Journal of Imaging Science and Technology® 63(6): 060405-1–060405-9, 2019.

Comparison of Detectability Index and Contrast Detection Probability

Robin B. Jenkin▲
NVIDIA Corporation, Santa Clara, California, USA
E-mail: rjenkin@nvidia.com

Abstract. *Autonomous vehicles rely on the detection and recognition of objects within images to successfully navigate. Design of camera systems is non-trivial and involves trading system specifications across many parameters to optimize performance, such as f-number, focal length, CFA choice, pixel, and sensor size. As such, tools are needed to evaluate and predict the performance of such cameras for object detection. Contrast Detection Probability (CDP) is a relatively new objective image quality metric proposed to rank the performance of camera systems intended for use in autonomous vehicles. Detectability index is derived from signal detection theory as applied to imaging systems and is used to estimate the ability of a system to statistically distinguish objects, most notably in the medical imaging and defense fields. A brief overview of CDP and detectability index is given after which an imaging model is developed to compare and explore the behavior of each with respect to camera parameters. Behavior is compared to matched filter detection performance. It is shown that, while CDP can yield a first order ranking of camera systems under certain constraints, it fails to track detector performance for negative contrast targets and is relatively insensitive.* © *2019 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2019.63.6.060405]

1. INTRODUCTION

The pursuit of the creation of autonomous vehicles has come to the fore in the past decade. While the debate concerning the optimum combination of sensor modalities among optical imaging, lidar, radar, etc. to achieve this continues, cameras are likely to remain the backbone of current solutions. The push toward self-driving vehicles probably represents the first mass deployment of safety critical systems that rely on optical imaging in the public arena. As such, metrics and tools need to be developed to aid design and predict the performance of such systems with respect to headline camera system parameters. The IEEE P2020 Image Quality for Autonomous Vehicles standard group has undertaken to adapt existing and develop new metrics for such purposes [1].

Of critical importance is an ability to estimate the capability of a camera system to detect objects at distance across a wide array of illumination conditions. Apart from obvious safety considerations, how else will be it possible to ascertain the performance that a deep neural network (DNN) *should* be capable of delivering if the basic objective performance of the input imaging device cannot be determined? At present most of the literature detailing DNN development and training provide little detail beyond image size and the numbers of images used for training and inference.

The challenge of creating such performance metrics is not trivial. Many objective measures exist, such as the modulation transfer function (MTF), noise power spectrum (NPS), signal-to-noise ratio (SNR), or quantum efficiency (QE). However, all these measure a single dimension or aspect of image quality, and analysis often falters when a prediction of *overall* system performance is demanded. To combine multiple objective measures requires some knowledge of their relative importance to the task at hand and the manner in which they trade. Metrics also need to track some basic measure of machine vision performance if intended to predict the suitability of the system for such a purpose. Complicating this further is the proliferation of non-linear image processing, rendering image signal processing (ISP) pipelines adaptive to scene input, predominantly via demosaic, sharpening, and noise reduction that is edge-aware and further the use of local tone mapping.

Aside from the difficulties above, automotive imaging also represents a wide range of challenging illumination conditions which need to be accounted for. Low light scenarios are typically first thought in these cases and while they undoubtedly exist, high-signal low-contrast detection problems are also prevalent, such as light-colored vehicles in fog, snow, and glare. High dynamic range (HDR) scenes are commonplace, especially at night or in tunnels. HDR sensor technologies combined with flicker avoidance and a need to keep motion blur to a reasonable level often places further constraints on the practical range of exposure times and system configurations that may be used.

Signal detection theory (SDT) is well known within the medical imaging and defense fields [2–4] and provides analysis tools that statistically determine the probability that an object will be detected against a background. It is built on a spatial frequency weighted analysis of signal and noise propagation through the system in question and, for an idealized case, yields the separation of the object and background in terms of mutual standard deviations or the signal-to-noise ratio of an optimized detection process. Known as detectability index [2], Kane has recently written

▲ IS&T Member.

Received Aug. 6, 2019; accepted for publication Oct. 13, 2019; published online Jan. 9, 2020. Associate Editor: Chunghui Kuo.
1062-3701/2019/63(6)/060405/9/$25.00

advocating the application of SDT to automotive imaging [5] and is used as a starting point for the analysis herein.

Contrast detection probability (CDP) is an empirical metric proposed by Geese et al. [6] as an IEEE P2020 metric to predict computer vision performance for autonomous vehicles. It is based on the premise that it is the ability of an imaging system to record contrast between a target and background and its interaction with noise that predominantly determines the ability to detect objects. By examining a distribution of contrasts, CDP estimates the spread of contrast due to noise in the system and calculates the probability that measured contrasts will fall within given bounds [6]. It is suggested by Geese et al. that the bounds may be set according to the application and desired level of visibility [6]. Further details on detectability index and CDP are given in the following sections.

This article compares the performance of matched filter detection for a simple object in simulated systems against predictions made by both CDP and an idealized observer derived from SDT. A basic imaging model is used to link metric performance and camera parameters to illustrate A–B comparison and ranking of systems that could be incorporated into a design process. Monochrome RAW images are simulated across a variety of illumination conditions and system configurations and analysis occurs prior to processing by an ISP to establish the baseline performance of the metrics. Analysis of the influence of an ISP is a topic for future work.

2. CONTRAST DETECTION PROBABILITY

Geese et al. define CDP as:

$$CDP_{K_{IN}} = P(K_{IN}(1-\varepsilon) \leq K_M \leq K_{IN}(1+\varepsilon)), \qquad (1)$$

where, K_{IN}, is input contrast, K_M, measured contrast, ϵ, contrast bounds, and P() probability. Stated simply CDP is the probability that measured contrast will fall between given bounds. Geese et al. suggest the use of Weber contrast, K_W, to perform the calculation, defined below:

$$K_W = \frac{E_{\mathrm{MAX}}}{E_{\mathrm{MIN}}} - 1, \qquad (2)$$

where E_{MAX} and E_{MAX} represent the maximum and minimum signals, respectively.

