https://doi.org/10.2352/ISSN.2470-1173.2020.16.AVM-041

Object Detection Using an Ideal Observer Model

Orit Skorka[1] and Paul J. Kane[2]
Intelligent Sensing Group, ON Semiconductor; [1]Santa Clara, CA, USA; [1]Rochester, NY, USA

Abstract
Many of the metrics developed for informational imaging are useful in automotive imaging, since many of the tasks – for example, object detection and identification – are similar. This work discusses sensor characterization parameters for the Ideal Observer SNR model, and elaborates on the noise power spectrum. It presents cross-correlation analysis results for matched-filter detection of a tribar pattern in sets of resolution target images that were captured with three image sensors over a range of illumination levels. Lastly, the work compares the cross-correlation data to predictions made by the Ideal Observer Model and demonstrates good agreement between the two methods on relative evaluation of detection capabilities.

Introduction
Much work has been done previously, particularly in medical imaging, to develop mathematical methods to quantify the signal detection performance of imaging systems in terms of their measured characteristics such as Noise Power Spectrum (NPS) and Modulation Transfer Function (MTF). Until recently, this work has not been widely applied in the field of automotive imaging. However, now that computer vision is finding wide application in automotive imaging, system designers are faced with the problem of selecting camera components that can support the detection and/or identification of objects at low error rates. As image sharpness and Signal to Noise Ratio (SNR) increase, the probability of false alarms decreases, and the probability of valid detections increases. The mathematical formalism of the Ideal Observer model can be used to relate detection performance to object and system characteristics. Detection performance depends on five key factors: the size and spatial structure of the object at the detector, the contrast of the object, the mean detected signal level, the Modulation Transfer Function (MTF) of the camera system, and the Noise Power Spectrum (NPS) of the sensor at the relevant signal level.

Ideal Observer SNR
The Ideal Observer is a Bayesian decision maker that maximizes the statistical precision of a hypothesis test where two possible outcomes are, for example, H2 (an object is present) and H1 (no object is present) when given a detected image [1]. The detected image is one realization from an ensemble of possible images that could arise depending on the noise in the detection process at the time of capture.

SNRI Formulation
In the case where both the signal and the background are known exactly, the only random fluctuations in the image are due to noise, and it can be shown that the SNR of the Ideal Observer (SNRI) is related to the imaging system characteristics as follows [1]:

$$SNRI^2 = \int\int \left(\frac{MTF^2(\nu_x, \nu_y) \cdot \mu^2}{NPS(\nu_x, \nu_y)} \right) \Delta S^2(\nu_x, \nu_y)\, d\nu_x\, d\nu_y. \quad (1)$$

Here ν_x and ν_y are the spatial frequencies, MTF is the imaging system MTF, NPS is the sensor Noise Power Spectrum, and μ is the mean signal level. ΔS is the Fourier transform of the difference object, which is the difference between the signals (objects) input to the system under the two hypotheses being tested in the general case and, for object detection, this is the difference between the object and a uniform background (object not present). In a previous paper [2], we showed how Eq. (1) could be applied to problems of interest in automotive imaging, such as detecting a small object on the road as a function of distance and illumination, or choosing a pixel size for a detection task given the performance characteristics of the lens and sensor.

SNRI Characterization Parameters
This section briefly refers to MTF, and then elaborates on NPS. It describes characterization procedure, presents example results, and proves that the NPS is, approximately, the mean noise power.

Modulation Transfer Function
Ideally, the point-spread-function (PSF) of the system should be used for SNRI calculation. However, due to the complexity of procedures to characterize PSF, MTF is calculated here using the line-spread-function (LSF), as described in the ISO-12233 standard [3]. Because this is a well-established method, it is not discussed here.

Noise Power Spectrum
The Noise Power (or Wiener) Spectrum represents the noise variance in each spatial frequency interval [4], and is related to the visual appearance of the noise pattern. In analogy with MTF, an NPS curve that decreases sharply with increasing spatial frequency (low-pass spectrum) indicates a noise pattern that has a soft or blurry appearance. Conversely, a flat NPS indicates a noise pattern with equal noise variance at all spatial frequencies, having sharply defined spatial fluctuations.

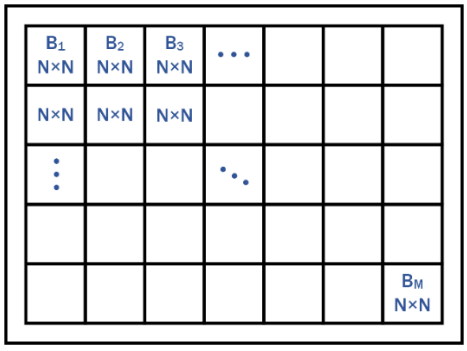


Figure 1. For NPS characterization, the image of a uniformly illuminated array is divided into M non-overlapping N×N pixel blocks.

