SAND2008-7508C

Haar Wavelet Compression and Empirical Variance Scaling of Large Hyperspectral Images

EAS 2008 Paper 334

11/18/08

Michael R. Keenan and Mark H. Van Benthem
Sandia National Laboratories, Albuquerque, NM

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Overview

- Motivation for work
- Importance of noise scaling
- The Haar wavelet
- Estimation of noise in hyperspectral images
- Performance comparison of empirical with analytical noise estimate
- Results and Conclusions
- Summary

Motivation

- Proper scaling of data prior to analysis is critical to obtaining the best unbiased estimate of factors
- Empirical estimates of noise do not assume an underlying noise distribution
- The Haar wavelet is easy to compute
  - Simple to understand compression
  - Appropriate for hyperspectral chemical images
  - Reduces data analysis time (PCA, MCR, PARAFAC)
  - Improves signal-to-noise ratio (SNR)

Importance of Noise Scaling

- **Least squares procedures assume that errors are normally distributed and this noise variance is independent and identically distributed (*iid*)**
- **If this assumption is invalid, the method allows larger error variances to have undue influence on the results**
- **For data with non-*iid* noise, (*e.g.*, Poisson, binomial, etc) noise scaling (or weighting) attempts to restore the normal, *iid* assumption**

Sandia National Laboratories

Methods of Noise Scaling

- Maximum likelihood (MLE)
  - Scales each variable and observation individually
    - Difficult to get a good estimate of the uncertainty
    - A poor variance estimate can lead to a worse result than doing nothing
  - Time and computationally intensive
- Optimal scaling
  - Based on explicit form of the nature of noise
    - Assumes a rank-1 noise structure
  - Like ML, weights each datum by its uncertainty
  - Noise variance can be estimated in various ways
    - Analytical distributions (Gaussian, Poisson, binomial)
    - Empirical estimates of variance

Noise Estimates from Hyperspectral Images

- Min/Max Autocorrelation Factors*
  - Estimates the noise covariance matrix
  - Relies on nature of remote sensing images
    - Signal in adjacent image pixels: strongly correlated
    - Noise in adjacent image pixels: uncorrelated
- Shift Difference*
  - Vertical or/and horizontal neighbor difference
- Poisson† and Binomial‡ Scaling
  - Data and errors distribution-based

*Green, et al., IEEE Trans. Geosci. & Remote Sens., **26**(1), 1988

† M. R. Keenan and P. G. Kotula, *Surf. Interface Anal.*, **36(3)**, 203 (2004)

‡ M. R. Keenan , et. al., *Surf. Interface Anal.* **40**(2), 97-106 (2008)

The Haar Wavelet

- Simple wavelet transform

$$H = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

  - Apply transform to each two-row, -column pair
  - Basically, a summation of the data elements
- For odd or non-radix-2 dimensioned data, the data can be zero-padded to desired size
- In images, transform is applied to each color
- Could also apply to wavelength or other data modalities
- Haar wavelet is orthogonal and symmetric
  - Preserves data structure (order of application unimportant)
  - Treats adjacent pixels in equivalently
  - Follows the linear-additive model

Sandia National Laboratories

How the Haar Transform Works

Given the matrix below. One pass of the transform yields…

$$H \times \begin{bmatrix} 5 & 1 \\ 6 & 2 \end{bmatrix} \times H$$

5	1	6	2	1	7	1	4
6	2	4	2	1	0	3	6
0	1	5	5	1	8	1	0
6	1	4	6	6	9	0	8
3	0	1	4	5	9	9	5
0	6	4	2	0	8	3	8
4	2	8	9	9	7	2	3
1	5	8	0	7	5	3	4

3.5	3.5	2.25	3.5	2	1.5	-1.25	-1.5
2	5	6	2.25	1	-0.5	-2.5	-1.75
2.25	2.75	5.5	6.25	-0.75	-0.25	-3	-0.25
3	6.25	7	3	-0.5	1.75	1	-0.5
-0.5	0.5	1.75	-1	0	0.5	-1.75	0
-1.5	0	-1.5	-1.75	-1.5	0.5	-1	2.25
-0.75	-0.25	1.5	0.75	2.25	-1.25	1	2.25
0	2.25	1	-0.5	1.5	-2.25	0	0

Now, a second pass of the transform

With the matrix from one-pass. A second pass of the transform yields…

$$H \times \begin{bmatrix} 3.5 & 3.5 \\ 2 & 5 \end{bmatrix} \times H$$

Note the "detail" matrices in yellow hatch don't get transformed again! But you create a new set of compressed and detail images.

