

Estimation Error in Image Quality Measurements

Peter D. Burns, Carestream Health, Inc
Rochester NY, USA

ABSTRACT

The development and adoption of standard image quality measurement and analysis methods have helped both the evaluation of competing imaging products and in technologies. Inherent in the interpretation of results from any particular evaluation, however, are the variation of the method itself, the sampling of test images, equipment, and test conditions. Here we take a statistical approach to measurement variation, and interpret the objective as being the estimation of particular system or image properties, based on data, collected as part of standard testing. Measurement variation was investigated for two signal-transfer methods commonly used for digital camera and scanner evaluation: the ISO 12233 slanted-edge spatial frequency response and the dead-leaves method for texture-MTF evaluation being developed by the Camera Phone Image Quality (CPIQ) Initiative. In each case, the variation due to the selection of analysis regions was computed by repeated analysis. The slanted-edge methods indicated a relative error in the range of 1-3% depending on the nature of the region selection. For the dead-leaves method, the amplitude spectrum (square root of the noise-power spectrum) showed a relative error of approximately 4-6%, however, this can be reduced by applying spectral estimation methods commonly used in image noise analysis.

Keywords: image quality measurement, measurement variation, MTF, SFR, dead-leaves MTF

1. INTRODUCTION

The adoption of image quality performance standards has led to the common notion that imaging performance, be it for signal microstructure, image noise, color accuracy, or optical distortion is something to be measured and specified. When performance measures are derived using standard test targets and analysis software, many sources of variation can be reduced. This can lead to the conclusion that even minor differences in computed results are due to the equipment or process under test. An improved understanding of the origins and control of measurement variation, however, can be obtained by addressing performance measurement as an estimation task.

Practical systems can introduce error during color calibration and when acquiring image scene information. For large populations, it is often assumed that the error can be modeled as a random variable having a zero mean. This type of variation reduces system precision. In the case of a single color instrument, camera, or scanner, however, error due to deterioration of a physical standard, optical filter, or detector will introduce a predictable error into the signal. This type of error is often described as a bias error and reduces system accuracy. Both bias and variation errors are modified as the signals are transformed (processed) into their final form, and error-propagation methods^{1,2} are often used to describe the influence of the data-processing path on the magnitude of bias error. In this paper, we address the measurement errors for two image quality measurements: the ISO 12233 slanted-edge spatial frequency response,^{3,4} and the dead-leaves method for texture-MTF evaluation⁵ being developed by the Camera Phone Image Quality (CPIQ) Initiative.

Error propagation analysis is often used to predict the transformation of variation or error when a signal undergoes a transformation. Less-often considered, however, is the measurement error inherent in the calculation of several derived measurements, such as for visual color-difference, acutance or image sharpness, graininess, modulation transfer function (MTF), contrast-to-noise ratio, or noise-power spectrum.

2. MEASUREMENT VARIATION

Consider the signal variations that result from exposure and detector characteristics. When signals are combined, e.g., by color-matrix or spatial operations, so are the variations and this can be modeled as forming a function of random variables. After describing the technical basis for analysis of the estimation error in terms of component-parameter variations, we address the magnitude of measurement variation for particular measurements of interest. Based on this

approach, it is possible to model the propagation of the first- and second-order error statistics in terms of expected bias and variation error when applying such measurements in specific performance monitoring or production tasks.

Many image quality measurements follow four basic steps of Fig. 1. In this report, we are primarily concerned with the measures based on well-established imaging parameters or descriptors, such Modulation Transfer Function (MTF),⁵ noise power spectrum, mean detected signal. Others based on novel or higher-order statistics, however, also follow this path. The first step is the adoption and following of particular evaluation conditions and methods. For several international standards, this includes the use of specific test targets (charts) and lighting conditions. For other evaluations, this could involve extracting images or other data from collections of images.



Figure 1. Several steps used in image quality evaluation based on imaging-system or image parameters.

The second part of the image quality evaluation is aimed at the extraction of image regions (for test images) on which the imaging parameters will be based. This can include the selection of regions of interest (ROIs) of image features for particular parameters; for example, edge features used in the edge SFR measurements described in ISO 12233, or the circular sinusoidal elements used in the Siemens-star SFR evaluation.⁶

The next step is the computation of imaging parameters such as average signal value, noise amplitude, MTF or SFR, image frequency spectrum or noise-power spectrum. Based on one or more of these imaging parameters, the image quality metric is computed in the final operation. Often this step involves the combining of several measurements, e.g., when we apply a visual contrast-sensitivity function and integrate (weighted sum) over spatial frequencies or form a signal-to-noise ratio (SNR).