The NPS is obtained from an image capture of a flat field, and is dependent on the signal level. As shown in Figure 1, the image is divided into M non-overlapping blocks of $N \times N$ pixels.

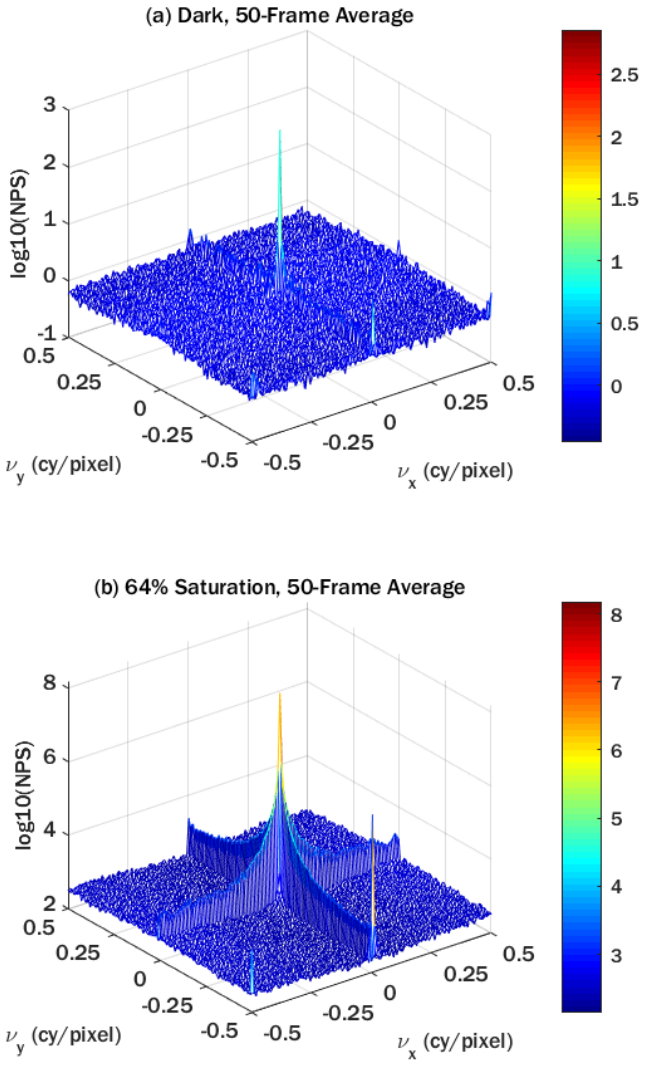


Figure 2. 3D NPS plots as obtained from 50-frame average images that were captured with a 1 MP monochrome image sensor and processed with 128×128 pixel blocks. Results are shown for images (a) in the dark and (b) when mean signal level was about 64% of saturation. In both signal levels, the NPS surface is mainly flat, except for a spike near zero spatial frequency.

Let the data array from block m be denoted B_m, and the grand mean of the captured image be denoted $\overline{B}$. Then the NPS is calculated as follows:

$$NPS = \frac{1}{M \cdot N^2} \sum_{i=1}^{M} |FFT(B_i - \bar{B})|^2, \quad (2)$$

where FFT denotes Fast Fourier Transform. The maximal spatial frequency is defined by pixel size, p, where Nyquist frequency is $\frac{1}{2p}$. Measurement resolution is $\frac{1}{N \cdot p}$, and a higher M (up to a certain limit) correlates with better statistical precision in the estimate at each frequency. Therefore, there is a trade-off between spatial frequency resolution and statistical precision.

The 3D surface plots of the NPS at two exposure levels, dark and 64% signal saturation, are shown in Figure 2. Both were calculated from 50-frame average images that were captured with a 1 MP monochrome CMOS image sensor, and processed with 54 pixel blocks of 128×128 pixels after excluding peripheral pixels due to illumination non-uniformity. Frame average was done in order to suppress temporal noise and emphasizes fixed-pattern noise (FPN). For the dark exposure in Figure 2(a), the surface is mainly flat, except for a spike near zero spatial frequency. This spike is due to low frequency non-uniformities that are nearly impossible to avoid in practice, and is not considered part of the 2D random noise pattern. The 64% signal saturation capture in Figure 2(b) also shows ridges along ν_x and ν_y that are indicative of row and column 1D random fixed-pattern noise.

