Statistical Interpretation of ISO TC42 Dynamic Range: Risky Business

Don Williams, Peter D. Burns, Michael Dupin
Eastman Kodak Company, Rochester, NY, USA, 14650-1925

ABSTRACT

Recently, two ISO electronic imaging standards aimed at digital capture device dynamic range metrology have been issued. Both ISO 15739 (digital still camera noise) and ISO 21550 (film scanner dynamic range) adopt a signal-to-noise ratio (SNR) criterion for specifying dynamic range. To resiliently compare systems with differing mean-signal transfer, or Electro-Optical Conversion Functions (OECF), an incremental SNR (SNRi) is used. The exposure levels that correspond to threshold-SNR values are used as endpoints to determine measured dynamic range. While these thresholds were developed through committee consensus with generic device applications in mind, the methodology of these standards is flexible enough to accommodate different application requirements. This can be done by setting the SNR thresholds according to particular signal-detection requirements. We will show how dynamic range metrology, as defined in the above standards, can be interpreted in terms of statistical hypothesis testing and confidence interval methods for mean signal values. We provide an interpretation of dynamic range that can be related to particular applications based on contributing influences of variance, confidence intervals, and sample size variables. In particular, we introduce the role of the spatial-correlation statistics for both signal and noise sources, not covered in previous discussions of these ISO standards. This can be interpreted in terms of a signal's spatial frequency spectrum and noise power spectrum (NPS) respectively. It is this frequency aspect to dynamic range evaluation that may well influence future standards. We maintain that this is important when comparing systems with different sampling settings, since the above noise statistics are currently computed on a per-pixel basis.

Keywords: dynamic range, metrology, image quality, standards

1. INTRODUCTION

The term *dynamic range* is as old as signal analysis itself. Ask anyone actively involved in the optical or imaging sciences to define it for image capture and you are sure to get an opinion. It may be qualitatively articulate and include words like *minimum, maximum, tonal range, accurately detect*, or *reliably detect*. Others will be quantitatively clinical and probably offer the ubiquitous "20 times log10 of the maximum signal to dark noise ratio" (…or is it 10 times?). The digital dilettantes are even identifiable with definitions that cite bit depth alone. It is truly a cluttered landscape and evokes an observation that this is science in action, at its messy best.

The two ISO electronic imaging standards of interest here are aimed at capture device dynamic range metrology. They have emerged from TC42/WG18 and begin to make sense of the confusion. Both ISO 15739 (digital still camera noise) and ISO 21550 (film scanner dynamic range) adopt signal-to-noise ratio (SNR) threshold criteria for determining dynamic range endpoints; specifically, where the incremental signal-to-noise ratio, SNR_i, is reduced to a value of 1.0. But how is one to interpret the dynamic range utility of a SNR_i of 1.0? At the extreme, does it mean that one can measure single pixel signal levels with 100% confidence? Or are there some implied constraints set by the standard but not specifically articulated by it? This paper is intended to address exactly those questions and suggest future thinking on the subject.

There is little doubt that signal and noise are appropriate specification criteria, but beyond this, these standards provide an opportunity for understanding alternative meanings for "signal" and "noise" as related to their spatial and possibly temporal components, as well as guidelines on their metrology. The concepts in the standards are easily extendable to signal types other than those in the optical sciences and, though imperfect, they act as a foundation for the questions posed in this paper. We attempt to catalyze thought by challenging the reader to think of a statistical interpretation of dynamic range, vis-à-vis SNR; a flexible yet unified vision that allows adaptations for not only varying levels of risk aversion but also specific spatial frequency bands and noise correlation statistics.

Proc. SPIE-IS&T Electronic Imaging Symposium, SPIE vol. **6059**, 207-214, 2006.

2. DEFINITIONS AND APPROACH

To start, we offer the following qualitative definition of dynamic range.

> **Dynamic Range –** the extent of energy over which a digital capture device can reliably detect signals: reported as either a normalized ratio (xxx:1) or in equivalent log optical density units.

