Toward a digital camera to rival the human eye

Orit Skorka
Dileepan Joseph

Toward a digital camera to rival the human eye

Orit Skorka
Dileepan Joseph
University of Alberta
Electrical and Computer Engineering
Edmonton, Alberta, T6G2V4 Canada
E-mail: orsk@ece.ualberta.ca

Abstract. *All things considered, electronic imaging systems do not rival the human visual system despite notable progress over 40 years since the invention of the CCD. This work presents a method that allows design engineers to evaluate the performance gap between a digital camera and the human eye. The method identifies limiting factors of the electronic systems by benchmarking against the human system. It considers power consumption, visual field, spatial resolution, temporal resolution, and properties related to signal and noise power. A figure of merit is defined as the performance gap of the weakest parameter. Experimental work done with observers and cadavers is reviewed to assess the parameters of the human eye, and assessment techniques are also covered for digital cameras. The method is applied to 24 modern image sensors of various types, where an ideal lens is assumed to complete a digital camera. Results indicate that dynamic range and dark limit are the most limiting factors. The substantial functional gap, from 1.6 to 4.5 orders of magnitude, between the human eye and digital cameras may arise from architectural differences between the human retina, arranged in a multiple-layer structure, and image sensors, mostly fabricated in planar technologies. Functionality of image sensors may be significantly improved by exploiting technologies that allow vertical stacking of active tiers.* © *2011 SPIE and IS&T.* [DOI: 10.1117/1.3611015]

1 Introduction

Despite significant progress in the area of electronic imaging since the invention of the CCD more than 40 years ago, digital cameras still do not rival the human eye. Consider that, in his keynote address to the 2010 International Solid-State Circuits Conference, Tomoyuki Suzuki, the Senior Vice-President of Sony, said "In developing the CMOS image sensor, the goal is exceeding human vision."[1] Toward that end, this work introduces a method to evaluate the performance gap between a digital camera and the human eye. A clear definition and quantification of limiting factors will help design engineers realize a digital camera to rival the human eye.

The large diversity in design and fabrication technologies for electronic image sensors encouraged many research groups worldwide to develop performance evaluation methodologies for digital cameras or image sensors. Franz et al.,[2] for example, suggested a method that mainly considers the modulation transfer function (MTF) and the signal-to-noise ratio (SNR). Rodricks and Venkataraman[3] introduced a method that includes metrics such as dark noise, linearity,

SNR, and MTF, and compared the response of a camera with a CCD sensor to that of a camera with a CMOS sensor at various integration time and ISO speed values. The European Machine Vision Association (EMVA) developed the EMVA Standard 1288,[4] for "Characterization and Presentation of Specification Data for Image Sensors and Cameras," to unify the way image sensors are tested and evaluated. Spivak et al.[5] analyzed high dynamic range image sensors, while focusing on SNR, dynamic range (DR), and sensitivity. Janesick[6] compared the performance of a back-illuminated CCD image sensor to that of linear CMOS image sensors with photodiode and pinned-photodiode configurations. Unlike the other works, he does define a figure of merit for an image sensor, which is determined by the SNR. This approach, however, is limited because it discounts other important factors, such as sensor dimensions and power consumption, that also affect the overall performance. Most importantly, none of the above methods uses a benchmark to evaluate electronic image sensors or imaging systems.

In research on image quality assessment, rendering techniques, and display technology, a benchmark is very well defined. Systems have always been evaluated according to how they match the characteristics of the human visual system (HVS).[7] Brémond et al.,[8] for example, presented a method for evaluation of still images that is composed of three indices of HVS image quality metrics: visual performance, visual appearance, and visual attention. Ma et al.[9] proposed a method for quality assessment of still images and video frames that is based on human visual attention. With any nonartificial image that is displayed in a digital format, the process starts when a scene is captured by an electronic imaging system. This is followed by digital signal processing (DSP), which includes steps such as tone mapping and compression. If the performance of the imaging system is such that a large amount of information is lost, even when sophisticated DSP algorithms are used, a high quality reconstruction of the original scene is infeasible.

In this work, digital cameras are evaluated with respect to the human eye. A demand for a digital camera that can successfully compete with the human eye exists in a large range of applications, varying from consumer electronics to machine vision systems for robotic modules. The work reviews the structure and operating principles of the human eye, and discusses performance measures and testing techniques that are used with human observers. For each one of the identified parameters, the work specifies the performance of the

eye in healthy adults. This value is used as the performance benchmark for digital cameras. To fairly compare between imaging systems of different types, the evaluation is always referred to the scene, i.e., the imaging system input. The process is concluded with a single figure-of-merit (FOM), which is given in decibel (dB) for an easy representation of large numbers.

Section 2 discusses considerations in the selection of parameters for the evaluation process. It also explains how each one of the chosen parameters is assessed for the human eye and for digital cameras. Section 3 presents the results obtained after the proposed method was applied to 24 modern electronic image sensors of diverse types and sources, assuming an ideal thin lens is used to form a complete digital camera. Section 4 discusses past trends in the area of electronic imaging, using them to predict which current trends are likely to boost the performance of digital cameras, by overcoming their most limiting factors, and to become a future trend in electronic imaging. Finally, Section 5 concludes the work.

2 Method

Various parameters can be used for characterization of imaging systems. They include cost, physical properties such as weight and size, power consumption, visual field, spatial resolution, temporal resolution, parameters related to signal and noise power, and color mapping. The performance evaluation presented here considers only some of these properties, but the method may be readily extended in the future.

Cost is excluded because it is not fixed for a digital camera. The price varies (usually drops) with time, and also depends on the place of purchase. Moreover, the human eye is priceless, and so all digital cameras are infinitely cheaper by comparison. Weight and size are excluded because they depend

on the design of the camera body and on the lens used. There are too many optional combinations for a given image sensor, and so these physical properties are excluded for simplicity. To narrow the scope of this work further, color mapping is not included in the performance evaluation method. All image sensors are treated as monochrome ones, i.e., they are evaluated according to their response to a varying intensity of white light. In general, color mapping of a digital camera depends not only on the image sensor, but also on the algorithm used for image processing and display. Therefore, various options can be applied to the same image sensor.

The eight parameters that are considered for performance evaluation are: power consumption (PC), visual field (VF), spatial resolution (SR), temporal resolution (TR), SNR, signal-to-noise-and-distortion ratio (SNDR), DR, and dark limit (DL). Two parameters, VF and SR, are related to the geometry of the imaging system, and the last four parameters are related to signal and noise power. To guarantee that performance is not limited by lens imperfections, and for simplicity, an ideal thin lens is assumed to complete the digital camera. In cases where the imaging system includes DSP, the properties of the image that is read after the DSP is applied are considered for the evaluation.

The evaluation process involves mathematical operations and, most importantly, calculation of ratios. Therefore, it requires explicit values to be defined for the parameters of all imaging systems, including the human eye. Consequently, the performance of the human eye is reviewed below, with reference to experiments done using human cadavers and observers.

2.1 Power Consumption

Figure 1 shows diagrams of (a) the human eye and (b) a digital camera. There are many similarities between the two
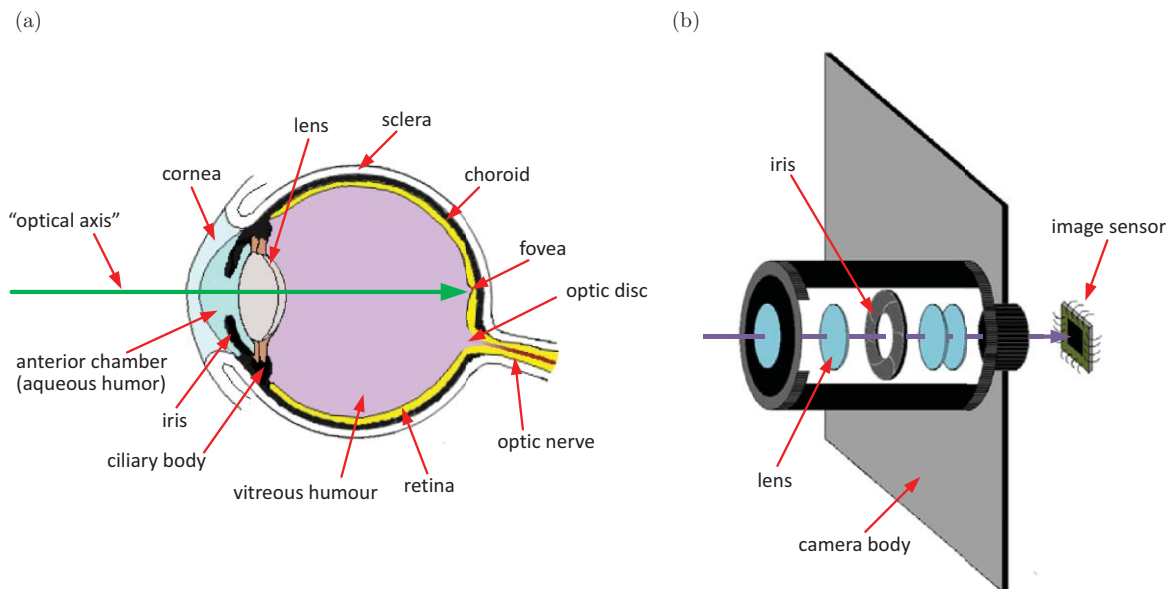

Fig. 1 (a) Horizontal cross-section of the right eye (top view). Light enters the eye through the iris aperture (the pupil). It is refracted by the cornea and the lens before reaching the retina, which is a light-sensitive tissue. The fovea is rich in cone photoreceptors; it is responsible for sharp central vision. (b) In digital cameras, light enters the camera body through the aperture in the iris diaphragm. It is refracted by a system of lenses before reaching an image sensor located in the focal plane.

Table 1 Approximate basal metabolic rates of different tissues that compose the human eye (Ref. 11).

Type	Tissues	Metabolic rate (W/m^3)
Passive	cornea, lens, sclera, choroid, vitreous humor, aqueous humor	0
Muscular	lens zonules, scleral muscle, ciliary muscle, iris dilator muscle, iris sphincter muscle	690
Neural	retina (including fovea)	10,000

systems. Both include an iris with an aperture, which is called a pupil. The sclera and the camera body prevent light from entering the imaging system from any direction other than the pupil. Refraction of light in the human eye is performed by the cornea and the lens. While the cornea has a fixed focal length, the focal distance of the lens can be varied by muscles that change its shape. In digital cameras, refraction of light is done by a system of lenses. The retina is a light sensitive tissue; it is composed of a complex layered structure. Optics of the eye form an image that appears upside down on the retina. The retina converts the image into electrical signals that are sent to the brain through the optic nerve. The image sensor plays an equivalent role in digital cameras. It converts photons into electrical signals, which are then read out through electrical wires.

