Using visible SNR (vSNR) to compare image quality of pixel binning and digital resizing

Joyce Farrell[a], Mike Okincha[b], Manu Parmar[ac], and Brian Wandell[ac]
[a]Dept. of Electrical Engineering, Stanford University, Stanford, CA. 94305 U.S.A.
[b]Omnivision Technologies, Santa Clara, CA 95054 U.S.A.
[c]Dept. of Psychology, Stanford University, Stanford, CA. 94305 U.S.A.

ABSTRACT

We introduce a new metric, the visible signal-to-noise ratio (vSNR), to analyze how pixel-binning and resizing methods influence noise visibility in uniform areas of an image. The vSNR is the inverse of the standard deviation of the S-CIELAB representation of a uniform field; its units are 1/ΔE. The vSNR metric can be used in simulations to predict how imaging system components affect noise visibility. We use simulations to evaluate two image rendering methods: pixel binning and digital resizing. We show that vSNR increases with scene luminance, pixel size and viewing distance and decreases with read noise. Under low illumination conditions and for pixels with relatively high read noise, images generated with the binning method have less noise (high vSNR) than resized images. The binning method has noticeably lower spatial resolution. The binning method reduces demands on the ADC rate and channel throughput. When comparing binning and resizing, there is an image quality tradeoff between noise and blur. Depending on the application users may prefer one error over another.

Keywords: sensor design, image quality, pixel binning, imaging pipeline

INTRODUCTION

Modern image sensors increasingly serve two functions: high quality still image capture and fast video capture. Some image sensors accomplish these different tasks by switching between two operating modes. In the still image mode, the sensor captures single frames at high resolution. In the video mode, the sensor captures frames at lower spatial resolution and higher temporal frame rates. This video mode can also be used to provide data for the camera viewfinder.

The engineering requirements for high-speed image capture differ significantly from still image capture. One important difference is that the video stream requires a high-sustained-throughput channel to carry sensor data. Transferring sensor data quickly can be energy intensive and require high channel capacity. When operating in video mode it is desirable to optimize performance by reading out only the amount of data needed by the target display system. A second difference is that video operation has a modest upper bound on the exposure duration. Thus, it is important to optimize the quantum efficiency of the sensor and signal-to-noise under moderate and low light level conditions.

Pixel-binning is a way to improve image sensor video performance. This method combines data from nearby sensor photo-sites prior to analog-to-digital conversion and read-out. By combining the charge prior to read-out, the signal is increased, the read-out noise is reduced and the ADC rate is reduced. Further, the output channel rate is decreased. Pixel binning improves low light sensitivity and high speed video capture in exchange for image spatial resolution.

Image resizing is an alternative to pixel binning. Resizing is typically implemented by filtering and sub-sampling the digital data. Some of the positive effects of pixel binning can be obtained by digitally resizing the data after readout and quantization. One advantage of digital resizing is that the filter can be optimized to match the resolution to the specific target display. A second advantage is that the high resolution pixel data remain available for storage and future use. The disadvantage of resizing method is that the demands on the analog-to-digital converter (ADC) and channel throughput remain high.

In this paper we analyze how pixel-binning and resizing methods affect system noise and spatial resolution. To evaluate the effects on noise we introduce a new metric: visible signal-to-noise ratio (vSNR). This measure evaluates the combination of sensor data, display processing, and viewer sensitivity to assess image noise. The vSNR metric extends conventional measures of image noise, such as SNR and pSNR, which do not account for display rendering or viewing circumstances. To evaluate sharpness we use the modulation transfer function. This measure can be computed using the methods defined in the ISO 12233 standard [1, 2].

Image sensors operate over a wide range of conditions and there are many possible sensor parameters. To understand how these factors combine to influence image-quality we simulate the imaging pipeline using the Image Systems Evaluation Toolbox (ISET) [3, 4] and focus the analysis on the effects of scene luminance, sensor pixel size, read noise and viewing distance.

