

Imaging Applications of Noise Equivalent Quanta

Brian W. Keelan
ON Semiconductor
San Jose, CA

Abstract

This paper explores how noise equivalent quanta (NEQ) can be estimated conveniently in digital imaging systems, and provides two very different examples of its application. A principal result is that, if an imaging system has a flat noise power spectrum (NPS) prior to down-sampling, and if pre-filtration is used to control aliasing while maintaining reasonable sharpness, the NPS will again become roughly flat after down-sampling, with the magnitude reduced by approximately the square of the down-sampling factor. This result allows the NEQ of a digital system to be conveniently estimated as the square of the product of the capture modulation transfer function (MTF), the linear pixel signal to noise ratio (SNR), the resampling factor, and the pre-filter MTF. Two examples of applications of this approximation are described: (1) understanding likely performance of pedestrian detection algorithms as a function of automotive image sensor properties; and (2) developing a new usability metric for mobile imaging, digital zoom factor, that combines the information contained in the commonly used parameters of megapixels and SNR_{10} .

1. Introduction

There are many cases in imaging where performance is traded off between signal fidelity, which can be measured using modulation transfer function (MTF), and noise, which is best characterized by noise power spectrum (NPS). In such cases, if both MTF and NPS can be determined, then it is useful to calculate a powerful Fourier metric, the noise equivalent quanta (NEQ). This metric combines MTF and NPS to quantify the square of the signal to noise ratio (SNR) as a function of spatial frequency [1a]. Although used heavily and to great advantage in medical imaging optimization [2,3], NEQ has infrequently been applied elsewhere in imaging. Examples of non-medical imaging studies using NEQ include Refs. [4 – 7].

Several barriers to the more widespread use of NEQ exist. First, NPS is less familiar to most people than is MTF, making NEQ a somewhat non-intuitive quantity. Second, although the propagation of MTF through a digital imaging system is well known, that of NPS is less so, with the Doerner equation [8] not explicitly covering sampled systems. Although the extension does not require any new concepts, the aliasing of two-dimensional noise can be hard to picture. I hope that this paper will lower both of these barriers somewhat, through presentation of a small amount of tutorial material, and the derivation of an approximate expression for digital system NEQ.

This paper is organized as follows. Section 2 provides background information on NEQ, including a defining equation and its role in signal detection. Sections 3 and 4 quantify the effect of image resampling on MTF and NPS, respectively. A surprisingly simple relationship is found for resampled NPS, derived under reasonable constraints. Section 5 uses this result to write an approximate equation for NEQ in an imaging system with

resampling, which then allows systems with different numbers of capture pixels to be compared fairly and with some rigor. Section 6 applies this equation to analyze imaging system performance versus distance and light level, as it relates to algorithmic detection of pedestrians. Finally, Section 7 describes a rather different application of the same equation, in which a usability metric, digital zoom factor, is defined and used to predict the optimal number of capture pixels for a fixed optical format.

2. Noise Equivalent Quanta (NEQ)

NEQ is a Fourier metric combining MTF and NPS as shown in Eq. 1 [2].

$$NEQ(v_x, v_y) = \frac{MTF^2(v_x, v_y)}{NPS(v_x, v_y) / \mu^2} \quad (1)$$

Here μ is the mean linear signal and v is spatial frequency, which can have a variety of units, such as cycles per mm in the sensor plane, cycles per pixel (equivalent to cycles per sample), or cycles per degree at the retina. There are two simplifications that occur when choosing to express spatial frequency in cycles per pixel, in which units the monochrome Nyquist frequency is one-half. First, NEQ is more easily interpreted (see next paragraph), and second, this choice yields a simple relationship between noise variance and flat NPS magnitude. Specifically, the noise variance is always equal to the integral of the NPS over all frequencies. If the NPS is flat (constant noise power, independent of frequency, corresponding to white noise), then the variance is equal to that constant noise power times the frequency integral. For cycles per pixel frequency, this integral is trivial; it runs from $-1/2$ to $+1/2$ in x and y , and so has a value of unity. Thus, the NPS magnitude exactly equals the variance, which is convenient.

Returning to Eq. 1, the numerator represents fractional signal modulation squared. By dividing the NPS by the square of the mean signal μ , the denominator becomes fractional noise power, which is fractional noise amplitude squared per unit frequency squared (i.e., per unit area in an x - y frequency plane). Now fractional modulation divided by fractional noise amplitude is just the signal to noise ratio (SNR) at the corresponding frequency, so the NEQ is SNR^2 per reciprocal unit frequency squared. With our choice of frequency units, the NEQ conveniently becomes SNR^2 per pixel area, meaning the SNR^2 that would be achieved by integrating signal over an area equaling one pixel. Had we instead chosen the frequency to be cycles per mm, the NEQ would be SNR^2/mm^2 , the SNR^2 achieved by integrating signal over an area equaling one square mm.