3.5	3.5	2.25	3.5	2	1.5	-1.25	-1.5
2	5	6	2.25	1	-0.5	-2.5	-1.75
2.25	2.75	5.5	6.25	-0.75	-0.25	-3	-0.25
3	6.25	7	3	-0.5	1.75	1	-0.5
-0.5	0.5	1.75	-1	0	0.5	-1.75	0
-1.5	0	-1.5	-1.75	-1.5	0.5	-1	2.25
-0.75	-0.25	1.5	0.75	2.25	-1.25	1	2.25
0	2.25	1	-0.5	1.5	-2.25	0	0

3.5	3.5	-0.75	0.625	2	1.5	-1.25	-1.5
3.5625	5.4375	-0.9375	0.8125	1	-0.5	-2.5	-1.75
0	-0.625	0.75	-1.25	-0.75	-0.25	-3	-0.25
-1.0625	0.4375	0.6875	-1.1875	-0.5	1.75	1	-0.5
-0.5	0.5	1.75	-1	0	0.5	-1.75	0
-1.5	0	-1.5	-1.75	-1.5	0.5	-1	2.25
-0.75	-0.25	1.5	0.75	2.25	-1.25	1	2.25
0	2.25	1	-0.5	1.5	-2.25	0	0

You can make many passes of the transform to further compress the matrix or image.

A Picture Example

A typical American family on the Rio Grande. (The dog stayed at home.)
Size: 2048 × 1536 pixels

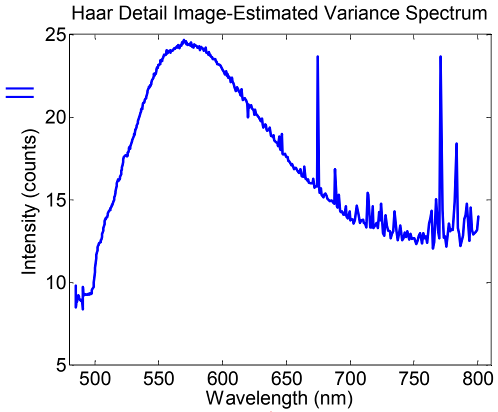
Picture, in upper left quadrant, after one pass of Haar transform (4x compression). The detail images, in the other three quadrants, have been amplified to show effect.

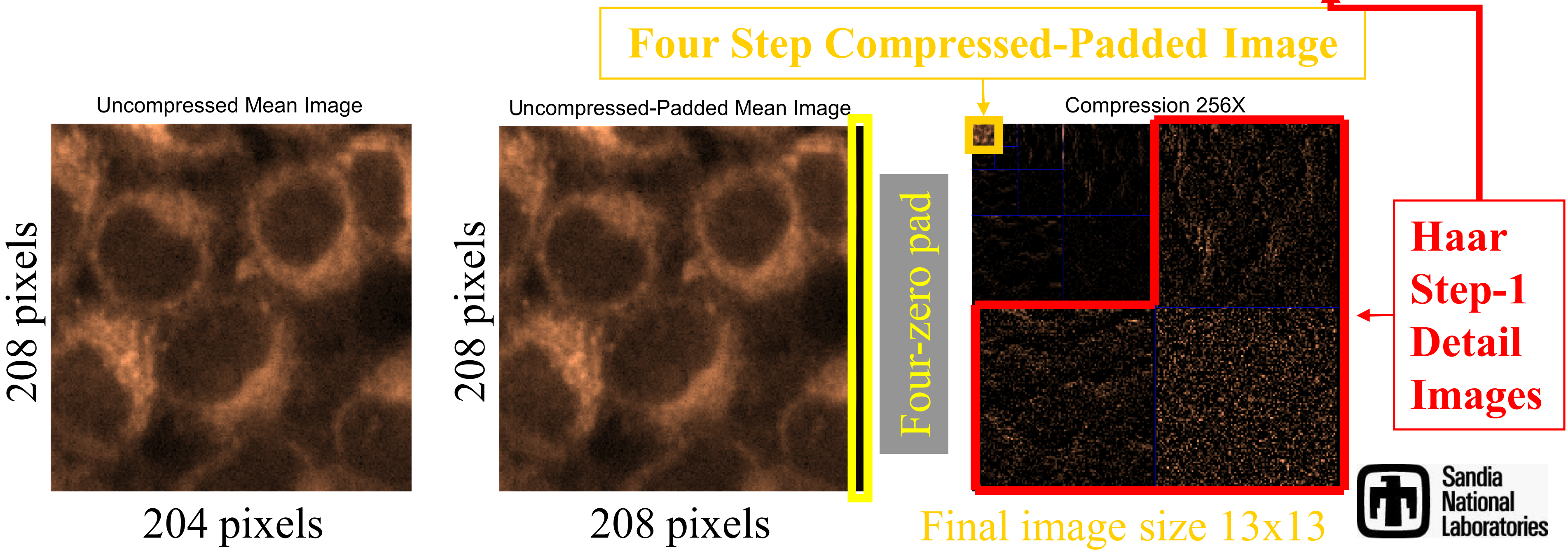
Keep in mind that this is a three color (wavelength) photograph. So each color, RGB, underwent the same compression.