The calculation of CDP is straightforward and practical implementation is detailed by Ebbert [7]. Two uniform tone patches, representing the brightest and darkest components of a desired contrast level, are recorded in chosen illumination conditions by the imaging system under analysis. The patches should be large enough that a reasonable statistical sampling of the noise processes of the imaging system are captured. Typically, 10×10 pixels in the final image for each is sufficient. After transformation of the patch data into linear input units via the system tone curve, calculation proceeds by evaluating the contrast of every pixel combination between the two patches to estimate a distribution of contrasts, Figure 1. CDP for the contrast, illumination, and system

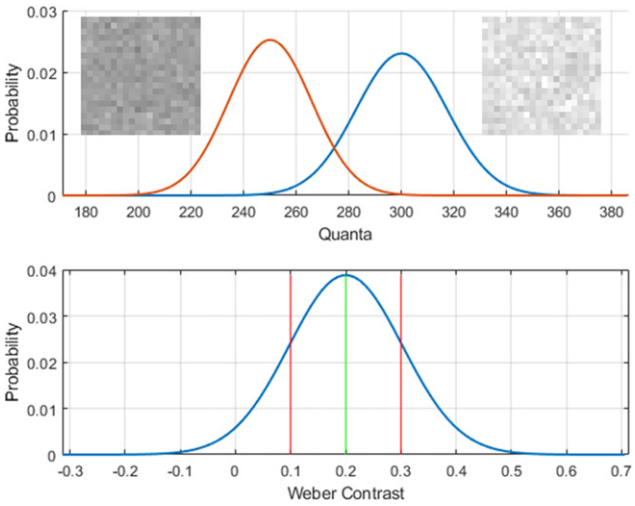


Figure 1. Distribution of Weber contrasts as calculated from recording bright and dark image patches and comparing each combination of pixels.

parameters used is then yielded by calculating the proportion of the distribution within the given limits. This procedure may then be repeated to calculate CDP values for different illumination and contrast combinations.

On limits setting, Geese et al. suggest that a bounds of 50% CDP is a good indicator for a threshold of visibility and detectability [6], though this has not been established with psychovisual calibration and may well change with display and illumination level. What is not clear is whether the bounds are intended to be calculated as 50% *of input contrast* or ±50% contrast as input contrast is varied. This is explored later.

2.1 *Detectability Index*

An accessible summary of the idealized observer result derived from statistical decision theory is given by Kane [5], and more extensive treatments of the topic may be found by Barrett and Myers [3], Beutel et al. [4] and in a report issued by the International Commission on Radiation Units and Measurements (IRCU) concerning the assessment of image quality for medical imaging [2]. Therefore, a brief overview is given here.

The task of object detection and classification may be described simply as an ability to discriminate whether or not an object is present and to which of a number of possible classes it belongs. For each of the possible outcomes, no object, object type one, object type two, etc., we can create a hypothesis, H_{K}, such that when given an image $g(x, y)$, the probability that hypothesis H_{K} is true given $g(x, y)$ may be written [5]:

$$\mathrm{p}(H_K|g(x,y)) = \frac{\mathrm{p}(g(x,y)|H_K)\mathrm{p}(H_K)}{\mathrm{p}(g(x,y))}, \qquad (3)$$

where $p(a \mid b)$ is the probability of an occurring given b and p(a) is the probability of a occurring as Kane writes [5]. Reducing this to the simplest case of determining whether an object is present or not we have two hypotheses and may

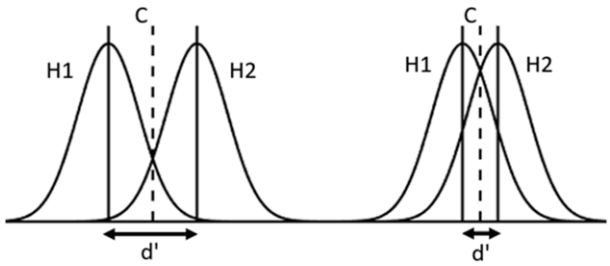


Figure 2. (a) and (b). Distribution of test values of given hypotheses H1 and H2 and threshold C, when (a) distributions are somewhat separated and (b) poorly separated.

write:

$$\mathrm{L} = \frac{\mathrm{p}(g(x,y)|H_2)}{\mathrm{p}(g(x,y)|H_1)}, \tag{4}$$

where L is the likelihood ratio [3, 5] and H_1 could be object present and H_2 object not present. Given an input f, the imaging process degrades the input by convolution with the point spread function (PSF), ∫ , of the system and the addition of noise, η, to yield g, an output image. For a specific detection test or operation, such as a matched filter for example, each of the hypotheses will create a distribution of test values given instances of g when the hypothesis is true, Figure 2(a).

Given an individual test value, either of the hypotheses is chosen by comparison to a threshold C. The threshold is chosen such that the likelihood ratio between the hypotheses is maximized. This is therefore a Bayesian decision maker [2, 5] and attempts to maximize the ratio of true positive to false positive results. As noise in the imaging system increases and the PSF degrades, the distributions associated with each hypothesis will increasingly overlap, Fig. 2(b). As this occurs it is ever more difficult to statistically separate the hypotheses and as such the ratio of true to false positives will go down. The detectability index, d', or $\mathrm{SNR_I}$, is the separation of the distributions expressed in mutual standard deviations and represents the signal-to-noise ratio of the detection test being performed to distinguish between hypotheses. It is noted by Barrett and Myers that if the distributions have unequal variances or are skewed, then d' is not sufficient to specify the system [3, p. 819].

An alternative way to write the $\mathrm{SNR_I}$ of the test process is [3, p. 819]:

$$\mathrm{SNR_I} = \frac{\langle t\rangle_2 - \langle t\rangle_1}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}, \tag{5}$$

where $\langle t\rangle_n$ is the expected value of the test and σ_n the standard deviation associated with each hypothesis.

It is shown in [2, p. 51] that the optimum test for the case where the signal and background are known exactly is a matched filter corresponding to the difference between them. Deriving expressions for the degradation of the signal due to the PSF of the imaging system and the addition of stationary uncorrelated Gaussian noise, then substituting them into Eq. (5) it is shown that the $\mathrm{SNR_I}$ of this idealized observer is [2, p. 51, 2]:

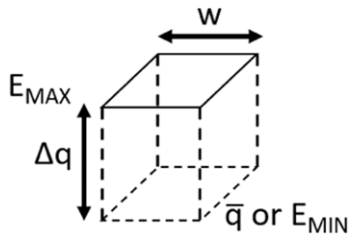


Figure 3. The object model as detailed by Kane [5].

$$\mathrm{SNR_I}^2 = K^2 \int \frac{|G(\nu)|^2\, MTF_{SYS}^2(\nu)}{NPS(\nu)}\, \mathrm{d}\nu, \tag{6}$$

where K is the large area signal transfer, $G(\nu)$, the Fourier transform of the difference between the signal and background, MTF, the system modulation transfer function, and NPS the noise power spectrum with respect to spatial frequency, ν [5]. It should be noted that other observers may be derived that account for correlated noise or when the target or background spectra are not known exactly [2]. Models of the human visual system may also be accounted for [2]. The idealized observer is used in this study as it is the most fundamental treatment of a detection process and should be considered a basic test of any metrics claiming to be proportional to machine vision performance.

