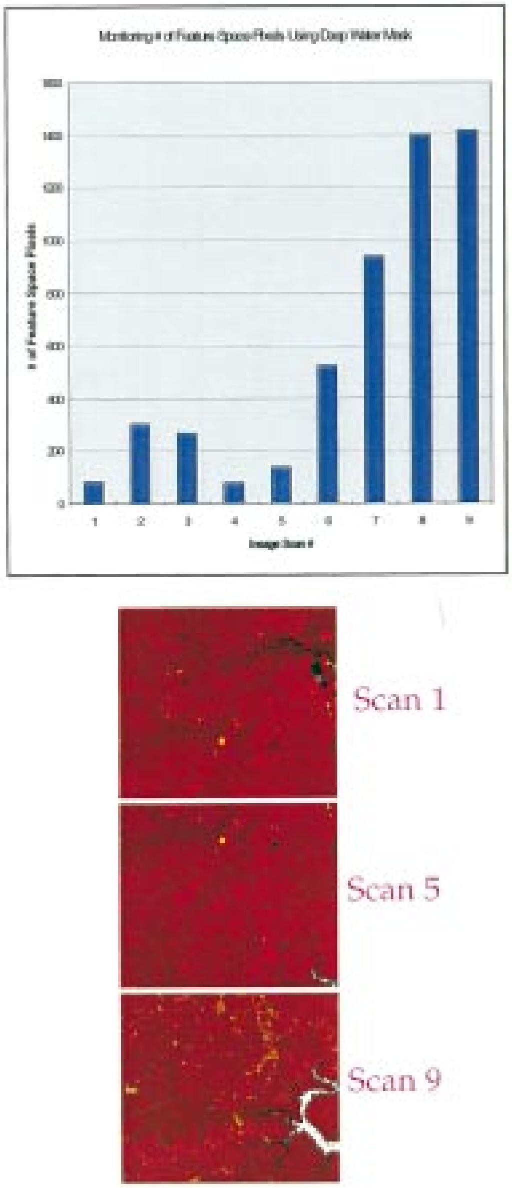

Figure 9. Bar chart of total number of pixels belonging to deep water areas through scans 1−9 with highlighted spatial locations of these feature pixels in the image space for scans 1, 5, and 9.

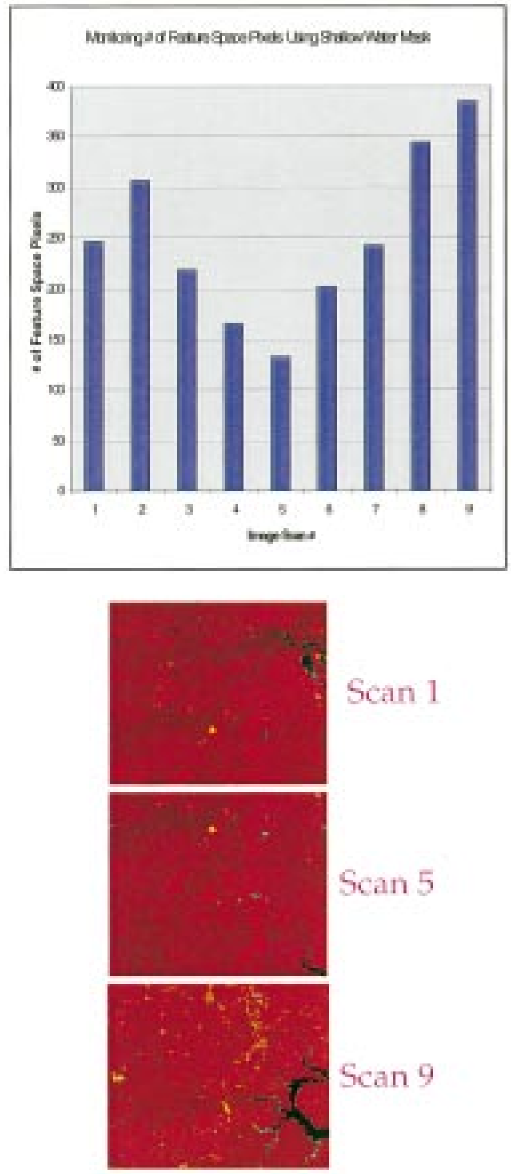

Figure 10. Bar chart of total number of pixels belonging to shallow water areas thorugh scans 1−9 with highlighted spatial locations of these feature pixels in the image space for scans 1, 5, and 9.

