

How Small Should Pixel Size Be ?

Ting Chen¹, Peter Catrysse¹, Abbas El Gamal¹ and Brian Wandell²

¹Information Systems Laboratory

Department of Electrical Engineering, Stanford University, CA 94305, USA

²Department of Psychology, Stanford University, CA 94305, USA

ABSTRACT

Pixel design is a key part of image sensor design. After deciding on pixel architecture, a fundamental tradeoff is made to select pixel size. A small pixel size is desirable because it results in a smaller die size and/or higher spatial resolution; a large pixel size is desirable because it results in higher dynamic range and signal-to-noise ratio. Given these two ways to improve image quality and given a set of process and imaging constraints an optimal pixel size exists. It is difficult, however, to analytically determine the optimal pixel size, because the choice depends on many factors, including the sensor parameters, imaging optics and the human perception of image quality. This paper describes a methodology, using a camera simulator and image quality metrics, for determining the optimal pixel size. The methodology is demonstrated for APS implemented in CMOS processes down to 0.18μ technology. For a typical 0.35μ CMOS technology the optimal pixel size is found to be approximately $6.5\mu\text{m}$ at fill factor of 30%. It is shown that the optimal pixel size scales with technology, but at slower rate than the technology itself.

Keywords: Signal-to-Noise Ratio(SNR), Dynamic Range(DR), Modulation Transfer Function(MTF), ΔE , CMOS APS, image sensor

1. INTRODUCTION

Pixel design is a crucial element of image sensor design. After deciding on the photodetector type and pixel architecture, a fundamental tradeoff must be made to select pixel size. Reducing pixel size improves the sensor by increasing spatial resolution for fixed sensor die size. Increasing pixel size improves the sensor by increasing dynamic range and signal-to-noise ratio. Because changing pixel size has opposing effects on key imaging variables, for a given a set of process and imaging constraints, an optimal pixel size may exist. The purpose of this paper is to understand the tradeoffs involved and to specify a method for determining optimal pixel size.

In older process technologies, the selection of such an optimal pixel size may not have been important, since the transistors in the pixel occupied such a large area that the designer could not increase the photodetector size (and hence fill factor) without making pixel size unacceptably large. As process technology scales, however, the area occupied by the pixel transistors decreases, providing more freedom to increase fill factor while maintaining an acceptably small pixel size. As a result of this new flexibility, it is becoming more important to use a systematic method to determine the optimal pixel size.

It is difficult to determine an optimal pixel size analytically because the choice depends on sensor parameters, imaging optics characteristics, and elements of human perception. In this paper we describe a methodology for using a gray scale digital camera simulator¹ and the S-CIELAB metric² to examine how pixel size affects image quality. To determine the optimal pixel size, we decide on a sensor area and create a set of simulated images corresponding to a range of pixel sizes. The difference between the simulated output image and a perfect, noise-free image is measured using a spatial extension of the CIELAB color metric, S-CIELAB. The optimal pixel size is obtained by selecting the pixel size that produces the best rendered image quality as measured by S-CIELAB.

Correspondence: Email: tingchen@isl.stanford.edu, peter@kaos.stanford.edu, abbas@isl.stanford.edu, wandell@stanford.edu; Telephone: 650-725-9696; Fax: 650-723-8473