METHODS

System architectures

We assess quality of two imaging methods (Figure 1). We simulate a specific hybrid pixel binning method described by Eastman Kodak [5] that is implemented in a 4T-pixel architecture. In this method, the charge is summed from two nearby pixels in the same color channel and column prior to readout. This step reduces the column spatial resolution of the output image so that only half as many pixels values are read-out. After quantization, the digital values from nearby pixels in the same color channel and row are averaged. These operations occur on the imaging sensor (Figure 1A).

The second method, digital image resizing, is implemented in a digital signal processor after the pixel data are read from the sensor (Figure 1B). There are many different image resizing algorithms [6, 7]. Here we use a filter and sub-sampling algorithm to digitally resize an image. The digital averaging improves SNR, but the data contain twice as much read-noise as the hybrid method.

Simulation Methods

We simulate the imaging pipeline using a set of Matlab routines called the Image Systems Evaluation Toolbox (ISET) [3, 4, 8]. The simulation software is designed to trace a physically accurate representation of the signals starting with the original scene and tracing through the optics, sensor and image processing to a final image.

Scene. A scene is a multidimensional array describing the spectral radiance (photons/sec/nm/sr/m^2) at each pixel in the sampled scene. The simulated scenes are Lambertian, with no directionally selective rays. Synthetic scenes of a uniform field are used to compare the visibility of noise. Synthetic scenes of a slanted bar are used to compare the sharpness of the image. A sweep frequency target is used to illustrate the combined effects of blur and noise; a calibrated spectral image whose contents include people and colored objects is used to provide a subjective impression of the processing. The mean scene luminance was varied for different computational experiments.

Optics. The scene radiance data are converted to optical image irradiance data at the sensor surface. The simulated lens is diffraction-limited (f# = 2.8; focal length = 3 mm). The simulations are carried out for a region near the center of the image, so that the effect of the lens is calculated as a wavelength dependent point spread function derived from a diffraction-limited shift-invariant model. In addition to the lens, we assume that there is an optical diffuser on the surface of the sensor. These diffusers are used for infrared cutoff and to prevent aliasing. We assume the diffuser has a point spread function of 1 um (full-width, half-maximum).

Sensor. The simulated sensor has a conventional RGBG Bayer color filter array (CFA). Pixel size, read noise, and other properties are varied. A typical pixel size, used in many of the simulations, is 1.4 microns with a well capacity of 6500 electrons. We assumed the fill factor is 95%, representing the effects of a microlens that focused 95% of the light on the photosensitive region of the pixel. Voltage swing is assumed to be 1.8 volts and conversion gain is calculated by dividing voltage swing by well capacity.

Other sources of sensor noise (dark voltage, DSNU, and PRNU) are zeroed so that at low light levels the main source of noise is read noise and at high light levels shot noise is the main source of variance. Exposure duration is fixed at 15 milliseconds to emphasize the comparison with images obtained at video frame rates (60Hz video).

Image Processing and Rendering. The image processing pipeline consists of (a) bilinear demosaicking, performed separately in each color channel (b) color correction, (c) image resizing (blurring and decimating). Display size is varied, but display resolution is fixed at 100 dpi.

Simulation. The image quality metrics we report do not require simulating a large field of view. The vSNR can be calculated from a small uniform field, and the MTF is measured using the ISO 12233 method that only requires identifying the response near the edge of a slanted bar. Hence, in most cases we simulate the central five degrees in the visual field of view for computational efficiency. When illustrating the natural image, we perform the simulations on a larger field of view and simulated sensor (Figure 10).