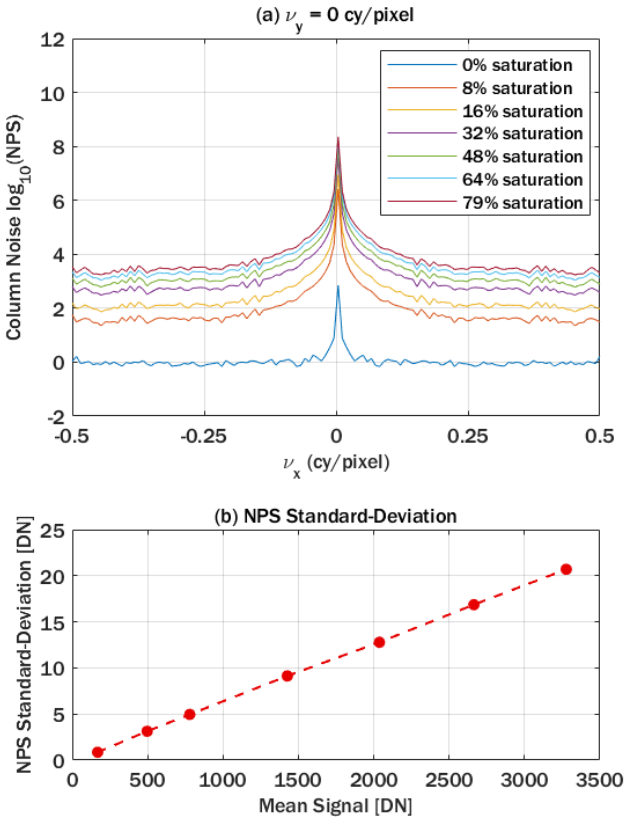


Figure 3. (a) Cross-section of 50-frame average images with different mean signal level that were captured with the same sensor as in Figure 2. Absence of temporal noise is very clear from the plots because the fluctuations in the estimate are maintained despite the change in mean signal level. (b) Mean NPS standard-deviation as calculated for the conditions that are specified in (a) for $|\nu_x| > 0$ vs mean signal level. There is linear relationship between the two parameters. That is in agreement with the characteristics of an image sensor FPN, as this noise is expected to be proportional to mean signal level.

Figure 3(a) presents the cross-section of the 3D NPS plots that were captured with mean signal level that varied from 0 to 79% signal saturation at $\nu_y = 0$ cy/pixel. One may observe that the variations in the NPS estimate are preserved despite the change in mean signal level, indicating that the noise is truly dominated by FPN. Figure 3(b) shows NPS mean standard deviation (STD) for $|\nu_x| > 0$ vs mean signal level, and one may observe that there is a linear relationship between the two properties. In a typical image sensor,

FPN is proportional to mean signal, therefore, this confirms that STD of the NPS of multi-frame average image represents the sensor FPN.

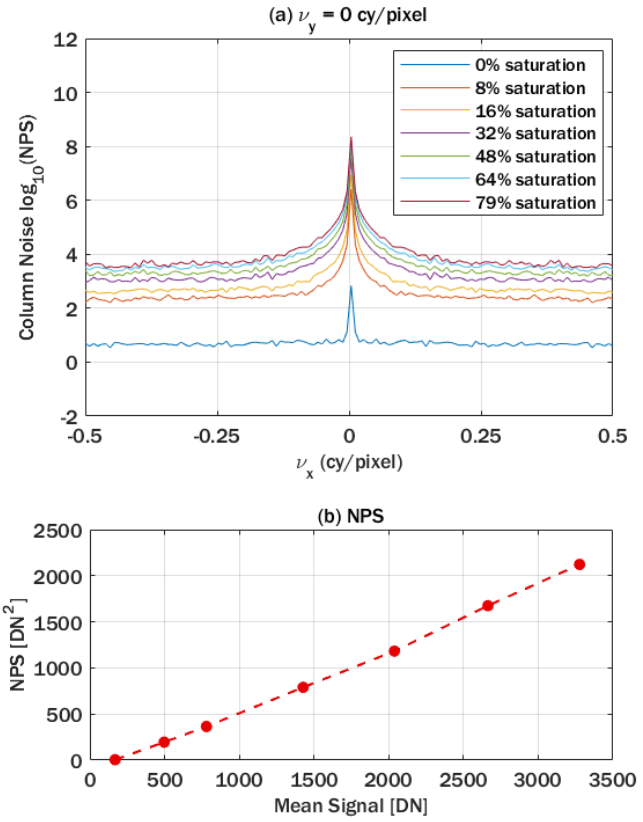


Figure 4. (a) Cross-section of a single frame images with different mean signal level that were captured with the same sensor as in Figure 2. Noise levels are clearly higher than in Figure 3(a), which is expected as these images include temporal and fixed-pattern noise. (b) Mean NPS as calculated for the conditions that are specified in (a) and as calculated for $|\nu_x| > 0$ vs mean signal level. The plot shows linear relationship between the two parameters, which is in agreement with the characteristics of a shot-noise limited system.

The NPS plots in Figure 4(a) were obtained using the same procedure that was used with the plots in Figure 3(a), however, this time single frame images were processed. Therefore, the plots include temporal noise and FPN. Figure 4(b) shows mean NPS for $|\nu_x| \geq 0$ vs mean signal level, and one may observe that there is a linear relationship between the two properties. In a system that is shot-noise dominant, noise power is proportional to mean signal level.