An interesting point is that a linear operation such as convolution with a space-invariant kernel (to sharpen, or denoise) does not change NEQ, because the numerator and denominator are

affected equally, the MTF squared being cascaded with the kernel MTF squared, and the NPS being shaped by that same kernel MTF squared. This accords with the general intuition that such an operation does not fundamentally change image utility. In contrast, SNR as usually calculated (mean signal divided by the root-mean-square noise, with no frequency dependence), will improve markedly with the application of a blurring filter, reflecting the decrease in noise, but not the commensurate decrease in sharpness.

NEQ is closely related to another Fourier metric, detective quantum efficiency (DQE), which is the NEQ divided by exposure in quanta. Because an ideal detector has only Poisson (shot) noise, the highest possible SNR is equal to the square root of the number of quanta, and thus the maximum value of DQE is unity. Although DQE is convenient, being normalized in this way, NEQ is the more relevant quantity when predicting the imaging system performance under a specific set of circumstances (that is, a particular exposure).

Quantification of task performance falls in the realm of signal detection theory, where detectability index, d' , is a commonly employed metric [9]. It is essentially the number of standard deviations of separation between distributions of two outcomes; a high d' corresponds to a high likelihood of detection. NEQ can be used to estimate the d' of a signal having a mean-normalized frequency spectrum, S , via Eq. 2. [3a].

$$d'^2 = \iint NEQ(v_x, v_y) \cdot S^2(v_x, v_y) \cdot dv_x \cdot dv_y \quad (2)$$

This equation and more advanced analogues based on model observers are commonly used to design and to optimize performance of medical imaging instrumentation for particular diagnostic tasks [3b]. A philosophically similar but simpler approach will be used in Section 5 to estimate image suitability for pedestrian detection.

To recap those points that will be used later, with spatial frequency in cycles per pixel, NEQ can be interpreted as the pixel SNR^2 as a function of frequency, and a flat NPS has a magnitude equal to the noise variance.

3. MTF and Resampling

It is commonly the case today that a higher number of pixels are captured than are ultimately displayed or analyzed algorithmically in computer vision. Two means of reducing numbers of pixels need to be distinguished: cropping and down-sampling. Cropping is simply discarding capture pixels to leave a smaller image corresponding to a reduced field of view (which need not be centered with respect to the original field of view). It is the principal operation employed in digital zooming, which emulates the reduced field of view of a longer focal length lens. Down-sampling involves resampling the image to reduce the number of pixels, while maintaining the same field of view. The simplest type of down-sampling is decimation, in which rows and columns of the original image are periodically removed. For example, a 2x decimation could be accomplished by deleting even-numbered rows and columns.

Digital zoom can have great compositional advantages through the removal of distracting and/or irrelevant content and the increased angular subtense of the main subject in the final displayed image. However, the cropping operation does not fundamentally change the quality of the individual pixel data (though it will generally lead to more challenging viewing conditions). In contrast, down-sampling has no effect on

composition (as the field of view is preserved), but it can have a substantial effect on the quality of the image data, by changing sharpness and noise, and introducing aliasing and other artifacts associated with reconstruction error [10a].

For purposes of pixel book-keeping, it is helpful to define the quantities M_c (capture megapixels), M_d (display or final image megapixels), digital zoom factor Z (where $Z = 2$ reduces the number of pixels to $1/4$ of the original number), and resampling factor R (where $R = 2$ down-samples to $1/4$ of the original pixels). These quantities are related by Eq. 3:

$$M_c = M_d \cdot Z^2 \cdot R^2 \quad (3)$$

The effect of linear resampling on MTF is well explained elsewhere [11]. In brief, a resampling operation can be viewed as comprising two steps: (1) an interpolating step that formally produces a continuous image with values at all points within the field of view; and (2) a sampling step that extracts discrete points, creating a new sampled image. The interpolating step involves a filter with an associated MTF, which modifies the system MTF up to that stage by being cascaded with it. The sampling step does not have an associated MTF, but the resampling factor R is essentially a magnification that scales the frequencies of the system MTF.

There are a number of common interpolators, such as linear, cubic spline, bicubic, and replication (sample and hold). Better filters have high response in the pass-band, for good sharpness, and low response in the stop-band, to prevent aliasing (if down-sampling) and reconstruction artifacts. Aliasing is the mapping of high-frequency signal and noise to lower frequencies during down-sampling, producing artifacts and increasing noise [10a]. Conversely, reconstruction error is the mapping of low frequency content to higher frequency, again creating artifacts [10a]. For the best trade-off of sharpness and artifacts/noise, the transition between the pass-band and the stop-band should occur near the Nyquist frequency of the resampled image. Consequently, regardless of resample factor, the MTF of good filters have a relatively modest range of variation of shape when plotted against cycles per pixel in the resampled image plane. This is convenient because it allows us to select a representative filter for computations and expect that similar results would be found for other good interpolators.