While it is common to consider the physical characteristics of image capture and signal readout as a starting point for discussions of variation in imaging performance, we will start by considering the variation in the derived measurements of imaging performance. The reason for this is that the observed variation in an imaging measurement, such as MTF is influenced by the details of data acquisition (test target and testing protocol), data selection (image sampling – ROIs in test images), and the processing methods used (data smoothing, parametric fitting, etc.). We start by considering several common types of signal transformations that are used when estimating image-quality related measures.*

If signal x , is subject to variation or error, we can model this as having a random component

$$x = \mu_x + e_x$$

where e_x is a zero-mean random variable with a probability density function and corresponding variance, σ_x^2 . We can define a bias of x by comparing the mean value with the true value, K

$$b_x = \mu_x - K,$$

which is zero if x is unbiased.

If we have an understanding of the magnitude and nature of measurement variation, under certain conditions, it is possible to model the influence of this variation on the resulting estimated image quality parameters. If we limit our attention to first- (bias) and second-order (variance) statistics, then useful approximations are available by casting the estimation as a transformation of random variables. When several random signals (variables) are related, a matrix-vector notation is often used. Examples of related signals include the values of a (sampled) MTF, noise-power spectrum, or CIELAB coordinates.

A general multivariate transformation can be described as

* Often the measured value is actually estimated as the mean value of observed data, but for our purposes, we will refer to image quality metrics estimated from measurements of the system or image.

$$\mathbf{y} = \mathbf{f}(\mathbf{x}), \quad (1)$$

where

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T, \quad \mathbf{y} = [y_1, y_2, \dots, y_m]^T,$$

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_m(x_1, x_2, \dots, x_n) \end{bmatrix}.$$

The superscript T indicates matrix transpose. If the bias in each component signal of \mathbf{x} is written as a vector

$$\mathbf{b}_x = [b_{x_1}, b_{x_2}, \dots, b_{x_n}]^T,$$

the output bias vector is

$$\mathbf{b}_y \cong \mathbf{J}_f \mathbf{b}_x, \quad (2)$$

where

$$\mathbf{J}_f = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \ddots & & \\ \vdots & & \ddots & \\ \frac{\partial y_m}{\partial x_1} & & & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

Each element of \mathbf{J}_f is evaluated at $(\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_n})$.

The variation in \mathbf{y} can be expressed in the form of the covariance matrix, Σ_x , where the diagonal elements are the variance values, e.g., $\sigma_{11} \equiv \sigma_{x_1}^2$. The resulting covariance matrix for \mathbf{y} is¹

$$\Sigma_y \cong \mathbf{J}_f \Sigma_x \mathbf{J}_f^T. \quad (3)$$

Common matrix transformations used in image quality estimates are those applied to red, green, and blue signals are part of color correction, illuminant transformation and the like. The propagation of bias error for a simple matrix transformation of a set of color signals can be understood as a special case of Eq. (3). For example,

$$\mathbf{y} = \mathbf{M}\mathbf{x}. \quad (4)$$

Since each element of the matrix, \mathbf{M} , is a constant, the resulting bias and covariance in \mathbf{y} are,

$$\mathbf{b}_y \cong \mathbf{M}\mathbf{b}_x, \quad (5)$$

$$\Sigma_y \cong \mathbf{M}\Sigma_x \mathbf{M}^T. \quad (6)$$

Equations (5) and (6) provide us with a way to express (or specify) the bias error and variations in computed image quality measures in terms of the underlying component measurement uncertainty. We now address this variation for two measurement methods used for evaluation of image capture systems.

3. SLANTED-EDGE SFR MEASUREMENT

The MTF is a measurement of the transfer of image (signal) information. In the past, methods based on test features, such as edges, lines, random fields, and periodic signals, such as square-waves and sinusoids, have been used for image quality evaluation. Choosing a standard method for a particular application calls for a balance of several factors, such as ease of testing and analysis and minimizing (or including) the influence of other system characteristics, such as lens-field variation, image noise, and more recently, certain adaptive image processing operations.