To conclude, NPS of a typical image sensor is constant for spatial frequencies that are greater than 0 cy/pixel and, in a very good approximation, it is equal to the mean noise power of the image sensor. Therefore, one may obtain rather accurate results if the frequency-dependent NPS is replaced by a simple average noise power estimate.

Experimental Results

To correlate results from SNRI calculation and pattern detection in actual images, image sets of the same target were captured at varied luminance conditions with three monochrome image sensors that differed by array size, pixel size, and fabrication process. A basic algorithm for matched-filter detection was applied for relative evaluation of detection capabilities. SNRI was calculated for the same sensors and the same pattern according to the procedure that was explained in the previous section. Results from both evaluation methods are compared.

Pattern Detection in Captured Images

Image sets of a chrome-on-glass USAF 1951 target with 100% contrast were captured using a camera module with a Sunex DSL945D lens (F#2.5, 70° field of view, 650 nm IR-cur filter) that was placed at a distance of about 100 mm from the target. The target was place at the output port of an OL-462 motorized integrating sphere system with a 3,000K illuminant. Figure 5(a) shows a photo of the setup.

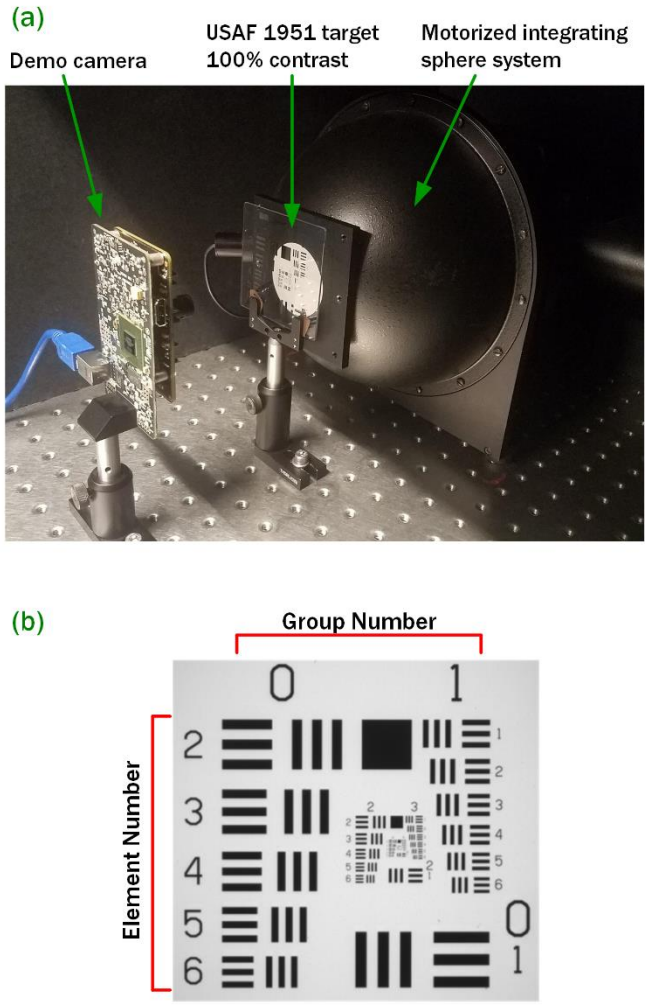


Figure 5. (a) The test setup included a demo camera, a motorized integrating sphere system with illuminant correlated color temperature of 3,000K, and a USAF 1951 target with 100% contrast. (b) Central region of the target. Each element includes a vertical and horizontal tribar patterns. Group and element numbers are the horizontal and vertical numbers, respectively.

The three image sensors that were used in this study were (a) 1 MP, 3.75 µm pixel size, global shutter, front-side illuminanted (FSI) imager, (b) 1 MP, 3.0 µm pixel size, global shutter, FSI imager, and

(c) 5 MP, 2.2 µm pixel size, rolling shutter, back-side illuminanted (BSI) imager.

