Spectral Density-based Statistical Measures for Image Sharpness and the Related Software

Nien Fan Zhang, Andras Vladar, Michael Postek, and Robert Larrabee

National Institute of Standards and Technology
Gaithersburg, MD 20899, USA
1. Introduction

2. Two-dimensional stationary processes

3. Moments of a multivariate distribution

4. Use generalized variance and bivariate kurtosis to measure image sharpness

5. Software to measure SEM image sharpness

6. Conclusions
1. Introduction

• In industry applications, such as automated on-line semiconductor production, there is a growing realization of the need for the development of a procedure for periodic performance testing of the image sharpness for scanning electron microscope (SEM).

• Postek and Vladar (1996) and Vladar, Postek and Davidson (1998) proposed a procedure to test image sharpness based on the spatial Fourier transform of the SEM images.

• It is observed that when an SEM image is visually sharper than a second one, the high spatial frequency components of the first image are larger than those of the second.

• We proposed two statistical image sharpness measures based on the spectral density for a 2-d stationary process.
Figure 1. Two SEM Images and their 2-D periodograms
2. Two-dimensional stationary processes

• We consider a continuous 2-dimensional processes \( \{X_{t,\tau}\} \)

The process is called weakly stationary if

1. \( E[X_{t,\tau}] = \mu_x \), a constant for all \( t \) and \( \tau \)
2. \( Var[X_{t,\tau}] = \sigma_x^2 \), a constant for all \( t \) and \( \tau \)
3. The (two-dimensional) autocovariance function \( Cov[X_{t,\tau}, X_{t+s,\tau+u}] \) only depends on \( s \) and \( u \) and thus denoted by \( R(s,u) \)

The two-dimensional autocorrelation is defined as

\[
\rho(s,u) = \frac{R(s,u)}{\sigma_x^2}
\]
• The 2-dimensional power spectral density function \( f(\omega, \nu) \) is given by

\[
f(\omega, \nu) = \left( \frac{1}{2\pi} \right)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(s, u) e^{-is\omega - iu\nu} \, ds \, du
\]

The spectral density function has the following properties:
1. \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\omega, \nu) \, d\omega \, d\nu = 1 \)
2. \( f(\omega, \nu) \geq 0 \)
3. \( f(-\omega, -\nu) = f(\omega, \nu) \)

Based on 1 and 2, the spectral density function is like a 2-d probability density function of a bivariate random vector.
3. Moments of a multivariate distribution

(1) Generalized variance

Let \( W \) be a \( p \)-dimensional random vector with up to 4\(^{th}\) finite moments. Let \( \mu_w \) be the mean vector and \( \Sigma_w \) be the covariance matrix. The generalized variance is \( |\Sigma_w| \). When \( p=2, W = (W_1, W_2)' \)

\[
\Sigma_w = \begin{bmatrix}
\sigma_{W_1}^2 & \rho_{12} \sigma_{W_1} \sigma_{W_2} \\
\rho_{12} \sigma_{W_1} \sigma_{W_2} & \sigma_{W_2}^2
\end{bmatrix}
\]

where \( \rho_{12} \) is the correlation between \( W_1 \) and \( W_2 \). The generalized variance is

\[
GV_W = \begin{vmatrix}
\sigma_{W_1}^2 & \rho_{12} \sigma_{W_1} \sigma_{W_2} \\
\rho_{12} \sigma_{W_1} \sigma_{W_2} & \sigma_{W_2}^2
\end{vmatrix} = \sigma_{W_1}^2 \sigma_{W_2}^2 (1 - \rho_{12}^2).
\]

The generalized variance is a scalar measure of dispersion of \( W \).
(2) Multivariate kurtosis

For a univariate random variable $Z$ with mean $\mu_z$ and finite moments up to the 4th, the kurtosis is defined as

$$\beta_2 = \frac{E[(Z - \mu_z)^4]}{\{E[(Z - \mu_z)^2]\}^2}$$

Assuming the probability density function $f_z(x)$ exists, the kth central moments is

$$m_k = E[(Z - \mu_z)^k] = \int_{-\infty}^{\infty} (x - \mu_z)^k f_z(x)dx$$

Thus,

$$\beta_2 = \frac{m_4}{(m_2)^2}$$
• The kurtosis of a Gaussian distribution is 3. The value of a random variable can be compared with 3 to determine whether its distribution is "peaked" or "flatted-topped" relative to a Gaussian distribution.