2.2 *Object Model*

A simple object model as detailed by Kane [5] is used in this study, Figure 3. The object is a square cobblestone of width, w, at a level Δq above a background level $\bar{q}$ [5]. The difference between the object and the background is then [5]:

$$\Delta \mathrm{g(x,\,y)} = \Delta \mathrm{q}.\prod\left(\frac{x}{w}\right).\prod\left(\frac{y}{w}\right), \tag{7}$$

where $\prod$ represents a rectangle function. The Fourier transform of Eq. (7) is then as below to yield the difference spectra [5]:

$$\mathrm{G}(v_x, v_y) = w^2.\Delta \mathrm{q}.\mathrm{sinc}(\pi w v_x).\mathrm{sinc}(\pi w v_y). \tag{8}$$

Substituting Eq. (8) into 6, assuming the simple case of uncorrelated shot noise and a unity MTF over the range of spatial frequencies of interest yields, [5]:

$$\mathrm{SNR_I} = \mathrm{c}.\sqrt{\bar{q}A}, \tag{9}$$

where A is the area of the object and c is contrast given by $\Delta q/\bar{q}$. It may be simply shown c is equivalent to Weber contrast:

$$\Delta q = E_\mathrm{MAX} - E_\mathrm{MIN} \tag{10}$$

$$\bar{q} = E_\mathrm{MIN} \tag{11}$$

$$c = \frac{\Delta q}{\bar{q}} = \frac{E_\mathrm{MAX} - E_\mathrm{MIN}}{E_\mathrm{MIN}} = \frac{E_\mathrm{MAX}}{E_\mathrm{MIN}} - \frac{E_\mathrm{MIN}}{E_\mathrm{MIN}} = \frac{E_\mathrm{MAX}}{E_\mathrm{MIN}} - 1. \tag{12}$$

As mentioned in the introduction, automotive imaging demands operation in extremes of illumination conditions. Burgess [8] notes that for simplicity, the original derivation of SNR_I by Rose defines SNR as:

$$\mathrm{SNR} = \frac{\text{mean signal}}{\sigma_b}, \tag{13}$$

where *mean signal* is the mean level of the signal and σ_b the standard deviation of the background quanta. The definition assumes that noise is the same in the case when the target is present and when it is not. This is not valid when considering shot noise that behaves according to Poisson statistics [8]. Burgess also notes the derivation is based on a Gaussian rather than Poisson distribution [8]. This limits the use of the result in Eq. (9) to low-contrast signals with high enough photon counts to consider them Gaussian [8]. For the special case described, it is straightforward to modify the equations to account for the difference in the noise levels when the signal is present. The expected value of the matched filter for the background and signal will be given by the mean value of each multiplied by the area of the filter, or:

$$\langle t\rangle_1 = \langle E_{\mathrm{MIN}}\rangle.A = \langle \bar{q}\rangle.A \tag{14}$$

$$\langle t\rangle_2 = \langle E_{\mathrm{MAX}}\rangle.A = \langle \bar{q} + \Delta q\rangle.A, \tag{15}$$

where $\langle\rangle$ denotes the expected value. As $\langle\bar{q}\rangle$ and $\langle\bar{q}+\Delta q\rangle$ are calculated across area A their respective variance becomes:

$$\sigma_1^2 = \frac{\bar{q}}{A} \tag{16}$$

$$\sigma_2^2 = \frac{\bar{q}+\Delta q}{A}, \tag{17}$$

and the standard deviation of the difference between the matched filter when applied to the signal and background, σ_M, is given by:

$$\sigma_M = A\sqrt{\frac{1}{2}\frac{\bar{q}}{A} + \frac{1}{2}\frac{\bar{q}+\Delta q}{A}}. \tag{18}$$

Calculating $\frac{\langle t\rangle_2 - \langle t\rangle_1}{\sigma_M}$ we find that SNR_I becomes:

$$\mathrm{SNR_I} = \frac{\sqrt{2}\Delta q}{\sqrt{\frac{\Delta q + 2\bar{q}}{A}}}. \tag{19}$$

2.3 *Imaging Model and Simulation*

For cobblestone targets of various reflectance, R, and size, the photoelectron signal model detailed by Jenkin and Kane [9] is modified to estimate the number of quanta, PH_{PIXEL}, collected in a given integration time, T_{INT}. Ambient light level and color temperature are first specified in lux and kelvin, L_{AMB}, and CT_{AMB}, respectively. Scaling for the radiometric properties of the illumination falling outside of the sensitivity of the human visual system is included by scaling a blackbody curve, $W(\lambda)$ with units $\mathrm{Wm^{-2}\mu m^{-1}}$ [10].

$$W(\lambda) = \frac{C1}{\lambda^5\left[e^{\frac{C2}{\lambda CT_{AMB}}} - 1\right]}, \tag{20}$$

where *C1* is the first radiation constant, $3.74 \times 10^8 \mathrm{Wm^{-2}\,\mu m^{-4}}$, and *C2* the second radiation constant, $1.44 \times 10^4\ \mu\mathrm{mK^{-1}}$ [11]. The relative spectral luminous efficiency curve, $V(\lambda)$, of the CIE is scaled by the peak luminous efficacy of human vision (683 lumens per watt at 555 nm) [12] and multiplied by the blackbody curve above then integrated to yield the total lux, L_{SOURCE}, represented by the illumination curve generated:

$$L_{\mathrm{SOURCE}} = 683.\int_{\lambda_{\mathrm{MIN}}}^{\lambda_{\mathrm{MAX}}} W(\lambda).V(\lambda)d\lambda, \tag{21}$$

where λ_{MAX} and λ_{MIN} are the maximum and minimum wavelengths of interest.