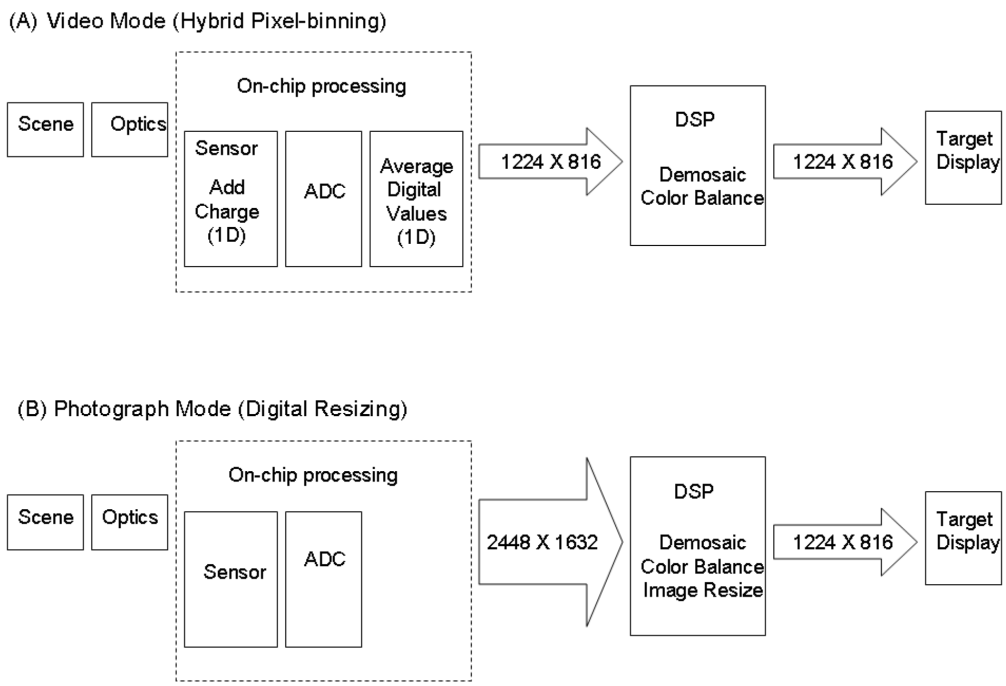

Figure 1. Comparison of the binning and resizing pipelines. (A) The binning method combines the charge from two nearby row pixels with a common color filter, converts the data to digital form, and then averages the digital values from another two nearby column pixels. This binning, which is a hybrid of analog and digital combination, takes place on the sensor. (B) The resizing method illustrates a conventional sensor with the resizing managed after the data are off the chip and in a digital signal processor (DSP). The binning method reduces the ADC rate and the channel output rate compared to the resizing method. In the simulations below we assume the data from both methods are displayed on a common output display

Image Quality Metrics

We use two image quality metrics to assess whether the differences between these two methods are likely to lead to significant differences in image quality, and if so under which imaging conditions. These metrics assess two important elements of image quality: the visible noise and the image sharpness. These metrics are clearly relevant to the objectives of the pixel binning architecture.

Perceived image quality depends on the conditions of the original scene, the capture system, and the viewing conditions. For example, the same image data rendered for a small, handheld mobile device and a large flat panel LCD television will have different image quality. At close distances chromatic noise may be highly visible, but this noise becomes invisible as viewing distance increases. Also, the effect of a system component can vary with imaging conditions. A factor such as read noise may be significant at low illumination levels, but unimportant in bright scenes where the read noise is a small fraction of the photon noise.