The important words in this definition are *reliably detect*. The reliability (think probability or confidence) in detecting any given signal is a function of not only the signal amplitude, but also the ambiguity that noise introduces in determining that signal. Adopting the ISO 21550 criterion, the mean level where the incremental signal-to-noise ratio, SNR_i, is equal to 1.0 defines the dynamic range endpoints. Because it adopts an incremental signal definition, its utility lies in quantifying how well a given object intensity, I_o can be distinguished from another intensity of an arbitrarily small difference, ΔI. In the context of a photographic image metrology, it can, for example, answer questions like, 'How reliably can this capture device distinguish between an optical density of 2.90 and 3.00?' The incremental signal, ΔD, in this case would be 3.00 – 2.90 = 0.10. When this incremental signal is divided by the standard deviation of the image noise (in the same physical units) to yield SNR_i , an image processing resilient metric is achieved.[1]

The mathematics of what follows is not difficult and is found in many statistical texts. We draw heavily from one introductory text.[2] What can be frustratingly difficult, though, is formulating the correct imaging question to ask of the statistics, and selecting the appropriate statistical tests to execute. The primary question addressed in this paper is, " How is one to interpret dynamic range limits as defined in the TC42 standards and what, if any, assumptions make that interpretation vague?"

We have formulated three separate but subtly different questions on image signal detection. To avoid abstraction, we use examples with numerical values consistent with the practices for which the standards were intended.

Consider the following population statistics in units of optical density,

incremental signal: $\Delta D = 0.10$,
noise: $\sigma_D = 0.10$,
high density dynamic range endpoint: $\mu_{DR} = 3.0$,
leading to: $SNR_i = 1.0$.

These values are intended for clarity, not to limit generality. The three questions we address are:

1) What observed incremental signal, $\Delta \overline{D}$, (i.e., $\Delta \overline{D} = \overline{D} - \mu_{DR}$) is required to exclude $\overline{D}$ from the dynamic range endpoint's population, μ_D ?

2) What sample size is required to estimate a population mean to within a prescribed uncertainty, δ_D , at a specified confidence level ?

3) Wishing to detect a specified difference in two population means ($\Delta\mu = \mu_1 - \mu_2$) what is the required sample size for given alpha and beta risks?

For completeness the following assumptions, except as noted, are made through the remainder of the paper. They are:

- Gaussian distributed, signal independent noise sources of known value.
- Equally divided two-tail risk assignment.
- Examples can be applied equally to low-density dynamic range endpoint.

3. EXAMPLES AND DISCUSSION

The utility of SNR_i is in analyzing the detection of small difference in the context of noise. Though not specifically addressed in the standards, statistical testing of a single mean value estimate, $\overline{D}$, is within the spirit of the standards and, is implicitly accommodated. We address both types of estimates and apply classical hypothesis testing and confidence interval methods to demonstrate the statistical signal detection implications of the standards. This paper is not intended as a review of statistical test methods. Rather it is meant to show how these methods apply to a variety of image signal detection tasks related to dynamic range measurement.

3.1 Mean Level Testing

Once a dynamic range endpoint for a device has been established, a question that arises is, ‘what is the neighboring density level one can confidently exclude as being equal to the dynamic range endpoint?’ More generally,

1) What observed incremental signal, $\Delta\overline{D}$, (i.e., $\Delta\overline{D} = \overline{D} - \mu_{DR}$) is required to exclude $\overline{D}$ from the dynamic range endpoint’s population, μ_{DR}?

The good news is that the answer is found by applying simple hypothesis testing of the sample mean relative to the dynamic range endpoint. The bad news is that the answer is qualified by alpha risk, sample size, and noise variables. Even if one constrains the alpha risk and sample size levels, we find that SNR_i alone is insufficient to arrive at an answer. One must know the noise level leading to a particular SNR_i. The first step to an answer is found in a statement of the null and alternate hypotheses, followed by a calculation of the test statistic, *t*.

Null hypothesis: $H_o : \mu_D = \mu_{DR} = 3.0$

Alternate hypothesis: $H_1 : \mu_D \neq \mu_{DR}$

$$t = \frac{\Delta\overline{D}}{\sigma_D / \sqrt{n}} \quad \text{, where } n = \text{sample size for } \overline{D} \qquad (1)$$

If the absolute value of the calculated *t* statistic in Eq. (1) is greater than the critical *t* value associated with a given alpha risk (see Fig. 1), one can reject the null hypothesis and conclude that $\mu_D \neq \mu_{DR}$. The statistical inter[pretation is that the test samples are taken as drawn from a population with a mean value that is different from that of the dynamic range engpoint μ_{DR}. Otherwise, there is not sufficient evidence to conclude that $\overline{D}$ is from a different endpoint population. Figure 2 illustrates the observed density differences needed to meet these requirements and shows the dependency on alpha risk and sample size. For instance, for the unity SNR_i conditions stated on LHS of Fig. 2, a rather large 0.20 (or greater) density difference estimated from a single pixel sample would be required to reject the null hypothesis with a 5% alpha risk. Only a 0.05 density difference would be required if sixteen pixels were used. If, however, the noise level is doubled so too are all required density differences, even though the SNR_i is still unity. This is shown on the RHS of Fig. 2 and indicates that SNR_i for this test is not unique and alone does not dictate statistical detection results.