Comparison between the power consumption of a biological system and an electronic system that serve similar roles is established in electronic engineering. It is noteworthy when biological systems can fulfill the same function while consuming less power. Mead, who did extensive research on biologically-inspired electronics, emphasizes a comparison of power consumption in his book "Analog VLSI and Neural Systems."[10] He went on to found Foveon, a digital camera company.

The power consumption of the human eye can be estimated using the basal metabolic rate of the different tissues from which it is composed. Some of the tissues, such as the sclera and the lens, have a very low metabolic rate and, therefore, can be categorized as "passive." There are also muscles that take part in the functionality of the eye, such as those that control the pupil and the lens. Their metabolic rate can be taken as that of typical muscle tissue in the human body. The retina is the most "power hungry" tissue in the human eye. Its metabolic rate is considered to be equal to that of brain gray matter. The metabolic rates of various tissues in the human eye are given in Table 1, as reported by DeMarco et al.[11]

To estimate the power consumption of the human eye, one needs to assess the volume of the different tissues. Straatsma et al.[12] report statistical details about the dimensions of the retina based on measurements done with 200 eyes from human cadavers. A mean value of 1340 mm^2 was calculated for the internal surface area of the retina from the reported data. The retinal thickness in healthy adults is 220 μm.[13,14] This shows that the volume of the retina

is approximately 295 mm^3. Therefore, the power consumption of the retina is approximately 3 mW. The metabolic rate of the muscle tissue is about one-tenth of the retinal metabolic rate. Assuming the overall volume of muscle tissue in the human eye is less than 30 times the retinal volume, the power consumption of the eye is less than 10 mW.

The power consumption of image sensors depends on the technology used for device fabrication and on the circuit design. In general, it increases with frame size and frame rate because more power is consumed with an increase in capacitance that needs (dis)charging, and with an increase in cycle frequency. The power consumption of image sensors is obtained from datasheets or other publications for one or more frame rates chosen by the manufacturer or author.

2.2 *Visual Field*

The visual field of an imaging system is the overall volume "viewed" by the system. The clinical method used to evaluate the human visual field is called perimetry. Various techniques and instruments have been developed for this purpose; they can be categorized in several ways. The Goldmann perimeter is an example of an instrument for manual perimetry, while the Octopus perimeter and the Humphery visual-field analyzer are examples of computer-controlled instruments. In kinetic perimetry, a stimulus of a known and constant luminance moves at a steady speed from an area outside the patient's visual field to an area inside. The patient is asked to report when the stimulus has been perceived. In static perimetry, there are multiple stimuli with fixed locations. However, the luminance may be either constant or varied.[15] Threshold strategies determine the threshold luminance of the patient at different locations by gradually increasing the luminance intensity at fixed points until they are perceived. In suprathreshold strategies, the patient is presented with stimuli luminance above normal threshold values at various locations in his or her visual field.[16]

The monocular visual field of a healthy adult extends approximately 50 deg to 60 deg superiorly and 70 deg to 75 deg inferiorly in the vertical direction,[15,16] as shown in Fig. 2(a). Horizontally, it extends about 60 deg nasally and 90 deg to 100 deg temporally, as shown in Fig. 2(b). The optic disk, which is the exit point of the optic nerve from the eye to the brain [see Fig. 1(a)], lacks photoreceptors. This causes a "blind spot" in the visual field that is located between 10 deg and 20 deg temporally. The binocular visual field extends about 200 deg horizontally because there is an overlap in the nasal monocular visual field of the two eyes, and each eye covers about 100 deg temporally. Vertical extent of binocular vision is similar to that of monocular vision.

To allow a fair comparison between the visual field of imaging systems of different types, the solid angle subtended by the imaging system is calculated. Moreover, an expression of the visual field as a solid angle emphasizes the fact that an imaging system captures a three-dimensional (3D) spatial volume and not a two-dimensional (2D) area. To simplify the calculations, which are required to estimate the solid angle subtended by the human eye, its visual field is treated as a right elliptical cone [see Fig. 2(c)]. The solid angle subtended by an elliptical cone with opening angles $\theta_\perp$ and $\theta_\parallel$ ($\theta_\perp \geq \theta_\parallel$) is given by[17]
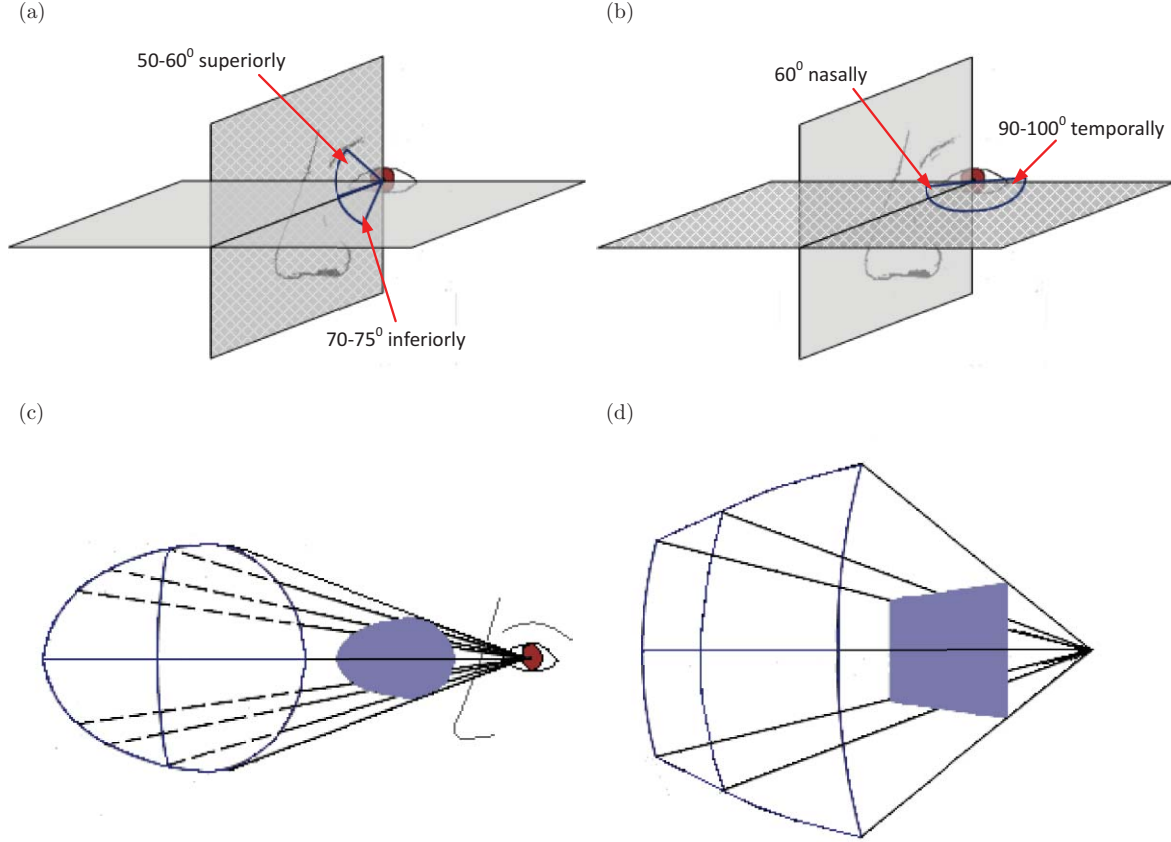
Fig. 2 (a) The visual field of a healthy adult extends, in the vertical direction, approximately 50 deg to 60 deg superiorly and 70 deg to 75 deg inferiorly. (b) It extends, in the horizontal direction, about 60 deg nasally and 90 deg to 100 deg temporally. (c) For simplicity, the solid angle viewed by the human eye is taken as the one seen by a right elliptical cone. (d) The solid angle viewed by a digital camera is the one seen by a right rectangular pyramid. It varies with image sensor dimensions and lens focal length.

$$\Omega = 2\pi \left[1 - \Lambda_0(\varphi, \alpha) \right], \tag{1}$$

where

$$\varphi \equiv \frac{1}{2}(\pi - \theta_\parallel), \tag{2}$$

$$(\sin\alpha)^2 \equiv 1 - \left[\frac{\cos(\theta_\perp/2)}{\cos(\theta_\parallel/2)} \right]^2. \tag{3}$$

Λ_0 is Heuman's lambda function, which is given by[18]

$$\Lambda_0(\varphi, \alpha) = \frac{2}{\pi} K(\alpha) E\left(\varphi, \pi/2 - \alpha\right) -$$

$$\frac{2}{\pi} \left[K(\alpha) - E(\alpha)\right) F\left(\varphi, \pi/2 - \alpha\right], \tag{4}$$

where $K(\alpha)$ and $E(\alpha)$ are complete elliptic integrals of the first and second kind, respectively, and $F(\varphi, \phi)$ and $E(\varphi, \phi)$ are incomplete elliptic integrals of the first and second kind, respectively.

To estimate the solid angle subtended by the monocular visual field of the human eye, the vertical opening angle, $\theta_\parallel$, is taken as 127.5 deg, and the horizontal opening angle, $\theta_\perp$, is taken as 155 deg. The blind spot in the visual field is ignored. A calculation performed using MATLAB functions for the elliptic integrals shows that the human eye captures a solid angle of 4.123 sr, i.e., $\Omega_{\text{eye}} \approx 4.1$ sr.

Conventional image sensors are rectangularly shaped. Therefore, a digital camera views a solid angle of a right rectangular pyramid, as shown in Fig. 2(d). Its visual field depends on the width, w, and length, ℓ, of the sensor array, and on the focal length of the lens, f_ℓ. In this pyramid, the image sensor is considered as the base, and the apex is located at a distance f_ℓ from the center of the base. The solid angle subtended by the image sensor, Ω_s, is given by[19]

$$\Omega_s = 4 \arcsin\left[\sin(\alpha_w/2) \cdot \sin(\alpha_\ell/2) \right], \tag{5}$$

where $\alpha_w = 2 \arctan\left[w/(2f_\ell) \right]$ and $\alpha_\ell = 2 \arctan\left[\ell/(2f_\ell) \right]$ are the apex angles of the right rectangular pyramid.

