Principal Components Analysis of Multispectral Image Data

Brent Neal* and John C. Russ**

*Reindeer Graphics, Inc., and **Materials Science and Engineering Dept., North Carolina State University

brent@reindeergraphics.com

Principal components analysis of multivariate data sets is a standard statistical method that was developed in the early half of the 20th century. It provides researchers with a method for transforming their source data axes into a set of orthogonal principal axes and ranks. The rank for each axis in the principal set represents the significance of that axis as defined by the variance in the data along that axis. Thus, the first principal axis is the one with the greatest amount of scatter in the data and consequently the greatest amount of contrast and information, while the last principal axis represents the least amount of information.

This technique is particularly useful for extracting the maximum contrast and structural information from a set of images, as well as finding correlations between the source variables and for determining if any are degenerate in their contribution. Except for remote sensing applications, the technique has not been widely applied to the field of imaging. This paper will illustrate several uses of principal components analysis (PCA) for processing and analyzing microscope images.
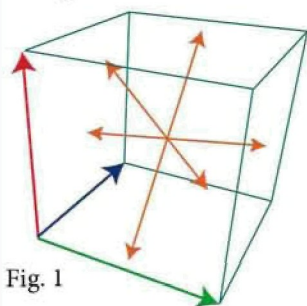

Fig. 1

Principal Components in Images

Contrast in images is represented by the variance of intensity values. For a typical RGB, 3-channel image, transforming to a principal component space (Figure 1) creates three new channels in which the first (most significant) contains the most structural contrast and information. Noise and image artifacts (including those from JPEG compression) are typically relegated to the least significant channel. Processing the least significant channel before transforming back to RGB can be used to remove image noise. It is also useful to examine the information that PCA produces about the source channels.

X-ray maps produced by an SEM provide an example. Typically, more than three elements are of interest, both in terms of their distribution and their co-occurrence (also called colocalization), since this gives information about the compounds or phases present. Further, while we can pick many combinations of three elements to display as RGB channels, we would like to be able to delineate all the phase boundaries in a single image. By definition, the principal components transform provides a way to do both of these things. In Figure 2, we have 9 SEM x-ray maps, presented 3 at a time in the red, green, and blue channels of 3 different images. It is not possible to display more than three channels in unique colors. This limitation arises from the fact that human vision has three kinds of cones, and so displays use three kinds of phosphors (red, green and blue); you cannot assign another color (e.g., yellow) to a fourth element because yellow will appear wherever the red and green elements are co-located. Principal components produce much more useful images by combining the basic elements to show structural and phase locations. Figure 3 shows the 9 principal components of the x-ray maps, in order of significance.