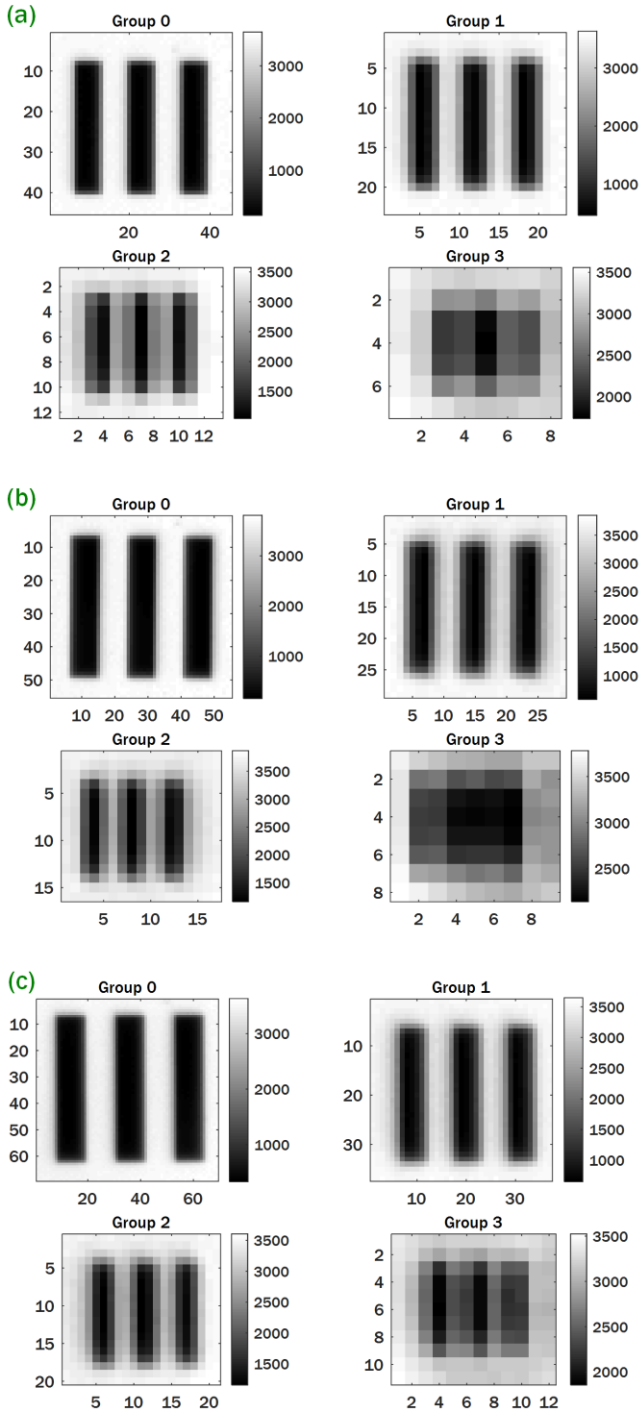


Figure 6. 50-frame average images of element #4 in groups 0–3 as captured with the (a) 3.75 µm, (b) 3.0 µm, and (c) 2.2 µm pixel sensors in high luminance conditions.

Figure 5(b) presents a photo of the central region of the target. Each element is composed of a horizontal and vertical tribar patterns. Group number of each element is represented by the horizontal number, and its element number is represented by the vertical number [5]. Relative spatial frequency of each element is given by:

$$\nu = 2^{Group\#+(Element\#-1)/6}. \tag{3}$$

Therefore, the size of an element decreases by a factor of 2 in the transition from one group number to the next one.

Figure 6 presents multi-frame average images of the vertical tribar pattern of Element #4 in Groups #0 to #3 as captured by the three image sensors with high scene luminance. Image of the tribar pattern in Groups #0 and #1 can be easily resolved in all cases, in Group #2 it is rather blurred in the images that were captured with image sensors (a) 3.75 µm pixel and (b) 3.0 µm pixel, and in Group #3 it is blurred in all cases.

The MATLAB built-in function normxcorr2 was used for a basic matched filter detection algorithm, and the images in Figure 6 were used as references. Normalized cross-correlation values, as calculated by this function at the location of the pattern, are shown in dashed lines in Figure 10 for the patterns in the four groups at varied background luminance level. Background luminance was measured on transparent regions the target. One may conclude from these plots that, with all image sensors, there is degradation in performance with decrease in feature size and in scene luminance, which is expected. Results also show that image sensor (c), 2.2 µm pixels, outperforms the other two image sensors and that, at high luminance levels and with large pattern size, performance of image sensor (a), 3.75 µm pixels, is comparable to that of (c).

SNRI Evaluation

To model the detectability of the tribar pattern under the assumptions of the Ideal Observer, SNRI was evaluated for image sensors (a), (b), and (c) using Eq. (1) formalism. However, because the calculation was done numerically, this equation was re-written in a discrete form as follows:

$$SNRI^2 = \Delta\nu_x\Delta\nu_y \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \frac{MTF^2(i,j)\cdot\mu^2}{NPS(i,j)}\Delta S^2(i,j) \tag{4}$$

Here i and j are spatial frequency indicies, $\Delta\nu$ is the spatial frequency spacing in cycles/pixel, and N_x, N_y are the number of spatial frequency samples in the x and y directions. Considering only the 2D random noise component, the quantity μ^2/NPS(i,j) reduces to μ^2/σ^2, where σ^2 is the noise power, which is recognized as the SNR^2 – in other words, this quantity can be taken outside the sum as a scale factor of SNR^2, as expressed in Eq. (5):

$$SNRI^2 \cong SNR^2 \cdot \Delta\nu_x\Delta\nu_y \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} MTF^2(i,j)\cdot\Delta S^2(i,j). \tag{5}$$

The SNR can be measured as a function of illumination by computing mean signal and noise in trasparent and opaque regions of the USAF 1951 target. MTF includes sensor MTF and lens MTF. Sensor MTF was measured by the slanted-edge method with a narrow band filter with central frequency of 550 nm. Lens MTF was calculated at 550 nm assuming a diffraction-limited lens.