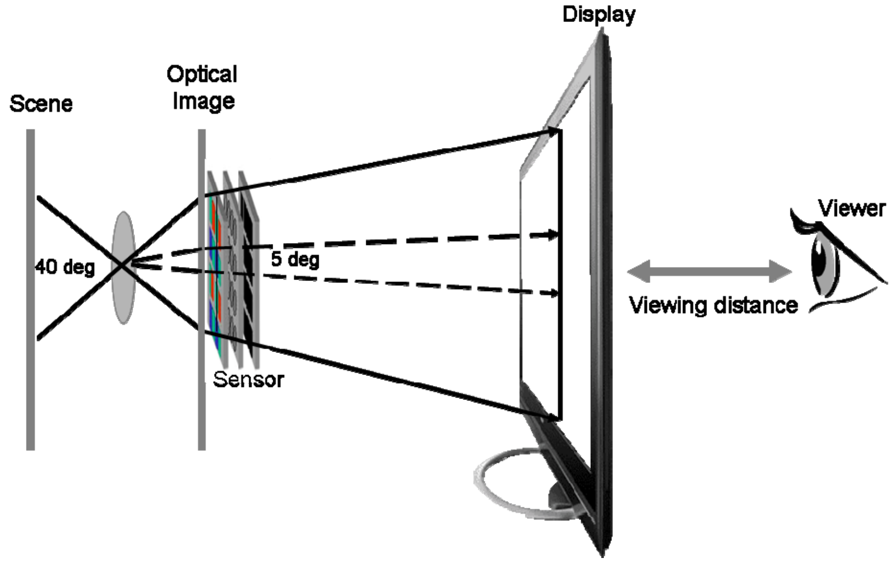

Figure 2. Simulation overview. The scene is assumed to be in a single depth plane 1m from the camera. In most simulations the mean scene luminance is 33 cd/m^2. The lens is diffraction-limited, F#2.8, 3mm focal length. The sensor has a Bayer color filter array with RGB filters found in commercial cameras. The sensor data are shown on a display with 100 dpi spatial resolution seen by a viewer at 0.5 m. The simulated images are equated prior to display for mean luminance. The size of the image on the target display is equal to the demosaicked binned sensor output; hence, the conventional output is resized to have one-fourth the number of pixels in the demosaicked image.

The list of factors that can be varied in the scene, optics, sensor, image processing, display, and viewing conditions is formidable; investigating the tradeoffs between all of these parameters in a brief paper is not practical. Hence, we focus this analysis on conditions in which we have discovered image quality differences between the outputs from the binning and resizing methods.

Visible SNR

A common metric to assess image noise is the signal-to-noise ratio (SNR): This is the ratio of the mean and standard deviation of the sensor response in a uniform image region (converted to a decibel scale). SNR measures the size of the signal relative to unwanted noise. This is a valuable metric for assessing a device, but the SNR metric does not account for other system components such as the observer or display properties. This is a problem because the visibility of sensor noise differs substantially when the image is presented on different displays or viewed from different distances. Hence, we need a measure of SNR that accounts for the display, the viewer, and the viewing geometry.

We introduce a modification of the conventional SNR, which we call the visible SNR (vSNR). This measure is derived from the Spatial CIELAB (S-CIELAB) representation [9, 10], a perceptual image representation in which distance measures perceptual difference. This representation accounts for both observer color sensitivity and spatial sensitivity.

The vSNR value is derived from the standard deviation of the S-CIELAB representation in a uniform image region. Unlike conventional SNR, there is no need to divide by the mean or to convert the measure into a decibel scale – these are already part of the S-CIELAB representation. The standard deviation of the S-CIELAB representation has units of ΔE. In theory, one ΔE is the threshold of visibility under ideal viewing conditions. The vSNR is the inverse of the standard deviation of the S-CIELAB representation of a uniform field and has units of $1/\Delta E$.

The S-CIELAB representation has one luminance and two chrominance dimensions. Hence, it is possible to estimate the noise contributions of the luminance and chrominance components separately and derive further insight into the appearance of the unwanted image noise.

System MTF

The modulation transfer function (MTF) measures the degradation in signal contrast at increasing spatial frequencies. It is possible to specify the spatial frequency of the image at different locations in the system, such as at the sensor surface (lines/mm) or as seen by the viewer on a display surface (cycles per degree). The ISO 12233 standard [1] is a method for specifying the MTF of a sensor output with respect to the sensor surface. We use this method to compare the spatial resolution of the binning and resizing methods.

RESULTS

Comparing binning and resizing: noise and resolution

The binning architecture is designed to reduce noise visibility at low illumination levels. Indeed, under these conditions, the binning method has higher vSNR than the resizing method (Figure 3A). In the simulation, image noise arises both from the photon variance and system components. These values vary between images, so that we calculate the mean and standard deviation of the vSNR in each condition. The vSNR of the simulated binning method is reliably higher than that of the resized image. When seen up close (0.3 meters), the noise is visible in the binning and resizing methods. Beyond 0.8 meters viewing distance, the noise is invisible in both methods.

The modulation transfer function for the resized method is superior to that of the binning method. Measured using the ISO 12233 slanted bar target, the contrast of the binned image is reduced to one half the maximum (MTF50) at a spatial frequency of 65 lines/mm; the MTF50 from the Matlab *imresize* method, which includes some spatial sharpening, is approximately 90 lines/mm.