Haar Image Compression and Noise Estimates

- Data from hyperspectral imaging microscope
  - Original image zero-padded
- Detail images give empirical estimate of spectral variance
- Spectral variance is obtained from the sum of squares of the first Haar step detail images

V =





Sandia National Laboratories

Analysis Scheme for Haar Wavelet-Based Empirical Noise Scaling

Data Model

$$\mathbf{D} = \mathbf{A}\mathbf{S}^{\mathrm{T}}$$

Haar Transform

$$\underset{\sim}{\mathbf{D}} = H\left[\mathbf{D}\right]H = \begin{array}{c|c} \mathbf{D} & \dot{\mathbf{D}}_v \\ \hline \dot{\mathbf{D}}_h & \dot{\mathbf{D}}_d \end{array}$$

Step-1 detail image

$$\dddot{\mathbf{D}} = \begin{bmatrix} \dot{\mathbf{D}}_v \\ \dot{\mathbf{D}}_d \\ \dot{\mathbf{D}}_h \end{bmatrix}$$

Empirical Noise Estimate

$$\overline{\mathbf{v}} = \mathbf{1}^{\mathrm{T}}\dddot{\mathbf{D}}$$

Scaling Matrix

$$\mathbf{K} = diag\left(\frac{1}{\sqrt{\overline{\mathbf{v}}}}\right)$$

Factor Analysis (PCA, MCR, etc.)

$$\tilde{\mathbf{D}} = \mathbf{D}\mathbf{K}$$

$$Factor: \tilde{\mathbf{D}} = \mathbf{A}\tilde{\mathbf{S}}^{\mathrm{T}}$$

$$\mathbf{S} = \mathbf{K}^{-1}\tilde{\mathbf{S}}$$

Can also generate image-mode scaling matrix.

Model-based Poisson and Binomial Scaling

Poisson Noise†

$$\bar{\mathbf{v}}_{row} = \frac{\mathbf{1}_c^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}}{c}$$

$$\bar{\mathbf{v}}_{col} = \frac{\mathbf{1}_r^{\mathrm{T}}\mathbf{D}}{r}$$

Binomial Noise‡

$$\tilde{\mathbf{V}} = \frac{\mathbf{D}}{\mathbf{1}_{r,c} - \frac{\mathbf{D}}{m}}$$

$$\bar{\mathbf{v}}_{row} = \frac{\mathbf{1}_c^{\mathrm{T}}\tilde{\mathbf{V}}}{c}$$

$$\bar{\mathbf{v}}_{col} = \frac{\mathbf{1}_r^{\mathrm{T}}\tilde{\mathbf{V}}^{\mathrm{T}}}{r}$$