2.3 Spatial Resolution

The spatial resolution of an imaging system represents the finest detail or the highest spatial frequency that can be perceived by the system. It may be derived from the system's response to varying spatial frequency. For simplicity, the spatial resolution is examined here only for one-dimensional (1D) patterns. The spatial response, for each spatial frequency, is defined through the Michelson contrast ratio, C, which is given by

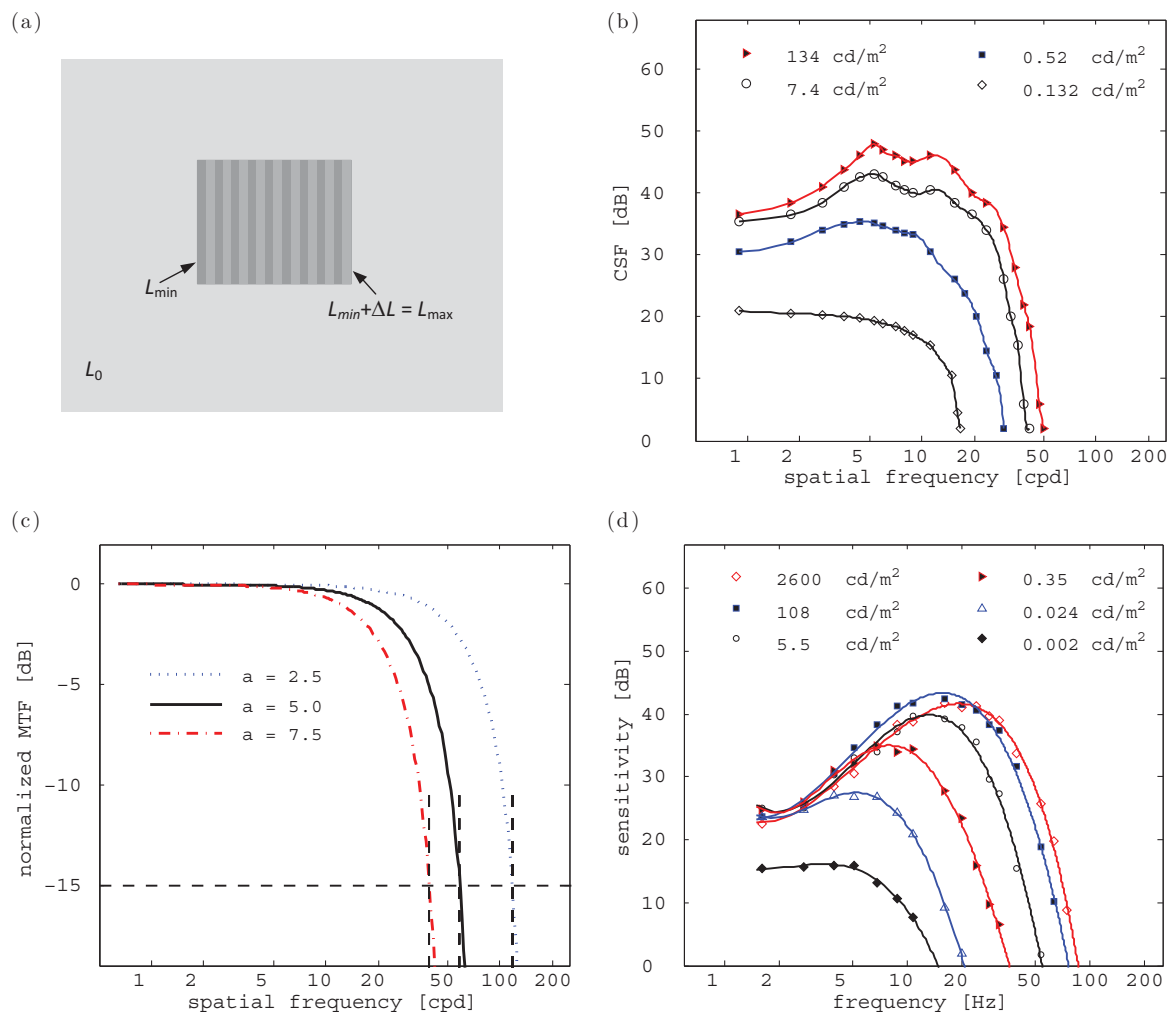$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \tag{6}$$

Fig. 3 (a) A pattern to test the spatial response of human observers. L_0 is the adaptation luminance. The examiner varies ΔL until a threshold level, ΔL_{th}, can be determined. (b) The CSF of an observer with good vision at four adaptation levels.[21] (c) The normalized MTF of a digital camera with $f_\ell = 20$ mm, $p = 10\,\mu$m, and $a = 2.5$, 5, and $7.5\,\mu$m. Spatial frequencies at 15 dB deterioration are indicated. (d) The temporal response of an observer with good vision, using data from Kelly (Ref. 22).

This is a general definition, where the term I can refer to various quantities. Although the spatial response is defined for sinusoidal-wave spatial patterns, testing is often performed with square-wave spatial patterns.

The spatial resolution of the eye is not uniform across the whole visual field. It is sharpest at the center and gradually declines toward the periphery or, as described by Traquair, it is "an island of vision or hill of vision surrounded by a sea of blindness."[20] This evolves from the nonuniformity of the retina. The fovea is the retinal area that is responsible for sharp central vision. Sharp vision of the whole visual field is obtained thanks to saccades, which are constant fast eye movements that bring different portions of the scene to the fovea.

A test pattern to characterize the spatial response of human observers is shown in Fig. 3(a), where L_0 is the adaptation luminance and $\Delta L = L_{\max} - L_{\min}$ is the tone difference. For a central pattern with a constant spatial frequency, $f_{\text{x-sc}}$, the examiner varies ΔL and the observer is asked to specify whether the tone difference is perceived. This is repeated

until a threshold value, ΔL_{th}, usually defined as the level with 50% probability of detection, can be determined. For a comprehensive characterization of an observer's spatial response, tests need to be repeated with different $f_{\text{x-sc}}$ and L_0 conditions.

The threshold values can be used to construct a plot of the contrast sensitivity function (CSF), which is the inverse of the threshold contrast, C_{th}, as given by

$$CSF = \frac{1}{C_{\text{th}}} = \frac{L_{\max} + L_{\min}}{\Delta L_{\text{th}}}. \tag{7}$$

Figure 3(b) shows the CSF of a young adult observer at four different adaptation levels. The spatial frequency is given in units of cycles per degree (cpd). The plot was constructed according to data presented by Patel.[21] The experiment was performed with an artificial pupil, whose diameter was smaller than the natural pupil diameter even at very bright conditions. Although original results were given in Td (Trolands), they were converted to cd/m². One Td, which represents the retinal illuminance, equals the scene lumi-

nance in cd/m^2 multiplied by the area of the pupil's aperture in mm^2. Conversion was done according to interpolation of data presented by De Groot and Gebhard[23] on the relationship between pupil diameter and scene luminance. The retinal illuminance (Td) was calculated from the expected pupil diameter (mm) at the corresponding scene luminance (cd/m^2). Figure 3(b) shows that spatial response of the human eye improves with increasing L_0.

To define a benchmark for the spatial response of the HVS, Snellen visual acuity charts are considered. These charts have been used since the 19th century as an evaluation standard for spatial response of human observers. During examination, the observer is asked to identify letters, or other patterns, of different sizes from a distance of 20 ft (or 6 m). Those who can clearly perceive the letters in the row that matches visual acuity of 20/20 are considered to have good acuity. The letter E in that row corresponds to a spatial frequency of 30 cpd. One may conclude from Fig. 3(b) that, at 30 cpd, the spatial response of the HVS in typical office luminance, i.e., about 100 cd/m^2, is about 15 dB lower than its maximal value. Therefore, for the performance evaluation, the spatial frequency at which the spatial response of a digital camera deteriorates by 15 dB is compared to the HVS benchmark of 30 cpd.

Unlike testing methods with human observers, which have to be based on threshold values, with digital cameras one may gradually change the spatial frequency of the scene stimulus and measure the system response directly. However, as results of such measurements are often not published, the spatial response is estimated here by calculations based on system geometry. The overall spatial response of a digital camera is affected by the response of the image sensor and the response of each component that is placed in the path of the light, such as lenses and filters.

The image sensor is usually composed of a 2D array of rectangularly shaped pixels. The pattern of light formed on the image plane is sampled in each pixel by the photodetector. Because image sensors may be considered as linear and space-invariant (isoplanatic) systems,[24] the sampled image is a two-dimensional convolution of the pattern on the image plane with the impulse response of the array, also called the point spread function. The impulse response is derived from the sampling pattern of the array.

The Fourier transform of the impulse response is called the optical transfer function (OTF). It defines the image sensor response to varying spatial frequencies along the x and y axes on the image plane, f_x and f_y, respectively, which are expressed in units of line-pairs per mm (lp/mm). The MTF is the magnitude of the OTF.

To simplify calculations, the MTF is given here for one-dimensional patterns and does not consider aliasing. Yadid-Pecht[25] shows that, for an array with pixel pitch p and photodetector pitch a, the 1D MTF is given by

$$\text{MTF}(f_x) = \frac{a}{p} \text{sinc}(af_x), \tag{8}$$

where

$$\text{sinc}(x) = \begin{cases} \sin(\pi x)/(\pi x), & x \neq 0; \\ 1, & x = 0. \end{cases} \tag{9}$$

The lens MTF needs to be multiplied by the MTF of the image sensor to calculate the spatial response of the digital camera. However, the lens is considered here to have an ideal spatial response, i.e., to have a unity MTF. While this is a simplification, the actual spatial resolution of the digital camera would not surpass the value taken here for the performance evaluation.

To refer the spatial response of the digital camera to the scene, the spatial frequency f_x in lp/mm on the image plane needs to be converted to $f_{x\text{-}sc}$ in cpd of the scene. Because an opening angle of 1 deg corresponds to a length of $d = 2f_\ell \tan 0.5$ deg on the image plane, $f_{x\text{-}sc} = f_x d$.

The MTF is often normalized in order to represent the contrast ratio, C, of the frequency response. Theoretically, one concludes from Eq. (8) that the bandwidth of the spatial response improves without limit as the photodetector pitch diminishes. However, practically, the photodetector must be large enough to absorb an adequate number of photons to generate a detectable electrical signal. Figure 3(c) shows the normalized MTF as calculated for a digital camera, where the sensor has a pixel pitch of $p = 10\,\mu m$ and the lens has a focal length of $f_\ell = 20$ mm. Calculations were performed for three values of photodetector pitch. The spatial frequency at 15 dB deterioration, which is considered here as the spatial resolution, is shown for each case.

To conclude, the spatial response of the HVS can be extracted from experiments performed with human observers. The benchmark is determined using a commonly used standard—the 20/20 row of the Snellen chart. With digital cameras, although the spatial response can be measured directly, experimental results are usually not provided. Therefore, it is estimated here according to the image sensor geometry and the given focal length.

2.4 Temporal Resolution

The temporal response of the human eye has been tested by different groups worldwide, who have experimented with human observers. In general, the published works agree that the response improves with retinal illuminance, and that temporal changes with frequencies greater than 80 to 90 Hz cannot be detected even at high luminance levels.

