Photometric limits for digital camera systems

Michael Schöberl
Andreas Brückner
Siegfried Foessel
André Kaup

Photometric limits for digital camera systems

**Michael Schöberl,[a] Andreas Brückner,[b]
Siegfried Foessel,[c] and André Kaup[a]**
[a]University of Erlangen-Nuremberg, Multimedia
communications and Signal Processing, Cauerstrasse 7,
91058 Erlangen, Germany
[b]Fraunhofer Institute for Applied Optics and Precision
Engineering, Albert-Einstein-Strasse 7, 07745 Jena,
Germany
[c]Fraunhofer Institute for Integrated Circuits,
Am Wolfsmantel 33, 91058 Erlangen, Germany
E-mail: schoeberl@LNT.de

Abstract. *Image sensors for digital cameras are built with ever decreasing pixel sizes. The size of the pixels seems to be limited by technology only. However, there is also a hard theoretical limit for classical video camera systems: During a certain exposure time only a certain number of photons will reach the sensor. The resulting shot noise thus limits the signal-to-noise ratio. In this letter we show that current sensors are already surprisingly close to this limit.* © 2012 SPIE and IS&T. [DOI: 10.1117/1.JEI.21.2.020501]

1 Introduction

The steady progress in semiconductor technology allows the manufacturing of smaller and smaller structures and image sensors with ever shrinking pixel sizes. One can get the impression that the pixel size is just limited by the technology and even smaller pixels are desirable. Today, consumer products with pixel sizes $d_p = 1.4 \; \mu$m are already on the market and devices with $d_p = 1.1 \; \mu$m are in production.[1] In comparison, photo receptors in the human eye are reported to be larger than 3 μm.[2]

The general modeling of light is well understood,[3] and simulation with commercial tools like ISET[4] is possible. In contrast, this letter addresses parameters like aperture and pixel size and their photometric consequences for modeling the amount of light that is available in a digital video camera system. One of the design parameters is the resulting image quality. With small pixels only a few photons will hit a single pixel during an exposure period and the signal-to-noise power ratio (SNR) will be poor due to shot noise.[5] Apart from all technological limitations, this physical boundary limits the performance of today's video cameras.

2 Image Acquisition Model

The scene radiates a certain amount of light. This is described by an average radiance in object space L_{obj}. The sensor sees an effective amount of light equivalent to the cone with a solid angle Ω as shown in Fig. 1(a). This
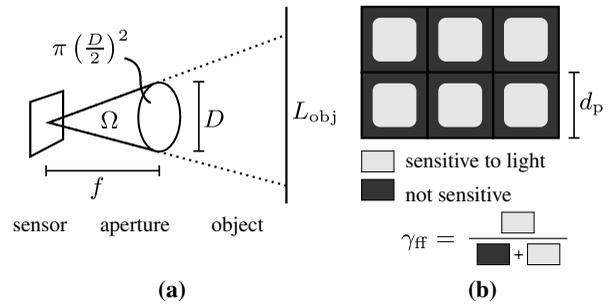
Fig. 1 Parameters of (a) focal length f, aperture diameter D and resulting solid angle Ω and (b) quadratic pixels with fill factor γ_{ff}.

cone is defined by the sphere of radius equal to the focal length f and a circular aperture disk with diameter D. The solid angle thus calculates to[6]

$$\Omega = \frac{\pi \cdot (D/2)^2}{f^2} = \frac{\pi}{4 \cdot (f/D)^2} \; [\text{sr}]. \tag{1}$$

The sensor receives an irradiance I

$$I = \eta_{lens} \cdot \Omega \cdot L_{obj} \; [\text{W} \cdot \text{m}^{-2}], \tag{2}$$

for a lens with optical transmittance η_{lens}.