Scaling Matrices

$$\mathbf{G} = diag\left(\frac{1}{\sqrt{\bar{\mathbf{v}}_{row}}}\right)$$

$$\mathbf{K} = diag\left(\frac{1}{\sqrt{\bar{\mathbf{v}}_{col}}}\right)$$

Factor Analysis

$$\tilde{\mathbf{D}} = \mathbf{GDK}$$

$$Factor : \tilde{\mathbf{D}} = \tilde{\mathbf{A}}\tilde{\mathbf{S}}$$

$$\mathbf{A} = \mathbf{G}^{-1}\tilde{\mathbf{A}}$$

$$\mathbf{S} = \mathbf{K}^{-1}\tilde{\mathbf{S}}$$

$$\mathbf{D} = m\left(\mathbf{1} - e^{-\frac{1}{m}\mathbf{AS}^T}\right) \rightarrow \underset{\sim}{\mathbf{D}} = -m\ln\left(\mathbf{1} - \frac{\mathbf{D}}{m}\right) = \mathbf{AS}^T$$

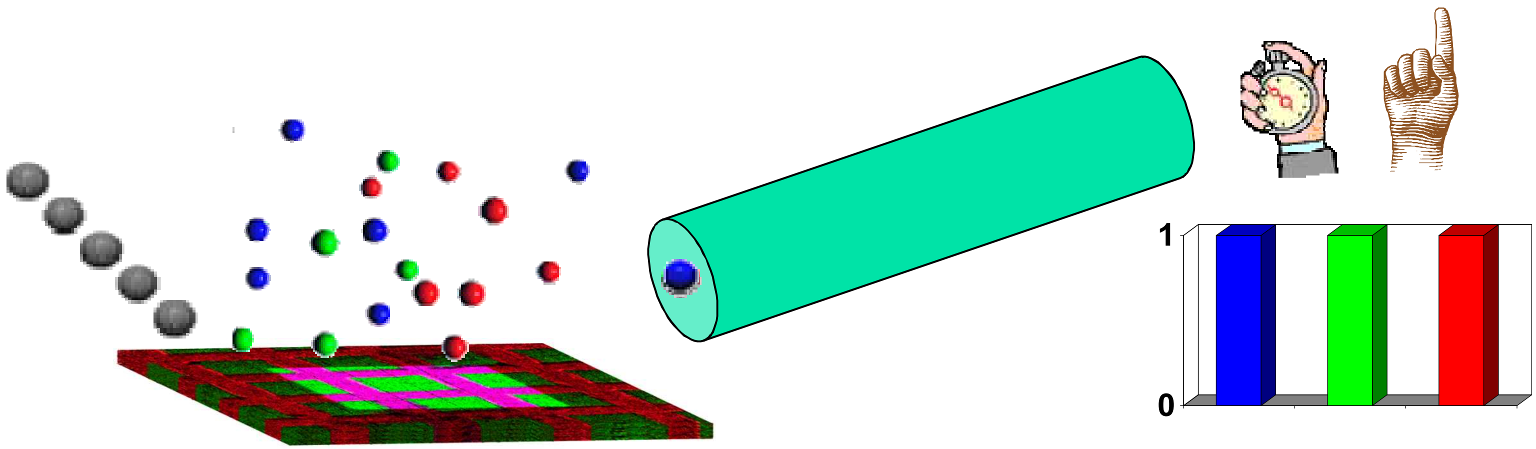
Binomial Data transformation

†**M. R. Keenan and P. G. Kotula, Surf. Interface Anal. 36, 203 (2004)**

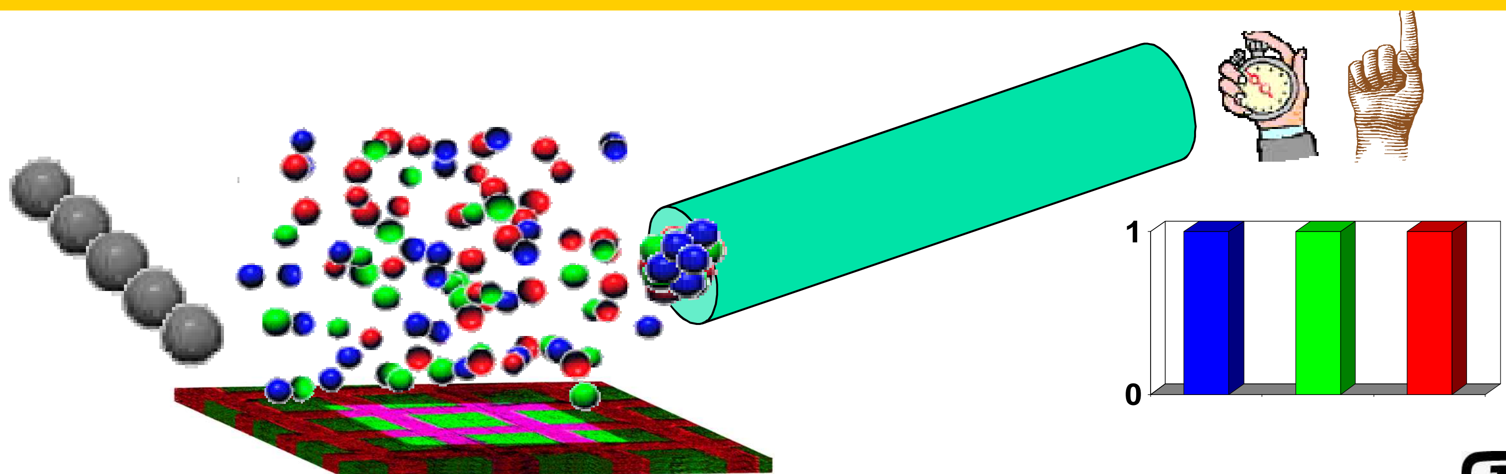
‡M. R. Keenan , et. al., *Surf. Interface Anal.* **40**(2), 97-106 (2008)