Considering the simplest case of a monochrome system prior to spatial image processing, we can now write the system MTF after resampling as:

$$MTF(v_r) = MTF_c(v_r / R) \cdot MTF_r(v_r) \quad (4)$$

where v_r is cycles per pixel in the resampled image plane, MTF_c is the capture MTF (lens plus sensor), MTF_r is the resample MTF, and R is the resampling factor.

4. NPS and Resampling

The behavior of the NPS during resampling is more complex. The capture NPS (prior to image processing operations) in most digital cameras at most light levels is very close to being flat, although exceptions occur when there is significant structured noise and/or at very low signal levels where shaped noise sources (like $1/f$ noise) may not be washed out by shot noise, which has flat NPS. We hereafter assume that the capture NPS is indeed flat. Fig. 1 shows the changes that occur in the NPS during resampling.

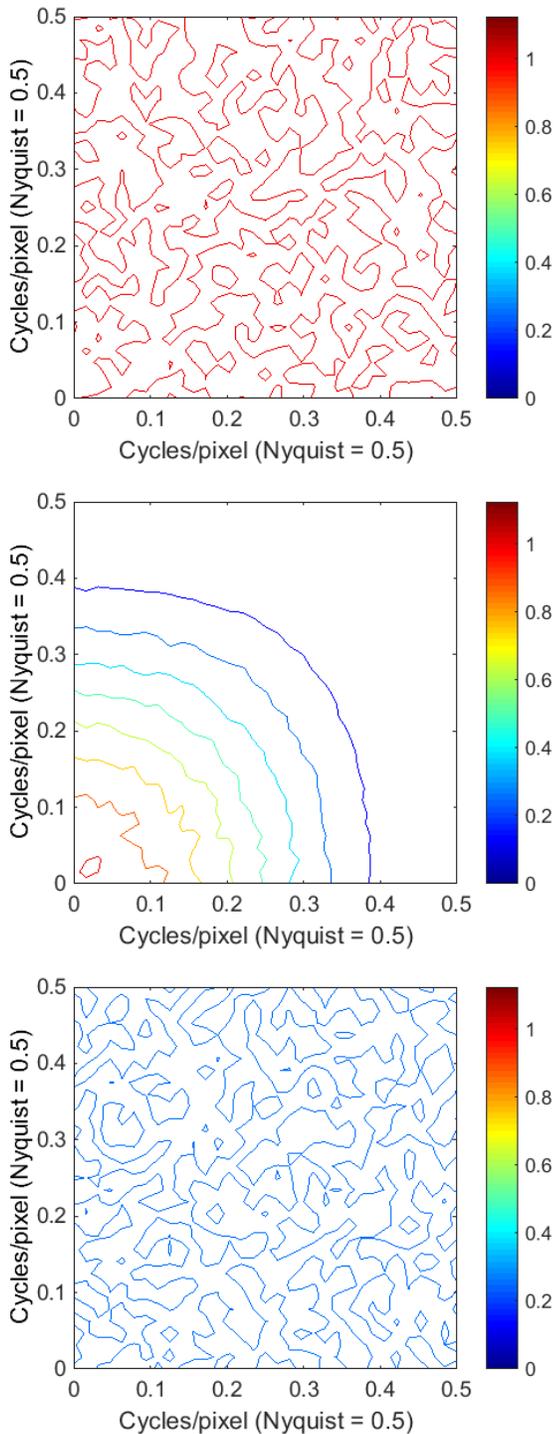


Figure 1. Contour plots of two-dimensional noise power spectra show the effect of down-sampling on noise. The x- and y-axes run from zero to Nyquist (capture Nyquist in the top and middle plots, resampled Nyquist in the bottom plot). The three color codings of z-data are the same in the three plots. Top: The NPS is initially flat, with only small fluctuations about the mean power, which has a value of 1.00. Middle: After digital low-pass pre-filtering, the NPS is peaked and band-limited, with most high-frequency noise suppressed to mitigate aliasing. Bottom: After 2x decimation, the NPS again is flat, with a mean power of ≈ 0.25 , about $\frac{1}{4}$ as much power as initially. Thus, the effect of an ideal resampler on white noise is approximately to reduce noise power by the down-sampling factor squared, without changing the NPS shape.

Each of the three contour plots have the same x and y axes, running from zero to Nyquist frequency, and the same color coding over the same range of z-data, from which the contours are drawn. The top plot shows the NPS in two dimensions, computed from a synthetic image. The NPS is initially flat, with only minor random fluctuations about the mean noise power value of 1.00 (red on the color bar), arising from the 4% standard error in NPS measurement expected for the 1024 synthetic noise realizations [1b]. The middle plot shows the NPS after a simple interpolation filter, corresponding to a 2x average-down operation, has been applied. The NPS is now more or less bell-shaped (the plot shows only one of four symmetric quadrants), being peaked at low frequency, and dropping close to zero at the highest frequencies, which have been very effectively smoothed. It can be seen that much of the power below the resampled Nyquist (0.25 units because of the 2x sample-down) has been retained, as desired in the stop-band, whereas most of the power in the stop-band has been suppressed, also as desired. Better filters, such as bicubic, would have a somewhat sharper transition from pass-band to stop-band.