The analysis of the principal components data starts with the covariance matrix generated by the transform. The matrix in Table 1 shows the result for the 9 elements in Figure 2. Each row shows the significance of the principal channel (first column) and the contributions of each source channel to that principal vector. In the case of the x-ray maps, the significance values indicate the number of distinct phases, and measure the area fraction of each phase. The components of the source channels are a function of the phase composition. This numerical data can be visualized using colocalization plots. Plotting colocalization uses axes representing the intensity values in different channels and shows the frequency with which various combinations of those values occur as a grayscale intensity. Dark clusters of points
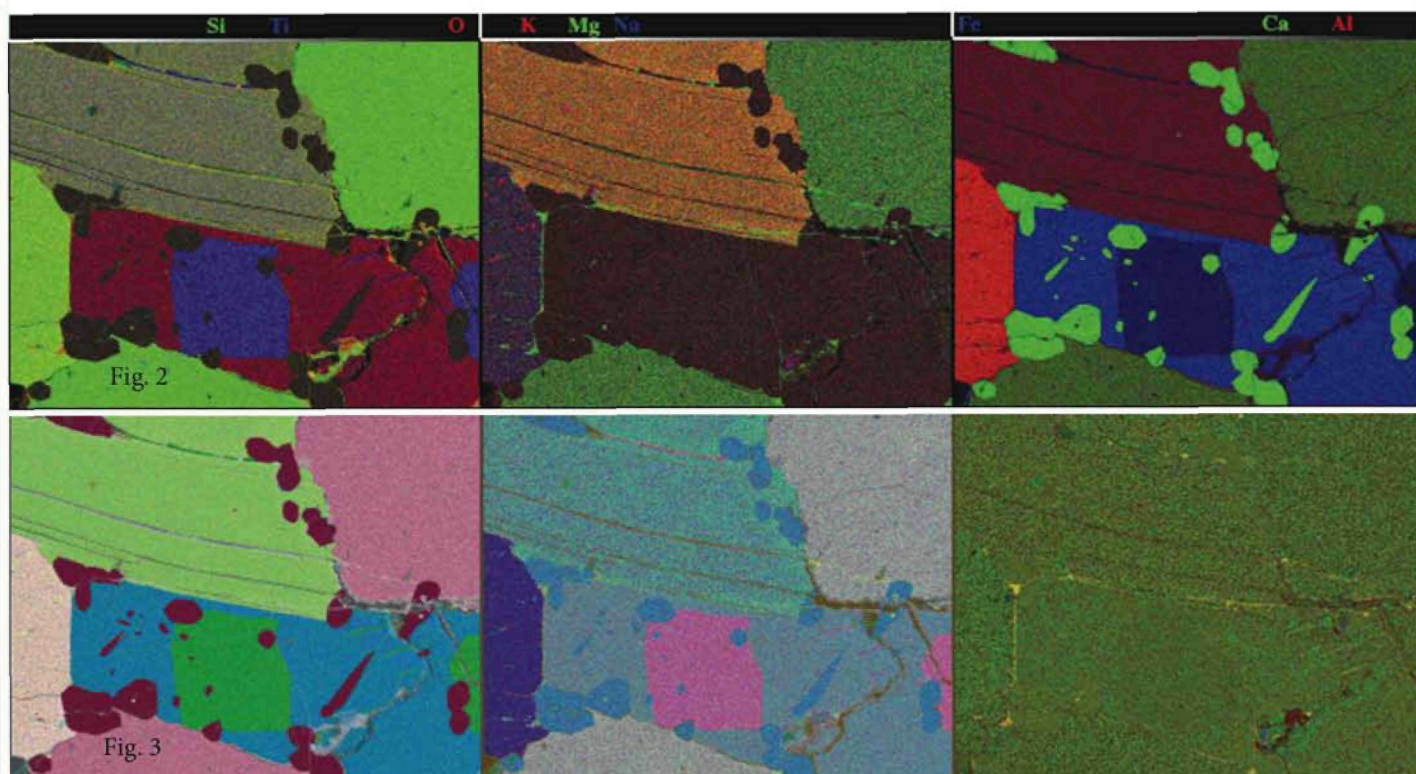