Figure 7 gives a graphic description of the patterns that are used to calculate the difference object, ΔS. Figure 7(a) is the tribar pattern with the dimensions of the object on the image plane as calculated from actual pattern size on the target and the demagnification factor of the lens, and Figure 7(b) is the uniform background with luminance level that is the average luminance of the dark and bright regions of the pattern. Because the target had 100% contrast, the uniform background here is 50% gray.

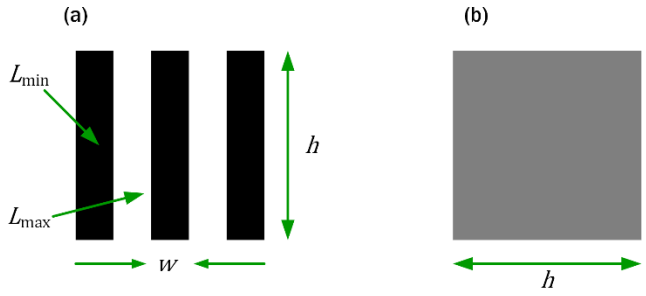


Figure 7. The difference object is a tribar pattern with 100% contrast (a) minus a square with an average signal level (b). Parameters h and w represent the height and the width of each stripe, respectively, and L_{min} and L_{max} represent dark and bright regions, respectively, for contrast calculation.

A mathematical description of ΔS can be written as follows:

$$g(x,y) = \left[\Pi\left(\frac{x}{w}\right) + \Pi\left(\frac{x-2w}{w}\right) + \Pi\left(\frac{x+2w}{w}\right)\right] \cdot \Pi\left(\frac{y}{h}\right), \quad (6)$$

where $\Pi(x) = \text{rect}(x)$ is the usual rectangle function [6]. Using this expression, the frequency domain model for the difference between the tribar pattern and a rectangle function of the same size takes the form:

$$\Delta S(\nu_x, \nu_y) = C_M wh \cdot sinc(w\nu_x) \cdot 2\cos(2\pi w\nu_x) \cdot sinc(h\nu_y) \quad (7)$$

where

$$C_M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (8)$$

is the Michelson contrast of the pattern.

Therefore, we see that the value of SNRI is linearly proportional to the SNR and the object contrast C_M (from Eq.(8)). In addition, SNRI scales with the square root of the area under the product of the squared MTF and difference object spectrum curves. If more blur is present, the MTF will fall off more rapidly with spatial frequency, reducing the value of the sum. As the object shrinks in size at the sensor plane, spanning fewer pixels, the difference object spectrum will spread out in frequency space, meaning the the MTF will affect a broader range of frequencies. This is illustrated in Figure 8. Figure 8(a) shows an intensity map of the function ΔS as calculated for the vertical tribar pattern of Group #0 and Element #4 for sensor (b), 3 µm pixel size. The majority of the energy is concentrated in the central region of the spatial frequency plane. Figure 8(b) shows the intensity map for the same pattern in Group #3, which is 8 times smaller in size. We see that the energy is much more spread out, since the pattern scales by the same factor of 8 in the spatial frequency domain.

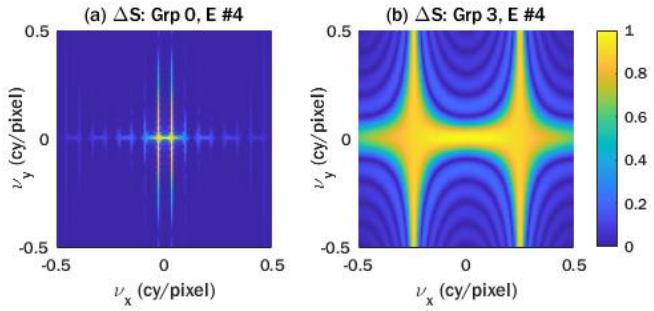


Figure 8. Intensity map of the function ΔS for (a) Group #0, Element #4 and (b) Group #3, Element #4 as calculated for the 3 µm pixel sensor.

Figure 9(a) presents the MTF of the three sensors, and Figure 9(b), shows their SNR as a function of illumination level, where the SNR includes both temporal noise and FPN. We note that the SNR of the 3.75 µm pixel sensor is high for luminances greater than 2 cd/m^2, thanks to its large pixel size, although it does not outperform the 2.2 µm sensor at any brightness level. All three SNR curves are monotonic, and linear on a log-log scale at high brightness. It is clear from the MTF plots, here shown on an absolute frequency scale, that the MTF increases with decreasing pixel size, as expected.