The cobblestone object is considered a Lambertian reflector with reflection, *R*, and thus the light scattered, L_{REF}, by the object in units of lux $\mathrm{m^{-2}str^{-1}}$ is [11]:

$$L_{REF} = \frac{R.L_{AMB}}{\pi}. \tag{22}$$

A factor, L_{SCALE}, by which to multiply $W(\lambda)$ may then be calculated, Eq. (23), to yield the blackbody curve correctly scaled to the wattage required to yield the lux reflected from the object. Multiplying by the absolute quantum efficiency curve of the sensor, $QE(\lambda$, and absolute transmission of an infrared filter, $I(\lambda)$, yields the spectrum of light available to the sensor in $\mathrm{Wnm^{-1}m^{-2}str^{-1}}$ before lens and pixel geometry are considered, $P(\lambda)$.

$$L_{\mathrm{SCALE}} = \frac{L_{REF}}{L_{\mathrm{SOURCE}}} \tag{23}$$

and

$$P(\lambda) = L_{\mathrm{SCALE}}.W(\lambda).I(\lambda).Q(\lambda). \tag{24}$$

The solid angle, Ω, of the lens collecting the signal reflected from the projected pixel area is [11]:

$$\Omega = \frac{\pi D_{\mathrm{OPTICS}}^2}{4r^2}, \tag{25}$$

where, D_{OPTICS}, is the effective diameter of the lens and *r* is range as previously. Multiplying by the solid angle and transmission of the lens, T_{OPTICS}, yields the power per nm per square meter, P_{SENSOR}, captured by the sensor:

$$P_{\mathrm{SENSOR}}(\lambda) = L_{\mathrm{SCALE}}.W(\lambda).I(\lambda).Q(\lambda).\Omega.T_{\mathrm{OPTICS}}. \tag{26}$$

Multiplying by the area of the pixel, A_{PIXEL}, yields the power per nm per pixel.

$$P_{\mathrm{PIXEL}}(\lambda) = L_{\mathrm{SCALE}}.W(\lambda).I(\lambda).Q(\lambda).\Omega.T_{\mathrm{OPTICS}}.A_{\mathrm{PIXEL}} \tag{27}$$

The energy per photon, $E(\lambda)$, is calculated using:

$$E(\lambda) = \frac{hc}{\lambda}, \tag{28}$$

where h is Plank's constant, 6.62×10^{-34} m^2 kg s^{-1}, and c is the speed of light, 299792458 ms^{-1}. Dividing $P_{\mathrm{PIXEL}}(\lambda)$ by $E(\lambda)$, multiplying by the integration time, T_{INT}, and integrating yields the total number of photoelectrons captured by the pixel, PH_{PIXEL}:

$$\mathrm{PH_{PIXEL}} = \int_{\lambda_{\mathrm{MIN}}}^{\lambda_{\mathrm{MAX}}} \frac{T_{INT}.P_{\mathrm{PIXEL}}(\lambda)}{E(\lambda)}\, \mathrm{d}\lambda. \tag{29}$$

The calculation is repeated for signal and background reflectance at a given illumination level. For the purposes of this study shot noise is simply modeled as $\sqrt{PH_{\mathrm{PIXEL}}}$, and no other noise sources are considered. The image height, h_{i}, of the cobblestone at range, r, is estimated using [9]:

$$h_i = \frac{h_o f}{r}, \tag{30}$$

where, h_o is the object height and f, focal length. The latter is estimated as [9]:

$$f = \frac{pN_H}{2\tan\left(\frac{FOV_H}{2}\right)}, \tag{31}$$

where, p, is pixel size, N_H, the number of pixels in the horizontal direction of the sensor, and FOV_H, the horizontal field of view. For a given image height the area of the image determined in squared pixels, which is equivalent to the area of the matched filter, A, in Eqs. (9) and (14)–(19), may then be calculated:

$$\mathrm{A} = \frac{h_i^2}{p^2}. \tag{32}$$

For given illumination conditions, camera geometries, and target sizes, the value of PH_{PIXEL} for background and target reflectance may be used to estimate Δq, $\bar{q}$ in the calculation of predicted $\mathrm{SNR_I}$ directly (Eq. (19)) and further E_{MAX} and E_{MIN} to generate patches of size $w \times w$ pixels. Poisson noise may then be added to patches and CDP calculated directly from them. In addition, a matched filter of the same size as the cobblestone target is generated and the output value of this filter evaluated for the background and target. Repeating the evaluation of the matched filter for many instances of the input images allows the actual $\mathrm{SNR_I}$ of the detection process to be evaluated via the mean and the standard deviation of the output and compared to the predicted values. For emphasis, $\mathrm{SNR_I}$ does not predict the image quality of the camera system directly, rather its ability to distinguish a target against a background. For the idealized observer, the optimum detection process is shown to be a matched filter and hence used here; however, others do exist.

3. RESULTS AND DISCUSSION

Figure 4 shows CDP calculated for 10×10 pixel target. The target size is chosen as it represents a suitable size to gather

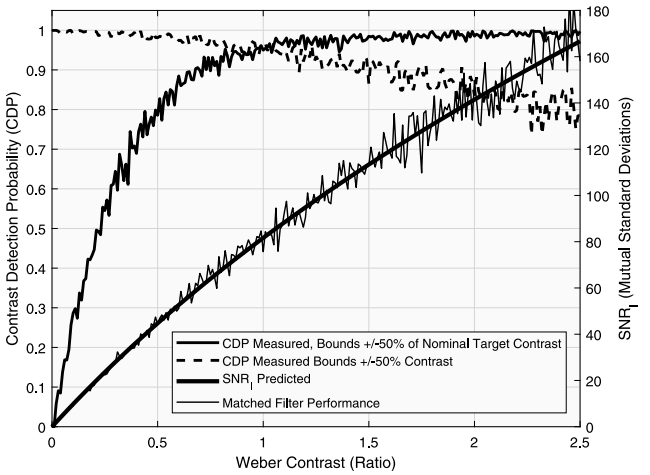


Figure 4. CDP calculated with constant and percentage contrast bounds, predicted $\mathrm{SNR_I}$, and matched filter performance plotted against Weber contrast for a 10×10 pixel cobblestone target with a background exposure of 100 photoelectrons. The target exposure is swept from 100 to 350 photoelectrons. $\mathrm{SNR_I}$ is estimated from matched filtering of 100 images per data point.