Fig. 2

Fig. 3
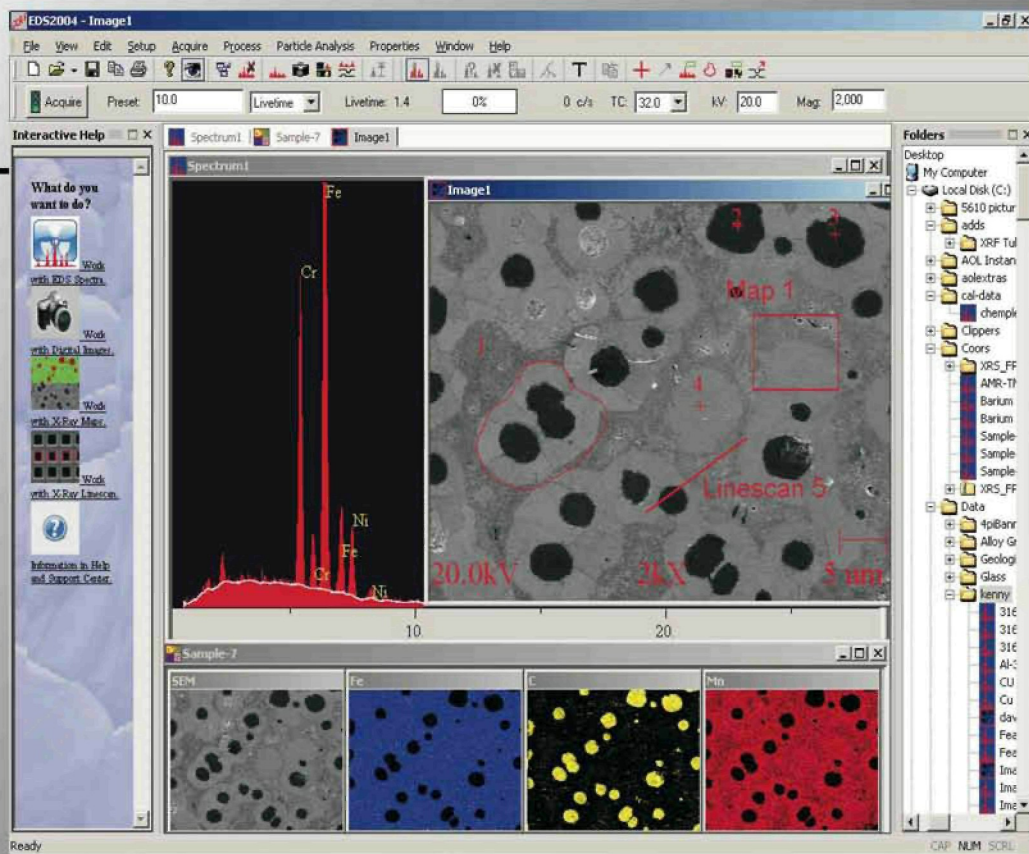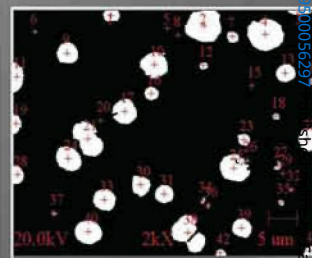
EDS 2004

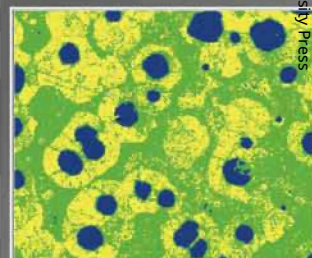*NEW Architecture Featuring Microsoft.NET™*

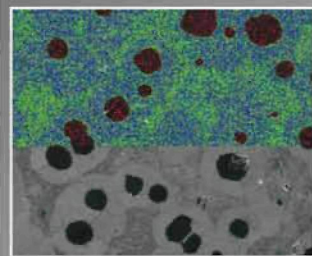## Expect More from your EDS Company
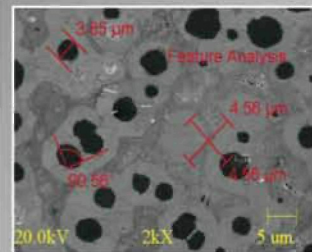
Image Morphology

Phase Analysis

Composition Mapping

Image Processing

INTERACTIVE ONLINE HELP

FULLY FEATURED, NO ADD-ON OPTIONS

Working with Spectra

Working with Images

Working with Maps

Working with Linescans

Review Linescan Properties

Select Elements

Acquire Linescans

View Linescan Data

Overlay Linescans

Export

EDS2004 includes all of these features and more standard, *NOT* as options:

~ High Resolution Imaging   ~ Spectral Matching
~ Spectral Mapping   ~ ZAF Quantitation
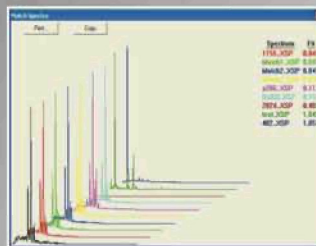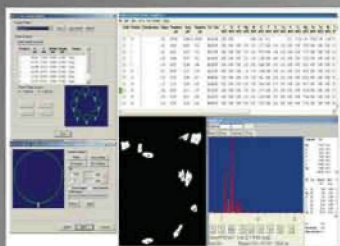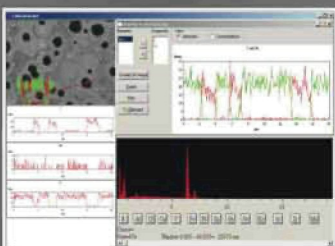~ Automated Particle Analysis   ~ Interactive Online Help

IXRF offers complete new systems for nearly any SEM or TEM, and for upgrades, we can interface to any existing EDS detector.