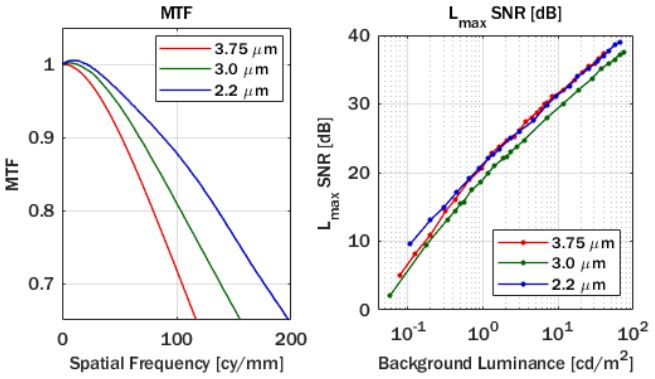


Figure 9. (a) MTF at 550 nm and (b) SNR curves of the three image sensors. MTF was measured using the slanted-edge method, and SNR was measured in transparent region of single frame images of the USAF 1951 target,

Cross-Correlation and SNRI

Figure 10 shows a series of plots comparing the normalized cross correlation with modeled SNRI, both as a function of illumination level, for each of the three pixel sizes. Figure 10 (a) to (d) show data for the vertical tribar pattern of Element #4 in Groups #0 to #3, respectively. Therefore within each graph the size of the tribar pattern on the sensor is the same, and shrinks as we proceed from (a) to (d). Each plot includes a horizontal line at SNRI = 5. This level of SNRI represents the conditions in which 99% of the decisions are correct, which is also the conditions of 5 standard deviations between the means of the true and false positive distributions [1].

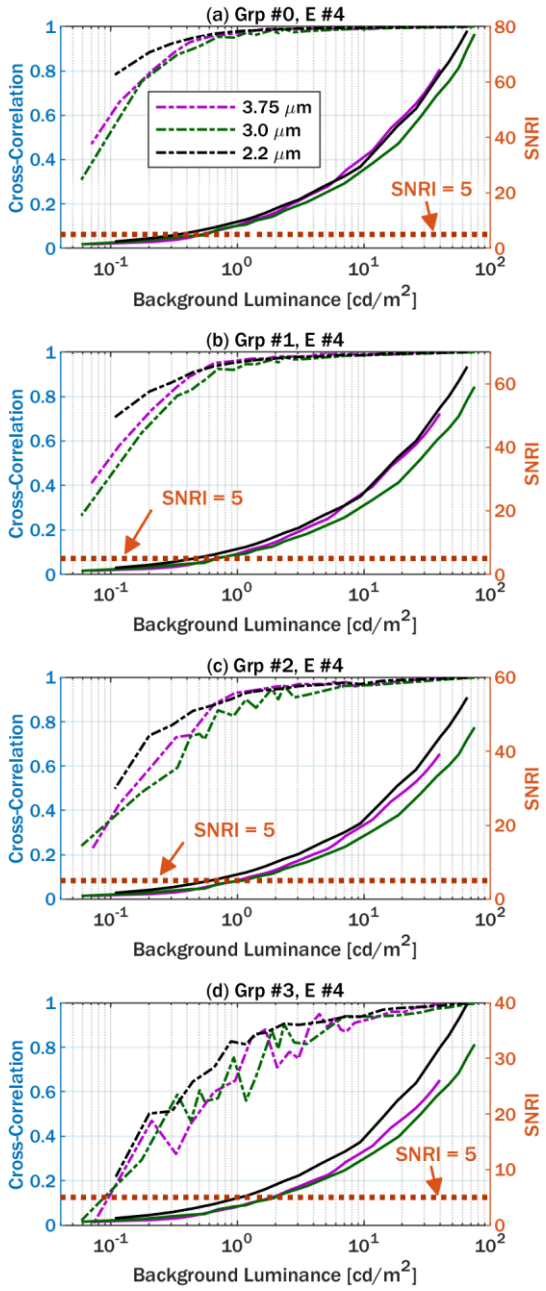


Figure 10. Plots (a) through (d) compare normalized cross-correlation performance and SNRI predictions for the vertical tribar pattern in Element #4 of the 1951 USAF pattern in Groups #0 through #3, respectively. Detection capabilities decrease with pattern size and with scene brightness, which is the background luminance of the target. Both evaluation methods agree that sensor (c), with 2.2 µm pixels, outperforms the other two image sensors and that at bright scene conditions with big pattern size, performance of sensor (a) with 3.75 µm pixels is comparable to that of sensor (c).