The bottom figure shows the NPS after the sampling step. The Nyquist frequency now corresponds to the resampled image, rather than the capture image, as in the preceding two plots. The sampling operation creates spectral replicates of the bell-shaped NPS centered at all integer pairs of cycle per pixel frequencies, and this grid of bells adds up to produce the final noise distribution. Referring to the middle plot of Fig. 1, and ignoring smaller contributions from more distant spectra, this is like adding inwardly pointing bell quadrants at the top left, top right, and bottom right corners of the plot. Remarkably, though composed of a sum of bell-shaped noise power spectra, the final result is again essentially flat. Essentially, the noise aliasing during sampling has filled in the higher noise frequencies that were suppressed by the filtering operation. But with the change in frequency to resampled cycles per pixel in the bottom plot, the mean power value is reduced to approximately 0.25 (blue on the color bar), a factor of R^2 lower than the starting noise power. In summary, if the NPS before resampling is fairly flat, then the NPS after resampling can be approximated by the NPS before resampling, divided by the square of the resampling factor, a very convenient result!

5. NEQ After Resampling

In this section we combine results from the previous three sections to write a simple equation for estimating NEQ after resampling. As mentioned in Section 2, when frequency is in cycles per pixel, a flat NPS will have a constant noise power equal to the noise variance. Assuming this to be true, the denominator of Eq. 1, which is fractional noise power, becomes equal to the variance divided by the square of the mean. This is exactly $1/\text{SNR}^2$, where SNR is measured in the usual fashion, namely, as mean raw sensor signal divided by the root-mean-square noise, with no frequency dependence. Combining this result, Eq. 1, Eq. 4, and the conclusion from the preceding section (resampled NPS \approx original NPS divided by the resample factor squared), finally yields the desired equation:

$$NEQ(v_r) \approx R^2 \cdot \text{SNR}^2 \cdot \text{MTF}_c^2(v_r/R) \cdot \text{MTF}_r^2(v_r) \quad (5)$$

Eq. 5 is quite informative. As mentioned earlier, good quality reconstruction MTFs do not vary widely when expressed in the resampled frequency space, so it is the first three terms that cause the principal variations in NEQ. SNR is universally used in the

digital imaging industry and could be considered the baseline metric, from which NEQ represents an improvement. In that light, the principal extensions to SNR in Eq. 5 are the squares of the resample factor and the capture MTF at rescaled frequency.

The former addresses a first key SNR limitation, namely, that pixel SNR is not suitable for comparison of sensors having different numbers of pixels, without further adjustments. For example, at equal optical format and technology, lower megapixel sensors will always have better pixel SNR, but are likely to yield lower final image quality (discussed further in Section 7). The R^2 term allows sensors with different numbers of pixels to be resampled by different factors to yield the same number of pixels for display or analysis, allowing rigorous comparison.

The second extension to SNR, the square of the capture MTF, addresses a second limitation of SNR, namely, that pixel SNR can be improved to arbitrarily high values by averaging operations such as convolution with a blur filter. As mentioned in Section 2, NEQ is unaffected by such operations if they are linear, because the filter MTF is cascaded with the other components of the capture MTF (which includes everything prior to resampling).

The two terms above also reflect the two ways in which down-sampling improves NEQ. As seen in the argument to the capture MTF, down-sampling reduces the frequencies in the capture plane that “matter”. As MTFs generally decrease with increasing frequency, this translates to improved modulation transfer and NEQ at a given frequency in the resampled image plane, which could be the display plane for visual applications or the analysis plane for a computer vision algorithm.

The R^2 term represents noise improvement, arising from the filtering step having an effect similar to the averaging of R^2 pixels. This degree of averaging provides a good balance between sharpness, noise, aliasing, and other artifacts. This statement is equivalent to the previous assertion that a good interpolator should have an MTF that transitions from the pass-band to the stop-band around the Nyquist frequency of the resampled image. It is also equivalent to the geometrical view that the spatial extent of the filter kernel should be similar to that of one resampled image pixel, projected back to the capture plane.

Before proceeding to the two sample applications of NEQ, it is appropriate to review the primary underlying assumptions of Eq. 5: (1) the original NPS is flat (corresponding to white noise), and (2) the interpolator filter has a “reasonable” response, in which the transition from pass-band to stop-band is approximately centered on the Nyquist frequency of the resampled image. The latter assumption can be violated but should be met in imaging pipelines of reasonable quality. The former assumption can be violated at very low light levels or in sensors with significant amounts of structured (fixed pattern) noise. Furthermore, in color sensors, even the luminance NPS is likely to be shaped somewhat by the demosaic operation, especially if it involves adaptive denoising. Despite these limitations, Eq.5 is conceptually very useful, and when comparing systems, estimation errors are likely to be quite correlated, so that predicted differences will be fairly accurate.