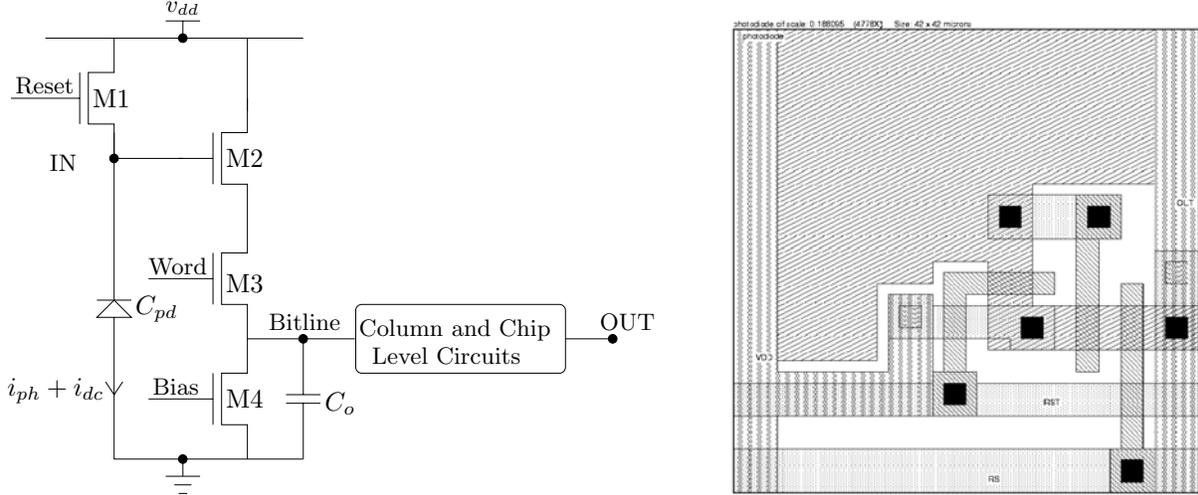


Figure 1. APS circuit and sample pixel layout

We illustrate the methodology by applying it to CMOS APS, using key parameters for CMOS process technologies down to 0.18μ . The APS pixel under consideration is the standard n+/psub photodiode, three transistors per pixel circuit shown in Figure 1. The sample pixel layout³ achieves 35% fill factor and will be used as a basis for determining pixel size for different fill factors and process technology generations.

The remainder of this paper is organized as follows. In section 2 we analyze the effect of pixel size on sensor performance and system MTF. In section 3 we describe the methodology for determining the optimal pixel size given process technology parameters, imaging optics characteristics, and imaging constraints such as illumination range, maximum acceptable integration time and maximum spatial resolution. In section 4 we explore this methodology using the CMOS APS 0.35μ technology. In section 5 we use our methodology and a set of process parameters to investigate the effect of technology scaling on optimal pixel size.

2. PIXEL PERFORMANCE, SPATIAL RESOLUTION, AND SYSTEM MTF VERSUS PIXEL SIZE

In this section we demonstrate the effect of pixel size on sensor dynamic range (DR), signal-to-noise ratio (SNR), and camera system modulation transfer function (MTF). For simplicity we assume square pixels throughout the paper and define pixel size to be the length of the side. The analysis in this section motivates the need for a methodology for determining an optimal pixel size.

2.1. Dynamic Range and Signal-to-noise Ratio versus Pixel Size

Dynamic range (DR) and *signal-to-noise ratio* (SNR) are two useful measures of pixel performance. Dynamic range quantifies the ability of a sensor to image highlights and shadows; it is defined as the ratio of the largest non-saturating current signal i_{max} , *i.e.* input signal swing, to the smallest detectable current signal i_{min} , which is typically taken as the standard deviation of the input referred noise when no signal is present. Using this definition and the sensor noise model it can be shown⁴ that DR in dB is given by

$$DR = 20 \log_{10} \frac{i_{max}}{i_{min}} = 20 \log_{10} \frac{q_{max} - i_{dc} t_{int}}{\sqrt{\sigma_r^2 + q i_{dc} t_{int}}} \quad (1)$$

where q_{max} is the well capacity, q is the electron charge, i_{dc} is the dark current, t_{int} is the integration time, σ_r^2 is the variance of the temporal noise, which we assume to be approximately equal to kTC , *i.e.* the reset noise when correlated double sampling (CDS) is performed.⁵ For voltage swing V_s and photodetector capacitance C the maximum well capacity is $q_{max} = CV_s$.