imaging statistics and also a point at which most DNNs start to have some difficulty with the detection of objects. The background exposure is held at 100 photoelectrons and the target exposure swept from 100 to 350 photoelectrons to create Weber contrasts from 0 to 250%. We will name these positive contrast targets as the target signal is greater than that of the background. Photoelectrons are specified directly in this case to evaluate the relative merits of the metrics rather than their performance when compared to the imaging model. The imaging model is used to generate results in Figures 9–14 to compare performance with respect to actual camera parameters. CDP is calculated using $\pm 50\%$ of the nominal Weber contrast as limits and also a constant $\pm 50\%$ contrast from the nominal Weber contrast. Also shown is the $\mathrm{SNR_I}$ measured from the matched filter detection process and that predicted using Eq. (19). Immediately it may be observed that using constant contrast bounds from the nominal Weber contrast causes CDP to start at a value of one at zero contrast and falls, contrary to contrast and measured $\mathrm{SNR_I}$ increasing. Any contrast measured at the nominal input contrast of zero is due to noise and, in this case, is clearly within the $\pm 50\%$ Weber contrast bounds. CDP calculated in this manner then continues to fall because, as the target signal increases, the absolute variance of the noise in the target also increases despite the signal-to-noise ratio improving. Therefore, the width of the distribution of Weber contrasts will continue to increase as contrast increases and will eventually exceed any limits that are kept constant. Limits for the calculation of CDP should therefore be set as a ratio of the nominal input contrast and this approach is used throughout the remainder of this article. Measured $\mathrm{SNR_I}$ is estimated from 100 images per data point and seen to increase with Weber contrast as would be expected. Eq. (19) predicts the idealized observer matched filter performance accurately across a very wide range of contrast.

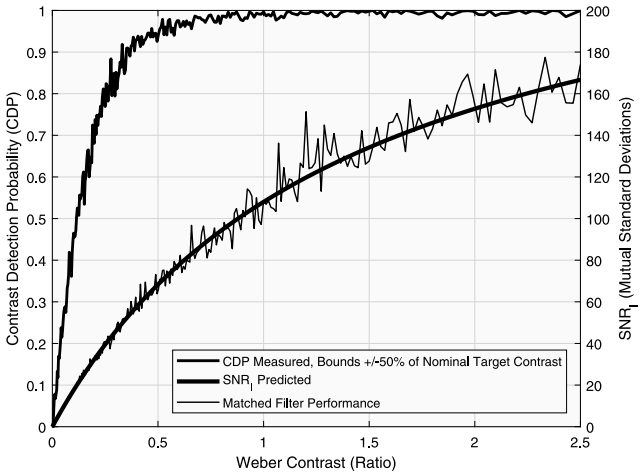


Figure 5. As per Fig. 4 though sweeping the target from 350 to 100 photoelectrons to create negative contrast targets.

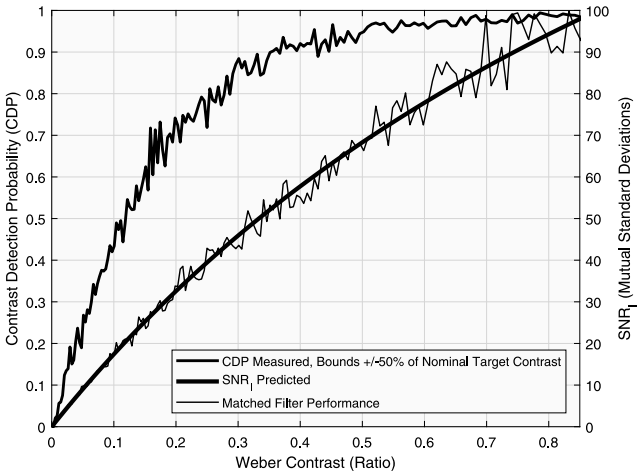


Figure 6. A portion of Fig. 5, plotted such that SNRI and CDP are scaled the same vertically. The relative over prediction of mid-contrast performance by CDP is apparent.

Comparing only the shape of the CDP curve to measured SNR$_I$, Fig. 4, CDP appears to over-estimate the performance of the imaging system at low contrasts relative to those that are higher. CDP rises much faster than SNR$_I$ between contrast of 0 to 50% and then asymptotically approaches a value of one from 50% to 250%. The metric is predicting that the 200% increase in contrast between 50% and 250% explains just one fifth of the detection performance. This is highly non-linear with respect to the actual matched filter performance and it would be difficult to predict machine vision performance from such a curve. Geese et al. state that a CDP of 1 indicates that contrast is faithfully recorded within the bounds set and performance is sufficient [6]. When this happens however, CDP is unable to predict further performance increases or distinguish between systems that are above this threshold. An SNR$_I$ of five would indicate that test results for two hypotheses would be separated by five standard deviations and the detection success for such a system will be close to 100%. It may be seen in Fig. 4 that CDP does not start to approach one until SNR$_I$ has reached a value of approximately 100, at a Weber contrast of approximately 125%. In this sense CDP is relatively slow to respond and has only reached a value of approximately 0.12 when SNR$_I$ (a value of 5) is predicting 100% detection success.

Figure 5 is calculated as per Fig. 4 for negative contrast targets. The background is held at 350 photoelectrons and contrast increased by sweeping the target from 350 to 100 photoelectrons. Both CDP and SNR$_I$ predict performance as being slightly better in the mid-contrast range due to holding the background at a higher SNR. Again, when compared to actual performance, CDP appears to predict performance of lower contrasts as being relatively higher than high contrasts when compared with SNR$_I$. This may be seen by plotting the portion of the CDP curve that is below one and scaling the SNR$_I$ axis to occupy approximately the same vertical distance, Figure 6. Again, in Figs. 5 and 6, it may be seen that CDP requires 100 standard deviations to reach a value of one.