6. Example 1: Pedestrian Detection

Computer vision in automotive applications is currently an area of intense activity. Initiatives related to safety ratings in Europe particularly are bringing sensor performance in difficult conditions under scrutiny. Our first application of NEQ is to the problem of algorithmic pedestrian detection, which could provide input to warning, headlight steering, and braking systems. We expect that the computer vision efficacy would depend

significantly upon the sensor signal fidelity and noise, and so anticipate that NEQ could provide a useful predictor of algorithmic performance. (As a contrived example of when this would not be true, if all pedestrians were mauve, color reproduction accuracy might be more important than spatial image characteristics.) Signal fidelity and noise would be strongly dependent upon distance to the pedestrian (affecting pixels subtending the pedestrian) and light level (affecting SNR). The most rigorous treatment of the problem would be to use Eq. 2 to predict detectability by cascading NEQ (at different distances and light levels) with the square of the signal (pedestrian) spectrum and integrating.

Not having a robust estimate of the signal spectrum, we will instead perform a single-frequency analysis, at the highest frequency of importance in one published pedestrian detection algorithm [12], which was 5 cycles per pedestrian height. We choose to focus on the highest frequency of importance because that is where the NEQ will be lowest and so most limiting. To capture 5 cycles per pedestrian, we will need at least 10 pixels vertically subtending the pedestrian. Most algorithms performing object detection will search at a number of scales to find the objects at different distances, so it is reasonable to assume that a resample factor has effectively been selected by the algorithm to produce a region of interest on the appropriate scale for analysis. As already stated, this could correspond to a situation in which the pedestrian were vertically subtended by as few as 10 pixels (to meet the Nyquist criterion for 5 cycles per pedestrian). However, it could be advantageous to have a larger number of subtending pixels, which would improve the modulation transfer of 5 cycles per pedestrian, but would reduce pixel SNR. The best resample factor would depend upon the details of the algorithm, but we will proceed assuming that a good compromise would be to resample so that the pedestrian is subtended vertically by 20 pixels. In most cases this will involve down-sampling, but if the pedestrian were far enough away, it could require up-sampling.

NEQ at 5 cycles per pedestrian, in an image bicubically resampled so that the pedestrian subtended 20 vertical pixels, was computed for a range of light levels and distances. The latter, with camera lens focal length (3.1 mm), pixel pitch (3.75 and 3 μm), and pedestrian height (1.75 m), determined capture pixels vertically subtending the pedestrian, in turn implying the needed resampling factor R . This calculation was repeated for real 1-megapixel and 2-megapixel sensors having the same optical format (1/2”) and using the same automotive grade camera lens. Parameters for which representative values were assumed included integration time (66 ms), pedestrian reflectance (20%), and camera lens f-number ($f/2.2$) and transmittance (90%). Pixel SNR was computed from real sensor data for pixel sensitivity (electrons per lux-s), linear full well, and noise floor (read noise, dark current shot noise, etc.). The two sensors shared very similar technology so this comparison was realistic but was not biased by factors unrelated to the pixel pitch per se. Lens MTF was approximated using the standard diffraction-limited formula, with the f-number increased 10% above the actual f-number, emulating some degradation from aberrations. Sensor MTF was estimated based on pixel size, microlens fill factor, crosstalk, and sampling phase MTF. The latter reflects the random alignment of the pixel sampling grid with signals in a scene [13]. For example, the modulation transfer of a sine wave at Nyquist can vary from good (peaks and troughs align with pixel centers) to zero (they fall on the pixel boundaries). The sampling phase MTF is a sinc function of the pixel pitch and should be included when modeling capture MTF or using super-sampled slanted edge MTFs.