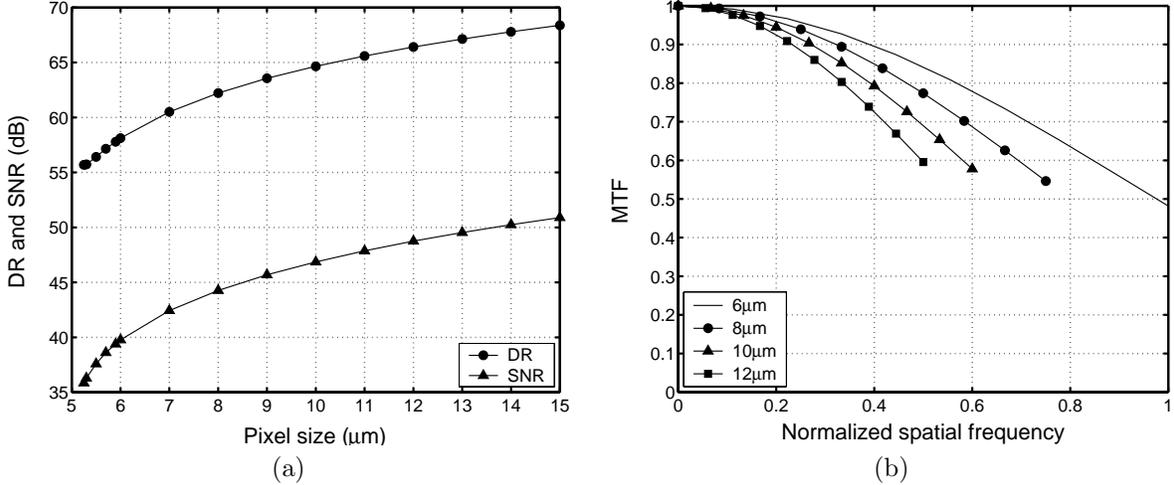


Figure 2. (a) DR and SNR (at 20% well capacity) as a function of pixel size. (b) Sensor MTF (with spatial frequency normalized to the Nyquist frequency for $6\mu\text{m}$ pixel size) is plotted assuming different pixel sizes.

SNR is the ratio of the input signal power and the average input referred noise power. As a function of the photocurrent i_{ph} , SNR in dB is⁴

$$\text{SNR}(i_{ph}) = 20 \log_{10} \frac{i_{ph} t_{int}}{\sqrt{\sigma_r^2 + q(i_{ph} + i_{dc}) t_{int}}} \quad (2)$$

Figure 2(a) plots DR as a function of pixel size. It also shows SNR at 20% of the well capacity versus pixel size. The curves are drawn assuming the parameters⁶ for a typical $0.35\mu\text{m}$ CMOS process, and integration time $t_{int} = 30\text{ms}$. As expected, both DR and SNR increase with pixel size. DR increases roughly as the square root of pixel size, since both C and reset noise (kTC) increase approximately linearly with pixel size. SNR also increases roughly as the square root of pixel size since the RMS shot noise increases as the square root of the signal. These curves demonstrate the advantages of choosing a large pixel. In the following subsection, we demonstrate the disadvantages of a large pixel size, which is the reduction in spatial resolution and system MTF.

2.2. Spatial Resolution and System MTF versus Pixel Size

For a fixed sensor die size, decreasing pixel size increases pixel count. This results in higher spatial sampling and a potential improvement in the system's *modulation transfer function* (MTF) provided the resolution is not limited by the imaging optics. For an image sensor, the Nyquist frequency is simply one half of the reciprocal of the center-to-center pixel spacing. Image frequency components above the Nyquist frequency cannot be reproduced accurately by the sensor and result in aliasing. The system MTF measures how well the system reproduces the spatial structure of the input scene below the Nyquist frequency and is defined to be the ratio of the output modulation to the input modulation as a function of input spatial frequency.^{7,8}

Under certain simplifying assumptions, the system MTF can be expressed as the product of the optical MTF, geometric MTF, and diffusion MTF.⁷ Each component causes low pass filtering, which degrades the response at higher frequencies. In our study, we only account for the optical and geometric MTF. Figure 2(b) plots system MTF as a function of the input spatial frequency for different pixel sizes. The results are again for the $0.35\mu\text{m}$ process mentioned before. Note that as we decrease pixel size the Nyquist frequency increases and MTF improves. The reason for the MTF improvement is that reducing pixel size reduces the low pass filtering due to geometric MTF.

In summary, a small pixel size is desirable because it results in higher spatial resolution and better MTF. A large pixel size is desirable because it results in better DR and SNR. Therefore, there must exist a pixel

size that strikes a compromise between high DR and SNR on the one hand, and high spatial resolution and MTF on the other. The results so far, however, are not sufficient to determine such an optimal pixel size. First it is not clear how to tradeoff DR and SNR with spatial resolution and MTF. More importantly, it is not clear how these measures relate to image quality, which should be the ultimate objective of selecting optimal pixel size.