The theory behind edge gradient analysis for MTF measurement is well established,^{6,8,9} and its application to digital capture devices via slanted edges is also well documented and part of an ISO standard.³ The method is based on the image (or system output) due to an input edge feature of high optical quality. Often the measured signal modulation can be taken as an estimate of the MTF of the system. In other cases, the output modulation is divided by the input edge modulation, frequency by frequency, to yield the measured system MTF. We will refer to the single output modulation, normalized to unity at zero frequency, as the spatial frequency response (SFR), which is consistent with the standard.

There are three basic operations: acquiring an edge profile from the (image) data, computing the derivative in the direction across the edge, and computing the discrete Fourier transform of this derivative array. If we interpret the slanted-edge SFR measurement as an estimation problem, several sources of error can be seen as introducing bias and/or variation into the estimated SFR.

For example, standard software programs do not require a precise alignment of the edge feature in scene with image sampling array. This requires estimating the edge location from the data. An error introduced into the computed slope propagates as a bias error in the resulting SFR or MTF measurement.⁴ Both bias error and variation are also introduced into practical measurements by image noise fluctuations.

Rather than investigate the various sources of systematic effort or variation that may be observed in a particular test or laboratory, our purpose is to focus on the variation that is intrinsic to a particular testing facility. As an example, we take the measurement of a camera SFR and the establishment of confidence limits around reported results. So for our purposes, it will be assumed that systematic, bias errors are not the objective, but rather an understanding of the variation within a single reported measurement. In the following examples, we have used test images of a standard test chart for a digital SLR camera typically used by the advanced amateur.

3.1 Observed variation

For a specific test facility,* we can evaluate the precision in the camera SFR results by gathering results of repeated measurements. To measure inter-image variation, several test images were captured. From these image files, 10 separate subimages of identical size (128×300 pixels) were used to compute the edge SFR. The results were then compared as shown in Fig. 2(a), where the mean and \pm one standard deviation are plotted. The variation as a function of spatial frequency is also shown as a relative error in Fig. 2(b). For the most important spatial frequency range, the relative error was found to be in the range of 1-3%.

Intra-image variation is of interest when we repeat the above experiment with a single test image file, but vary the ROI used for the evaluation. This situation applies when the ROI is manually selected by the user (via computer mouse) and variations in position occur. In this case, a single test image was used, and a fixed analysis region size was chosen. The edge SFR analysis was conducted repeatedly, but with random x- and y- translations introduced in the ROI selection. Note that in this situation, we are evaluating the measurement sensitivity to, e.g., the influence of the image sampling along the edge feature and image noise. Each different ROI in the set ($N = 100$) partially overlaps so that the variation across the target is minimized but differs sufficiently (approximately ± 20 pixels in each direction) that different edge sampling and noise statistics are used.

The error introduced into the SFR measurement at high frequencies can be compared by considering Figs. 2(b) and 3(c). These appear to indicate the extent to which the ROI selection influences measurement variation at higher frequencies. When independent image subarrays are chosen (Fig. 2), the image noise fluctuations are independent. When the analysis is performed for a set of partially overlapping ROIs from the same image, the variation is reduced.

* The testing facility would include: test target, lighting, testing protocol, such as camera settings and alignment, and analysis software.