v4.2c1. These average times vary between approximately 1.5 and 2.5 min using the above mentioned software and hardware. If desired, these average times may be considerably reduced upon using faster compiled software equipped with more efficient code. However, the above tabulated average times provide a benchmark for comparing future applications of feature monitoring schemes using MIA. Some industrial processes that gather data using on-line imaging sensors (e.g. steel rolling, paper production etc.) may require much faster cycle times between gathering images and calculating results. However, in many process situations, updating the monitoring plots every 1−2 min might be quite adequate for fault detection.

The information extracted from these multivariate images and plotted in Figures 9−12 can be used for statistical process control (SPC) or automatic feedback control. In a SPC scheme the height of the bars corresponding to the frequency of occurrence of each type of fault would be monitored in each successive image. Whenever any one of these bars exceeds a predetermined threshold or tolerance, an alarm would be given and the process operator or engineer would look for an assignable cause. This would be greatly aided by highlighting the spatial location of the faults in the image space as shown in the lower half of Figures 9−12. For example, if most of the faults occurred near one edge of a sheet coating process, one might logically investigate the feed system of the coating machine for that edge. The results could also be used in an automatic feedback control scheme if spatial actuators were present which could be used to affect the particular feature being monitored. For example, if the feature corresponded to the thickness of a particular layer in a multilayer sheet coating operation and spatial actuators were available to adjust the flow of coating material for that layer in the region of interest, then feedback control algorithms could be used to maintain the layer thickness uniform across the sheet. Another example might be in a situation where a camera is used to monitor combustion in a furnace or incinerator. An indication of incomplete combustion occurring in a given region of

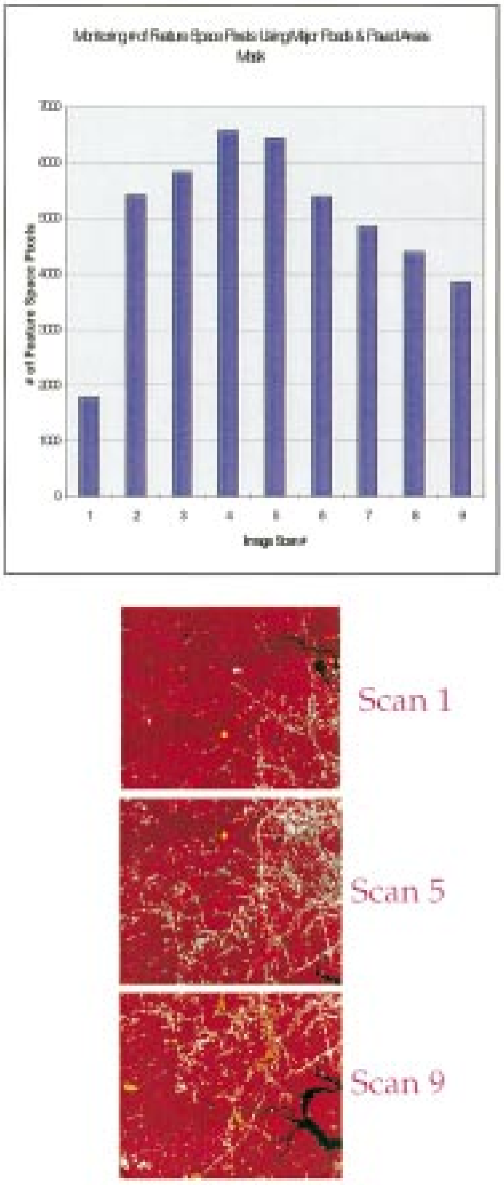

Figure 11. Bar chart of total number of pixels belonging to major roads and paved areas through scans 1−9 with highlighted spatial locations of these feature pixels in the image space for scans 1, 5, and 9.