3. METHODOLOGY FOR DETERMINING OPTIMAL PIXEL SIZE

In this section we describe a methodology for selecting the optimal pixel size. The goal is to find the optimal pixel size for a given process parameters, sensor die size, imaging optics characteristics and imaging constraints. We do so by varying pixel size and thus pixel count for the given die size, as illustrated in Figure 3. Fixed die size enables us to fix the imaging optics. For each pixel size (and count) we use our camera simulator with a synthetic Contrast Sensitivity Function (CSF)⁹ scene, as shown in Figure 4 to estimate the resulting image using the chosen sensor and imaging optics. The rendered image quality in terms of the S-CIELAB ΔE error is then determined. The experiment is repeated for different pixel sizes and the optimal pixel size is selected to achieve the highest image quality.

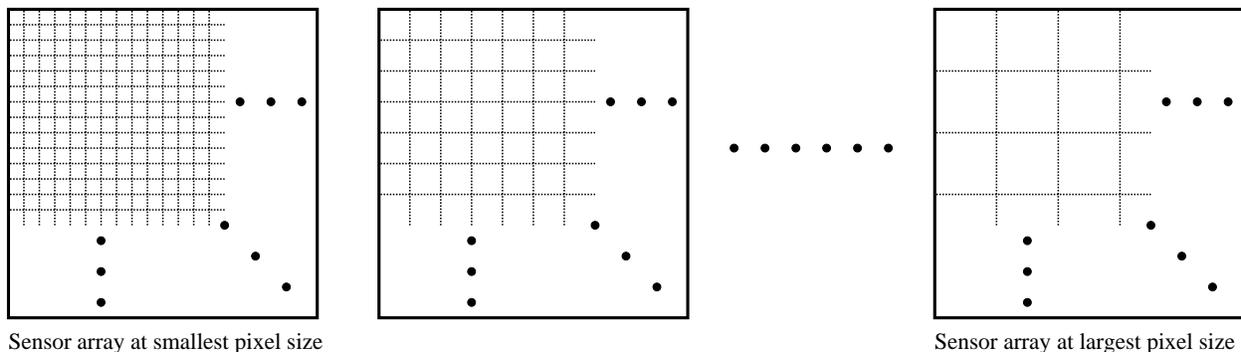


Figure 3. Varying pixel size for a fixed die size

The information on which the simulations and comparisons are made is as follows :

- A list of the sensor parameters for the process technology.
- The smallest pixel size and the pixel array die size.
- The imaging optics characterized by focal length f and $f/\#$.
- The maximum acceptable integration time.
- The highest spatial frequency desired.
- Absolute radiometric or photometric scene parameters.
- Rendering model including viewing conditions and display specifications

The camera simulator¹ provides models for the scene, the imaging optics, and the sensor. The imaging optics model accounts for diffraction using a wavelength-dependent MTF and properly converts the scene radiance into image irradiance taking into consideration off-axis irradiance. The sensor model accounts for the photodiode spectral response, fill factor, dark current sensitivity, geometric MTF, temporal noise, and fixed pattern noise (FPN). Exposure control can be set either by the user or by an automatic exposure control routine, where the integration time is limited to a maximum acceptable value. The simulator reads spectral scene descriptions and returns simulated images from the camera.

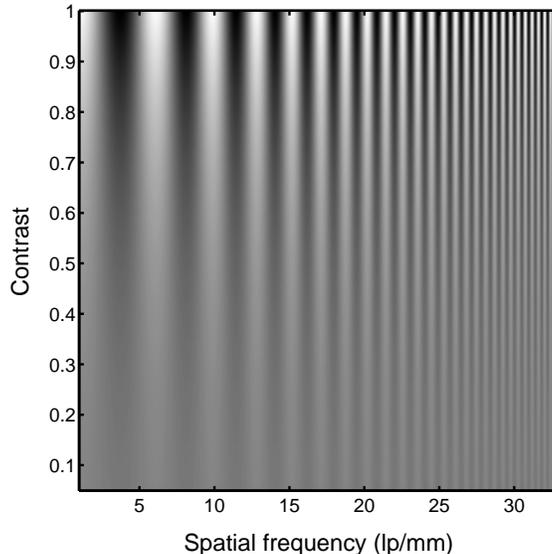


Figure 4. Synthetic Contrast Sensitivity Function (CSF) scene

For each pixel size, we simulate the camera response to the test pattern shown in Figure 4. This pattern varies in both spatial frequency along the horizontal axis and in contrast along the vertical axis. The pattern was chosen first because it spans the frequency and contrast ranges of normal images in a controlled fashion. These two parameters correspond well with the tradeoffs for spatial resolution and DR that we observe as a function of pixel size. Secondly, image reproduction errors at different positions within the image correspond neatly to evaluations in different spatial-contrast regimes, making analysis of the simulated images straightforward.