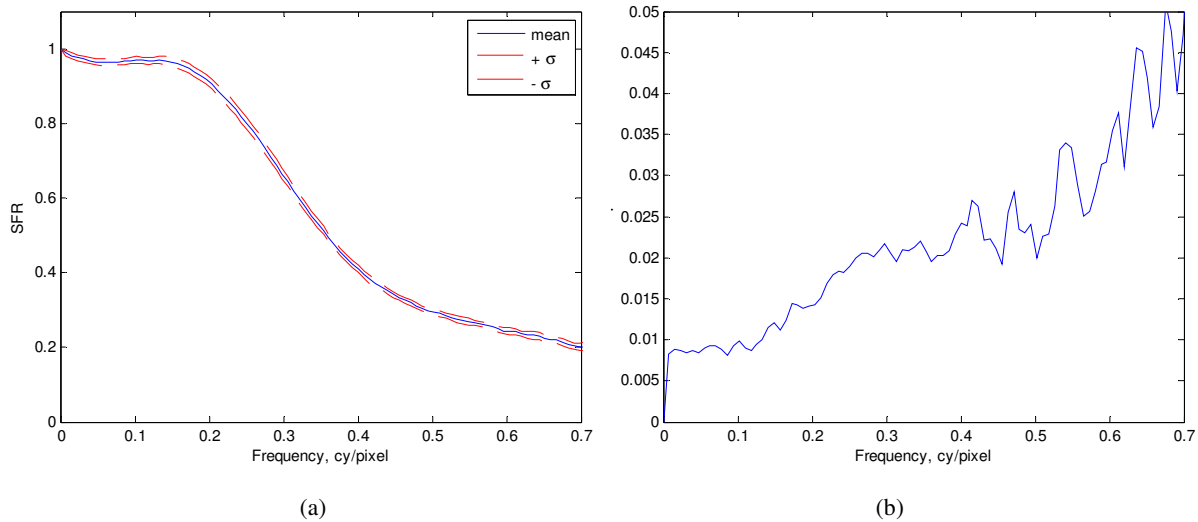


Figure 2.: Results of repeated slanted-edge SFR evaluation ($N = 100$) for different subregions (ROIs) (a) mean response and ± 1 standard deviation, and (b) relative error.

4. DEAD-LEAVES MTF MEASUREMENT

A recently proposed method to evaluate digital cameras and cell-phone cameras uses a computed image field comprising overlapping features (circles or rectangles).⁵ An example of this computed “dead leaves” target is shown in Fig. 4(a). The method is aimed at providing an effective MTF for image fluctuations (signals) that are influenced by adaptive or signal-dependent image processing. Many image processing operations, such as those for sharpening and noise reduction, apply spatial operations that vary with local image features. For these operations, a unique MTF description does not apply, but suitable classes of important signals can be used to measure an effective MTF. The “dead leaves” MTF method relies on several underlying statistical characteristics of this image field.

As with most signal-transfer descriptions such the MTF, the ratio of output (processed image) measurement with the corresponding, modeled or measured, characteristics for the input target image is called for. In this case the signal spectrum is computed as the amplitude of the discrete Fourier transform (DFT) of the two-dimensional luminance image array. The basic steps of the proposed method are:

1. Transform the captured image array of the target field to one encoded as proportional to luminance.
2. Compute the power-spectral density as the square of the amplitude of the two-dimensional DFT of the array.
3. Divide this array, frequency-by-frequency, by the modeled spectrum for the specific target to yield a two-dimensional array as the square of the effective MTF.
4. Compute the square-root, frequency-by-frequency, of this vector.
5. Compute the one-dimensional MTF vector by a radial-average of this array,

We should note that this method is currently under development and refinement by members of the Camera Phone Image Quality (CPIQ) Initiative, and it is likely that a variant of this method will be adopted. The initial description⁵ of the use of the dead-leaves target and method focused on the use of the computed target and its spatial attributes, such as scale and shift invariance. Less information was given for the particulars of the sampling and estimation to be used. McElvain *et al.*¹⁰ refer to a second method for estimating the effective “texture MTF” by reversing steps 4 and 5.

During the development and evaluation of this method for digital camera evaluations, several areas have been described. One of these is the deviation of the observed signal spectrum from that predicted by the model, as described in Ref. 10. They proposed two extensions to the method aimed at reducing variation (in this case, bias error) from the characteristics

assumed by the method. The first is to modify the assumed input signal spectrum based on the image sampling of the camera under test. While this may be a practical solution, spectra based on ideal computed arrays can be susceptible to alias errors even when combined with pre-filtering.