The method used to test the temporal response somewhat resembles the one used to test the spatial response of human observers. The main difference is that instead of working with spatial variations in frequency and contrast, these factors are now required to vary with time. Kelly,[22] for example, performed experiments using an apparatus that generates a stimulus in the form of

$$L(t) = L_0[1 + m \cdot \cos(2\pi f t)], \tag{10}$$

where L_0 is the adaptation luminance, m is the modulation amplitude, and f is the frequency. The modulation amplitude is an example of the Michelson contrast ratio defined in Eq. (6), i.e.,

$$m = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}, \tag{11}$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum of $L(t)$, respectively.

During the experiment, while L_0 and f are kept constant, the observer is asked to report whether temporal changes can

be detected for different *m* values. The threshold modulation amplitude, m_{th}, is defined as the minimal *m* that the observer requires to be able to notice temporal changes in the scene, and the ratio between L_0 and m_{th} defines the sensitivity. Experimental results are shown in Fig. 3(d). As before, the original data was given in Td and was converted to cd/m² based on the dimensions of the artificial pupil used in the experiment. Results show that, in typical office conditions, the human eye can detect temporal changes up to frequencies of 65 Hz. This value is used in the performance evaluation.

The temporal resolution of a digital camera is proportional to the frame rate of its image sensor, as indicated by the manufacturer. According to the Nyquist theorem, in order to reconstruct a signal, the sampling rate needs to be at least two times higher than the highest frequency it contains. Therefore, the highest temporal frequency that can be properly captured by a digital camera, and which determines its temporal resolution, equals half the frame rate of its image sensor in the best case.

2.5 *Signal and Noise Power*

The signal and noise power of an imaging system determines four important measures: the SNR, the SNDR, the DR, and the DL. Noise sources can be found in the imaging system and in the scene. For a fair comparison, all the noise is referred to the scene, i.e., the input of the imaging system.

Two types of noise affect the performance of imaging systems: temporal noise and fixed pattern noise (FPN). In this paper, the SNR considers only the temporal noise. The SNDR considers both temporal noise and FPN, where the two are assumed to be uncorrelated. In some works,[5,26] both temporal noise and FPN are included in the definition of SNR, i.e., their definition of SNR is equivalent to SNDR here. The SNDR of an imaging system cannot be greater than its SNR under the same operating conditions.

FPN exists in any array of analog or mixed-signal sensors due to inherent variability. With digital cameras, it is reduced by methods such as correlated double-sampling (CDS) and/or calibration. However, the residual FPN causes distortion. With the human eye, FPN is avoided because the retina is sensitive to the temporal derivative of the stimulus intensity and not to the intensity itself.[27] One proof of this mechanism is that static images formed on the retina without saccadic influence fade away to the observer. Although the best way to represent the ratio between signal and noise power is through the SNDR, both SNR and SNDR are considered here because only the temporal noise is specified for some image sensors.

The dark limit is the lowest luminance level at which the SNDR begins to exceed 0 dB. At this operating point, the signal and noise power are equal. The dynamic range is the maximal range of luminances that the imaging system can safely capture with SNDR greater than 0 dB. Therefore, it is limited by the DL at one end. The bright limit (BL) of the DR is determined by the luminance level that causes the SNDR to drop abruptly to zero, or that damages the imaging system, or that causes other undesirable conditions, such as a sudden increase in distortion.

An intrascene DR and an interscene DR can be defined for every imaging system, including the human eye. The intrascene DR is the maximal luminance range that the imaging system can capture using fixed operating conditions. It may
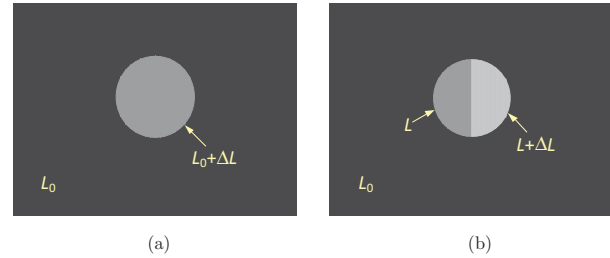

Fig. 4 Test patterns to examine (a) the interscene, and (b) the intrascene response of human observers. L_0 is the adaptation luminance. In both cases, ΔL_{th}, the minimum increment an observer requires to distinguish two luminances, is determined. Assuming that signal power equals total noise power at the threshold level, SNDR may be calculated.

depend on an adaptation point. With the human eye, the pupil size should be constant and, with digital cameras, parameters such as gain and exposure time should be constant. The interscene DR is the maximal luminance range that the imaging system can capture with adaptation included. With the human eye, time may be allowed for adjustment to new luminance conditions and, with digital cameras, operating parameters may be likewise varied. Only the intrascene DR is considered here for performance evaluation. Nonetheless, the interscene response of the human eye is explained below because it is needed to explain the intrascene response.

Measures related to signal and noise power of the human eye may be assessed using contrast-sensitivity test patterns. The test pattern shown in Fig. 4(a) can be used to examine the interscene response of human observers.[28] The observer is first given enough time to adapt to a new background luminance level, L_0, and is then asked to indicate whether the scene looks uniform or whether the central pattern is perceived. The test pattern shown in Fig. 4(b) may be used to examine the intrascene response of human observers. In this test, while L_0 and L are kept constant, ΔL is varied. The observer is asked to indicate whether the central pattern looks uniform or whether the tone difference between the two sections is perceived. Reported results are the threshold levels.

Assuming that a signal cannot be detected as long as its power is smaller than the noise power, the threshold luminance, ΔL_{th}, represents the luminance level for which the signal power and the noise power are equal. Therefore, the SNDR may be written as

$$\text{SNDR}_{\text{interscene}} = 20 \log \left(\frac{L_0}{\Delta L_{th}} \right) \qquad (12)$$

for the interscene response, and as

$$\text{SNDR}_{\text{intrascene}} = 20 \log \left(\frac{L}{\Delta L_{th}} \right) \qquad (13)$$

for the intrascene response.

Hecht[29] presents results of experiments done with human observers according to data reported by Koenig and Brodhun in 1889. The goal of these experiments was to find the interscene response of the human eye. The response was tested at luminance levels that cover the whole range in which the eye can function, from the dimmest perceptible luminance to intense levels that cause temporary blindness. The results, presented in Fig. 5(a), show that the interscene
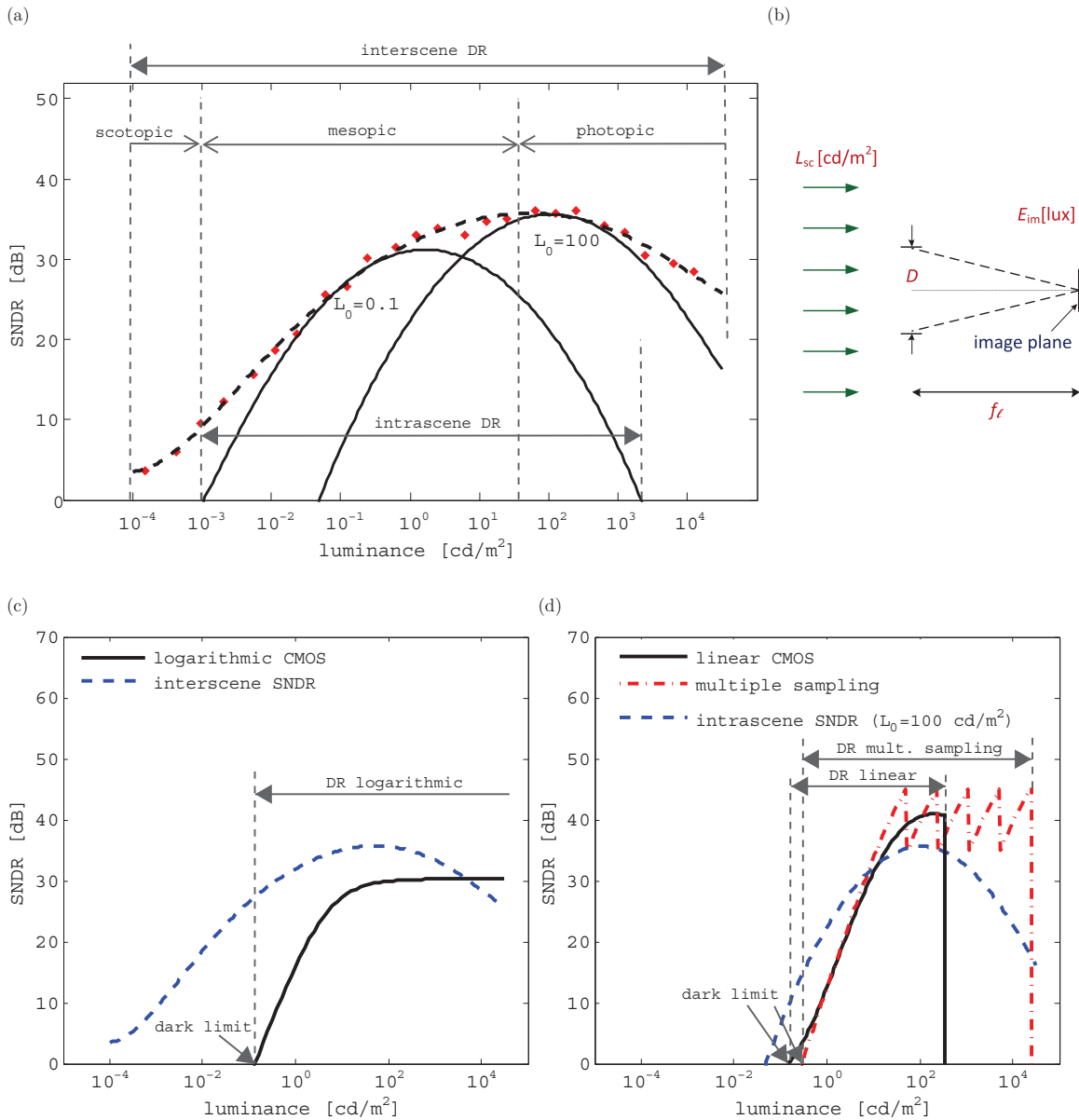
Fig. 5 (a) The interscene response obtained from experiments with human observers (Ref. 29). Using a model developed by Barten (Ref. 31) the intrascene response was calculated at adaptation levels of 0.1 and 100 cd/m² for a 4 deg ×4 deg visual field. (b) A simplified model of an imaging system that is used to refer the scene luminance, L_{sc}, to the image plane illuminance, E_{im}. (c) The interscene response of the human eye exhibits a high DR but a low SNDR, and resembles the intrascene response of logarithmic CMOS image sensors. (d) The intrascene response of CCD and linear CMOS image sensors is characterized by a low DR but a high SNDR. It is compared to the intrascene response of the human eye. When multiple sampling is used to increase the DR, the SNDR has a sawtooth pattern.

DR of the human eye extends at least 170 dB. They also show that the peak SNDR of the human eye equals 36 dB, which is reached in typical office luminance. This value is used for both peak SNR and peak SNDR benchmarks in the evaluation process.