• Figure 2 shows the probability density functions of the standard Gaussian distribution, the Student $t$ distribution with df = 5 and the density function $S(x)$ discussed in Kaplansky (1945). The $t$ distribution has kurtosis of 9. The distribution with the density function of $S(x)$ has a kurtosis of 2.667.

• It is clear that the distribution with a smaller kurtosis is more flat-topped or has a larger shoulder than that with a larger kurtosis. The distribution with the density function of $S(x)$ has the largest shoulder and smallest kurtosis among the three distributions.
Figure 2. Probability density functions of three distributions
Multivariate kurtosis has been proposed by Mardia (1970). Let $W$ be a $p$-d random vector with finite up to the 4th moments. The kurtosis of $W$ is defined as

$$\beta_{2,p} = E\{(W - \mu_w)'\Sigma_w^{-1}(W - \mu_w)\}^2$$

The bivariate kurtosis for $W = (W_1, W_2)'$ is calculated by

$$\beta_{2,2} = \frac{\gamma_{4,0} + \gamma_{0,4} + 2\gamma_{2,2} + 4\rho_{12} (\rho_{12}\gamma_{2,2} - \gamma_{1,3} - \gamma_{3,1})}{(1 - \rho_{12}^2)^2}$$

where $\rho_{12}$ is the correlation between $W_1$ and $W_2$ and $\gamma_{k,l}$ is

$$\gamma_{k,l} = \frac{E[(W_1 - \mu_1)^k (W_2 - \mu_2)^l]}{\sigma_1^k \sigma_2^l}$$
• The bivariate kurtosis of a 2-d Gaussian distribution is 8.

• Figures 3 and 4 show that the curvatures for the 2-d Gaussian distribution and the bivariate t distribution with a df = 9.

• It is clear that the 2-d Gaussian distribution with a smaller bivariate kurtosis (\(=8\)) has a larger shoulder than that for the bivariate t distribution with a df = 9 and kurtosis = 11.2.
Figure 3. Probability density function of a 2-d Gaussian distribution
Figure 4. Probability density function of a 2-d t distribution
4. Use generalized variance and bivariate kurtosis to measure image sharpness

For a given set of discrete 2-d lattice data, \( \{ w_{t,\tau}; t, \tau = 1, 2, \ldots, N \} \)

The 2-d discrete Fourier transform is

\[
\xi(\omega, \nu) = \frac{1}{2\pi N} \sum_{t=1}^{N} \sum_{\tau=1}^{N} w_{t,\tau} e^{-it\omega - i\nu\nu}
\]

for \( \omega, \nu = 2\pi \lambda / N; -[(N - 1) / 2] \leq \lambda \leq [N / 2] \)

The 2-d periodogram is

\[
I(\omega, \nu) = \left| \xi(\omega, \nu) \right|^2 = \frac{1}{(2\pi N)^2} \left| \sum_{t=1}^{N} \sum_{\tau=1}^{N} w_{t,\tau} e^{-it\omega - i\nu\nu} \right|^2
\]
A discrete density based on \( \{I(\omega_i, \nu_j), i, j = 1, 2, \ldots, N\} \) is obtained by normalization. Namely,

\[
h(\omega_i, \nu_j) = \frac{I(\omega_i, \nu_j)}{\sum_{m=1}^{N} \sum_{n=1}^{N} I(\omega_m, \nu_n)}
\]

The corresponding marginal means and variances are given by

\[
\mu_\omega = \sum_{i=1}^{N} \omega_i \sum_{j=1}^{N} h(\omega_i, \nu_j)
\]
\[
\mu_\nu = \sum_{j=1}^{N} \nu_j \sum_{i=1}^{N} h(\omega_i, \nu_j)
\]
The corresponding covariance and correlation can be obtained. Then by the definition, the generalized variance can be obtained.