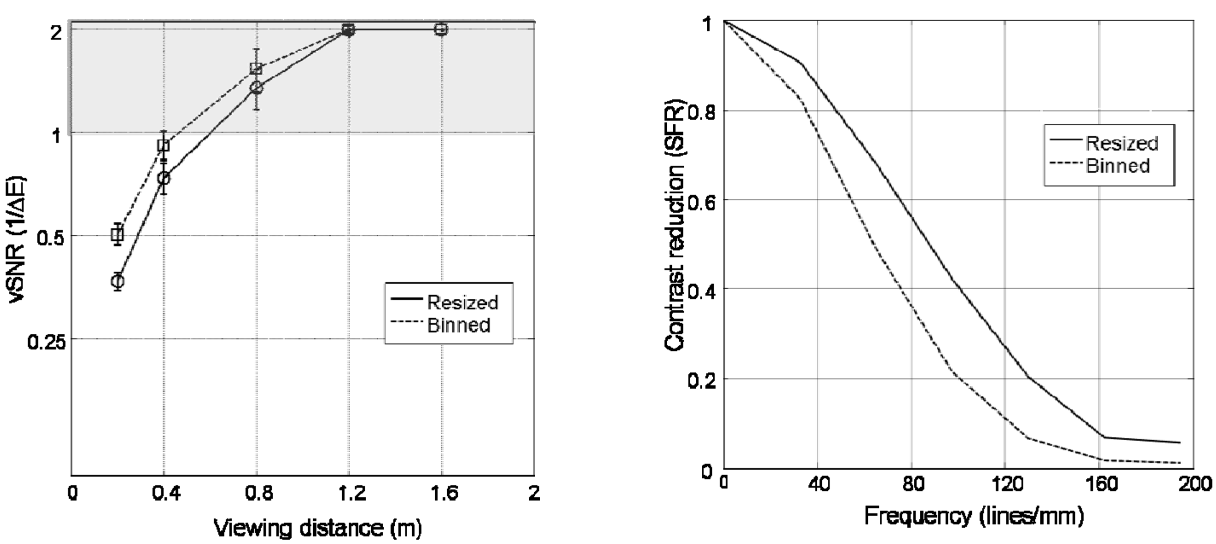

Figure 3. Visible SNR (vSNR) and modulation transfer function (MTF) of the binning and resizing methods. (A). The vSNR is calculated from the standard deviation of the response to a uniform image (mean luminance 33 cd/m^2). We assume the image is displayed on a 100 dpi sRGB display. The curves show that vSNR increases with viewing distance (i.e., noise visibility declines). Under these simulation parameters, the vSNR is higher for the binned than the resized method. A vSNR of 1 means the standard deviation of the noise is 1 ΔE and thus invisible (shaded region). Using this criterion the image noise becomes invisible at 0.4 m (binning) and 0.6 m (resizing). (B). The modulation transfer functions for the binning and resizing methods are compared in the same conditions. While the binning method has a vSNR advantage, the resizing method has an advantage for spatial resolution. The half-maximum spatial frequency contrast is at 60 lines/mm for the binning method and 90 lines/mm for the resizing method. The noise included in the simulation does not significantly impact these values when the simulation is repeated multiple times. The simulations are based on a 4 Megapixel sensor (1.4 um, quarter-inch) and rendered on a display matched to the binning spatial resolution.

The simulation conditions for this example were chosen to be favorable for the binning architecture: low scene luminance (33 cd/m2), a small pixel size (1.4 µm), and a large amount of read noise (15 electrons). The target display size was matched to the output of the binned sensor and had a spatial resolution of 100 dots per inch (dpi). Below we explore conditions under which there is no significant difference. Further, it is possible to apply a small smoothing filter to the resized image so that the vSNR and MTF50 match the binning method nearly precisely (not shown).