The retina includes two types of photoreceptors: cones and rods. Cones are responsible for color vision and operate in bright light; they are mostly concentrated in the fovea. Rods are responsible for vision in dim light. The interscene response of the human eye is composed of three regions of operation.[30] Color vision, or photopic vision, occurs at luminances greater than 3 cd/m². Mesopic vision occurs at luminances between 3 and 0.001 cd/m². In this range, the

response of the eye to color gradually deteriorates. Finally, vision at luminances lower than 0.001 cd/m², in which only rods are operative, is called dark vision, or scotopic vision. As the human eye can detect color until the luminance drops to 0.001 cd/m², this value is considered as its dark limit in the performance evaluation.

The intrascene response was calculated according to a mathematical model developed by Barten,[31] based on experimental work performed by Rogers and Carel.[32] It expresses the effect of background luminance on the CSF as a Gaussian function of the logarithm of L_0 divided by L. Although originally developed for the CSF, this model may be used to estimate the intrascene SNDR because SNDR tests

are equivalent to CSF tests at low spatial frequencies. In general, the performance of the eye is best when the background luminance is similar or somewhat lower than the object luminance. Calculation of the intrascene response was performed at two adaptation levels. Results are shown in Fig. 5(a), along with the inter-scene response. The response around $100\,\mathrm{cd/m^2}$ is blocked by the BL of the human eye. The DL of the intrascene response around $0.1\,\mathrm{cd/m^2}$ corresponds to the DL of the mesopic region. The intrascene DR at this adaptation luminance covers 126 dB, and this value is used in the performance evaluation.

To estimate the dark limit of a digital camera, one needs to refer to scene luminance, L_{sc}, which is expressed in $\mathrm{cd/m^2}$, to image plane illuminance, E_{im}, which is expressed in lux, as shown in Fig. 5(b). The two measures are related as follows:

$$L_{sc} = \frac{E_{im}}{T}\frac{1}{\pi}\left(\frac{2f_\ell}{D}\right)^2 = \frac{E_{im}}{T}\frac{4f_\#^2}{\pi}, \qquad (14)$$

where T is the lens transmission, D is the aperture diameter, and $f_\# = f_\ell/D$ is the f-number of the lens. This formula is derived from Smith.[33] The dark limit of an image sensor, $E_{im\text{-}DL}$, may be calculated for integrating sensors if the sensitivity, the dark signal, and the dark noise are given, as explained in Section 3.2

CMOS image sensors with logarithmic pixel circuits are characterized by a high DR but low SNR. Their response is somewhat similar to the interscene response of the human eye [see Fig. 5(b)]. The BL of logarithmic image sensors is determined by the image plane illuminance that generates photocurrent that causes a nonlogarithmic response of the pixel circuit. Above such an illuminance, residual FPN would increase substantially. With an ideal lens model, image plane illuminance (lux) may be referred to scene luminance $(\mathrm{cd/m^2})$, using Eq. (14), for comparison with the human eye.

Although CMOS image sensors with linear pixel circuits, as well as CCD image sensors, achieve high SNDR, they have a low DR. Their BL is determined by the image plane illuminance that causes a saturating response because the charge generated during the integration time is greater than the well capacity. At this point, the SNDR drops abruptly to zero. There are several methods to increase the DR of low DR image sensors. Those based on multiple readouts of each pixel to construct one frame are characterized by a sawtooth SNDR in the region where the bright limit of the DR is extended. Such methods include multimode sensors and multiple-capture methods, which can be based on either global or local control of integration time.[5] Figure 5(c) compares the response of linear CMOS image sensors, with and without extension of DR, to the intrascene response of the human eye around $100\,\mathrm{cd/m^2}$.

Although SNR and SNDR should be represented as a curve that covers the whole DR of an image sensor, a single value is usually given in datasheets. This value is either the peak SNR or the SNR at a certain operating point chosen by the manufacturer. Similarly, data for the photoresponse nonuniformity (PRNU), which allows calculation of SNDR, is normally specified only for certain operating conditions. Assuming an ideal lens, the peak SNR, peak SNDR, and DR of an image sensor are equal to those of the digital camera that accommodates it.

2.6 *Figure of Merit*

The performance evaluation method considers eight parameters: PC, VF, SR, TR, SNR, SNDR, DR, and DL. For a digital camera, each parameter is evaluated and compared to the corresponding benchmark of the human eye. The FOM is defined as the performance ratio, expressed in decibels, of the parameter that proves to be the weakest one according to such ratios. To rival the human eye, a digital camera needs to demonstrate performance that is equivalent to or better than that of the human eye at every single measure. If the FOM of a digital camera is negative, there is at least one parameter for which the performance of the human eye exceeds that of the digital camera. If the FOM is positive, the digital camera is superior to the human eye in all eight parameters.

To refer the VF, SR, and DL to the scene, the focal length, f_ℓ, and f-number, $f_\#$, of the lens need to be specified. For simplicity, the calculations are done using a thin-lens model, which means only the image sensor specifications are required. Moreover, to eliminate the effect of lens imperfections on overall performance, the lens is assumed to be ideal, i.e., with unity transmission and unity MTF.

There is a trade-off between VF and SR, which depends on the focal length of the lens. When f_ℓ is short, as with wide angle lenses, a large volume is captured by the imaging system. However, small details would not be distinguished. Therefore, short f_ℓ results in high VF but low SR. Similarly, when f_ℓ is long, as with telephoto lenses, one can clearly see small details but the volume captured is very limited. To compare digital cameras having the same lens model, the same f_ℓ and $f_\#$ values are used with all image sensors in the performance evaluation.

The reduced eye model treats the refracting surfaces in the human eye, i.e., the cornea and the lens, as a single equivalent lens.[34] Considering refractive indices and other physical properties, one concludes that the power of the equivalent lens is 60 diopters. This gives a focal length of 17 mm. In film cameras based on the full-frame format $(24\times36\,\mathrm{mm^2})$, lenses with focal lengths of 35 to 70 mm are considered normal lenses, i.e., neither wide angle nor telephoto.[35] Electronic image sensors are usually smaller than the full-frame format. Therefore, lenses with shorter focal lengths are typically used to achieve a similar VF. A focal length of 17 mm, the equivalent focal length of the human eye, is in the range of normal lenses for electronic image sensors. Therefore, it does not give priority to the VF over the SR, or vice versa, and is taken as f_ℓ in the performance evaluation.

At the DL benchmark $(0.001\,\mathrm{cd/m^2})$ the pupil diameter of the human eye is 6.5 mm.[23] This value, along with $f_\ell = 17$ mm, results in $f_\# = 2.6$. This f-number, which lies in the 2 to 16 range that is typical for photography, is used in the performance evaluation.

3 Results

The performance of 24 modern image sensors combined with an ideal lens is evaluated with respect to the human eye. The evaluation was based on data provided in commercial datasheets or academic publications that describe the image sensors. Therefore, the main criteria in the selection of image sensors for the survey was the sufficiency of details provided in the documentation. Example calculations are shown for two of the surveyed sensors. The FOM and the two most

limiting factors are summarized for all the suveyed image sensors.

3.1 Modern Image Sensors

A list of the surveyed image sensors is presented in Table 2. The zeroth row of the table is dedicated to the human eye, which sets the benchmark for the performance evaluation. Data given for frame size and pixel pitch in this row refers to the number of cone photoreceptors and their diameter, respectively. Values are taken from Wyszecki and Stiles.[60]

Table 2 includes raw data that characterizes the image sensors. Some of the details, such as pixel size and fill factor, are required for calculations related to the performance evaluation. Parameters that do not require any complicated processing for the evaluation, such as power consumption and frame rate, are not shown in the Table 2 because they are presented in a figure (Fig. 6). Details related to control and operation of an image sensor are excluded because they are not considered in the evaluation. Raw data for calculations of signal and noise power are also excluded because there is no uniformity in the way data was measured or reported, even for image sensors of similar types. Moreover, not all parameters are relevant to all image sensors. For example, the term "full well capacity" is meaningful for CCD and linear CMOS image sensors, but not for logarithmic ones.

The image sensors included in the survey are divided into three groups. Image sensors 1 to 7 were fabricated in CCD technology. They include full-frame, frame-transfer, and interline-transfer CCD image sensors. Image sensors 8 to 16 are commercial CMOS ones, where all are based on linear integration. Image sensors 17 to 24 are called academic CMOS image sensors because they were described in scientific publications rather than commercial ones. In each group, the image sensors are presented in chronological order, which was determined by the publication date.

The design of each academic image sensor focuses on boosting one or more performance measures. Lim et al.[36] and Dubois et al.[37] target temporal resolution. Matsuo et al.[38] aimed to achieve low temporal noise by using pixel circuits based on the pinned-photodiode configuration. Kitchen et al.[39] show a digital pixel sensor with pulse-width-modulation control that is used to improve the DR. Hoefflinger[40] and Storm et al.[41] use logarithmic pixel circuits to achieve high DR image sensors. Mase et al.[42] and Lulé et al.[43] also target the DR. However, they use methods that are based on multiple integration times to construct a single frame. The image sensor described by Lulé et al.[43] is the only one not fabricated in a planar technology. It is a vertically-integrated (VI) CMOS image sensor, in which hydrogenated amorphous-silicon photodetectors were deposited on top of CMOS circuits.

3.2 Example Calculations

Calculation examples for the signal and noise power parameters are presented here for two image sensors: a commercial CCD sensor and an academic CMOS sensor that uses two modes of operation to construct a single frame.

Image sensor 6 (Texas Instruments, TC237B[44]) is a CCD sensor. With linear sensors, changes in the response are proportional to changes in the scene luminance. The DR of sensor 6 with CDS is 64 dB, and its typical full-well capac-

ity, v_{sat}, is 30 k electrons. Therefore, the rms voltage of the temporal noise in the dark, $N_{dark\text{-}rms}$, may be derived from

$$DR = 20\log\left(\frac{v_{sat}}{N_{dark\text{-}rms}}\right), \qquad (15)$$

i.e.,

$$64 = 20\log\left(\frac{30000}{N_{dark\text{-}rms}}\right), \qquad (16)$$

which results in $N_{dark\text{-}rms} = 19\,e^-$ or 0.246 mV when using the charge conversion factor $13\,\mu V/e^-$.

With image sensors that are based on integration time, i.e., CCD and linear CMOS ones, $E_{im\text{-}DL}$ may be calculated as follows:

$$S \cdot T_{int} \cdot E_{im\text{-}DL} - DS \cdot T_{int} = N_{dark\text{-}rms}, \qquad (17)$$

where S is the sensitivity, given in V/(lux s), T_{int} is the integration time, and DS is the dark signal, which is expressed in V/s. This equation shows that if $DS \cdot T_{int} \gg N_{dark\text{-}rms}$, which can be achieved with long integration times, $E_{im\text{-}DL} \approx DS/S$. Therefore, the DL cannot be improved beyond a certain limit even when T_{int} is increased.