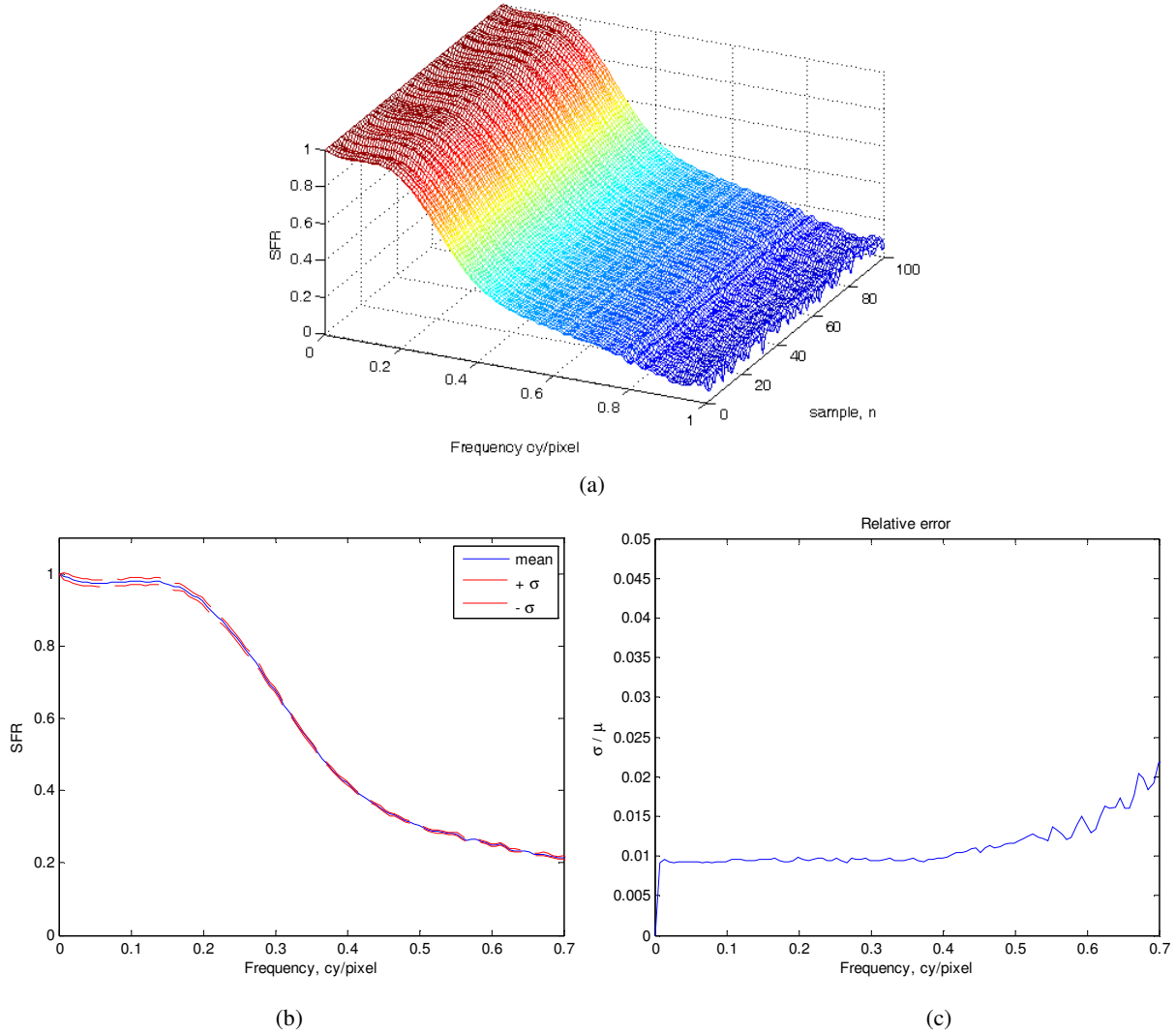


Figure 3. Results of repeated slanted-edge SFR evaluation ($N = 100$), varying the image analysis region for a single digital camera file. (a) SFR results, $N = 100$, (b) mean response and \pm one standard deviation, and (c) relative error.

A characteristic of the dead-leaves target that provides scale (image-sampling) independent evaluation is the rapidly falling signal spectrum as spatial frequency increases. This, however, can be a problem for the evaluation of many systems because image noise and image compression artifacts in processed images increased the measured spectra at high frequencies, as shown in Fig. 5(c) and (d). In some cases, these unwanted components can dominate the observed spectra and therefore bias the resultant texture-MTF. To mitigate this effect, a second suggested modification is to apply an empirical correction to the ideal signal spectrum based on the level of the (remaining) image noise, which is estimated from a uniform image area.

4.1 Observed variation

The first experiment was an investigation of the rotational variation of the spectrum of the sampled target. Figure 4(b) and (c) show the amplitude spectrum and cross-sections, respectively. The radial average spectrum is plotted in Fig. 4(c). We conclude that that the deviation from rotational symmetry is not due to an observable bias for this target, but it is

what might be expected for spectral estimation results from isotropic stochastic image arrays. This variation could be reduced by employing, e.g., a short-block periodogram spectral estimation method, common in image noise evaluation.

Another source of variation for this method is the computing of the power spectral density or its square root, the amplitude spectrum. As for the edge SFR results of Fig. 3, variations due to the placement of the ROI in the image field were investigated. An analysis region was chosen to under-fill the image of the dead-leaves target by 25 pixels on all sides. The signal spectrum was computed repeatedly, translating the ROI in random x- and y- directions for each measurement. Figure 5 shows the results in the form of the average (scaled) spectrum \pm one standard deviation and the corresponding relative error.