On the sensor some area is used for interconnects and transistors so that only some of the area is sensitive to light. Figure 1(b) shows pixels of size d_p. The ratio of active to total areas is expressed as an effective sensor fill factor γ_{ff}. Even with clever manufacturing like micro-lenses or back side illumination, $\gamma_{ff} < 1$ holds. A single pixel thus captures a certain amount of radiant power (radiant flux) Φ_{pix} of the sensor irradiance

$$\Phi_{pix} = d_p^2 \cdot \gamma_{ff} \cdot I \; [\text{W}]. \tag{3}$$

A single photon of wavelength λ has the energy $h\frac{c}{\lambda}$ with the speed of light c and Planck's constant h. The radiant flux Φ thus consists of N_{phot} photons

$$N_{phot} = \frac{1}{h\frac{c}{\lambda}} \cdot \tau_{exp} \cdot \Phi_{pix}, \tag{4}$$

during a certain time interval (exposure time) of τ_{exp}. In the photoreceptor only some of these photons are converted into electrons $N_{elec} = \eta_{qe} \cdot N_{phot}$ while others are not, due to reflection, recombination and other material interactions. The conversion rate is expressed as quantum efficiency η_{qe}.[7] The electrons are then collected in the pixel. Although we will see N_{elec}, on average, the charge is still quantized and the actual number of electrons is subject to shot noise due to the occurrence of random events. For N electrons the associated shot noise is of strength $\sqrt{N}$.[5] As $N_{elec} \propto \Phi_{pix}$, signal power is represented with N_{elec} and SNR thus calculates to

$$\text{SNR} = N_{elec} / \sqrt{N_{elec}} = \sqrt{N_{elec}}. \tag{5}$$

In CCD or CMOS technology there are further sources of sensor noise,[8] which are neglected in the ideal case.

SNR is a parameter that is directly visible in the final images. For answering the original question, we can combine the above equations. This leads to

$$d_{\text{p,min}} = \sqrt{\text{SNR}^2 \cdot \frac{h\frac{c}{\lambda}}{\eta_{\text{qe}} \cdot \gamma_{\text{ff}} \cdot \tau_{\text{exp}}} \cdot \frac{4 \cdot (f/D)^2}{\eta_{\text{lens}} \cdot \pi \cdot L_{\text{obj}}}}. \quad (6)$$

3 Results for Ideal System

At first we assume ideal technology. A typical indoor scene is illuminated with a luminance of $L_v = 100$ cd m^{-2}.[9] For the peak sensitivity of the human eye at a wavelength of $\lambda = 555$ nm the SI unit candela is defined[10] as radiant intensity of $1/683$ W sr^{-1}. The radiance in object space is then

$$L_{\text{obj}} = 100 \cdot \frac{1}{683} \approx 0.146 \text{ W sr}^{-1} \text{ m}^{-2}. \quad (7)$$

We further assume a perfectly transparent lens with $\eta_{\text{lens}} = 1$, a wide aperture $f/D = 2.8$, fill factor $\gamma_{\text{ff}} = 1$ and quantum efficiency $\eta_{\text{qe}} = 1$. For achieving typical video frame rates a maximum exposure time of $\tau_{\text{exp}} = 0.03$ s is used. For a human observer, images without visible noise are preferred. From psychophysical studies a thousand-photon limit is reported as the threshold for visibility of shot noise.[5] We therefore set SNR $= \sqrt{1000} \approx 32$. With green light with $\lambda = 555$ nm the minimum pixel size calculates to $d_{\text{p,min}} = 0.9$ μm.

The influence of different apertures is shown in Fig. 2. With larger aperture diameters, even smaller pixels can be used. A variation of luminance is also possible: In practice, the human color perception (photoptic vision) starts at $L_v = 3$ cd m^{-2}.[9] The luminance in daylight exterior scenarios is typically $L_v = 10^4$ cd m^{-2}.[9] The resulting minimum pixel sizes thus range from 5 to 0.09 μm as shown in Fig. 3.