FREE software upgrade via the internet for the life of the system means that your system will always have the latest feature developments from IXRF.

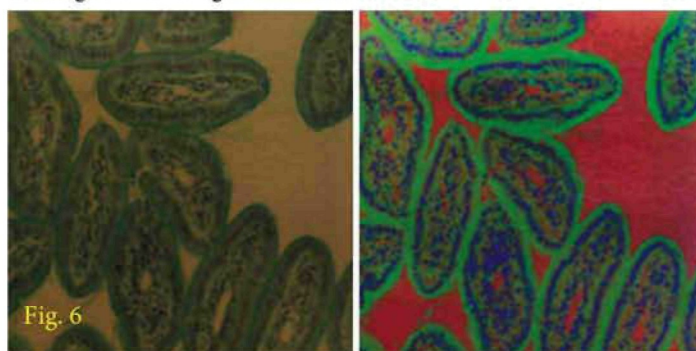*FastLinescans*

Automated Particle Analysis

Spectral Matching

Free software upgrades mean you will never have to replace your system!

IXRF Systems

					Table 1					
Channel	Significance	Al	Ca	Fe	K	Ng	Na	O	Si	Ti
1	49.2%	0.238	0.221	-0.601	0.041	0.255	0.112	-0.055	0.634	-0.221
2	20.9%	0.397	-0.681	0.063	0.402	0.161	0.080	0.217	0.174	0.321
3	12.8%	0.067	-0.155	0.563	-0.278	0.040	0.049	0.281	0.371	-0.597
4	7.0%	-0.395	-0.165	0.098	-0.521	0.259	-0.026	-0.148	0.405	0.531
5	4.8%	-0.525	0.143	0.160	0.548	0.588	-0.095	0.101	0.014	0.111
6	2.7%	0.246	0.622	0.323	0.105	-0.086	0.153	0.419	0.188	0.442
7	1.3%	0.127	-0.015	-0.271	-0.412	0.490	0.028	0.559	-0.432	-0.016
8	0.8%	0.464	0.150	0.322	-0.079	0.494	0.148	-0.592	-0.192	0.002

cal microscope image processed using the principal components technique.

PCA can also make processing and measuring samples with difficult or complex staining more tractable. In Figure 6, the original image (left) shows a stained intestinal tissue sam-

correspond to regions in the original image with characteristic values of the original data. The six clusters shown in the colocalization plots in Figure 4 represent the phases present in the sample. These are visible as the red, yellow, orange, green, blue, and purple regions in the left-most image in Figure 3.

There are many other common applications for which multispectral images are useful. For biological samples, the use of different excitation wavelengths for various stains and dyes produce multiple source channels suitable for principal components analysis. In addition, many of the most effective stains do not produce as much color contrast as desired for visualization or for subsequent thresholding and measurement. To make effective use of PCA techniques, the