Sandia National Laboratories

TOF-SIMS Provides Both Poisson and Binomial Data

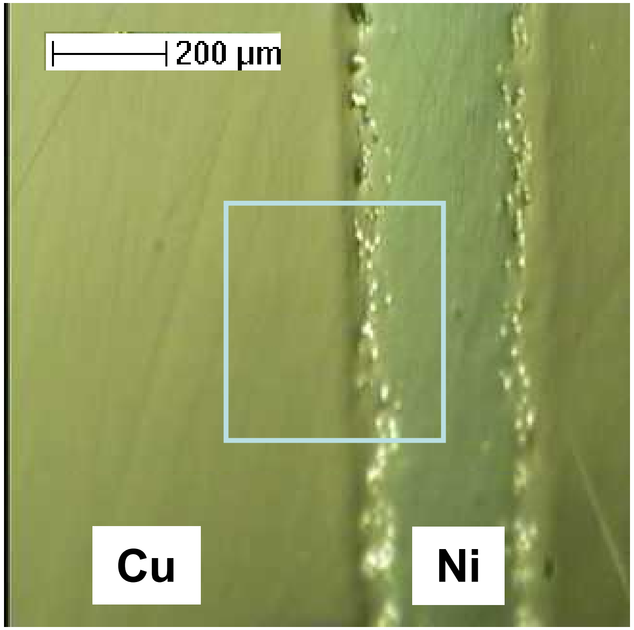


TDC detector is insensitive to coincident ion arrivals



Sandia National Laboratories

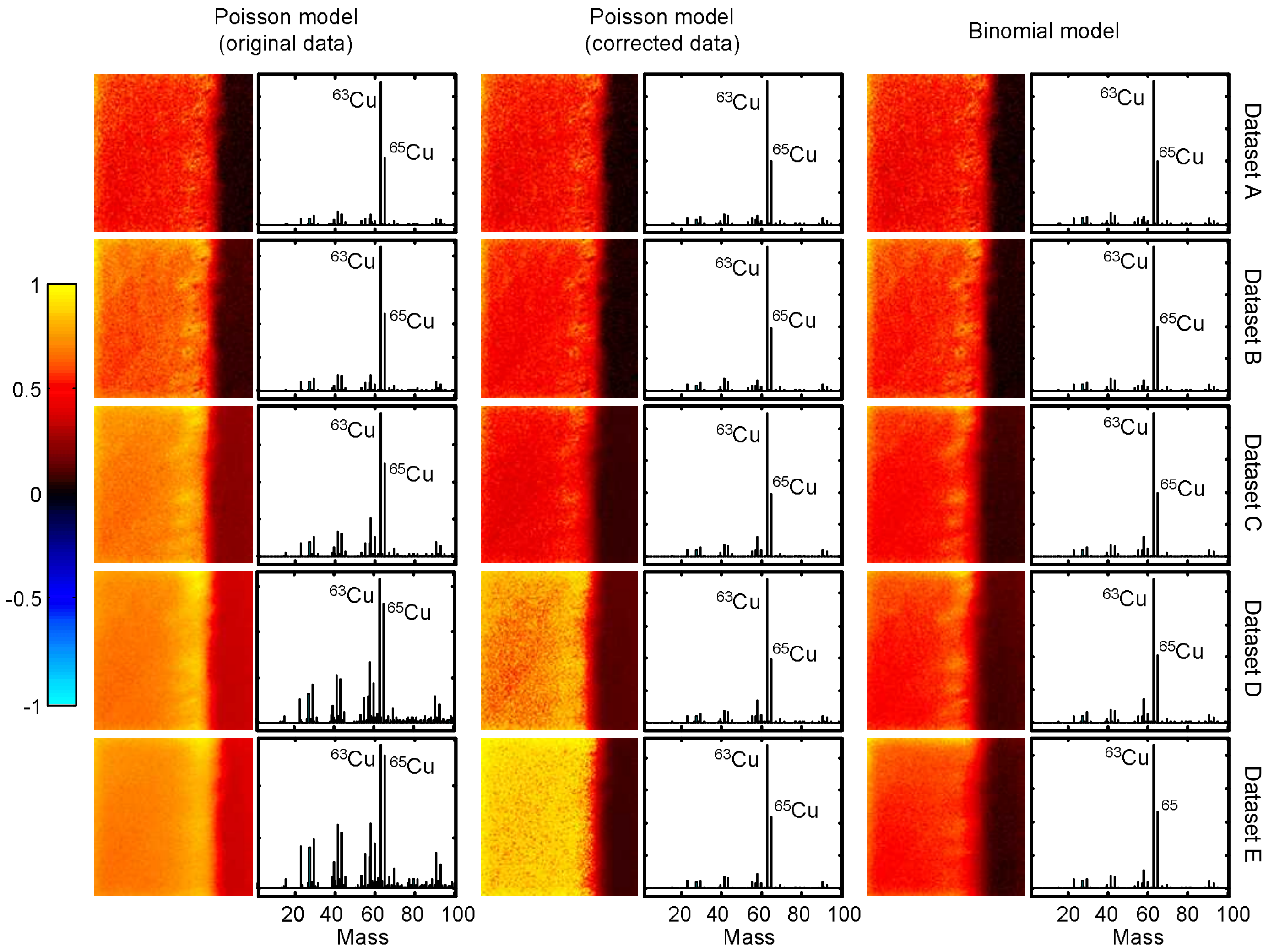
Experimental



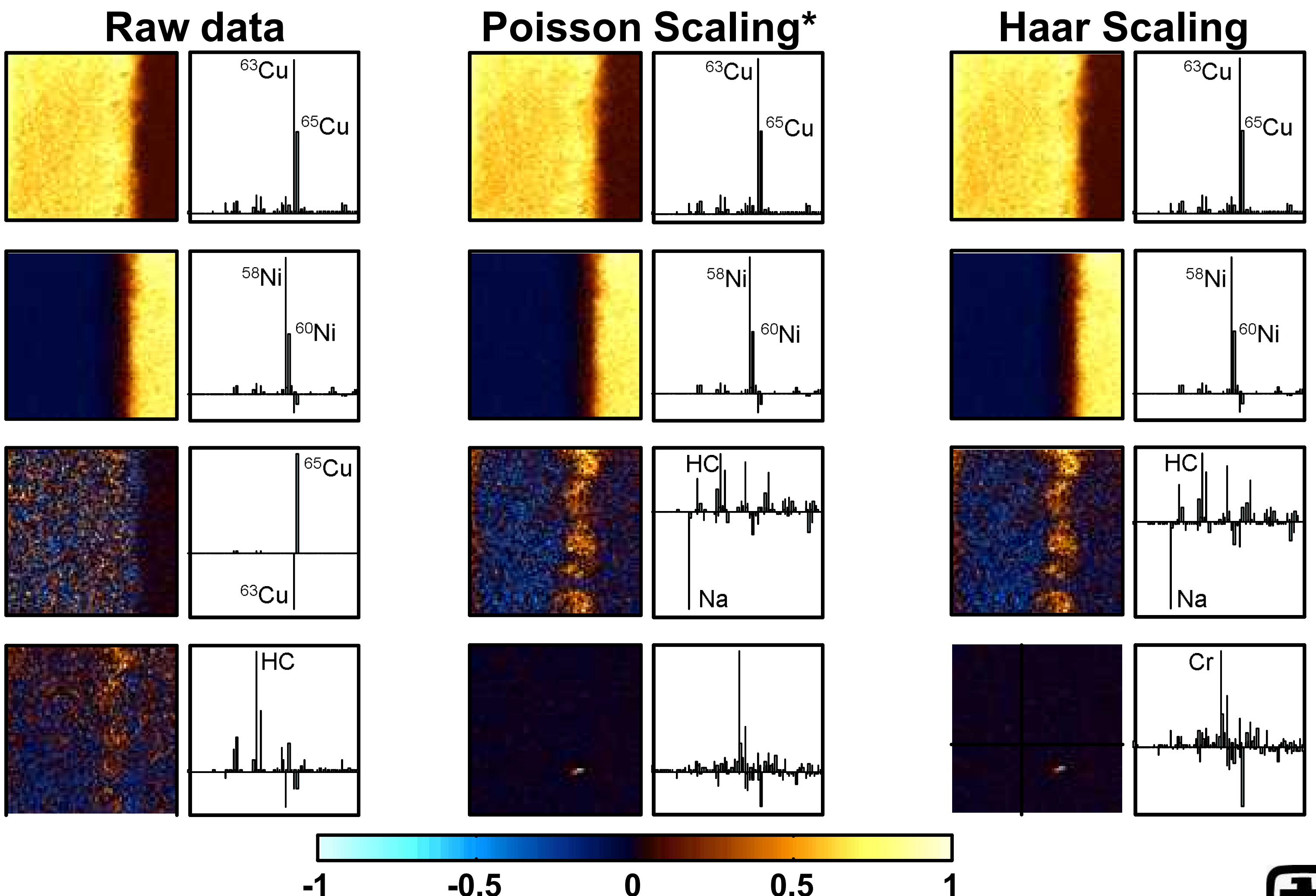
- Copper/Nickel diffusion couple
  - Zinc impurity in the copper
  - Kirkendall voids
- ION-TOF model TOF-SIMS IV
  - Bi-ion source
  - High-current bunched mode
  - 128 x 128 pixels
  - 199 mass channels: 1 – 100 amu
  - 100 raster frames (100 shots/pixel)
- 14 total data sets
  - Vary PI peak width and suppressor
  - Span factor of 27 in secondary ion production
  - Applied multivariate techniques to 5