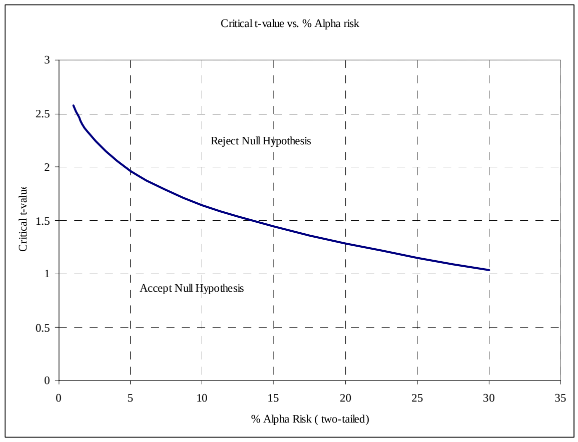

Figure 1: Critical Value vs. Two-tailed alpha risk

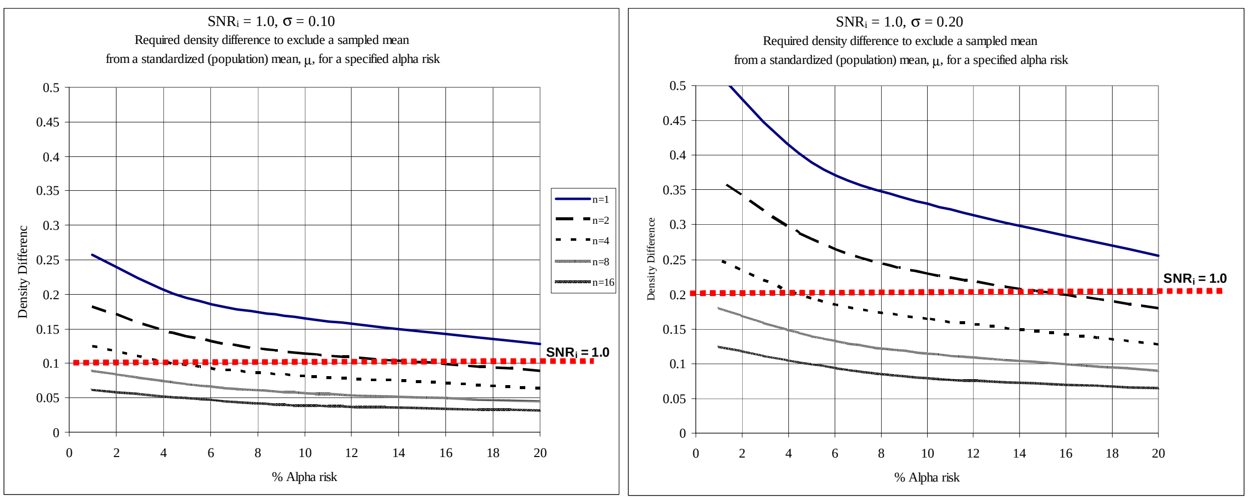

Figure 2: Hypothesis Testing results for the mean $\overline{D}$

The hypothesis test example of question (1) provides only a minimum of knowledge about an estimate $\overline{D}$. It merely answered the question of what measured density values are required to conclude that an observed mean was not operating at the dynamic range limit population, μ_{DR} = 3.0. A higher standard would require knowing the goodness of the estimate $\overline{D}$. For the assumed risk, we are secure in the knowledge that $\overline{D}$ can be excluded from dynamic range limit, $\mu_{DR} = 3.0$. Further details include knowing the population mean within confidence bounds. This implies an interval estimate and, unlike the point estimate of question (1), it better conveys the state of knowledge about the mean of a population. This brings us to question #2 which addresses confidence intervals of signals.

> 2) What sample size is required to estimate a population mean to within a prescribed uncertainty, δ_D, at a specified confidence level ?