We re-inspect Figure 1. The sharp image in Figure 1c and its cone in 1d is wider than that of image in Figure 1a. Thus, the generalized variance can be used to measure image sharpness.

\[ \sigma^2_\omega = \sum_{i=1}^{N} (\omega_i - \mu_\omega)^2 \sum_{j=1}^{N} h(\omega_i, \nu_j) \]

\[ \sigma^2_\nu = \sum_{j=1}^{N} (\nu_j - \mu_\nu)^2 \sum_{i=1}^{N} h(\omega_i, \nu_j) \]
The $\gamma_{k,l}$ is given by

$$\gamma_{k,l} = \frac{\sum_{j=1}^{N} \sum_{i=1}^{N} h(\omega_i, \nu_j)(\omega_i - \mu_\omega)^k(\nu_j - \mu_\nu)^l}{\sigma_\omega^k \sigma_\nu^l}$$

and the bivariate kurtosis corresponding to $h(\omega, \nu)$ can be obtained.
The marginal variances are used to measure the dispersions of the marginal distributions. The difference between the two marginal variances measures the difference between the dispersions of the marginal distributions.

The marginal kurtoses of the marginal distributions are $\beta_{2,\omega} = \gamma_{4,0}$ and $\beta_{2,\nu} = \gamma_{0,4}$. The marginal kurtoses are used to measure the shapes of the shoulders of the marginal distributions. Marginal variances and marginal kurtoses can be used to detect the asymmetry of the periodogram in vertical and horizontal sections corresponding to the astigmatism in image caused by possible instrument vibration. We use

$$r_v = \frac{|\sigma_{\omega}^2 - \sigma_{\nu}^2|}{\min(\sigma_{\omega}^2, \sigma_{\nu}^2)} \quad \text{and} \quad r_K = \frac{|\beta_{2,\omega} - \beta_{2,\nu}|}{\min(\beta_{2,\omega}, \beta_{2,\nu})}$$

to measure the relative difference of the marginal variances and marginal kurtoses. Therefore, checking the marginal kurtosis difference would be the first cut to eliminate anomalous data.
A real example demonstrates how the generalized variance and the bivariate kurtosis are used to measure the sharpness of SEM images. The sample is an etched silicon wafer with the artifact referred to as “grass”, an etching artifact that can occur on silicon wafers during processing.
Figure 5. Five micrographs and their 2-d and 1-d Periodograms
Figure 6. Generalized variances of five samples
Figure 7. Bivariate kurtosis of five samples
Figure 8. Relative differences of marginal variances and kurtoses
5. Software to measure SEM image sharpness

SEM Monitor is the software which provides the techniques developed by Andras Vladar, Mike Postek, and N. F. Zhang for SEM performance analysis. Several sharpness measures including the bivariate kurtosis are used.

Spectel Research Co. in collaboration with NIST and Hewlett-Packard developed the software package running on a workstation system.

NIST now is offering an SEM image sharpness reference artifact (RM 8091), which can be used successfully for these procedures.
Fourier Transform and Kurtosis Calculation

• Metrologia SEM Monitor of Spectel Research Co.
• Calculates image sharpness, eccentricity, and kurtosis
• A proper set of parameters ensure well-tracking results
• Both image sharpness value and kurtosis value can be used to set the electron microscope better, closer to ideal
• All these calculations can be done in real-time, which greatly help the operator in achieving good results
• Certain noises also can be found if method is implemented on a fully automatic system
Fourier Transform and Kurtosis Calculation
Fourier Transform and Kurtosis Calculation

Astigmatic image
6. Conclusions

While the variance is used to measure the dispersion of a probability distribution, the kurtosis of a probability distribution has been used to measure the shape and width of the shoulder of the distribution. These properties are exploited for the spectral density for a two-dimensional stationary process and then successfully applied to SEM image to measure image sharpness. The bivariate kurtosis and the corresponding relative difference of the marginal kurtoses are implemented in a workstation system and used in semi-conductor industry for SEM performance monitoring.