Sandia National Laboratories

First Principal Component



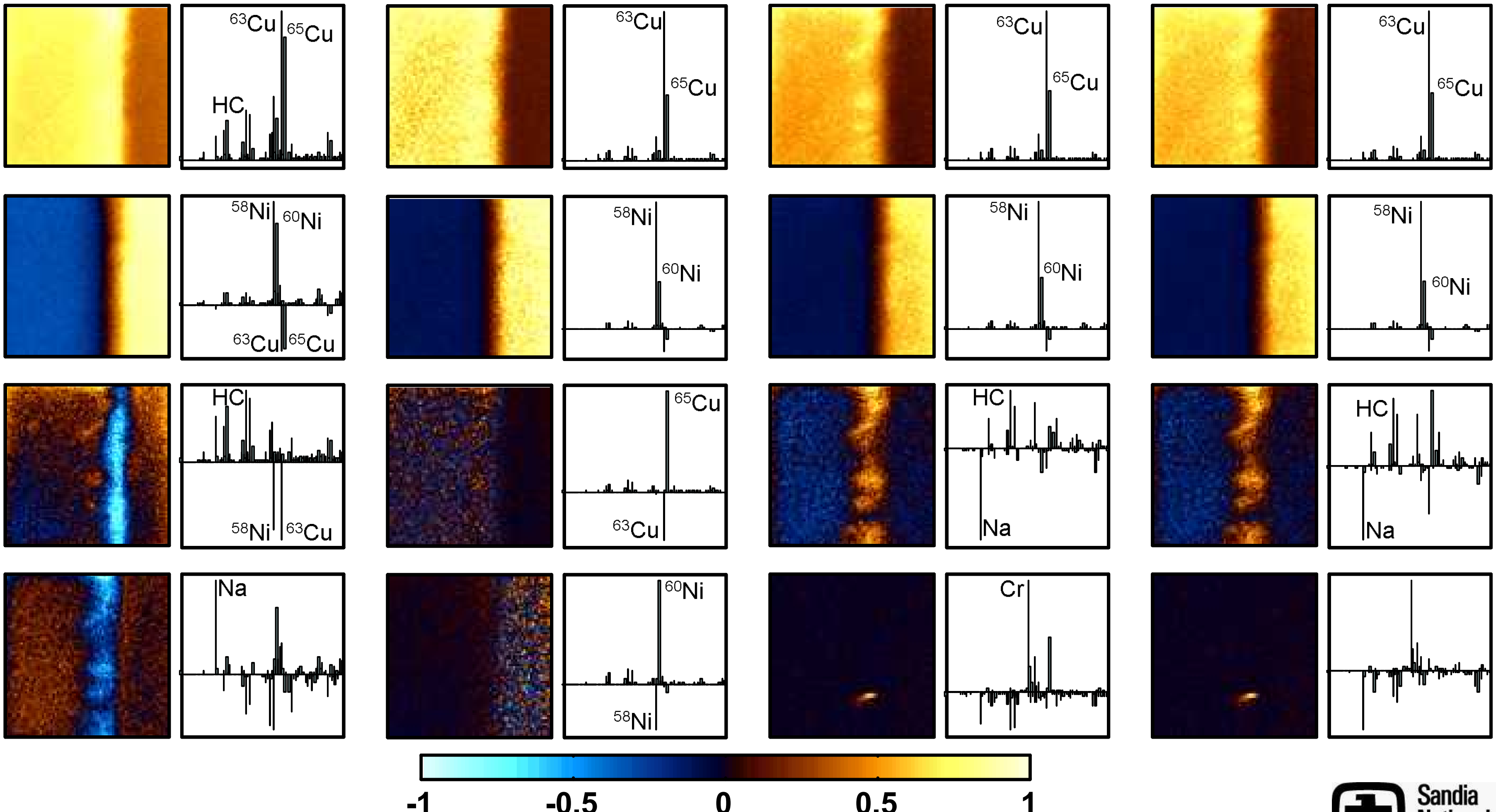
PCA of Poisson-distributed Data



*M. R. Keenan and P. G. Kotula, *Surf. Interface Anal.*, **36**, 203 (2004)

PCA of Binomially Distributed Data

Raw data	Corrected data	Binomial Scaling*	Haar Scaling



*M. R. Keenan, et. al., *Surf. Interface Anal.*, **40**, 97 (2008)

Sandia National Laboratories

Results and Conclusions

- **Noise scaling is an important (pre)processing step for non-*iid* data**
- **Employing the appropriate model is critical when using model-based scaling**
  - **Poisson scaling is unsuitable for binomially distributed data**
- **Using Haar wavelet to estimate noise is model or distribution independent**

Summary

- **Motivation for work**
  - **Seeking best estimate for noise scaling**
- **Importance of noise scaling**
  - **Transform noise to *iid* in least squares assumptions**
- **The Haar wavelet**
  - **Simple, effective wavelet transform**
- **Estimation of noise in hyperspectral images**
  - **Model-based versus empirical methods**
- **Performance comparison of empirical with analytical noise estimate**
  - **Empirical estimate from Haar detail images more flexible**

Acknowledgments

- **Vincent S. Smentkowski**
  - **General Electric Global Research**
    Niskayuna, NY 12309
- **James A. (Tony) Ohlhausen, Paul G. Kotula, Ryan Davis and Mike Sinclair**
  - **Sandia National Laboratories**
    Albuquerque, NM 87185-0886
- **Allan B. Brasier and Ping Liu**
  - **University of Texas Medical Branch**
    Galveston, TX

Sandia National Laboratories