Fig. 4
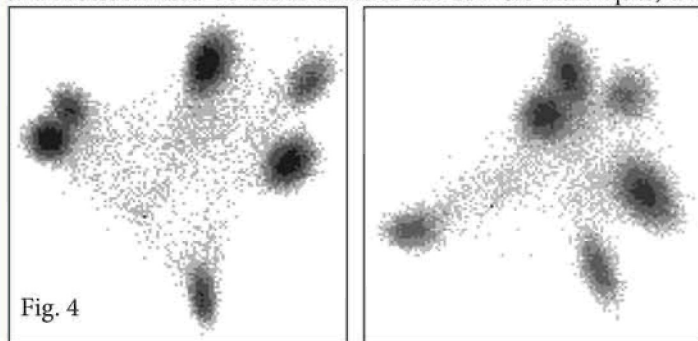
selection of specimen preparation and imaging techniques should be broadened to include as wide a variety of stains, wavelengths, bright and darkfield illumination, and other conditions as possible. The principal components transform will, by definition, maximize contrast in the most significant channels. Figure 5 shows an example: a confo-
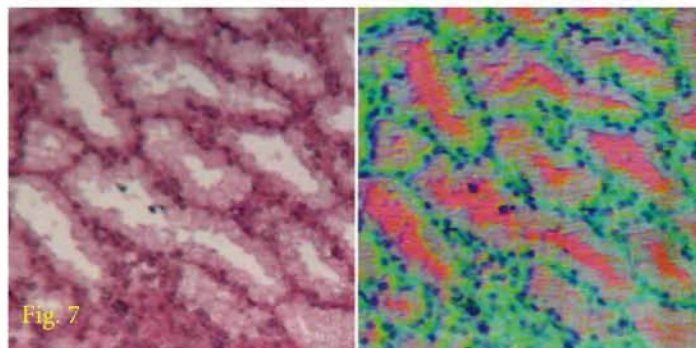

Fig. 5

ple. The brown background is a common effect of many stains, but makes thresholding the structure present difficult, since brown contains red, green, and blue. Applying a principal components transform to the image produces an increase in contrast-as shown. In the resulting image (right), the boundaries of the cells are distinct from the background, and the internal structure of the cell is distinct, making thresholding an easier task.

Fig. 6

One of the most difficult problems in measurement of stained tissue arises when the stain produces only subtle differences in saturation or brightness and very little color variation. Transforming to another color space, such as Lab or HSI, does not produce enough contrast or variation to threshold the different structures. Even visualization is difficult. By using a principal components transform and assigning the most significant channels to the red, green, and blue channels for display (Figure 7), the structure is revealed clearly and measurements can be made.

Conclusion

Principal components analysis, or PCA, a standard method for statistical analysis, provides a powerful method for processing and analyzing image data. While this technique has largely been restricted to remote sensing applications, there are many cases where microscopists can benefit from PCA. The principal channels with highest significance capture the most information in the image and provide maximum contrast- for viewing and measurement. The covariance matrix also provides useful data about the source channels. ■
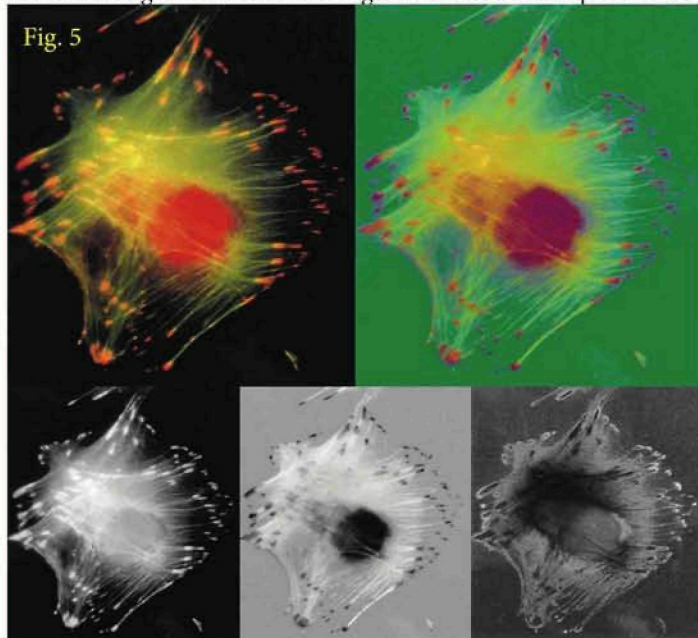

Fig. 7

System evolution in practice

Mother Nature has achieved some of the most impressive examples of seamless integration. We also make every effort to apply this strategy to our comprehensive software and hardware solutions for the fields of imaging and microscopy.