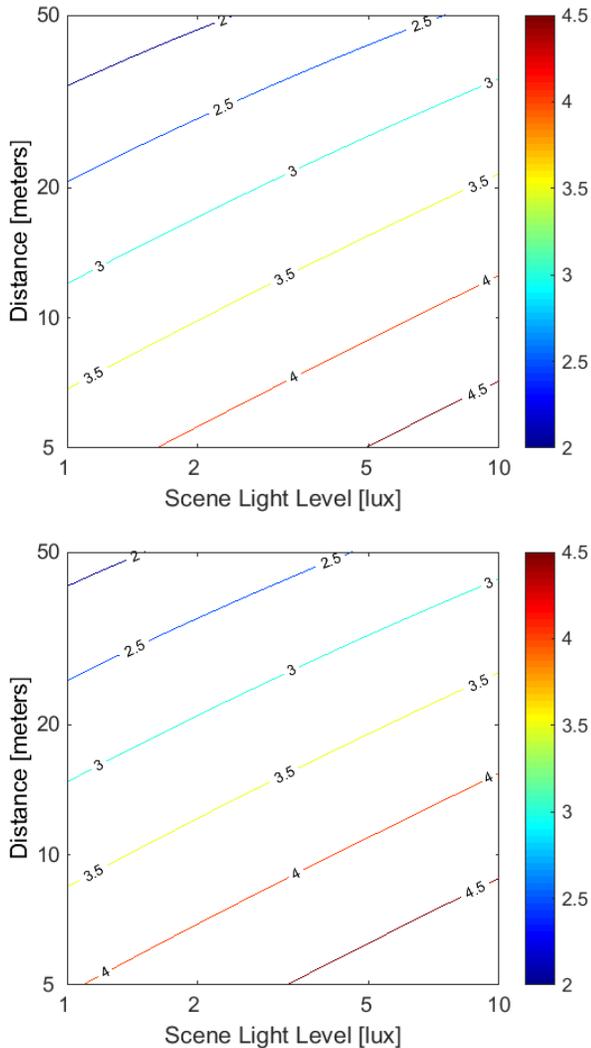


Figure 2. Contour plots of the common logarithm of noise equivalent quanta at the key frequency of 5 cycles per pedestrian, for 1-megapixel (top) and 2-megapixel (bottom) sensors with equal optical format. Good algorithmic detection of pedestrians is likely for values above approximately 2.5 log units. Performance drops at long distances (upper parts of plots) because there are not enough pixels subtending the pedestrian. Performance drops at low light levels (left parts of plots) because SNR is reduced. At constant optical format, the higher megapixel sensor provides slightly better results because of the increased number of pixels subtending a pedestrian.

Results are shown in Figure 2, with the top figure corresponding to the 1-megapixel sensor, and the bottom to the 2-megapixel sensor. The x-axes are logarithmic light level from 1 to 10 lux, corresponding to very low illumination. The y-axes are logarithmic distance from 5 to 50 meters. The contours show the common logarithm of NEQ. As expected, the lowest NEQ values occur at low light levels (producing low SNR) and long distances (producing poor MTF because of the small number of pixels subtending the pedestrian), which map to the top left corners of the plots.

Although we do not have algorithm performance data against which to “calibrate” the NEQ, it is plausible that an algorithm would perform adequately under conditions yielding good visual

image quality. A mid-tone linear SNR of 40 (32 dB) is often assumed to enable fine pictorial quality, and typical mid-frequency system modulation transfer for a stationary capture with good focus might be about $\frac{1}{2}$. So an NEQ of $(40 \times \frac{1}{2})^2 = 400$ would seem sufficient for good algorithmic performance, although adequate performance might be possible at substantially lower values. However, taking this as a guideline, we’d anticipate that a $\log_{10}(\text{NEQ})$ of 2.6 would support viable pedestrian detection, so most of the (logarithmic) space mapped in the figures would be covered.

Comparing the top and bottom plots, it is seen that the 2-megapixel sensor performs slightly better, providing coverage to longer distances and lower light levels. The reason for this result is that the 2-megapixel sensor always has $\sqrt{2}$ more pixels subtending the pedestrian, providing an MTF advantage, because the pixel aperture, essentially a point spread function, is always smaller when projected onto the pedestrian. Although the pixel SNR is poorer, after resampling to the same number of pixels in the resampled image used for algorithmic analysis, the SNR of the two sensors are very close, implying that the pixel performance approximately “scales”. This is a good achievement but not too surprising at the large pixel sizes involved (3.75 and 3 μm). We will see a different behavior with smaller pixels in the next example.

7. Example 2: Digital Zoom Factor

The two most common parameters cited in connection with sensor capabilities for mobile applications are, at the point of sale, number of capture pixels (hereafter, just “megapixels”), and in discussions between sensor vendors and handset manufacturers, SNR_{10} , the light level in lux at which a linear SNR of 10 is achieved [14]. Although both can be correlated with image quality, they each have major shortcomings. SNR_{10} is only useful for making comparisons between sensors having the same number of pixels. Because SNR_{10} is a capture pixel-based metric, taken literally, it predicts that the optimal number of pixels is always exactly one, which would produce a sensor that was a bit deficient in resolution. Conversely, capture megapixels provides a rigorous metric for comparison only when comparing sensors using the same pixel design. Taken literally, it implies that the best number of pixels is always the maximum number that can be manufactured. An example of a question that cannot be answered by either SNR_{10} , or megapixels, is the optimal number of capture pixels for a given optical format, a question of significant practical interest.