The sensitivity of sensor 6 with an infrared (IR) filter is 32 mV/lux, and its dark signal is 1 mV. Since S and DS are expressed in mV/lux and mV, respectively, the integration time is, actually, not required for calculation of the dark limit. The datasheet does specify that a 16.67 ms exposure time was used to estimate the smear. However, it is not clear whether a similar integration time was also used to find S and DS. The minimum detectable image plane illuminance is calculated by

$$(32\,mV/\,lux) \cdot E_{im\text{-}DL} - 1\,mV = 0.246\,mV, \qquad (18)$$

which gives $E_{im\text{-}DL} = 0.039$ lux and results in $L_{sc\text{-}DL} = 0.336\,cd/m^2$.

RMS value of the distortion, σ_v^d, and rms value of the temporal noise, σ_v^t, are needed to calculate the SNDR. The datasheet specifies that the spurious nonuniformity under illumination (usually called PRNU) is 15%. No data is given for the temporal noise under illumination. Therefore, the value of $N_{dark\text{-}rms}$ is used instead, which gives:

$$\begin{aligned}
SNDR &= 20\log\left[\frac{v_{sat}}{\sqrt{\left(\sigma_v^d\right)^2 + \left(\sigma_v^t\right)^2}}\right] \\
&= 20\log\left(\frac{30000}{\sqrt{4500^2 + 19^2}}\right) \\
&= 16.5\,dB. \qquad (19)
\end{aligned}$$

Image sensor 21 (Storm et al.[41]) has two modes of operation: linear and logarithmic. At the beginning of each readout cycle, the photodiode capacitance is precharged, and the pixels are set to work in linear mode. Afterward, the pixels are switched to operate in logarithmic mode. Therefore, the linear mode targets dim scenes, whereas the logarithmic mode targets increased DR in bright scenes.

Storm et al. report that the sensitivity of the image sensor is $S = 726\,mV/(lux\,s)$, and its dark noise is $N_{dark\text{-}rms} = 0.95\,mV$. The pixel dark current is 0.388 fA. Using the specified conversion factor of $15.35\,\mu V/e^-$, the dark signal is

Table 2 The 24 image sensors included in the performance evaluation: 1 to 7 are commercial CCD, 8 to 16 are commercial CMOS, and 17 to 24 are academic (VI) CMOS image sensors. The zeroth row refers to the retina and fovea of the human eye.

Image sensor		Technology (Supply voltage) (V)	Data conversion (ADC bits)	Sensor size (frame size) mm² (Mp)	Pixel pitch (fill factor) μm (%)
0.	Human retina	organic	pixel level	1341.5 (5–6.5)	4.5–9
	Human fovea	organic	pixel level	1.77 (0.11–0.115)	1–4
1.	Kodak (Ref. 45) KAF-50100, 2010	CCD full frame (15)	board level	49.1×36.8 (50)	6
2.	Dalsa (Ref. 46) FTF2021M, 2009	CCD full frame (24)	board level	24.5×24.4 (4.2)	12 (94)
3.	Kodak (Ref. 47) KAI-02050, 2008	CCD interline (15)	board level	8.8×6.6 (1.9)	5.5
4.	Atmel (Ref. 48) AT71200M, 2003	CCD full frame (15)	board level	35×23 (8.1)	10
5.	Sony (Ref. 49) ICX285AL, 2003	CCD interline (15)	board level	8.8×6.6 (1.4)	6.45
6.	Texas Instruments (Ref. 44) TC237B, 2001	CCD frame tr. (26)	board level	4.9×3.7 (0.3)	7.4
7.	Philips (Ref. 50) FXA 1012, 2000	CCD frame tr. (24)	board level	8.2×6.6 (2.1)	5.1
8.	Hynix (Ref. 51) YACD5B1S, 2010	CMOS linear (2.8)	chip level (10)	2.8×2.1 (1.9)	1.75
9.	Samsung (Ref. 52) S5K4E1GA(EVT3), 2010	CMOS linear (2.8)	column level (10)	3.7×2.7 (5.1)	1.4
10.	Cypress (Ref. 53) Lupa-4000, 2009	CMOS linear (3.5)	chip level (10)	24.6×24.6 (4.2)	12 (37.5)
11.	Aptina imaging (Ref. 54) MT9P031, 2006	CMOS linear (2.8)	chip level (12)	5.7×4.3 (5)	2.2
12.	Aptina Imaging (Ref. 55) MT9M001C12STM, 2004	CMOS linear (3.3)	chip level (10)	6.7×5.3 (1.3)	5.2
13.	Samsung (Ref. 56) S5K3A1EA, 2004	CMOS linear (2.8)	column level (10)	4.9×3.9 (1.3)	3.8
14.	STMicroelectronics (Ref. 57) VS6502, 2004	CMOS linear (3.3)	chip level (10)	3.6×2.7 (0.3)	5.6

Table 2 (*Cont.*)

Image sensor		Technology (Supply voltage) (V)	Data conversion (ADC bits)	Sensor size (frame size) mm² (Mp)	Pixel pitch (fill factor) μm (%)
15.	National Semiconductor (Ref. 58)	CMOS linear	chip level	6.2×7.7	6
	LM9638, 2002	(3)	(10)	(1.3)	(49)
16.	Hynix (Ref. 59)	CMOS linear	chip level	6.4×4.8	8
	HV7141D, 2001	(3.3)	(8)	(0.5)	(30)
17.	Lim et al. (Ref. 36)	CMOS linear	column level	1.8×1.3	5.6
		(2.8)	(10)	(0.1)	
18.	Matsuo et al. (Ref. 38)	CMOS linear	column level	17.3×9.1	4.2
		(3.3)	(14)	(8.9)	
19.	Dubois et al. (Ref. 37)	CMOS linear	board level	2.2×2.2	35
		(3.3)		(0.004)	(25)
20.	Hoefflinger (Ref. 40)	CMOS log.	chip level	7.7×5.0	10
	VGAy	(3.3)	(10)	(0.4)	(40)
21.	Storm et al. (Ref. 41)	CMOS lin-log	column level	2.0×1.6	5.6
		(3.6)		(0.1)	(33)
22.	Kitchen et al. (Ref. 39)	CMOS linear	pixel level	2.9×2.9	45
		(3.3)	(8)	(0.004)	(20)
23.	Mase et al. (Ref. 42)	CMOS linear	column level	6.6×4.9	10
		(3.3)	(12)	(0.3)	(54.5)
24.	Lulé et al. (Ref. 43)	VI-CMOS linear	board level	14.1×10.2	40
		(5)	(12)	(0.1)	(100)

DS = 37.173 mV/s. According to the description of the sensor activation, one may conclude that at a frame rate of 26 Hz, the integration time in linear mode is $T_{int} = 50\,\mu s$. Using Eqs. (14) and (17), one may find that the DL equals 226 cd/m².

However, this DL, which is very high, is inconsistent with the plot shown in Fig. 22 of Storm et al.[41] that presents SNR against image plane illuminance. A halogen light source is used for the measurement, and the light intensity is given in units of mW/m². Assuming a color temperature of 3200 K, which is typical for this type of light source, the luminous efficacy is 28 lm/W.[61] One may conclude from this plot that a digital camera with this image sensor can detect luminance levels of at least 3.4 cd/m², and this value is used for performance evaluation. It is not clear, though, whether the measurement was done at a frame rate of 26 Hz. If T_{int} is large enough, so that $E_{im\text{-}DL} \approx DS/S$, the DL is 0.43 cd/m², which is comparable to that of the other image sensors included in the survey.

The plot in Fig. 22 of Storm et al.[41] shows that peak SNR is achieved when the sensor is activated in linear mode and equals 44.5 dB. This value defines the SNR

for the performance evaluation. Data regarding distortion is provided only for the logarithmic mode, and it is referred to the sensor response. Therefore, the SNDR may be calculated only for logarithmic operation. The plot also shows that peak SNR in logarithmic mode equals 32.5 dB.

In logarithmic sensors, changes in the response are proportional to the changes in the logarithm of the scene luminance. The response y of a pixel operating in logarithmic mode to stimulus x is:[62]

$$y \approx a + b \cdot \ln x + \epsilon, \tag{20}$$

where a and b are temporally-constant spatially-varying parameters, and ϵ is temporally-varying noise with spatially-constant statistics.

The rms value of the temporal noise in the scene, σ_x^t, is calculated as follows:

$$SNR = 20 \log \left(\frac{x}{\sigma_x^t} \right) = 32.5 \, dB. \tag{21}$$

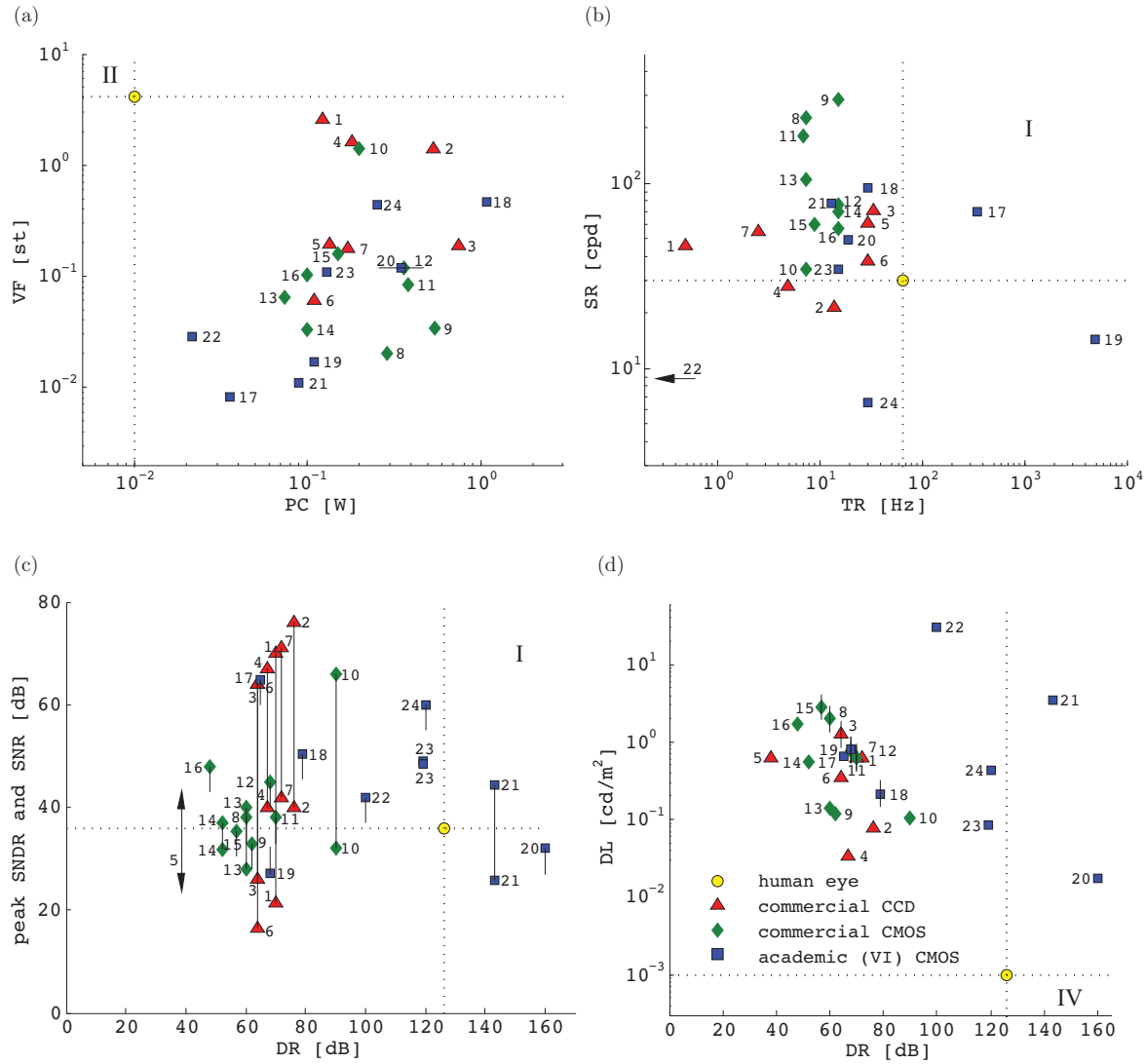Therefore, σ_x^t is 2.37% of the scene luminance.