4 Radiometric Modeling

Up to now, we used monochromatic light only. We now extend this and also include the spectral distribution of light. Again, we start with a scene with a luminance of $L_v = 100$ cd m^{-2}. Now, the light is made up of radiation from a light bulb. This is modeled as a black body at a certain color temperature T and a spectral radiance of

$$L_{\text{obj},\lambda}(\lambda, T) = L_0 \cdot \frac{2hc^2}{\lambda^5} \cdot \frac{1}{e^{\frac{hc}{\lambda kT}} - 1} \; [\text{W} \cdot \text{sr}^{-1} \cdot \text{m}^{-3}]. \quad (8)$$
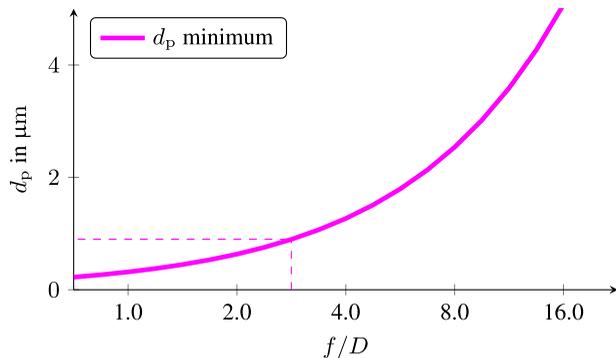

Fig. 2 Minimum pixel sizes for photon limited system with varying apertures, ideal system with SNR $= \sqrt{1000}$ and scene with luminance of $L_v = 100$ cd m^{-2}, dashed line for $f/D = 2.8$.
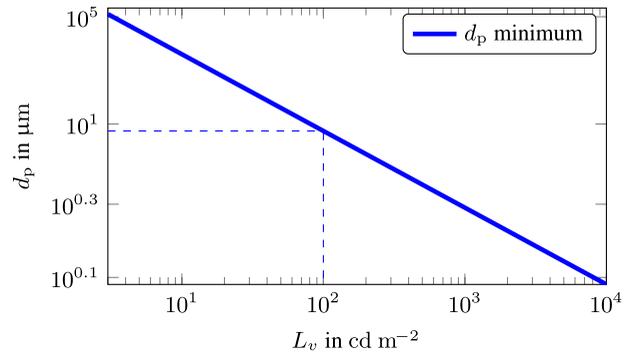

Fig. 3 Minimum pixel sizes for photon limited system with varying luminance, ideal system with SNR $= \sqrt{1000}$ and aperture $f/D = 2.8$, dashed line for $L_v = 100$ cd m^{-2}.

With the photoptic luminous efficiency function[11] V_m, we set

$$K_m \cdot \int_0^\infty V_m \cdot (\lambda) \cdot L_{\text{obj},\lambda}(\lambda, T) \, d\lambda \overset{!}{=} L_v, \quad (9)$$

with $K_m = 683$ lm W^{-1}. The resulting normalized spectral radiance $L_{\text{obj},\lambda}(\lambda, T)$ is now perceived by the human eye as a luminance of $L_v = 100$ cd m^{-2}. Figure 4 shows the resulting set of normalized spectral radiances for typical color temperatures.

Today, most cameras are used to capture scenes for later viewing by a human. The camera should therefore create a representation of the scene that is similar to that of the human visual system. We simulate an ideal camera with the spectral sensitivity curves based on the Stockman and Sharpe cone measurements of the human eye.[12] The corresponding spectral sensitivity functions for long (L), medium (M) and short (S) wavelengths are shown in Fig. 5. However, we assume an ideal camera with ideal color filters and material without any attenuation ($\eta_{\text{qe}} = 1$) at peak efficiency.

In Table 1, the resulting minimum pixel sizes are shown for the radiometric simulation. The luminosity case with monochromatic light at $\lambda = 555$ nm corresponds to the ideal simulation from above. There is less than 10% error for the simulation with L and M cones compared to the luminosity. This is plausible from the high similarity of the respective sensitivity curves. However, the capturing of blue light (short wavelengths with cone S) requires larger pixels. At short wavelengths, the individual photons have a higher energy and thus, there are fewer for a given radiant flux. This explains the problem of inferior performance of
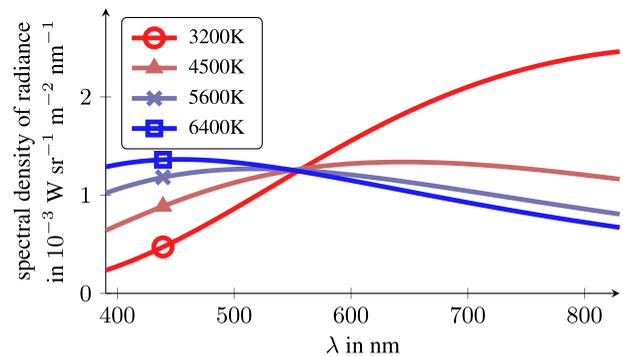