In addition to the simulated camera output image, the simulator also generates a “perfect” image from an ideal (*i.e.* noise-free) sensor with perfect optics. The simulated output image and the “perfect” image are compared by assuming that they are rendered on a CRT display, and this display is characterized by its phosphor dot pitch and transduction from digital counts to light intensity. Furthermore, we assume the same white point for the monitor and the image. With these assumptions, we use the S-CIELAB ΔE metric to measure the point by point difference between the simulated and perfect images.

S-CIELAB² is an extension of the CIELAB standard.¹⁰ In this metric, images are converted to a representation that captures the response of the photoreceptor mosaic of the eye. The images are then convolved with spatial filters that account for the spatial sensitivity of the visual pathways. The filtered images are then converted into the CIELAB format and distances are measured using the conventional ΔE units of the CIELAB metric. In this metric, one unit represents approximately the threshold detection level of the difference under ideal viewing conditions.

4. SIMULATION RESULTS

Figure 5 shows the simulation results for an $8\mu\text{m}$ pixel, designed in $0.35\mu\text{m}$ CMOS process, assuming a scene luminance range from 25 to 1000 cd/m^2 and maximum integration time of 100ms. The test pattern includes spatial frequencies up to 33 lp/mm, which corresponds to the Nyquist rate for a $15\mu\text{m}$ pixel. Shown are the perfect CSF image, the output image from the camera simulator, the ΔE error map obtained by comparing the two images, and a set of iso- ΔE curves. Iso- ΔE curves are obtained by connecting points with identical ΔE values on the ΔE error map. Remember that larger values represent higher error (worse performance).

The largest S-CIELAB errors are in high spatial frequency and high contrast regions. This is consistent with the sensor DR and MTF limitations. For a fixed spatial frequency, increasing the contrast causes more

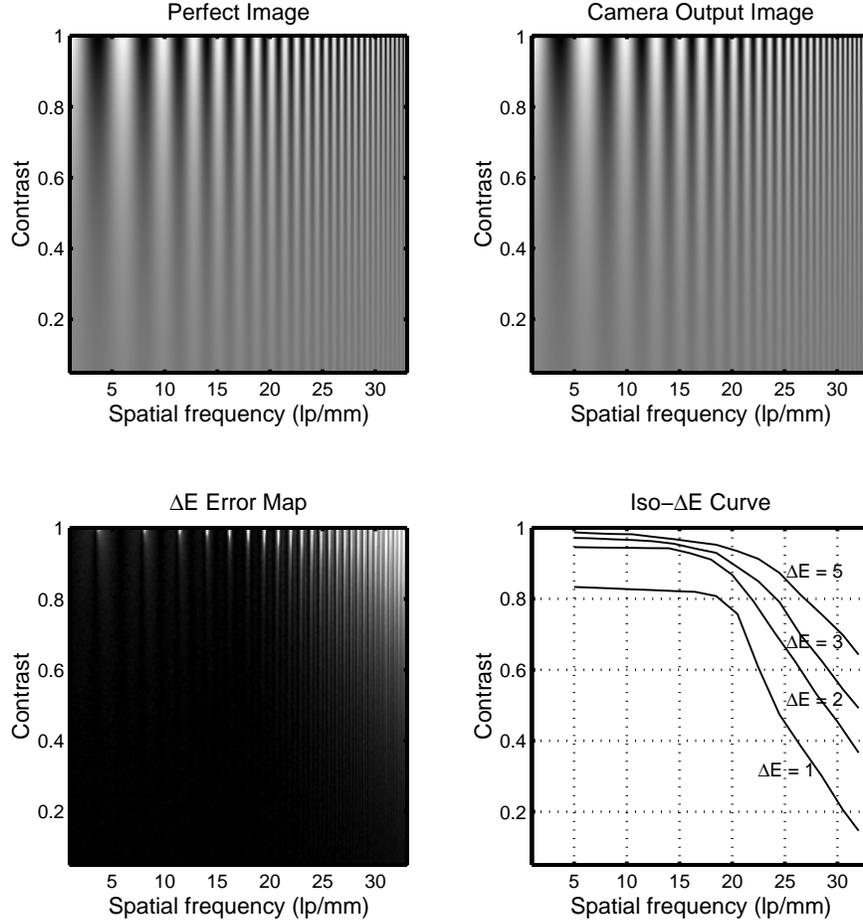


Figure 5. Simulation result for a 0.35μ process with pixel size of $8\mu\text{m}$

errors because of limited sensor dynamic range. For a fixed contrast, increasing the spatial frequency causes more errors because of geometric MTF limitations.