Fig. 6 Performance evaluation results: (a) visual field versus power consumption, (b) spatial resolution versus temporal resolution, (c) peak SNDR and SNR versus dynamic range, and (d) dark limit versus dynamic range. The quadrant in which performance is superior to the human eye is indicated by the quadrant label.

Storm et al. report that, after 2-parameter calibration, the FPN is 2% of the logarithmic response, i.e., $b \cdot \ln(10)$, which equals 77 mV per decade. To refer the rms value of the distortion in the sensor response, σ_y^d, to the rms value of the distortion in the scene, σ_x^d, one needs to express changes in the stimulus in terms of changes in the response, which may be done with a derivative:

$$\frac{dy}{dx} = \frac{b \cdot \ln(10)}{x \cdot \ln(10)} = \frac{\sigma_y}{\sigma_x}. \tag{22}$$

Therefore,

$$\frac{\sigma_x^d}{x} = \frac{\sigma_y^d}{b} = 0.02 \cdot \ln(10) = 0.046, \tag{23}$$

or σ_x^d is 4.6% of the scene luminance.

The SNDR of sensor 21 may now be calculated as follows:

$$\text{SNDR} = 20 \log \left[\frac{x}{\sqrt{\left(\sigma_x^d\right)^2 + \left(\sigma_x^t\right)^2}} \right]$$

$$= 20 \log \left(\frac{1}{\sqrt{0.046^2 + 0.0237^2}} \right)$$

$$= 25.7 \, \text{dB}. \tag{24}$$

This section presented example calculations of the signal and noise power properties for two modern image sensors. These sensors differ by their technology and by their operating principles. The first example considered a CCD sensor, which is linear, and its calculation process was straightforward. The second example considered a CMOS image sensor that operates in two modes, linear and logarithmic, to construct a single frame. It required a more complicated calculation process. Therefore, this section demonstrates how the

performance evaluation method of Sec. 2 can be applied to extract performance properties in a unified way that makes it possible to compare two significantly different imaging systems.

3.3 *Performance Evaluation*

Results obtained for all the parameters considered in the performance evaluation are shown in Fig. 6. Image sensors located in the second quadrant of plot (a), first quadrant of plots (b) and (c), and fourth quadrant of plot (d) perform better than the human eye in both parameters shown. In cases where information was not available for a certain parameter, the FOM was used instead of the missing parameter to determine the point in the plot, and a line is drawn parallel to the missing parameter axis. For example, image sensor 20 in Fig. 6(a) is missing information regarding power consumption. If the actual point is located to the right, the missing parameter is the most limiting factor (MLF) and determines the FOM, which would be lower. If the actual point is to the left, it would not change the FOM.

Figure 6(a) shows that, with an ideal lens, none of the surveyed image sensors is superior to the human eye in terms of PC and VF. The VF depends on sensor dimensions. In general, one sees a deterioration in the PC with improvement in the VF. However, this does not apply to all cases because there are other factors that affect the PC, such as frame rate and circuit design. This plot also shows that commercial CCD and CMOS image sensors with similar dimensions have comparable power consumption. However, one should consider that CCD sensors operate under higher voltage levels (see Table 2), and also require more complicated power systems to activate the charge transfer mechanism. Although these issues are significant from a camera cost and complexity perspective, they are outside the scope of this paper.

The PC might be misleading for image sensors with off-chip analog-to-digital converters (ADCs). ADCs are an integral part of any digital camera, and the actual PC of a system that includes such an image sensor and an external ADC is higher. Likewise, the very low VF of several academic CMOS image sensors might underestimate their potential. These image sensors were designed to prove a concept and, therefore, only a small array was fabricated to save resources.

The TR is shown in Fig. 6(b) along with the SR. While the TR obtained for CCD image sensors ranges from 0.5 to 34 Hz, the TR of commercial CMOS image sensors lies in a much narrower band that ranges from 7 to 15 Hz. The academic CMOS image sensors have, in general, higher TR than commercial ones, and sensors 17 and 19 perform even better than the human eye at this measure. All this may indicate that commercial CMOS image sensors are targeting video applications rather than still imaging, and that high video rate imaging is one of the current trends in CMOS image sensors. Sensor 22 is represented by an arrow because no data was specified for its frame rate, and using its FOM resulted in a point located out of the range of the plot.

The SR depends on pixel pitch and fill factor. In cases where the fill factor was not mentioned (see Table 2), reasonable values were assumed for the calculation. Full frame and frame transfer CCDs can have a fill factor as high as 100%, while interline transfer CCDs can have a fill factor as low as 20%.[63] To estimate the SR, fill factors of 80% and 40% were assumed for the frame transfer and interline transfer CCDs, respectively, in Table 2. A fill factor of 40% was also assumed for the CMOS image sensors missing this data. This value is close to the average fill factor of the other CMOS image sensors in Table 2, excluding the vertically-integrated one.

The pixel pitch of sensor 24 is relatively large because of a bond pad for vertical integration and a complicated circuit that are placed in the pixel to increase the DR. This manifests in a low SR. In general, Fig. 6(b) shows that the SR that can be achieved with modern electronic image sensors is at least comparable to that of the human eye, even when a normal focal-length lens, rather than a long focal-length one, is used.

Figure 6(c) shows the peak SNDR and SNR of the surveyed image sensors versus their DR. If the specifications were given with and without an IR-cut filter, the performance with the filter is considered. With some of the linear CMOS image sensors, e.g., sensor 16, the magnitude of the temporal noise was specified only for dark conditions. This value was used to calculate the peak SNR, which resulted in this measure being equal to the sensor's DR. However, the actual peak SNR would be lower. Image sensors for which both SNR and SNDR data was provided are represented by a line with two end points. The top one refers to the SNR, and the bottom to the SNDR. Cases where only the SNR could be calculated, such as sensor 18, are represented by a downward line with one end point, and cases where only the SNDR could be calculated, such as sensor 19, are represented by an upward line with one end point. Sensor 5 is represented by a double arrow because its datasheet does not provide any information related to noise, and use of its FOM resulted in a negative value.

Commercial CCD image sensors can have better SNR than CMOS ones. However, the DR of both is comparable and rarely exceeds 70 dB. Sensor 10 has an operating mode that allows an increased DR by multiple exposure times. Five of the academic CMOS image sensors have a DR of at least 100 dB, among which sensors 20 and 21 demonstrate a DR that is superior to the benchmark defined by the human eye. This indicates that a current trend in electronic imaging is improved DR. Image sensor 24 achieves both high DR and high SNR. However, this comes at the expense of large pixels [see Fig. 6(b)].

Figure 6(d) presents the DL versus the DR. The DL could not be extracted for image sensors 1, 3, 8, 11, 12, 15, 18, and 19. The FOM is used instead with a line parallel to the DL axis. The documentation of sensor 20 specifies $E_{\text{im-DL}}$ in lux. For sensor 22, $E_{\text{im-DL}}$ was determined from a plot [Fig.11(a) of Kitchen et al.[39]] for the longest integration time shown. One may conclude from the figure that none of the image sensors has a DL that is lower than the benchmark set by the human eye. The DL of commercial CCD image sensors is somewhat better than that of CMOS ones.

Table 3 summarizes the performance evaluation results for each image sensor. The number of entries is out of the eight parameters considered for the evaluation. Also included are the first MLF, which determines the FOM, and the second MLF. Table 3 shows that with each one of the surveyed image sensors, at least one parameter is significantly weak, even with the benefit of an ideal lens. The DR proved to be the most limiting factor in most cases, and it is followed by the DL and the VF. Currently, no digital camera rivals the human eye.

Table 3 Summary of results for the 24 image sensors in Table 2. "Entries" refers to how many of the eight parameters in the performance evaluation could be determined. The FOM is the performance gap of the MLF of a digital camera, composed of the corresponding image sensor and an ideal lens, with respect to the human eye.

Sensor	Entries	FOM (dB)	1st MLF	2nd MLF
1.	7	−56	DR	TR
2.	8	−50	DR	DL
3.	7	−62	DR	PC
4.	8	−59	DR	DL
5.	6	−88	DR	DL
6.	8	−62	DR	DL
7.	8	−56	DL	DR
8.	6	−66	DR	VF
9.	7	−64	DR	DL
10.	8	−40	DL	DR
11.	6	−56	DR	VF
12.	6	−58	DR	PC
13.	8	−66	DR	DL
14.	8	−74	DR	DL
15.	6	−69	DR	VF
16.	7	−78	DR	DL
17.	7	−61	DR	DL
18.	6	−47	DR	PC
19.	6	−58	DR	VF
20.	6	−31	VF	DL
21.	8	−71	DL	VF
22.	6	−90	DL	VF
23.	8	−39	DL	VF
24.	7	−53	DL	PC

4 Discussion

Section 3 demonstrated not only the room for improvement but also the large variability in electronic image sensors. They differ by properties such as fabrication technology, readout mechanism, architecture, and pixel design. Past trends in the area of electronic imaging may suggest which of the present trends will lead to a significant improvement in performance and, therefore, have the potential to become dominant in the future.