Eq. 2 shows how a confidence interval is constructed.

$$\overline{D} - \delta_D < \mu_D < \overline{D} + \delta_D \qquad (2)$$

$$\text{where} \quad \delta_D = \frac{1.96\,\sigma_D}{\sqrt{n}}$$

The equation for calculating sample size, *n*, for a confidence interval is

$$n = \left(z_{\alpha/2} \frac{\sigma_D}{\delta_D} \right)^2 \qquad (3)$$

This calculation can be applied to a very practical problem for densitometry measurement using scanners. A common rule of thumb for macro densitometry has been an uncertainty δ_D = 0.02. This rule is commonly applied equally to all density levels. Assuming σ_D = 0.10 for a candidate scanner and adopting 95% confidence interval (i.e., $z_{\alpha/2}$ = 1.96) one finds that n = 96 to reliably detect a mean signal difference of 0.02. In other words, for a single scanner realization, roughly a 10 x10 pixel area is required. For a scanner at 300 dpi this roughly translates to an area equivalent to a 1mm circular aperture. The only way to maintain this level of uncertainty with smaller sample sizes (i.e. smaller equivalent apertures and higher spatial frequencies) is to increase ones risk or by reducing the effective noise.

3.2 Incremental Signal Tests

The previous examples concentrated on testing of a singular mean, μ_D against a pre-established standard, $\mu_{DR.}$ Frequently, two sets of population means, μ_1 and μ_2 are presented that need to be tested against each other. With respect to the dynamic range standards considered here, this is probably a more appropriate and practical test; the testing of two sample averages whose difference is incrementally small and on the order of the noise level, i.e, incremental signal to noise ratio, SNR_i .

One could apply a simple hypothesis test, as in question (1), to the two means but, as we have seen, this provides only marginal insight. Instead, we will use an approach that involves addressing both alpha and beta risks to two observed means. The consideration of both of these risks is of practical importance when one can be specific about the difference that one wishes to detect. Recall that alpha risk (α) governs the probability of concluding that a process change has occurred when none has (false detection). The beta risk (β) governs the probability of failing to decide a process change when one has indeed occurred (missed detection). For a given sample size, these two are played off one another. But when the sample size is a variable these two risks can be satisfied simultaneously. We address the latter with the next question since it is of more practical utility in every day use.

3) Wishing to detect a specified difference in two population means ($\Delta\mu = \mu_1 - \mu_2$), what is the required sample size for given alpha and beta risks?

This has all of the components of an incremental signal detection problem where one has freedom in selecting the number of pixels used for detection. We use typical values of α=β = 0.05. The sample size calculation for detecting a density difference, $\Delta\mu$, of two observed means is

$$n_1 = n_2 = n = 2\left[\left(z_{\alpha/2} + z_{\beta} \right) \frac{\sigma}{\Delta\mu} \right]^2 \qquad (4)$$

Figure 3 illustrates the tradeoff between sample size and incremental signal, $\Delta\mu$, for α = β = 0.05 and σ_D = 0.10. The confidence limits on the difference in population means are also plotted in Figure 4. For an incremental signal difference of $\Delta\mu = 0.02$, a total of $n_1 + n_2 = 2$ x 180 = 360 samples are required. This is nearly double the number of samples needed for δ_D = 0.02 of #2 and demonstrates the increase in sample size requirements for α and β combinatorial risks imposed by incremental signals; the kind precisely defined in the cited standards. Implicitly, this increase in sample size lowers the spatial frequency limits which one can draw signal detection conclusions.

The ISO/TC42 standard appropriately chooses large sample sizes for characterizing population statistics leading to dynamic range. However, it is left to the user of the device to translate these values for more practical single realization captures for signals and noise of varying spatial frequency content. Briefly, we discuss the philosophy behind this spatial frequency approach.