Fig. 4 Spectral radiance of black bodies with temperatures T, intensity scaled to be perceived as luminance of 100 cd m^{-2}.
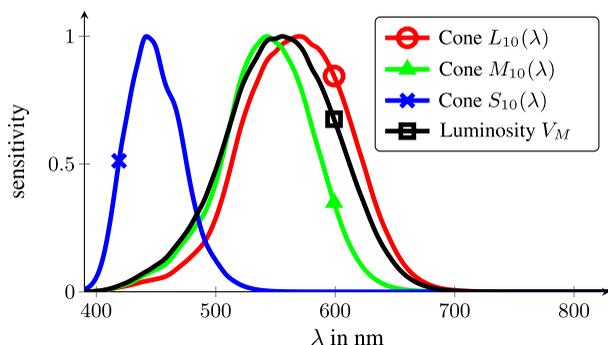
Fig. 5 Sensitivity functions of 10-deg cone fundamentals for *L*, *M* and *S* cone and luminous efficiency function V_m.

Table 1 Minimum pixel sizes (in μm) based on radiometric calculations for light sources with black body radiation of temperature *T* and monochromatic light source.

Light source	Cone *L*	Cone *M*	Cone *S*	Luminosity
$T = 3200$ K	0.86	1.04	2.13	0.89
$T = 4500$ K	0.89	1.02	1.64	0.90
$T = 5600$ K	0.90	1.01	1.45	0.90
$T = 6400$ K	0.91	1.01	1.36	0.90
$\lambda = 555$ nm	0.92	0.92	26.00	0.90

blue color channels in typical digital cameras. The extreme case of observing monochromatic green light with a short wavelength sensitivity leads to even fewer photons and would require pixels with 26 μm. In general, the monochromatic calculation is only slightly optimistic but gives a good approximation to a radiometric computation.

5 Results with Current Technology

The above numbers represent the theoretical limit for ideal sensors. In practice, a real world camera does not achieve these numbers. For example, a highly optimized three layer stacked image sensor is reported by Hannebauer et al.[13] For pixels of size $d_p = 4.8$ μm a high fill factor of $\gamma_{ff} = 0.95$ and quantum efficiency of $\eta_{qe} = 0.8$ is possible with many (costly) optimizations. In current 1.4 μm consumer grade sensors the *backside illumination* (BSI) technology enables close to 100% fill factor.[14] For color imaging, the spectral sensitivity is not without attenuation and peak quantum efficiencies of about $\eta_{qe} \approx 0.5$ are reported by OmniVision[14] and Aptina.[15] In scientific CMOS sensors, the combined sensor readout noise is reported as low as 1.3 electrons/pixel[16] and can thus be neglected among 1000 electrons. The combined assumption of $\eta_{lens} = 0.95$, $\gamma_{ff} = 0.95$ and $\eta_{qe} = 0.5$ leads to a minimum pixel size of $d_{p,min} = 1.34$ μm. With mass-market sensors and additional noise,[8] larger pixels are required.

These small pixels also reach another technological limit of decreasing full well capacity. For example Aptina reports[15] $C = 5000$ electrons, which leaves only a dynamic range of $5{:}1$ from noise visibility[5] to overexposure. As a result, most of the image will still look noisy. However,

this is a technological challenge that could be addressed with multiple readouts during the exposure.[17]

Another limitation comes with optical diffraction. Even in ideal optics the achievable resolution of a camera system is limited. The Sparrow criterion suggests[3] that there is no gain in resolution below a critical pixel size of $d_{p,crit} = \frac{\lambda}{2} \cdot f/D$. For our example of $f/D = 2.8$ and $\lambda = 555$ nm, we obtain $d_{p,crit} = 0.78$ μm. Achieving this limit, however, is challenging, especially in the off-axis field, and leads to expensive optics. A further decrease in aperture requires a dramatic increase of the technological efforts and smaller tolerances for optics manufacturers.