The purpose of this section is to use NEQ to define a metric that combines some of the content of SNR_{10} and megapixels, so that the above “optimal number of pixels” question can be answered. It is desired that the metric be something that can be understood by end consumers, and so possibly useful at the point of sale. One way of achieving this is to have the metric relate to a particular use case of interest. An example of such a benchmark metric is enlargeability factor, which is the maximum optical magnification to which a frame of film can be enlarged, without the overall image structure quality falling below a threshold [10b]. This metric could tell an end consumer how large a print could be made from a particular size of negative, if well exposed.

The metric proposed here, digital zoom factor, has some similarities to enlargeability factor. Mobile devices are desired to be thin and so are heavily constrained with respect to optical zoom, which creates a distinct difference from the single lens reflex experience. A primary, arguably the primary, benefit of increased

megapixels is the possibility of supporting higher levels of cropping for compositional improvement, as discussed in connection with, and quantified by, the digital zoom factor Z of Eq. 3. Physically, digital zoom factor can be interpreted as follows: if the user places their fingertips on a touchscreen or touch-pad, and then spreads them apart to zoom into an image, until the quality is just acceptable, then Z is the ratio of their final to initial separation. This metric could be explained to end users with a simple graphic, which relates to a use case that is likely of significant relevance to them.

Because digital zoom factor relates quality to degree of zooming, the important factors to incorporate into the metric are those that vary with zoom. This includes those factors that are frequency-dependent, such as sharpness and noise, but excludes those that are not, such as color and tone reproduction. Given the sharpness and noise dependence, NEQ could provide a reasonable basis for determining the quality threshold. As discussed in the previous section, a final (resampled) image NEQ of around 400 might correspond with a pretty high quality threshold, and that value is used in the analyses that follow.

As in the previous section, a single frequency was analyzed for simplicity. In this example, there is no single scene component determining that choice (viz., a pedestrian), so instead the choice was made based on display capability. Specifically, the NEQ was computed at the mid-frequency of one-half the resampled image Nyquist frequency, an arbitrary but not unreasonable choice. Representative mobile imaging values of input parameters were assumed; these included: light level (500, 50, and 5 lux); scene reflectance (20%); lens aperture ($f/2.1$), transmittance (90%), and MTF (like $f/2.3$ diffraction-limited lens); exposure criterion (ISO saturation exposure if not limited by maximum integration time of 66 ms); display addressability ($1080p\ 16:9 = 2.1$ megapixels); resampling (bicubic); and optical format ($1/2.3''$, $1/3''$, and $1/4''$). Realistic estimates were used for sensor MTF and pixel performance (linear full well, noise floor, and sensitivity) versus pixel size, for a similar pixel technology level.

To find the digital zoom factor for a given case, Eq. 5 was solved iteratively for the value of resample factor R that produced an NEQ of 400 at display half-Nyquist. That value of R was then used in Eq. 3 to determine the digital zoom factor Z . Before considering the question of optimal megapixels at a fixed optical format, the predictions were tested in cases where the correct behavior was known. Table 1 shows Z values for the three optical formats at the three light levels. As must be true, Z decreased (corresponding to lower quality) as photosensitive silicon area decreased (left to right) or light level decreased (top to bottom). Z values of less than one correspond to cases where the NEQ = 400 threshold could not be achieved even without cropping; in fact, the imagery would actually have to be minified (down-sampled) to meet the NEQ criterion, producing a Z value below unity.

	1/2.3"	1/3"	1/4"
500 lux	2.31	1.79	1.35
50 lux	1.23	0.96	0.72
5 lux	0.40	0.31	0.23

Table 1. Digital zoom factor at constant pixel size ($1.1\ \mu m$) for three optical formats (columns) and three light levels (rows). As expected, digital zoom factor (and spatial quality) decreases with decreasing photosensitive silicon area (left to right) and decreasing light level (top to bottom). This is essentially a sanity check that digital zoom factor behaves sensibly in well-understood cases. Values below unity imply that an image does not clear the NEQ threshold even if no cropping is done; rather, minification of the image would be needed to meet the NEQ criterion.

	1.75 μm	1.4 μm	1.1 μm	0.875 μm
500 lux	1.87	2.10	2.31	2.47
50 lux	1.20	1.22	1.23	1.22
5 lux	0.44	0.42	0.40	0.37

Table 2. Digital zoom factor at constant optical format ($1/2.3''$) for four pixel sizes (columns) and three light levels (rows). At high light level (500 lux), digital zoom factor increases as there are a larger number of smaller pixels (left to right) because of improved MTF. At low light level (5 lux), the opposite trend is observed, because the noise floor does not scale down with pixel size, and becomes a larger fraction of the diminished signal (which does approximately scale with pixel area). This effect more than compensates for the MTF advantage of the smaller pixels, which is still present. At intermediate light level (50 lux), the MTF and noise floor effect just about cancel out, so the digital zoom factor is essentially independent of pixel size.