The effects of read noise and scene luminance

At increasing levels of read noise, vSNR declines. The significance of this decline, however, depends upon the scene luminance. When the scene luminance is relatively high, binning and resizing vSNR levels are high and the noise should be invisible (Figure 4A). At lower scene luminance, the vSNR is higher for binning than resizing. This difference is very small when the read noise is less than 8 electrons. For sensors with more than 10 electrons of read noise, the difference becomes significant. Hence, the advantage of pixel binning is mainly for images acquired under a low scene luminance and with relatively noisy pixel read-out.

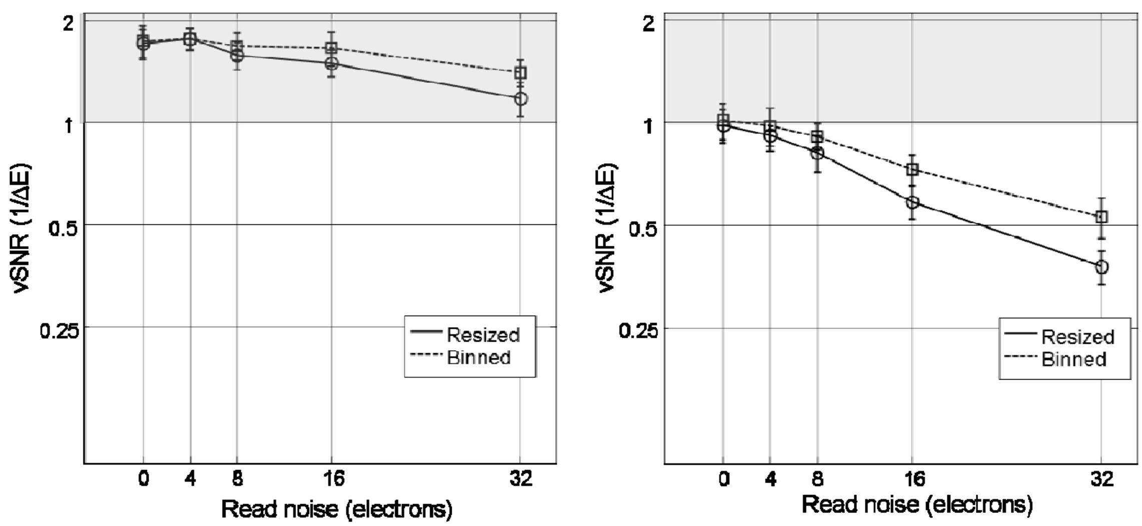

Figure 4. Binning is effective under low illumination and high read noise conditions. The two panels estimate the vSNR as a function of pixel read noise. In bright scenes (100 cd/m^2, left) the effect of the read noise is small compared to the photon noise. Hence, the vSNR for both methods is very high and not different. Under lower scene luminance conditions (33 cd/m^2, right) the read noise is significant. At extremely high levels of read noise, the binning method has an advantage. At more typical and modest levels, the two methods are very similar and the noise itself is barely visible. Viewing distance: 0.5 m. Display resolution: 100 dpi. Display samples matched to binned samples.

The effects of pixel size

As pixel size scales, the photon catch is reduced; in this sense, the effects of smaller pixels are comparable to operating under low luminance conditions. The binning method increases the vSNR of images acquired with small pixels, just as it increases the vSNR under low luminance conditions. The simulation in Figure 5 was carried out using a 33 cd/m^2 scene and a pixel a read noise level of 15 electrons. The noise becomes visible for pixel sizes below 1.6 um, and in this small pixel regime the binning method has reliably less visible noise than the resizing method.

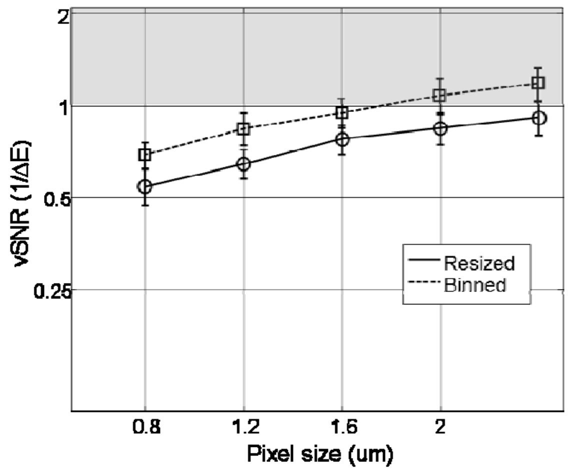

Figure 5. The effect of pixel size on vSNR. At very small pixel size, the reduced noise from the binning improves the vSNR. Simulation parameters: Read noise: 15 electrons. Viewing distance: 0.5m. Scene mean luminance 33 cd/m^2.