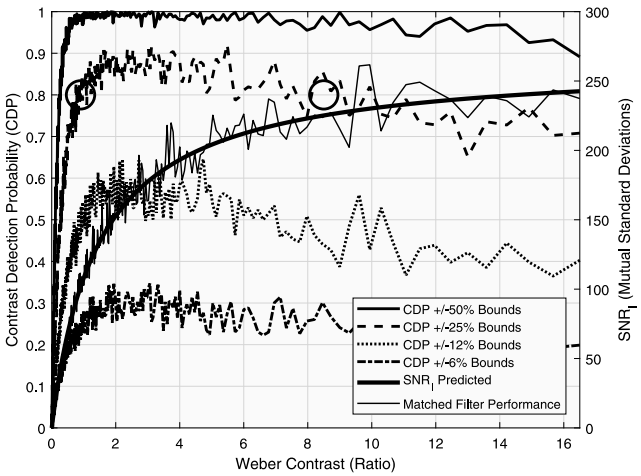
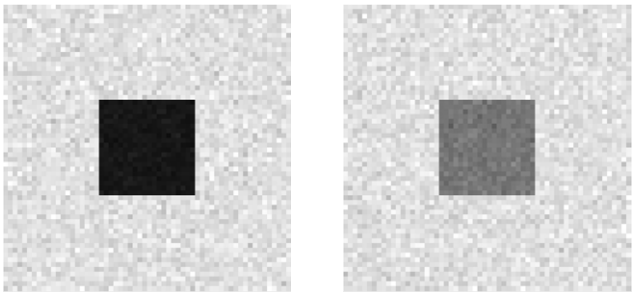


Figure 7. CDP calculated with a background level of 350 and target value swept from 350 to 20 photoelectrons. CDP is calculated using various bounds. CDP is seen to reach a maximum value and then fall as contrast increases. The circles indicate targets that have the same value of CDP despite different contrasts and numbers of photoelectrons at approximately 37 and 185 in this case.

The target value is swept from 350 to 10 photoelectrons to examine very high negative contrast signals, Figure 7. The figure shows that CDP, calculated for a number of different bounds, peaks and then starts to decrease as contrast increases. CDP is predicting lower imaging performance as contrast increases and this is contrary to the measured SNR$_I$. In the case of these negative contrast targets, the SNR of the target is decreasing as the contrast between the background and the target is increasing. This is adversely affecting the calculation of contrast, because of the low value of the denominator in the Weber contrast calculation, broadening the distribution of contrasts calculated. At the same time, however, the actual separation between the background and target levels remains high. Despite some correlation, it is signal separation with respect to noise that is more significant for object detection than contrast with respect to noise.



(a) (b)

Figure 8. A target of, (a), 37 and, (b), 185 photoelectrons against a background of 350 photoelectrons. Full system scale has been set to 400 photoelectrons. This would represent approximately 16x gain in a typical 2 μm system. CDP is calculated as 0.8 for both combinations with limits as 25% of nominal input contrast.

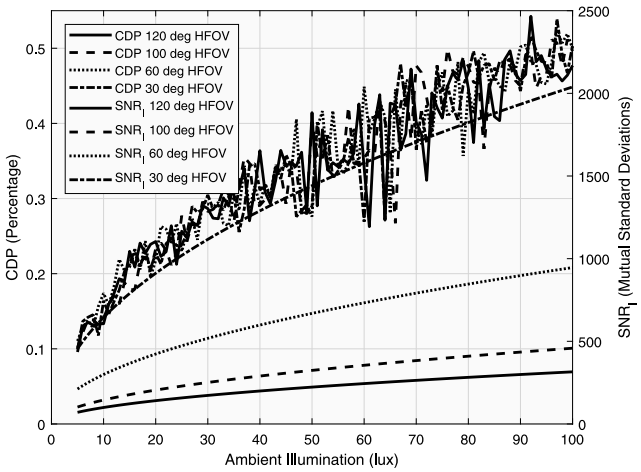


Figure 9. CDP and predicted SNR$_I$ calculated from the output of the imaging model for a system with a 2 μm pixel and $f/2.2$ lens at various horizontal field of views (HFOV), imaging a 0.5 × 1.8 m object of 10% reflectance against a background of 1% at a distance of 50 m with an exposure time of 10 ms for various ambient light levels at a color temperature of 5500 K. CDP does not account for the magnification of the target image.

The maxima observed in the CDP results will cause it to predict the same result for targets of different reflectance with the same background as indicated by the circles in Fig. 7. For the example shown, CDP has the same value, 0.8, for targets with approximately 185 and 37 photoelectrons against a background of 350, Figure 8. In fact, any point on the curve beyond the maximum may be paired with a target with a different photoelectron count and changing calculation bounds does not change this behavior. Because of this caution should be exercised when comparing negative contrast targets of different values using CDP.

Fig. 9 illustrates CDP and predicted SNR$_I$ calculated from the output of the imaging model for a system with a 2 μm pixel and f/2.2 lens at various horizontal field of views (HFOV) imaging a 0.5 × 1.8 meter object of 10% reflectance against a background of 1% at a distance of 50 m with an exposure time of 10 ms for various ambient light levels at a color temperature of 5500 K. A typical monochrome

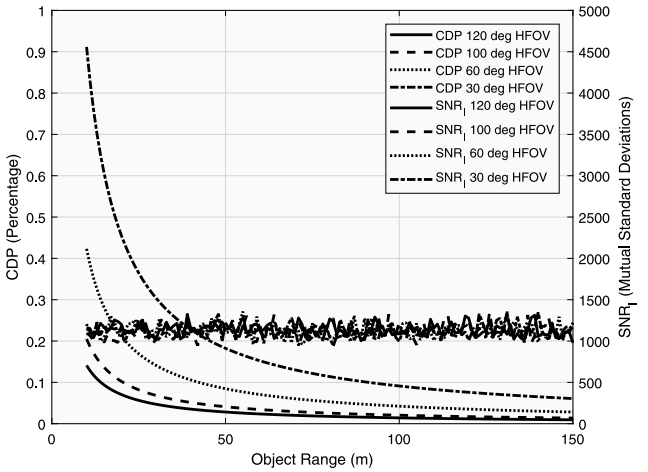


Figure 10. As per Fig. 9 with respect to distance at 20 lux ambient light level.

quantum efficiency curve is used. The object is intended to represent the approximate dimensions and reflectivity of an adult in dark clothing. Measured SNR$_I$ is omitted for clarity as contrast is 900% or below and predicted SNR$_I$ is shown to have little error even at high-contrast levels. CDP is calculated using ±20% bounds. For this positive contrast target, CDP correctly predicts the trend of the imaging performance and the shape of the curve is correct when compared with the result predicted by SNR$_I$. The calculation of CDP is based on comparison of pixels; however, it fails to predict the effect of the lens HFOV on the performance. As the HFOV increases, target magnification is smaller and detection performance diminishes. Magnification is a primary parameter in machine vision performance and as such crucial to describe behavior of camera systems to serve this function.