Table 2 summarizes the results for the much more interesting case of optimal number of pixels for a fixed optical format, in this case, $1/2.3''$. The three rows are the same as in Table 1, representing high, medium, and low light levels. The columns correspond to four pixel sizes. Consider first the 500 lux results. Digital zoom factor increases with decreasing pixel size, which corresponds to a larger number of smaller pixels. The improvement here is for essentially the same reason as identified in the pedestrian detection case – as pixel size decreases, MTF improves. If the pixel performance “scales”, the resampled image SNR is approximately independent of pixel size, so there is a net benefit to increasing megapixels. However, at 5 lux, the opposite trend is observed: digital zoom factor decreases (slowly) as pixel size decreases. The MTF advantage of smaller pixels is still present, but it is more than compensated by the increased impact of the noise floor. As pixel size decreases, noise floor does not usually decrease proportionally (at least for smaller pixels), whereas sensitivity can vary almost proportionally. Thus, with decreasing pixel size, signal can decrease faster than noise floor, resulting in lower SNR and NEQ. At 5 lux, that is the dominant effect. Finally referring to the 50 lux results, the MTF and SNR effect almost perfectly cancel, leaving hardly any change in digital zoom factor over a factor of two change in pixel size.

So the optimal number of pixels for a given optical format is dependent upon light level, with a smaller number (larger pixel size) being optimal at lower light levels, and with the optimal number increasing as noise floor decreases.

8. Conclusion

A simple equation for estimating NEQ from SNR, capture MTF, resample factor, and resampling MTF (Eq. 2) is derived. It is approximately valid when capture NPS is reasonably flat (representing white noise) and the resample filter has a transition from pass-band to stop-band near the Nyquist frequency of the resampled image (which provides a good balance between sharpness, noise, aliasing, and other artifacts). NEQ permits rigorous comparison of sensor performance between sensors having different numbers of pixels, and provides a significant improvement over traditional SNR for evaluating spatial quality of digital imaging systems. Two very different examples of the application of NEQ are described, one regarding the computer vision task of pedestrian detection, the other involving the definition of digital zoom factor, a new usability metric of potential utility in mobile photography. The latter metric quantifies the degree of cropping possible from a camera, without falling below a spatial quality threshold.

Acknowledgments

The author would like to thank Yan Zhai, Rui Shen, Robin Jenkin, Paul Kane, and Michael Brading for helpful discussions.

References

- [1] J. C. Dainty and R. Shaw, *Image Science*, London, Academic Press, 1974. (a) Ch. 5. (b) Ch. 8.
- [2] I. A. Cunningham, "Applied Linear Systems Theory," In J. Beutel, H. L. Kundel, and R. L. Van Metter (eds.), *Handbook of Medical Imaging: Volume 1, Physics and Psychophysics*, Bellingham, SPIE Press, 2000. Ch. 2.
- [3] ICRU Report 54, *Medical Imaging – The Assessment of Image Quality*, Bethesda, MD, International Commission on Radiation Units and Measurements, 1996. (a) Eqs. 3.18 and 3.10. (b) Appendix E.
- [4] P. Burns, "Signal-To-Noise Ratio Analysis of Charge-Coupled Device Imagers", in Proc. SPIE Vol. 1242, p. 187, 1990.
- [5] E. Tsiang, J. Cole, M. DeWeert, A. Sparks, D. Yoon, J. Fisher, G. Sawai, and T. Glover, "Application of a figure-of-merit for optical remote sensors to an airborne hyperspectral sensor", in Proc. SPIE Vol. 5546, pp. 49–60, 2004.
- [6] J. Mendez, S. Balzer, S. Watson, R. Reich, and D. O'Mara, "A Multi-Frame, Megahertz, CCD Imager", in IEEE Transactions on Nuclear Science, Vol. 56, Issue 3, pp. 1188–1192, 2009.
- [7] B. W. Keelan, "Objective and subjective measurement and modeling of image quality: A case study", Proc. SPIE Vol. 7798, p. 779840, 2010.
- [8] E. C. Doerner, "Wiener Spectrum Analysis of Photographic Granularity", J. Opt. Soc. Am. Vol. 52, Issue 6, pp. 669 – 672, 1962.
- [9] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Los Altos, CA, Peninsula Publishing, 1966. p. 60.
- [10] B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction*, New York, Marcel Dekker, Inc., 2002. (a) Ch. 18. (b) Ch. 12.
- [11] G. Wolberg, *Digital Image Warping*, Washington, DC, IEEE Computer Society Press, 1990.
- [12] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian Detection Using Wavelet Templates", in Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition, Puerto Rico, IEEE, 1997.
- [13] D. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems*, Bellingham, WA, SPIE, 2001. Sec. 2.2.2.
- [14] J. Alakarhu, "Image sensors and Image Quality in Mobile Phones," in Proceedings of the International Image Sensor Workshop, Ogunquit Maine, 2007.