Now to select the optimal pixel size for the 0.35μ technology we vary pixel size as discussed in the section 3. The minimum pixel size, which is chosen to correspond to 5% fill factor, is $5.3\mu\text{m}$. Note that we are here in a sensor-limited resolution regime, *i.e.* pixel size is bigger than the spot size dictated by the imaging optics characteristics. The minimum pixel size results in a die size of $2.7 \times 2.7 \text{ mm}^2$ for a 512×512 pixel array. The maximum pixel size is $15\mu\text{m}$ with fill factor 73%, and corresponds to maximum spatial frequency of 33 lp/mm. The luminance range for the scene is again taken to be between 25 and 1000 cd/m² and the maximum integration time is 100ms.

Figure 6 shows the iso- $\Delta E = 3$ curves for four different pixel sizes. Certain conclusions on the selection of optimal pixel size can be readily made from the iso- ΔE curves. For instance, if we use $\Delta E = 3$ as the maximum error tolerance, clearly a pixel size of $8\mu\text{m}$ is better than a pixel size of $15\mu\text{m}$, since the iso- $\Delta E = 3$ curve for the $8\mu\text{m}$ pixel is consistently higher than that for the $15\mu\text{m}$ pixel. Similarly a $6\mu\text{m}$ pixel is better than an $8\mu\text{m}$ one. However, it is not clear whether a $6\mu\text{m}$ pixel is better or worse than a $5.3\mu\text{m}$ pixel, since their iso- ΔE curves intersect such that for low spatial frequencies the $6\mu\text{m}$ pixel is better while at high frequencies the $5.3\mu\text{m}$ pixel is better.

Instead of looking at the iso- ΔE curves, we simplify the optimal pixel size selection process by using the mean value of the ΔE error over the entire image as the overall measure of image quality. We justify our choice by performing a statistical analysis of the ΔE error map. This analysis reveals a compact, unimodal