Repeating the calculation with respect to distance at an ambient light level of 20 lux, Fig. 10, a similar situation is encountered. As would be expected, CDP does not vary with object distance when only ambient light is considered as the wattage per unit area of light falling on the sensor remains constant if there are no adverse weather conditions or MTF degradation. It is the diminishing size of the target that causes SNR$_I$ to fall with respect to distance and increasing HFOV. This again is not accounted for in the calculation of CDP by the absence of the consideration of target magnification.

Changing pixel size while keeping the pixel count the same allows the sensor area to grow and increases the total number of photoelectrons collected per pixel, Fig. 11. This is recognized both in the increasing values of CDP and predicted SNR$_I$ with respect to pixel size. CDP is able to predict the rank of the camera systems correctly in this case; however, it fails to correctly predict performance when the sensor size is kept the same as pixel size grows, Fig. 12. In this case the image size remains the same and thus the number of photoelectrons per unit area or photon flux at the image plane stays the same. Any improvement in per-pixel noise is counteracted by a reduction in the number of pixels across the target, and as such the detectivity of the target is

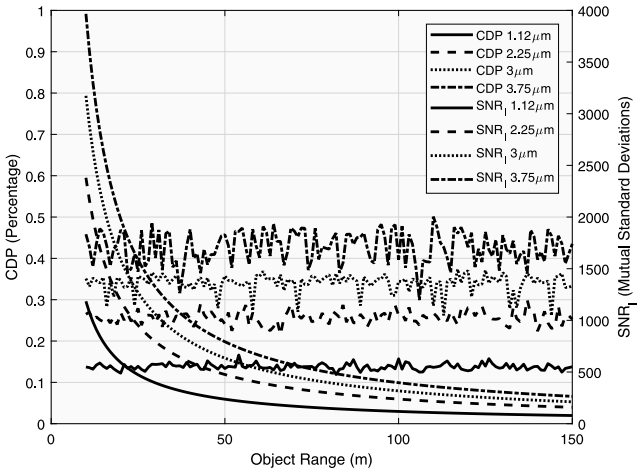


Figure 11. CDP and SNR_I calculated for systems with increasing pixel size. The pixel count is maintained at 3800 and the sensor area is allowed to grow proportionally. A 60 degree HFOV f2.2 lens is modeled. The light level is 20 lux with an exposure time of 10 ms.

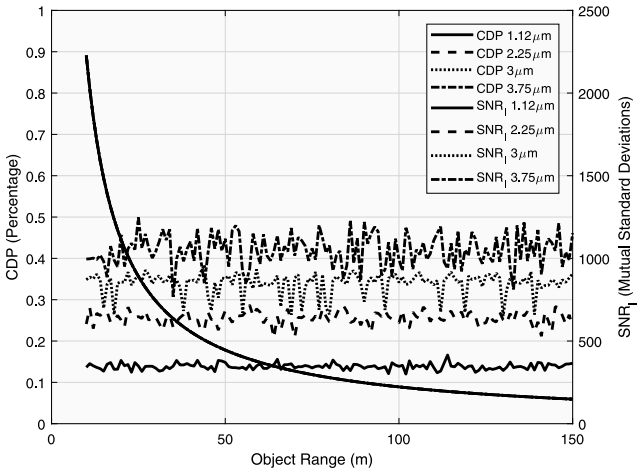


Figure 12. As per Fig. 11, though keeping the sensor format the same while increasing pixel size and reducing pixel count. The approximate horizontal size of the sensor is 8 mm, a typical 1/2 inch format. Any improvement in per-pixel noise is counteracted by a reduction in the number of pixels across the target, and as such the SNR_I of the target at a given distance remains constant with the change in pixel size. The SNR_I curves for each of the different pixel sizes are the same and superimposed on each other. In practice, the absolute sensitivity per μm, read noise and MTF of the sensor would be slightly different for different pixel sizes and cause differences in the SNR_I.

expected to remain constant. As there is no simulation of MTF degradation or NPS shaping in this study the SNR_I is therefore exactly the same for each pixel size modeled. In reality, the absolute sensitivity of pixels per square μm and MTF does change with respect to their size and we would expect to see some difference in practice. As expected, CDP increases with respect to an increase in aperture and is able to correctly rank cameras when all other parameters are kept the same, Fig. 13.

As a final example, Fig. 14 compares CDP and SNR_I calculations for two systems; (a) 2.1 μm pixel, 100 deg HFOV, $f/1.8$ and (b), 2.25 μm pixel, 70 deg HFOV, $f/2.2$. Both have the same horizontal pixel count, 3800, monochrome quantum efficiency curve as above and are imaging the previous 0.5×1.8 m target at a light level of 200 lux and 5500 K with an integration time of 10 ms. It is difficult to predict from these headline specifications which system is expected to perform best. System (a) has a better aperture than (b), though a wider HFOV and thus lower magnification. SNR_I in fact predicts system (b) will perform better and it appears that the system magnification wins over the improvement in aperture in (a) in this case. CDP however measures opposite, system (a) as performing better than (b). Despite being sensitive to aperture and pixel size changes when pixel count is kept the same, insensitivity to the HFOV difference has caused CDP to predict the opposite to SNR_I. This suggests that, while CDP may be able to rank systems based on the variation of a single system parameter if chosen carefully, it is not appropriate to use where multiple variables are to be optimized simultaneously. CDP does not account for magnification in the metric and therefore basic target size and FOV behavior is omitted. Balancing pixel size and count against FOV is a fundamental need when designing camera systems. Overly reducing FOV to increase effective resolution beyond that needed for the task at hand unnecessarily constrains the real-world space that may be searched in a single frame. Also shown in Fig. 14 are results

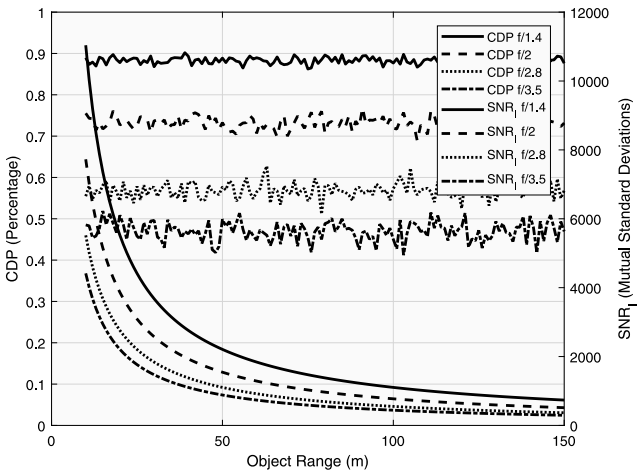


Figure 13. The effect of aperture on SNR_I and CDP for a pixel size of 2.1 μm and count of 3800, at 200 lux and an integration time of 10 ms.

of the matched filter detection for the two systems. It is seen that SNR_I tracks performance closely.