Visualizing examples: sweep frequency and natural images

The image noise (vSNR) and spatial resolution (MTF50) metrics provide useful quantitative assessments. Further insights, and confidence in the simulations, can be derived by creating visualizations of critical test targets and selected natural images.

First, we simulated an image of a sweep frequency pattern. In addition to the frequency sweep from left to right, the synthetic target was created with increasing contrast from the bottom to the top of the image. This target is helpful in clarifying the impact of binning (left) and resizing (right) on both noise and spatial resolution (Figure 6).

Notice that the harmonics in the sweep frequency pattern remain visible at finer resolution in the resizing method. The noise in the uniform region at the right of the pattern is also more visible. The visual impression is confirmed by the trace shown overlaid on the image. The trace measures the frequency sweep more closely in the resizing method compared to the binning method; the variance at the right of the image in the uniform region is also more visible. The low spatial frequency colored bars visible in the binned image reflects the fact that the binned method does not benefit from the addition of an anti-aliasing filter.

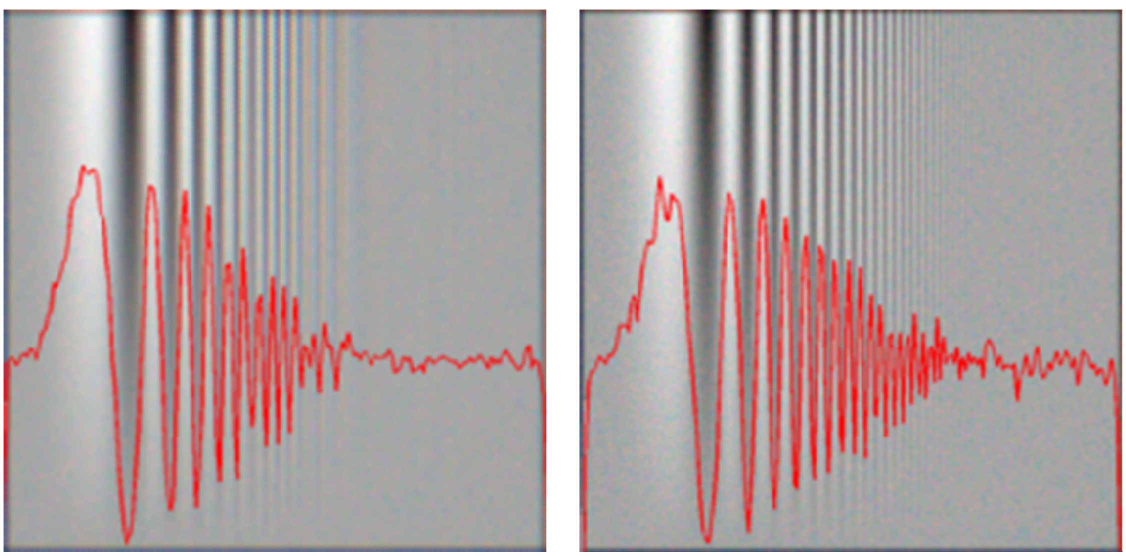

Figure 6. Sweep frequency images simulated using binning and resizing methods. In addition to the frequency sweep, the image contrast increases from the bottom to the top of the image. The image on the left was simulated using the binning method, and on the right using the resizing method. The traces measure the image intensity at the 10th row of the image. Aliasing is visible as low frequency vertical bars in the binned image. The increased noise and higher spatial resolution of the resizing method are visible in the image and the traces.