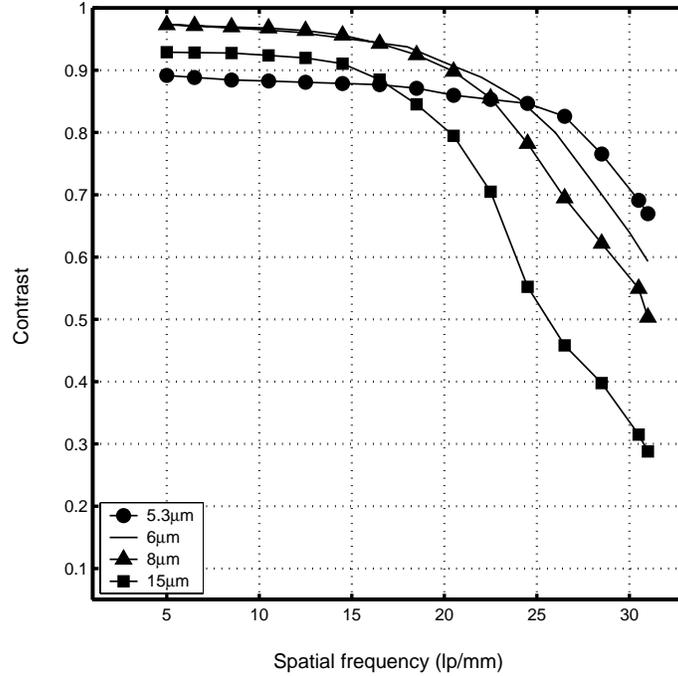


Figure 6. Iso- $\Delta E = 3$ curves for different pixel sizes

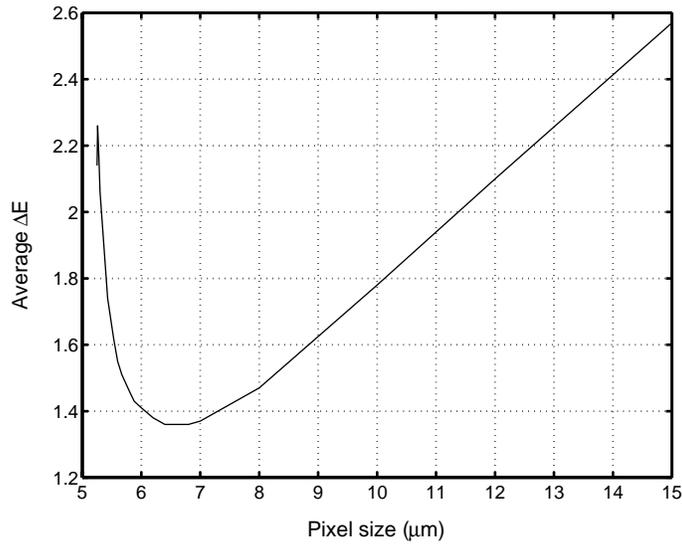


Figure 7. Average ΔE versus pixel size

distribution which can be accurately described by first order statistics, such as mean or maximum. Figure 7 shows mean ΔE versus pixel size and an optimal pixel size can be readily selected from the curve. For the 0.35μ technology chosen the optimal pixel size is found to be $6.5\mu\text{m}$ with a 30% fill factor.

5. EFFECT OF TECHNOLOGY SCALING ON PIXEL SIZE

How does optimal pixel size scale with technology? We repeat the simulations discussed in the previous section for technologies down to 0.18μ . The mean ΔE curves are shown in Figure 8. The optimal pixel size

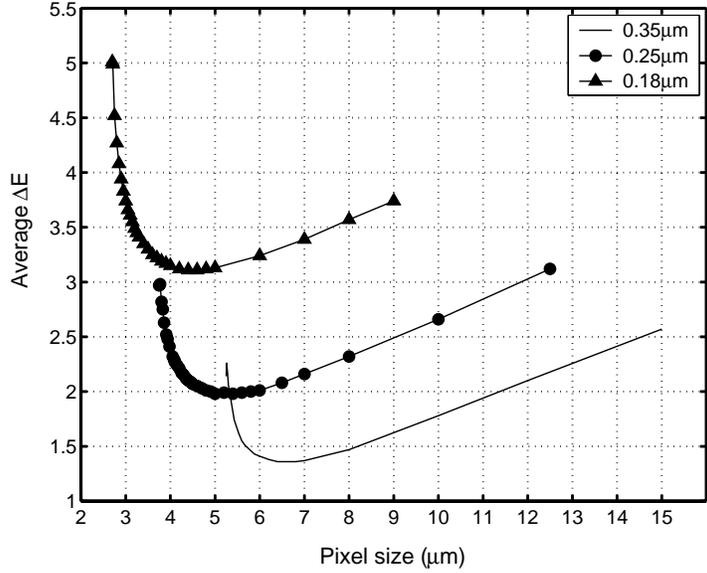


Figure 8. Average ΔE versus pixel size as technology scales

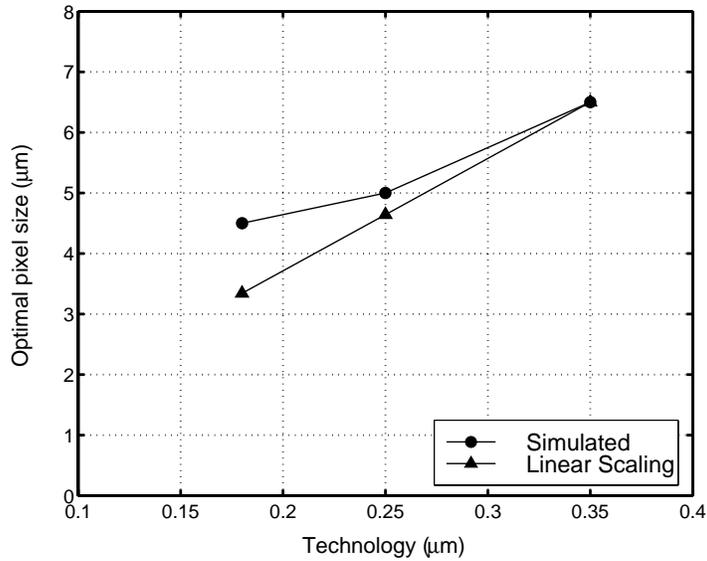


Figure 9. Optimal pixel size versus technology

shrinks, but at a slower rate than technology (Figure 9). Also, the image quality, as measured by the mean ΔE , degrades as technology scales.