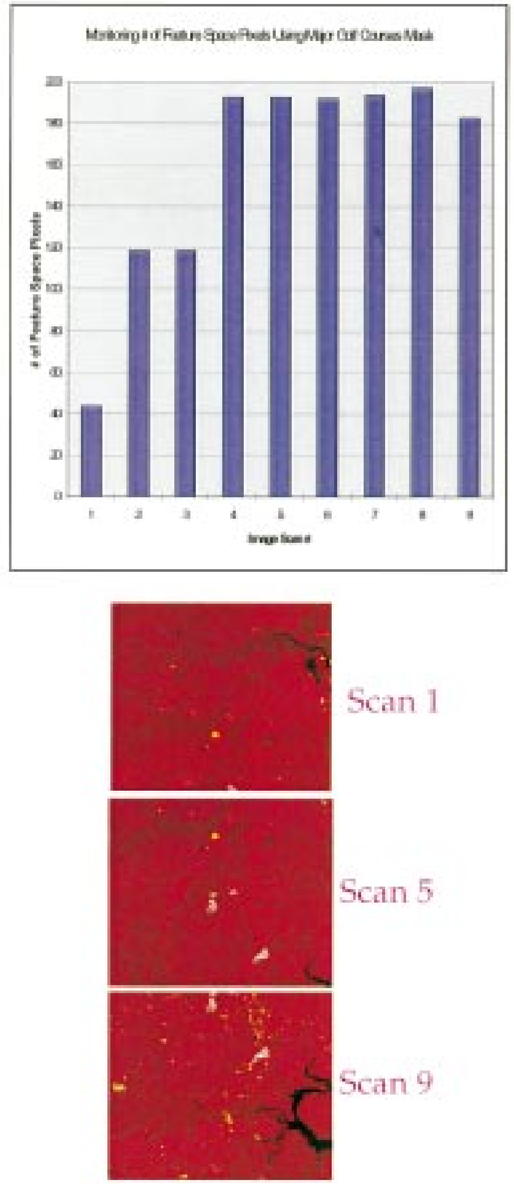

Figure 12. Bar chart of total number of pixels belonging to major golf courses through scans 1−9 with highlighted spatial locations of these feature pixels in the image space for scans 1, 5, and 9.

the furnace could be counteracted by increasing oxygen or fuel feed to that section. The above SPC and spatial feedback control strategies would rely upon the successful extraction of various features, evaluating their intensity of occurrence, and identifying their spatial locations in the image.

3.2. Alternative Approaches for the Detection of Infrequent Faults. The approach illustrated in the previous section to process monitoring with real-time imaging was based on the premise that under normal process operation the image will contain the whole range of features. For example, in each satellite scan of the earth's surface, most features such as deep and shallow water, paved areas, etc., were present to some extent in every image scan; only their relative frequencies changed. Therefore, it was logical to build the PCA model on an image (or composite image) that contained all features and use the concept of masking in the score space to monitor the changing frequencies of occurrence of each feature.

However, in many statistical process control (SPC) situations one is producing a product (e.g. sheet or film production) in which only common cause variation is present under normal operation. Faults or defect features appear only as infrequent special causes. In this situation, rather than using the previous approach and building a PCA model using a composite image consisting of common cause background plus representative features of all known defects, an alternative would be to build the PCA model on a composite image that is representative of only the common cause variation that is present in "good" product. A traditional multivariate SPC approach[21−23] could then be applied. An in-control region can be defined in the score space of the first few principal components. As new "good" product is scanned by the imaging sensor, the scores from this image should fall within this in-control region, and the prediction errors (elements of $\underline{\mathbf{E}}$ in Figure 3) should be small implying that the PCA model is still valid. If a defect fault not usually present in good product were to appear in the image, the scores corre-