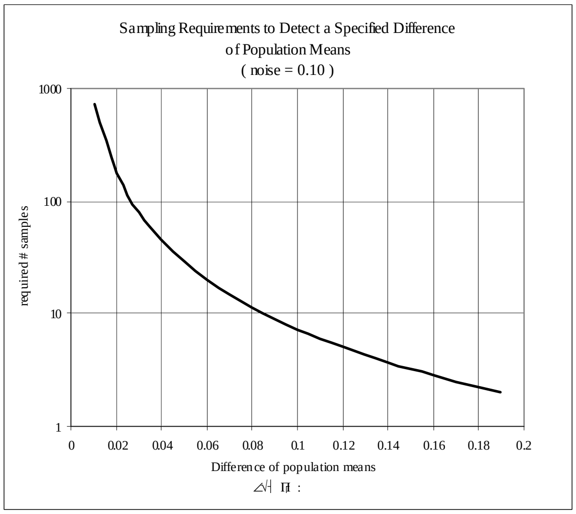

Figure 3: *Sampling requirements to detect a different population means*

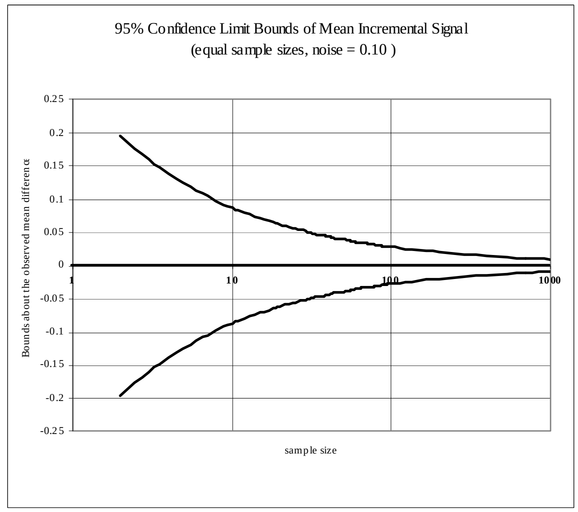

Figure 4: 95% confidence limit bounds for mean incremental signal ($n_1 = n_2$; rms noise = 0.10)

3.3 SNR and Spatial Correlation

The dynamic range and digital still camera standards generally adopt a pixel-centric view of both signal and noise characteristics. The signal to be detected is a small difference in an original input scene, and the observed rms noise is calculated for distributions of individual pixel values. The estimated SNR level is in terms of pixel-to-pixel fluctuations. As we have seen, however, a general statistical interpretation of these performance measures includes the detection of varying signal levels from the observed mean values. When we perform a statistical test based on observed values we normally expect that a sample size greater than $n = 1$ will be required, as shown in section 3.1, where a sample size of 96 was needed to detect a mean signal difference of 0.1. For the practical application of imaging performance, this means that the observed noise statistics are usually based on the observed pixel values in a local, nominally uniform, area. This is usually easy to accommodate with test targets that contain several uniform steps for which mean and standard deviation data can be computed.

Characterizing image noise by a single standard deviation, however, takes no account of spatial correlation in the noise fluctuations. Such correlation can arise from several sources. The signal readout and in-camera signal processing such as color-filter array interpolation can induce a correlation between the fluctuations in neighboring pixels, as can signal-mixing operation, e.g., white-point correction. [3, 4] A well-established method for measuring the extent of both the variation and spatial correlation of image noise involves estimation of the noise-power spectrum. [5] The noise-power spectrum (NPS) describes the noise variance fluctuation as a function of spatial frequency. This is analogous to the description of image signal transfer in terms of a modulation transfer function (MTF). Alternatively, the spatial auto-correlation function can be used if a spatial domain measure is preferred.

As an illustration of the visual impact of noise correlation, consider the examples in Fig. 5. At the top of the figure we see a noise-free input image. In terms of detection one could, for example, define the cross feature or the thin light lines as representative of an important signal. One the lower left uncorrelated noise has been added to the input scene with the resulting noisy appearance. Noise of the same standard deviation was added to the noise-free array and is shown in the lower right. While the appearance is again degraded, it is quite different from the uncorrelated case. In addition one could argue that specific image features (signals) present in the uncorrelated case, are not detectable for this case. It is likely that future developments in standard imaging performance evaluation will include the consequences of image processing and the resulting image noise correlation.