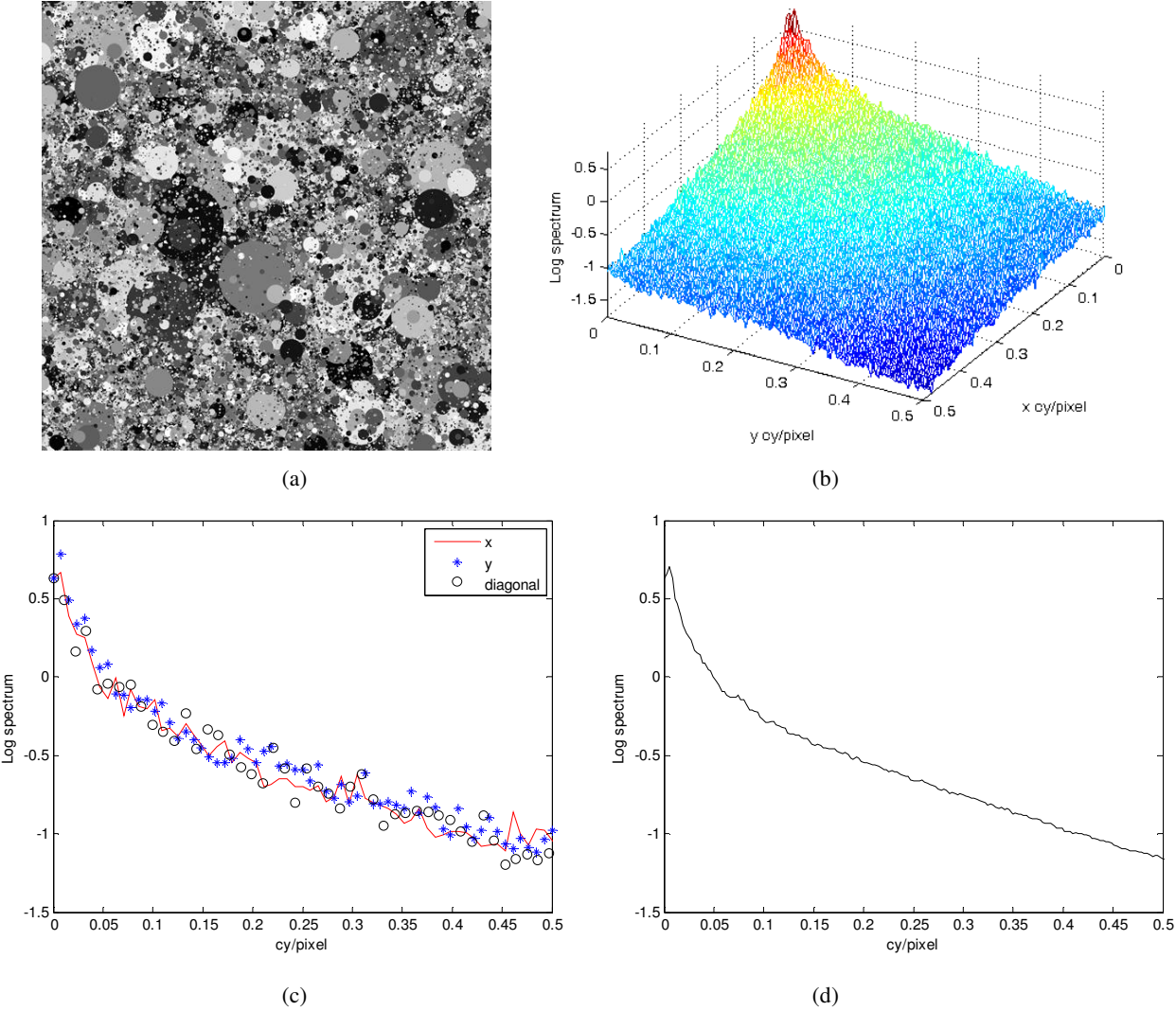


Figure 4. (a) Dead-leaves test target, with (b) amplitude spectrum, (c) corresponding cross-sections, and (d) radial average spectrum. Note the use of a log scale.