Table 1. Average Times (over 9 Scans) To Monitor 4 Features of Interest from a Satellite Multivariate Image Sequence

feature of interest	time to evaluate and plot scores (s)	time to count no. of pixels under mask (s)	time to overlay highlighted pixels on $\mathbf{T}_1$ image (s)	total time (s)
deep water areas	30.85	122.07	1.06	153.98
shallow water areas	30.12	96.04	1.01	127.18
major roads and paved areas	30.69	65.57	1.39	97.66
major golf courses	30.78	94.05	1.01	125.83

sponding to the pixel locations where the defect feature has occurred shall fall outside their control region in the score space. Furthermore, the squared prediction errors (SPEs) (i.e. the sum of the squared values of the elements in the vectors of the $\underline{\mathbf{E}}$ array at each pixel location—see Figure 2) should become large and exceed a statistically defined control limit at those same pixel locations.

Using contribution plots[24] one could then interrogate the underlying PCA model to see what spectral wavelengths were the greatest contributors to score and SPE deviations at these designated pixel locations. The spectral features of these contributions might be expected to contain a signature of the type of fault that has occurred in any specific pixel region. The spectral signatures could be obtained for all previously observed faults or defects and classification methods such as K-nearest neighbor (KNN)[25] or soft independent modeling of class analogy (SIMCA)[25,26] could be used to assign the current defect to one of the known causes.

4. Conclusions

In this paper we have investigated the extension of multivariate image analysis techniques based upon multiway PCA methods to real-time monitoring situations. Using a LANDSAT (MSS) satellite image sequence as an example these methods were shown to be capable of utilizing large amounts of digital image data that are available from on-line imaging equipment. The information gathered from these data can then be used to analyze it to detect and isolate faults or quality features in the process or product and to do so in a time frame that is reasonable for many industrial situations.

The key points in the approach are as follows. All modeling and model calibration (mask determination, etc.) is performed off-line on a representative training image. On-line monitoring is then achieved quite rapidly by using the fixed loading vectors for the primary principal components from this model to update the score vectors. All monitoring is then performed in the score space by enumerating the pixel densities in masked regions of the score space. The number of defects or product features present in the image space is proportional to these densities. Once the pixel density exceeds an upper control limit, one can toggle back to the image space to view the locations and structure of the defects. Depending upon the type, severity, and location of the faults one may stop the process to rectify any assignable cause, mark the image number and fault locations for rectification in the next stage of the process, or apply spatial feedback control strategies to compensate for the occurrence of certain features.

There are many different problems that can be addressed using multispectral imaging sensors and several different approaches to building and using multiway PCA models for feature extraction and process monitoring. One alternative SPC approach to real-time MIA was outlined for situations where defects or faults are infrequent.

In spite of the ready availability of multivariate imaging systems in off-line settings, there has been very little application of these systems to industrial processes for on-line monitoring. One reason for this has been the lack of suitable image analysis techniques capable of handling the large volumes of data and extracting the relevant information in a sufficiently short period of time. We hope that the on-line MIA approach proposed in this paper addresses some of these problems and will expand the use of imaging sensors for monitoring industrial processes.

Acknowledgment

The authors thank Professor Paul Geladi for providing the LANDSAT (MSS) multivariate image data which have been used as an example in this paper.

Literature Cited

(1) Glasbey, C. A.; Horgan, G. W. *Image Analysis for the Biological Sciences*; John Wiley & Sons: Chichester, U.K., 1995.

(2) Parham, R. A.; Wood Structure-Soft Woods. In *Pulp and Paper Manufacture, Vol. 1*; Kocurek, M. J., Stevens, C. F. B., Eds.; Joint Textbook Committee of Paper Industry: Montreal, 1983.

(3) Baxes, G. A. *Digital Image Processing: Principles and Applications*; John Wiley & Sons: New York, 1994.

(4) Pratt, W. K. *Digital Image Processing*; John Wiley & Sons: New York, 1978.

(5) Gonzalez, R. C.; Wintz, P. *Digital Image Processing*, 2nd ed., Addison-Wesley Publishing Co. Inc.: Reading, MA, 1987.

(6) Serra, J. *Image Analysis and Mathematical Morphology*; Academic Press Inc. (London) Ltd.: London, 1982.