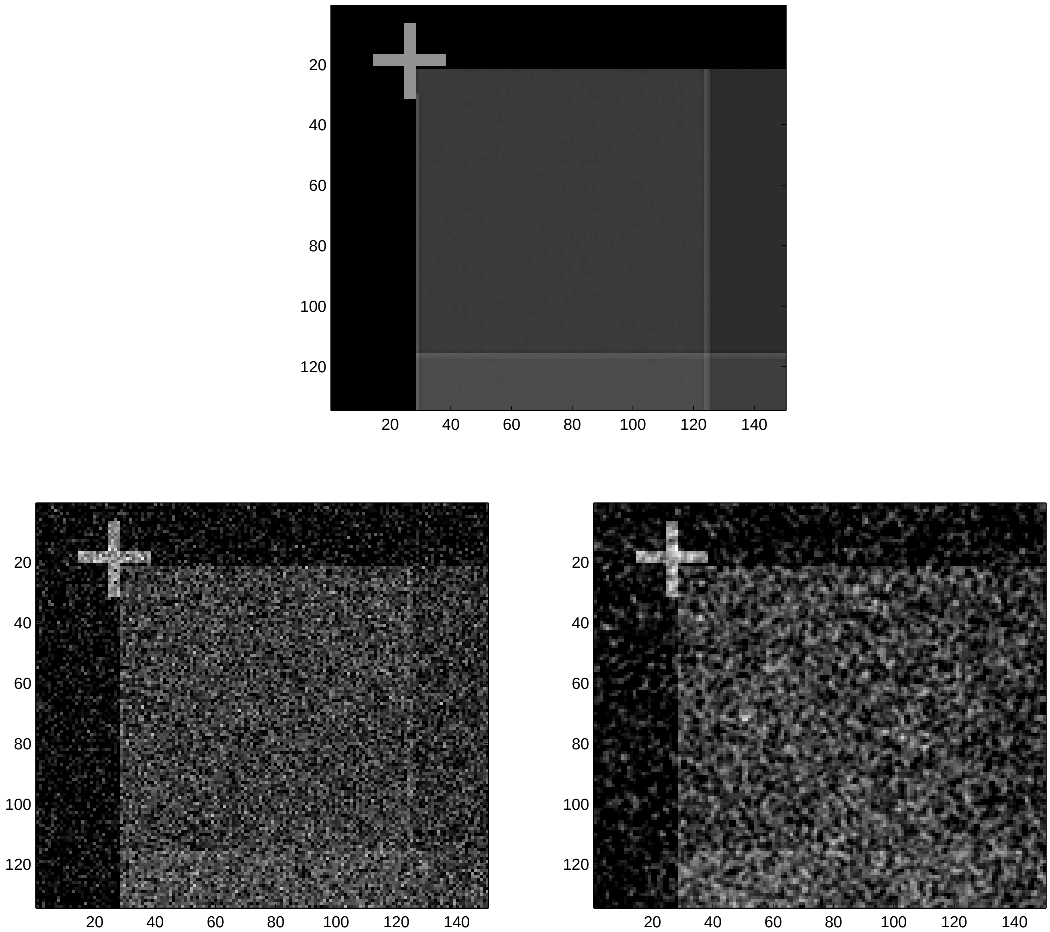

Figure 5: Example of spatially uncorrelated (lower left) and correlated (lower right) noise added to a noise-free test image

CONCLUSIONS

No doubt, the signal and noise component approach of ISO 15739 and ISO 21550 is an appropriate model for determining a device's dynamic range when a signal detection criterion is adopted. These standards use large area (i.e. low frequency) signal and pixel-to-pixel (i.e. high frequency) noise data and provide an excellent foundation for enhancements. For single realization data, we have demonstrated how the existing standards may be adapted for sample size and risk level variables if component signal and noise metrics are known. We encourage future revisions of these standards to more comprehensively accommodate for risk level and spatial frequency content of both signal and noise by way of sample size, Modulation Transfer Function (MTF), and Noise Power Spectrum (NPS) metrics. ISO TC42 has a rich history of providing well founded science based metrics to digital capture device imaging performance. The dynamic range metrics issued thus far are a good platform for a more unified approach that provides tuning for acceptable levels of risk and spatial frequency content.

REFERENCES

1. D. Williams, The Array Scanner as Microdensitometer Surrogate: A Deal with the Devil or... a Great Deal?, *Proc. SPIE , Vol. 5668,* , 241-246, 2005
2. A. D. Rickmers and H. N. Todd, Statistics; An Introduction., McGraw-Hill, 1967
3. P. D. Burns and R. S. Berns, Modeling Colorimetric Error in Electronic Image Acquisition, *Proc. Optics and Imaging in the Information Age, IS&T, 147-149*, 1997.
4. P. D. Burns, Variation and Calibration Error in Electronic Imaging, *Proc. PICS Conf*., 152-155, IS&T, 2002.
5. J. C. Dainty and R. Shaw, Image Science, Ch. 8, Academic Press, London, 1974