The features illustrated in the synthetic sweep frequency pattern are also evident in simulations of a natural image (Figure 7). The two images on the left use the binning method and the two images on the right use resizing. The resized images appear sharper. This can be seen both in the details within the face and the flowers. There is less image noise in the pixel binning simulation. This can be seen by comparing the uniform gray wall in the background.

Which condition is preferred, the sharper detail in the subject matter or the reduced noise in uniform regions, will depend on the viewer. This question is open to perceptual experimentation.

Figure 7. Natural image examples. The images on the left were simulated using the binning method, and the right using resizing. Other parameters as in the previous examples.

DISCUSSION

The binning architecture has a clear advantage for one important metric: The ADC rate is reduced by a factor of two, and the channel throughput is reduced by a factor of four.

Here we address the comparison of quality for images produced with the binning and resizing methods. Under low illumination conditions and for pixels with relatively high read noise, images generated with the binning method have less noise (high vSNR) than resized images. The binning method generally has noticeably lower spatial resolution. We have not investigated the range of conditions in which the lost resolution is visually important, though examining the sweep frequency pattern and natural image simulations show that the difference is noticeable.

In conditions when there is no significant ADC rate or channel bandwidth limitation, it is advantageous to read the entire high resolution image. In this case it will be possible to resize the image later to produce an image that is very similar to that obtained by binning. A resized image can be further blurred to match a binned image with respect to vSNR and

MTF50. There is an image quality tradeoff between noise and blur, and depending on the application users may prefer one error over another. This question is open to experimental investigation through user-preference studies.

The ability to simulate the two methods makes it possible to assess performance in a wide range of conditions prior to implementing the full hardware solution. Simulations based on quantitative descriptions of the imaging conditions and sensor parameters enable an early assessment of proposed technologies that includes measuring engineering properties as well as providing some glimpses of the likely image outputs. In this case, the simulations enable us to confirm that the binning architecture is effective under low scene luminance and high pixel read noise. The curves and image visualization suggest that the image quality difference between binning and resizing is generally quite modest, although the spatial resolution of the resizing method is noticeably better. The most important value of the binning architecture is the reducing in ADC rate and channel throughput. .

ACKNOWLEDGMENTS

This work was supported in part by a generous grant from Microsoft.

REFERENCES

1. International Standards Organization, *Photography-Electronic Still Picture Cameras-Resolution Measurements*. 1999: New York, New York, USA.

2. Williams, D., *Benchmarking of the ISO 12233 Slanted-Edge Spatial Frequency Response (SFR) Plug-in*, in *Proceedings of the 51st IS&T PICS Conference*. 1988. p. 133-136.

3. Farrell, J., F. Xiao, P. Catrysse and B. Wandell, *A simulation tool for evaluating digital camera image quality.* Proceedings of the SPIE, 2004. **5294**: p. 124-131.

4. Farrell, J., M. Parmar, P. Catrysse and B. Wandell, *Digital Camera Simulation*, in *Handbook of Digital Imaging*, M. Kriss, Editor. in press, Wiley.

5. Vasta, P., *CMOS Image Sensors Address High Quality, Functionality Requirements.* Ecnmag.com 2006.

6. Thévenaz, P., T. Blu, and M. Unser, *Image interpolation and resizing*, in *Handbook of Medical Imaging*, I.N. Bankman, Editor. 2000, Academic Press: Orlando. p. 393-420.

7. Unser, M., A. Aldroubi, and M. Eden, *Enlargement or reduction of digital images with minimum loss of information.* IEEE Transactions on Image Processing, 1995. **4**(3): p. 247-258.

8. Farrell, J., M. Okincha, and M. Parmar, *Sensor calibration and simulations.* Proceedings of the SPIE, 2008. **6817**.

9. Zhang, X. and B.A. Wandell, *A spatial extension to CIELAB for digital color image reproduction*, in *Society for Information Display Symposium Technical Digest*. 1996. p. 731-734.

10. Zhang, X. *S-CIELAB: A spatial extension to the CIE L*a*b* DeltaE Color Difference Metric*. 1998 [cited; Available from: http://white.stanford.edu/~brian/scielab/scielab.html.