(7) Hancock, M. S.; Belmont, M. R.; Buckingham, D. J. Development of an Image Capture and Analysis Technique for the Investigation of the Very Early Stages of Combustion in a Lean-Burn Engine and the Detection of a Novel Early Combustion Phase which Correlates with Subsequent Cycle Quality. *Proc. Inst. Mech. Eng.* **1990**, *204*, 125.

(8) Jain, R.; Kasturi, R.; Schunck, B. G. *Machine Vision*; McGraw-Hill Inc.: New York, 1995.

(9) Suresh, B. R.; Fundakowski, R. A.; Levitt, T. S.; Overland, J. E. A Real-Time Automated Visual Inspection System for Hot Steel Slabs. *IEEE Trans. Pattern Anal. Machine Intell.* **1983**, *6*, 29.

(10) Geladi, P.; Grahn, H. *Multivariate Image Analysis*; John Wiley & Sons: Chichester, U.K., 1996.

(11) Geladi, P.; Isaksson, H.; Lindqvist, L.; Wold, S.; Esbensen, K. Principal Component Analysis of Multivariate Images. *Chem. Int. Lab. Sys.* **1989**, *5*, 209.

(12) Esbensen, K. H.; Geladi, P. Strategy of Multivariate Image Analysis (MIA). *Chem. Int. Lab. Sys.* **1989**, *7*, 67.

(13) Grahn, H.; Szeverenyi, N. M.; Roggenbuck, M. W.; Delaglio, F.; Geladi, P. Data Analysis of Multivariate Magnetic Resonance Images I. A Principal Component Analysis Approach. *Chem. Int. Lab. Syst.* **1989**, *5*, 311.

(14) Wold, S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **1978**, *4*, 397.

(15) Jackson, J. E. *A User's Guide to Principal Components*; John Wiley & Sons: New York, 1991.

(16) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1.

(17) Golub, G. H.; Van Loan, C. F. *Matrix Computations*; The John Hopkins University Press: Baltimore, MD, 1983.

(18) Nikolov, S. G.; Hunter, H.; Grasserbauer, M. De-noising of SIMS Images via Wavelet Shrinkage. *Chem. Int. Lab. Sys.* **1996**, *34*, 263.

(19) Geladi, P.; Esbensen, K. Multivariate Image Analysis in Chemistry: An Overview. In *Applied Multivariate Analysis in SAR and Environmental Studies*; Devillers, J., Karcher, W., Eds.; Kluwer: Dordrecht, The Netherlands, 1991.

(20) Bharati, M. H. Multivariate Image Analysis for Real-Time Process Monitoring. M.Eng. Thesis, McMaster University, Hamilton, ON, Canada, 1997.

(21) Kresta, J. V.; MacGregor, J. F.; Marlin, T. E. Multivariate Statistical Monitoring of Process Operating Performance. *Can. J. Chem. Eng.* **1991**, *69*, 35.

(22) Kourti, T.; MacGregor, J. F. Multivariate SPC Methods for Process and Product Monitoring. *J. Qual. Tech.* **1996**, *4*, 409.

(23) Nomikos, P.; MacGregor, J. F. Monitoring Batch Processes Using Multiway Principal Component Analysis. *AIChE J.* **1994**, *8*, 1361.

(24) MacGregor, J. F.; Jaeckle, C.; Kiparissides, C.; Koutoudi, M. Process Monitoring and Diagnosis by Multiblock PLS Methods. *AIChE J.* **1994**, *5*, 826.

(25) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; John Wiley & Sons: New York, 1998.

(26) Wold, S.; Albano, C.; Dunn, W. J., III; Esbensen, K.; Hellberg, S.; Johansson, E.; Sjöström, M. Pattern Recognition: Finding and Using Regularities in Multivariate Data. In *Proceedings of the IUFOST Conference Food Research and Data Analysis*; Martens, H., Russwurm, H., Jr., Eds.; Applied Science Publishers: London, 1983.

Received for review June 1, 1998
Revised manuscript received September 14, 1998
Accepted September 15, 1998

IE980334L