6. CONCLUSION

We proposed a methodology using a camera simulator, synthetic CSF scenes, and S-CIELAB for selecting the optimal pixel size for an image sensor given process technology parameters, imaging optics parameters, and imaging constraints. We applied the methodology to photodiode APS implemented in CMOS technologies down to 0.18μ and demonstrated the tradeoff between DR and SNR on one hand and spatial resolution and MTF on the other hand due to the selection of pixel size. Using mean ΔE as an image quality metric, we

found that indeed an optimal pixel size exists, which represents the optimal tradeoff. For a 0.35μ process we found that a pixel size of around $6.5\mu\text{m}$ with fill factor 30% under certain imaging optics, illumination range, and integration time constraints achieves the lowest mean ΔE . We found that the optimal pixel size scales with technology, albeit at slower rate than the technology.

The proposed methodology and its application can be extended in several ways:

- The imaging optics model we used is oversimplified. A more accurate model that includes lens aberrations and optical vignetting is needed to find the effect of the lens on the selection of pixel size. This extension requires a more detailed specification of the imaging optics by means of a lens prescription and can be performed by using a ray tracing program.¹¹
- Microlens arrays are often used to increase the effective fill factor. Since the increase in fill factor heavily depends on the microlens design, we did not have a general way of modeling it. Given a specific microlens model, the effect on optimal pixel size can be readily performed using our methodology.
- The methodology needs to be extended to color.

ACKNOWLEDGEMENTS

The work reported in this paper was supported under the Programmable Digital Camera Project by Intel, Hewlett-Packard, Kodak, Interval Research, and Canon. The project was started as a class project in EE392b at Stanford University in spring of 1999, therefore the authors would like to acknowledge the initial work done by the students in the class. The authors would also like to acknowledge the help of J. DiCarlo, S.H. Lim, X.Q. Liu, K. Salama and H. Tian for valuable discussions. Peter Catrysse acknowledges the support from Hewlett-Packard through the Fellow/Mentor/Advisor Program at Stanford University. He is also an "Aspirant" of the Fund for Scientific Research - Flanders (Belgium) associated with the Applied Physics Department (TONA) of the Vrije Universiteit Brussel.

REFERENCES

1. P. B. Catrysse, B. A. Wandell, and A. El Gamal, "Comparative analysis of color architectures for image sensors," in *Proceedings of SPIE*, vol. 3650, pp. 26–35, (San Jose, CA), February 1999.
2. X. Zhang and B. A. Wandell, "A Spatial Extension of CIELAB for Digital Color Image Reproduction," *Society for Information Display Symposium Technical Digest* **27**, pp. 731–734, 1996.
3. S. K. Mendis, S. E. Kemeny, R. C. Gee, B. Pain, C. O. Staller, Q. Kim, and E. R. Fossum, "CMOS Active Pixel Image Sensors for Highly Integrated Imaging Systems," *IEEE Journal of Solid-State Circuits* **32**(2), pp. 187–197, 1997.
4. D. X. Yang and A. El Gamal, "Comparative analysis of SNR for Image Sensors with Enhanced Dynamic Range," in *Proceedings of SPIE*, vol. 3649, pp. 197–211, (San Jose, CA), February 1999.
5. H. Tian, B. A. Fowler, and A. El Gamal, "Analysis of temporal noise in CMOS APS," in *Proceedings of SPIE*, vol. 3649, pp. 177–185, (San Jose, CA), February 1999.
6. T. Chen, P. B. Catrysse, A. El Gamal, and B. A. Wandell, "How small should pixel size be?," *expanded version, in preparation*.
7. G. C. Holst, *CCD Arrays, Cameras and Displays*, JCD Publishing and SPIE, Winter Park, Florida, 1998.
8. B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., Sunderland, Massachusetts, 1995.
9. F. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology* **197**, pp. 551–566, 1968.
10. C.I.E., "Recommendations on uniform color spaces,color difference equations, psychometric color terms," *Supplement No.2 to CIE publication No.15(E.-1.3.1) 1971/(TC-1.3)*, 1978.
11. *CODE V.40*, Optical Research Associates, Pasadena, California, 1999.