6 Conclusion

In our photometric analysis, we discuss the number of photons per pixel. With small pixels the image quality is limited by shot noise, and for indoor scenarios the current video cameras are surprisingly close to this fundamental limit. We estimate that even with ideal technology, a pixel size below $d_p = 0.9$ μm will not capture enough light to generate visually pleasing videos any more. Current technology is far from perfect and with optimistic assumptions, the limit at $d_p = 1.34$ μm is close to current sensors. However, for other imaging scenarios like outdoor daylight still photography, there is plenty of room at the bottom.

References

1. R. Fontaine, "A review of the 1.4 um pixel generation," in *Int. Image Sensor Workshop (IISW)*, Hokkaido, Japan (2011).
2. J. B. Jonas, U. Schneider, and G. O. H. Naumann, "Count and density of human retinal photoreceptors," *Graefes Arch. Clin. Exp. Ophthalmol.* **230**, 505–510 (1992).
3. J. Goodman, *Introduction to Fourier Optics*, Roberts & Company Publishers, Englewood, Colorado, USA (2005).
4. J. Farrell et al., "A simulation tool for evaluating digital camera image quality," in *SPIE Electron. Imag.—Image Quality and System Performance* **5294**, 124–131 (2004).
5. F. Xiao, J. Farrell, and B. Wandell, "Psychophysical thresholds and digital camera sensitivity: the thousand photon limit," in *SPIE Electron. Imag.—Digital Photography* **5678**, 75–84 (2005).
6. R. Kingslake, *Optical System Design*, Academic Press, London **1** (Oct. 1983).
7. B. Fowler et al., "A method for estimating quantum efficiency for CMOS image sensors," in *SPIE Electron. Imag.—Solid State Sensor Arrays: Development and Applications II* **3301**, 178–185 (1998).
8. R. Gow et al., "A comprehensive tool for modeling CMOS image-sensor-noise performance," *IEEE Trans. Electron. Devices* **54**, 1321–1329 (June 2007).
9. W. Smith, *Modern Optical Engineering: The Design of Optical Systems*, Tata McGraw-Hill Education, Englewood, Colorado, USA (1990).
10. P. Giacomo, "News from the BIPM: resolution 3—definition of the candela," *Metrologia* **16**(1), 55–61 (1980).
11. L. Sharpe et al., "A luminous efficiency function, $V*(\lambda)$, for daylight adaptation," *J. Vision* **5**(11), 948–968 (2005).
12. A. Stockman and L. Sharpe, "The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype," *Vis. Res.* **40**(13), 1711–1737 (2000).
13. R. Hannebauer et al., "Optimizing quantum efficiency in a stacked CMOS sensor," in *SPIE Electron. Imag.—Sensors, Cameras, and Systems for Industrial, Scientific, and Consumer Applications XII* **7875**(1), 787505 (2011).
14. H. Rhodes et al., "The mass production of second generation 65 nm BSI CMOS image sensors," in *Int. Image Sensor Workshop (IISW)*, International Image Sensor Society (IISS) (2011).
15. G. Agranov et al., "Pixel continues to shrink . . . pixel development for novel CMOS image sensors: a review of the 1.4 um pixel generation," in *Int. Image Sensor Workshop (IISW)*, International Image Sensor Society (IISS) (2011).
16. B. Fowler et al., "A 5.5 mpixel 100 frames/sec wide dynamic range low noise CMOS image sensor for scientific applications," in *SPIE Electron. Imag.—Sensors, Cameras, and Systems for Industrial/Scientific Applications XI* **7536**, 753607 (Jan. 2010).
17. M. Schöberl et al., "Digital neutral density filter for moving picture cameras," in *SPIE Electron. Imag.—Computational Imag. VIII* **7533**, 75330L (January 2010).