The fruits of our labors: **iTEM** - the new TEM imaging platform for use in conjunction with the latest in TEM camera technology, our 10 MegaPixel side-mounted camera, **Morada**. **Scandium** - our new integrated SEM platform for use with all SEMs.

And finally, there's **analySIS FIVE®**, our universal imaging platform for use in combination with our wide range of light microscopy camera solutions.
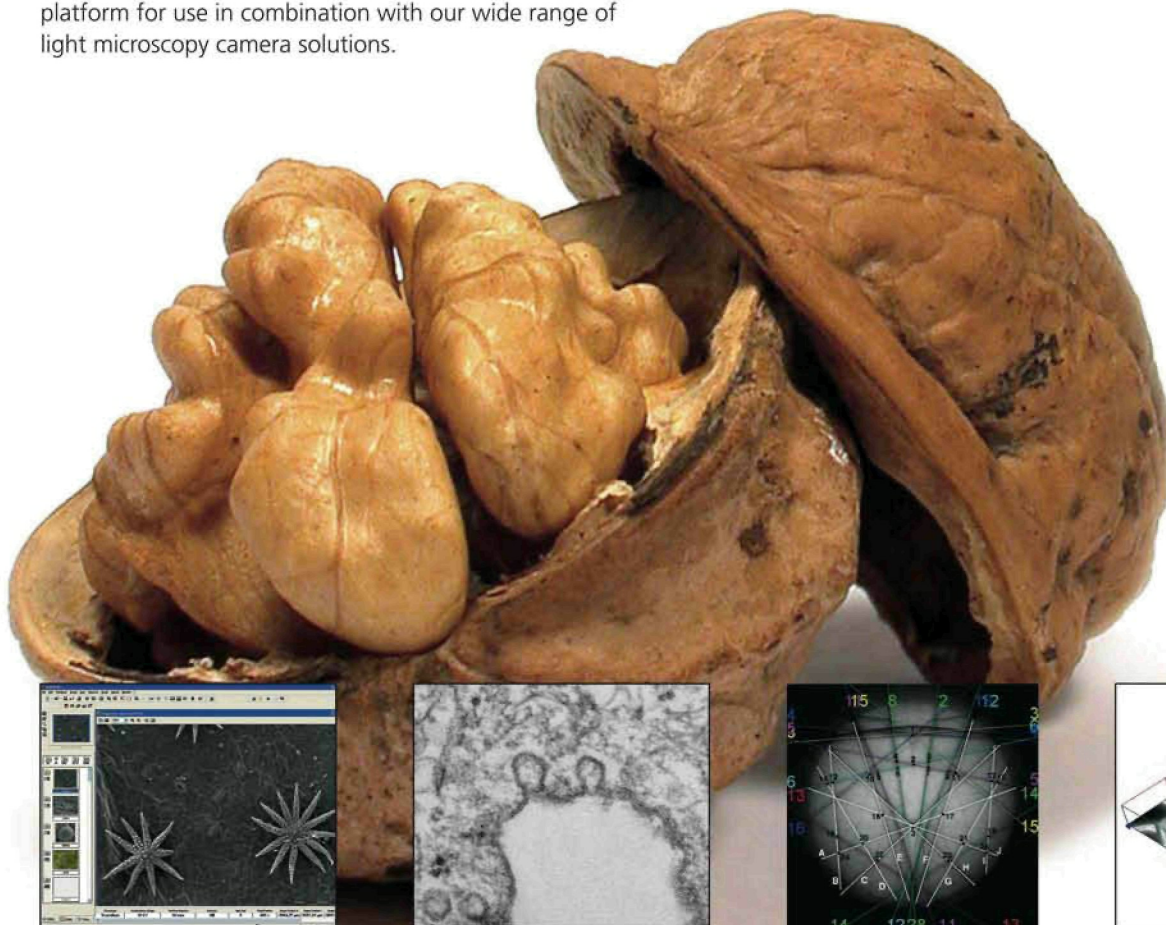
Turning Visions into Realities

Soft Imaging System

SIS