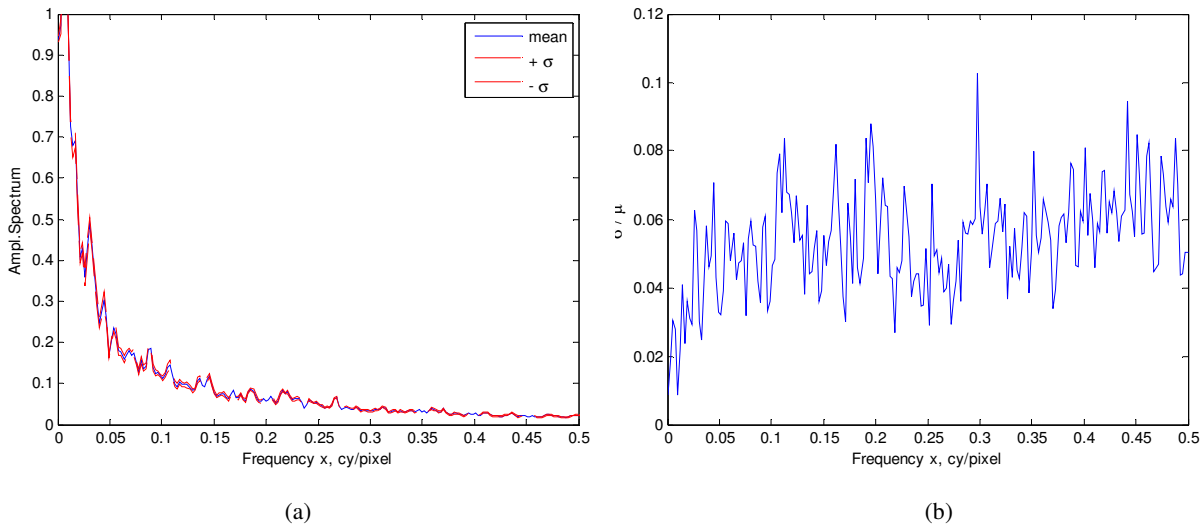


Figure 5. Results of repeated dead-leaves target signal (magnitude) spectrum evaluation ($N = 100$), varying the image analysis region for a single test image in normalized units. (a) mean response and \pm one standard deviation, and (b) relative error.

5. SUMMARY

The practical measurement of many image quality parameters can be thought of as an estimation problem, based on the gathered data, often from digital images. We have investigated two signal transfer methods commonly used for digital camera and scanner evaluation: the ISO 12233 slanted-edge spatial frequency response, and the dead-leaves method for texture-MTF evaluation. Rather than compare the methods, which measure different characteristics, the intent is to indicate how the development of quality measurements can be evaluated and refined, based on the observed variation during normal use. The variation due to the selection of analysis regions was computed by repeated analysis. The slanted-edge method indicated a relative error in the range of 1-3%, depending on the nature of the region selection. For the dead-leaves method, the amplitude spectrum (square root of the noise-power spectrum) showed a relative error of around 4-6%, however, this can be reduced by applying spectral estimation methods commonly used in image noise analysis.

REFERENCES

- [1] Burns, P. D and Berns, R. S., "Error propagation analysis for color measurement and imaging," *Color Res. Appl.* 22, 280-289 (1997).
- [2] Burns, P. D., "Variation and calibration error in electronic imaging," *Proc. PICS Conf.* 152-155, (2002).
- [3] ISO/TC WG18, *Photography – Electronic still picture cameras – Resolution measurements*, ISO, 1998
- [4] Burns, P. D., "Slanted-edge MTF for digital camera and scanner analysis," *Proc. PICS Conf.* 135-138 (2000).
- [5] Cao, F., Guichard, F., and Hornung, H., "Measuring texture sharpen of a digital camera," *Proc. SPIE 7250*, 72500H (2009).
- [6] Dainty, C. J. and Shaw, R., [Image Science], Academic, New York, 244-246 (1974).
- [7] Wueller, D., "Proposal for a standard procedure to test mobile phone cameras," *Proc. SPIE 6069*, 229-241(2006).
- [8] Scott, F., Scott, R. M., and Shack, R. V., "The use of edge gradients in determining modulation-transfer function," *Photogr. Sc. and Eng.*, 7, 64-68 (1963).
- [9] Jones, R. A., "An automated technique for deriving MTFs from edge traces," *Photogr. Sc. and Eng.* 11, 102-106 (1967).
- [10] McElvain, J., Campbell, S. P., Miller, J., and Jin, E. W., "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," *Proc. SPIE 7573*, 75370D (2010).