One may conclude from Figure 10 that the normalized cross correlation increases with illumination until it saturates. At the same time, the SNRI increases and crosses the threshold level of 5 near the same brightness where the cross correlation reaches about 0.95 in plots (a) and (b). As the pattern size decreases from (a) to (d), this transition point shifts towards higher illumination levels. This means that for a smaller pattern size, a higher SNR level is required to achieve the same probability of detection, which is the expected behavior. It is also clear that in general, sensor (a), the 2.2 µm pixel sensor, outperforms the other two sensors, although sensor (c), the 3.75 µm pixel sensor, approaches the same performance for the largest pattern at bright illumination levels. At high light levels, the increased resolution and lower noise floor of the smaller pixel sensor is not as critical to detection of the presence of an object.

Conclusion

This work demonstrated good agreement between Ideal Observer predictions and correlation statistics in real captured images. This shows that the SNRI metric is valuable as a tool to evaluate relative performance of electronic imaging systems for pattern detection. Furthermore, the SNRI metric can be computed from well-established performance metrics such as SNR and MTF, and is applicable in a wide range of imaging conditions, including those that are mostly relevant to automotive applications.

Acknowledgment

The authors thank Radu Ispasoiu, Bob Gravelle, and James Tornes for their help and advice on this work.

References

[1] International Commission on Radiation Units and Measurements, "Medical Imaging - The Assessment of Image Quality," 1996.

[2] P. J. Kane, "Signal detection theory and automotive imaging," in *Electronic Imaging*, 2019.

[3] *ISO 12233:2017 Photography -- Electronic still picture imaging -- Resolution and spatial frequency responses,* 2017.

[4] J. C. Dainty and R. Shaw, Image Science, Academic Press, 1974, p. 222.

[5] Edmund Optics, "1951 USAF Resolution Calculator," [Online]. Available: https://www.edmundoptics.com/knowledge-center/tech-tools/1951-usaf-resolution/. [Accessed 2 March 2020].

[6] R. N. Bracewell, The Fourier Transform and Its Applications, McGraw Hill, 2000.

Author Biographies

Orit Skorka joined Aptina in 2013 and is now with the Intelligent Sensing Group at ON Semiconductor working on pixel characterization and image quality of CMOS image sensors for automotive, security and other imaging applications.

Paul J. Kane received a B.S. in Physics from the University of Scranton and an M.S. in Optics from the University of Rochester. He was a scientist at the Kodak Research Laboratories for 28 years, working primarily in the areas of imaging science and optics. His projects there included imaging system modeling and simulation, image processing for OLED displays, 3D imaging and light scattering from small particles. In 2015 he joined ON Semiconductor, focusing on automotive and security applications. Mr. Kane holds 35 U.S. patents in the areas of algorithms, display technologies and solid state lighting, and is a Senior Member of SPIE.

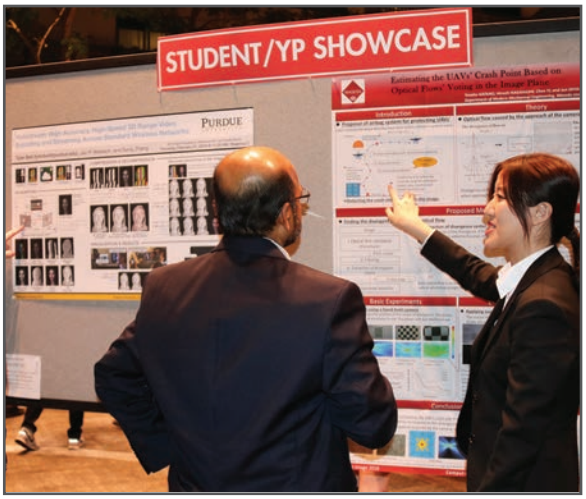
JOIN US AT THE NEXT EI!

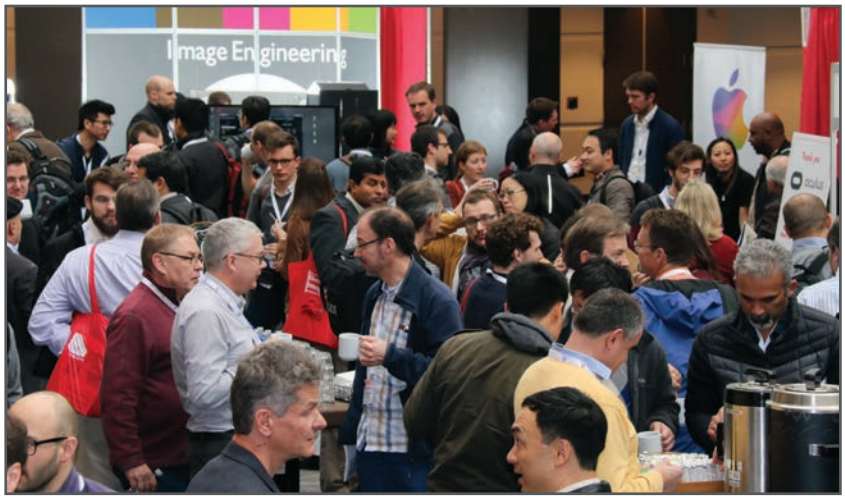
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org