It is proposed that SNR_I is a good candidate as an overall system metric that will relate camera and illumination specifications to machine vision performance in automotive and other applications. It may be measured simply via evaluation of the MTF and NPS of a system with basic knowledge of targets and backgrounds that are to be distinguished. Failing good knowledge of targets and background, Noise Equivalent Quanta (NEQ) may be calculated as suggested by Keelan [13]. NEQ provides a good first order rank of system performance and is similar to Eq. (6) without scaling for the frequency content of the

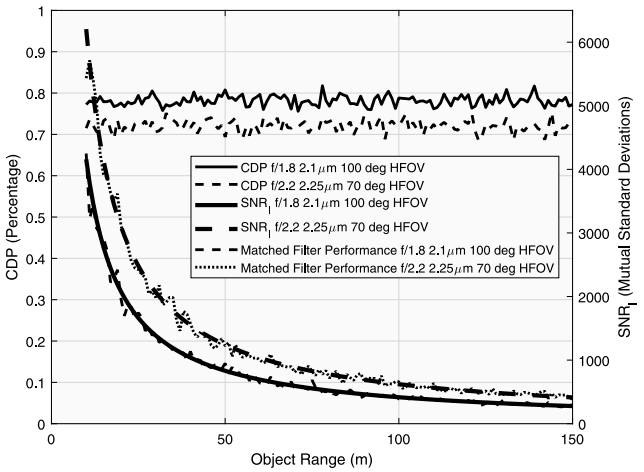


Figure 14. Comparison of CDP and SNR_I for two systems, (a) $f/1.8$ 2.1 μm 100 deg HFOV versus (b) $f/2.2$ 2.25 μm 70 deg HFOV where the pixel count is the same at 3800, and light level 200 lux imaging the target as for Fig. 9. Also shown is the measured matched filter performance.

target [13]:

$$NEQ(V_x, V_y) = \frac{MTF^2(V_x, V_y)}{NPS(V_x, V_y)/\mu^2} \tag{33}$$

where μ is mean linear signal. NEQ does not indicate suitability for a specific task as does SNR_I; however, it does provide curves that are consistent with detection theory and may be either integrated to provide first order performance or cascaded with target and background information, when known, to provide SNR_I. Regardless of the approach taken, it is clear that reliable measures of MTF and NPS are needed for robust prediction and measurement of system performance. While the use of detection theory and derivative measures is well established in the defense and medical communities, work is needed to adapt and ascertain its relevance in the automotive field. In particular, work to create links between SNR_I and DNN performance should be welcomed.

4. CONCLUSIONS

Predicted SNR_I has been shown to closely follow measured matched filter performance for a rudimentary imaging model and detection process. CDP does not track detection performance for this model and a lack of information concerning image magnification and pixel size is the primary cause for this. Mid-contrast performance tends to be over-estimated by CDP when compared with SNR_I. Comparison of diverse systems will lead to results that are not consistent with established detection theory and it is possible to generate CDP values that are identical for negative contrast targets of differing nominal contrast values, and further, predictions of camera system performance that contradict SNR_I. CDP is slow to respond and at bounds of 50% of nominal input contrast, only reaches a value of 100% at 100 standard deviations of matched filter response between the target and background. SNR_I has an open upper bound and can continue to provide estimation of system performance beyond the point where CDP reaches its upper limit of one and for a wide range of contrast and illumination conditions.

ACKNOWLEDGMENT

The author is grateful to Paul Kane, Chris Dainty, and Sean Pieper for many useful discussions and support leading to the completion of this work.

REFERENCES

[1] IEEE Autonomous Imaging White Paper, The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016–5997.
[2] ICRU Report 54, Medical Imaging – The Assessment of Image Quality, Bethesda MD: International Commission on Radiation Units and Measurements, 1996.
[3] H. H. Barrett and K. Myers, *Foundations of Image Science* (Wiley, Hoboken, NJ, 2004).
[4] J. Beutel, H. Kundel, and R. Van Metter, *Handbook of Medical Imaging* (SPIE Press, Bellingham, WA, 2000), Vol. 1.
[5] K. Paul, "Signal detection theory and automotive imaging," *IS&T Electronic Imaging: Autonomous Vehicles and Machines 2019 Proceedings* (IS&T, Springfield, VA, 2019), pp. 027-1–027-7(7).
[6] M. Geese, U. Seger, and A. Paolillo, "Detection probabilities: performance prediction for sensors of autonomous vehicles," *IS&T Electronic Imaging: Autonomous Vehicle and Machines 2018 Proceedings* (IS&T, Springfield, VA, 2018), pp. 148-1–148-14(14).
[7] L. Ebbert, *Implementation of CDPBachelor Thesis*, (University of Applied Sciences Dusseldorf, 2018).
[8] A. E. Burgess, "The Rose model, revisited," J. Opt. Soc. Am. **16**, 633 (1999).
[9] R. Jenkin and P. Kane, "Fundamental imaging system analysis for autonomous vehicles," *IS&T Electronic Imaging: Autonomous Vehicles and Machines 2018 Proceedings* (IS&T, Springfield, VA, 2018).
[10] M. Richardson, M. A. Richardson, I. C. Luckraft, R. S. Picton, A. L. Rodgers, and R. F. Powell, *Surveillance and Target Acquisition Systems, Brassey's Land Warfare Series, London* (Brasseys, London, UK, 1997).
[11] M. A. Richardson, "Electro-optical systems analysis Part 2," J. Battlefield Technol. **5**, 21 (2002).
[12] R. W. G. Hunt, *Measuring Color, Ellis Horwood*, 2nd ed. (Ellis Horwood, London, UK, 1991).
[13] B. W. Keelan, "Imaging applications of noise equivalent quanta," *IS&T Electronic Imaging: Image Quality and System Performance XIII Proceedings* (IS&T, Springfield, VA, 2016), pp. 1–7.