The first electronic image sensors were based on MOS devices. Various configurations fabricated in n-channel MOS (NMOS), p-channel MOS (PMOS), and bipolar technologies were demonstrated in the 1960s.[64] However, their image quality was poor, mainly due to a high FPN. The invention of the CCD by Willard Boyle and George Smith in the late 1960s, for which they were granted the Nobel Prize in Physics in 2009, enabled the first generation of electronic image sensors. CCD technology created the image sensor market because it showed a significant improvement in image quality when compared to MOS technology. CCD image sensors were almost free of FPN, and achieved a higher spatial resolution because CCD technology allowed fabrication of smaller pixel dimensions than MOS technology. For three decades, CCD was the dominant technology in the area of electronic image sensors.

The second generation of electronic image sensors emerged in the 1990s when various groups around the world, and mainly the NASA Jet Propulsion Laboratory,[65] decided to put more effort into the development and improvement of CMOS active pixel sensors. The advantages of these image sensors over CCD devices included increased chip-level functionality, lower cost, and the ability to operate with a simple power system. These advantages made CMOS image sensors suitable for mobile applications, where there was a demand for compact systems that are also low power. Other properties, which prevented CMOS image sensors from competing with CCD in the early days, improved with developments in the CMOS industry and, particularly, in the area of CMOS image sensors.

As they still do not rival the human eye, further work needs to be done to improve modern digital cameras. Using a design approach that is inspired by biological systems too literally is not the best way to accomplish comparable functionality. A better approach is to develop methods that use available technologies for implementation of systems that can compete with natural ones. Nonetheless, one should examine how structural differences between the human retina and electronic image sensors lead to a functional gap between the two. For example, the retina has a multilayered structure, whereas electronic image sensors are usually fabricated in single-tier technologies.

A schematic cross-section of the human retina is shown in Fig. 7(a). The retinal pigment epithelium delivers nutrients to the retina and disposes of the metabolic waste. It also prevents reflections of light rays that are not absorbed by photoreceptors.[60] Rod and cone cells absorb light, and convert it into a neural response in the form of impulses. The human retina contains 115 to 120 million rods and 5 to 6.5 million cones, which are not distributed uniformly. Impulses travel vertically through the bipolar cells and finally reach the ganglion cells. Each ganglion cell transmits the impulses through its nerve fiber (an axon) to the brain.

Both horizontal and amacrine cells form lateral connections between bipolar cells, the former at the receptor-bipolar synapse (a "node" where cells are connected electrically and chemically), and the latter at the bipolar-ganglion synapse. Horizontal cells have a role in retinal processing, which possibly involves chromatic interaction between photoreceptors. Amacrine cells play a role in modulation of signals that are transmitted to the ganglion cells.[66]

Conversion of light into impulses by the retinal photoreceptors is equivalent to data conversion that is done at pixel level. In image sensors, digitization of the analog signals generated by the photodetectors can be done at four different levels: board level, chip level, column level, and pixel
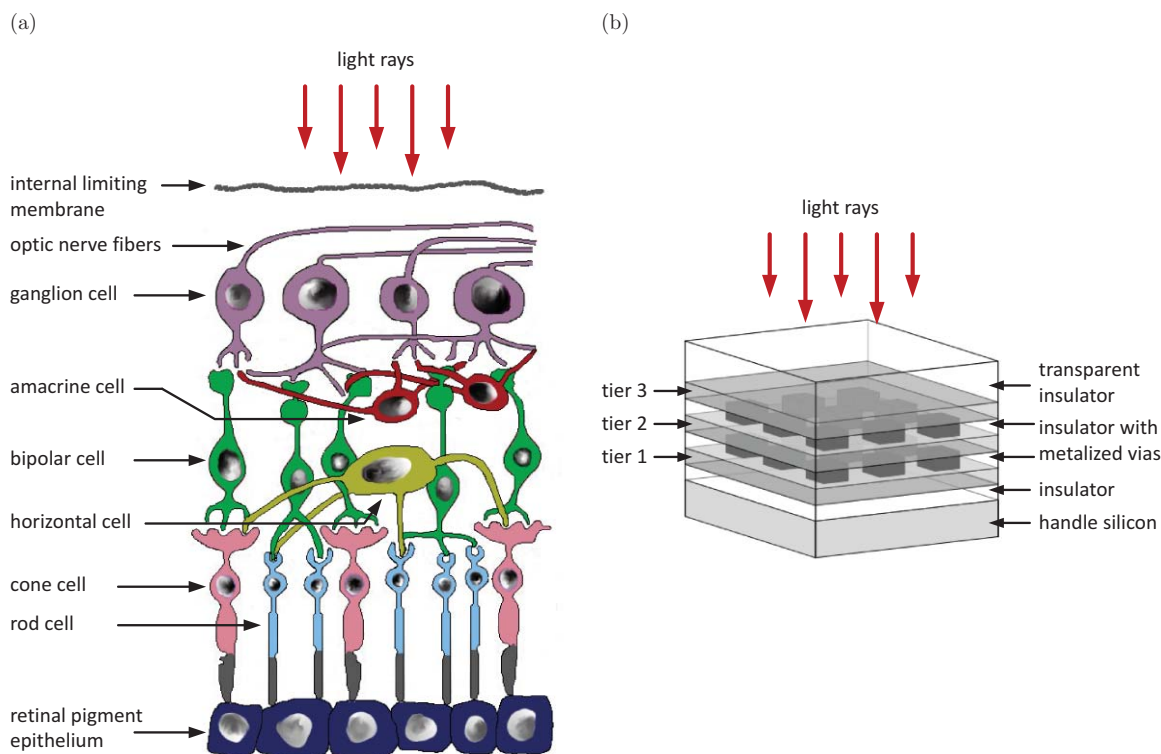
(a)

light rays

internal limiting
membrane

optic nerve fibers

ganglion cell

amacrine cell

bipolar cell

horizontal cell

cone cell

rod cell

retinal pigment
epithelium

(b)

light rays

tier 3
tier 2
tier 1

transparent
insulator

insulator with
metalized vias

insulator

handle silicon

Fig. 7 (a) The human retina has a vertically-integrated structure. Photoreceptors absorb light and generate electric pulses, which are transmitted to bipolar and ganglion cells, and then to optic nerve fibers. The retina also has horizontal and amacrine cells that form lateral connections. (b) A vertically-integrated CMOS image sensor, made by TSV technology, has multiple tiers. For example, photodetectors, analog circuits, and digital circuits may be optimized independently with three tiers.

level. The level of data conversion is mentioned in Table 2 for each of the image sensors. In CCD image sensors, where integration of electronics with photodetectors is, at large, impossible, the data conversion must be done at board level. In almost all CMOS image sensors in Table 2, the ADCs are integrated with the sensor array. The benefits of this integration include reduction in system size, power consumption, and noise. Noise reduction is essential for improvement of performance in parameters related to signal and noise power.

The longer the path an analog signal needs to travel to reach an ADC, the greater the noise power it accumulates. The travel distance is minimal when the data conversion is done at pixel level. Digital pixel sensors have been demonstrated.[67–69] However, ADCs are composed of a large number of transistors. In planar technologies, the pixel area is shared between photodetector and electronics. Working with deep submicron CMOS processes, which allows increased device density, is undesirable for image sensors due to poor optical performance.[70] Therefore, implementation of more circuits in the pixel area results in large pixel dimensions, which degrades the SR. See, for example, sensor 24 in Sec. 3.

Fabrication of ICs in 3D structures, where two or more dies are stacked to form a multiple-tier device, allows more pixel-level circuitry while reasonable pixel dimensions are maintained. Moreover, each tier can be fabricated in a process that best suits the type of devices it contains. Therefore, digital circuits can be fabricated in a high-density nanoscale

CMOS process, while photodetectors are prepared in a large scale process. Analog circuits can be either fabricated in the same process as the digital ones or in an intermediate scale process. Furthermore, vertical integration allows more degrees of freedom in the design of the photodetectors. In some fabrication methods, it is feasible to use materials that are other than crystalline silicon.

The DL, which proved to be an important limiting factor in modern digital cameras can be improved by photodetector optimization. Image sensors fabricated by planar technologies that include avalanche photodiodes for photodetection in dim light have been presented.[71] However, this architecture requires a large pixel area and, therefore, suffers from low SR.

The horizontal connections between the retinal cells allow mechanisms of feedback and control. A processed image rather than "raw data" is sent from the retina to the brain. This is equivalent to DSP that is used for image enhancement in camera systems. Image sensors that include analog circuits for implementation of image enhancement features have been demonstrated in planar technologies.[72] However, also in this case, the additional circuitry came at the expense of an enlarged pixel area and reduced SR. In general, the DSP is done at board level in modern digital cameras.

Among the various technologies for vertical integration of IC devices,[73] through-substrate via (TSV) technologies[74] are the only ones that allow fabrication of multiple-tier ICs with vertical integration at pixel level. In TSV technologies, via holes are etched through the substrate, and then filled

with metal to form electrical connections between tiers. The process also requires alignment and attachment at either die or wafer level. A VI-CMOS image sensor has been demonstrated in stacked-SOI technology,[75] which is one of the TSV methods. Figure 7(b) shows a schematic of a multiple-tier image sensor fabricated by this technology.

There are issues that need to be overcome before technologies for vertical stacking of ICs can be widely used. For example, alignment of small features at die or wafer level is a mechanical challenge. Moreover, heat dissipation is a major problem in such structures because of the increase in volumetric device density. Finally, to make these devices affordable for consumers, fabrication costs need to drop. Nonetheless, the interest in fabrication of 3D ICs has increased in recent years because these devices have widespread advantages.[76]

5 Conclusion

The human factor has been widely considered for evaluation of digital displays. However, it is rarely used as a reference point in the design and evaluation of digital cameras and electronic image sensors, although these are fundamental systems at the heart of every process of digital recording of scenes. This work introduces a method that evaluates performance of digital cameras with respect to the human eye. It is motivated by a wide range of applications in which digital cameras are expected to observe scenes the same way they are perceived by the human eye and, therefore, to rival the human eye.

After considering various parameters that can be used to characterize an imaging system, eight are taken for the evaluation. The process is concluded with a figure of merit, which represents the performance gap for the parameter that appears to be the weakest when compared to the human eye. Assessment methods for the eight parameters are covered for the human eye and for digital cameras. Experiments performed with human observers and cadavers are reviewed to determine and justify benchmark values for the human eye. Information given in datasheets or other detailed publications is needed to assess the performance of digital cameras.

The performance evaluation method has been applied to 24 modern CCD and CMOS image sensors, both commercial and academic ones, where an ideal lens is assumed to complete the electronic imaging system. In the majority of cases the dynamic range proved to be the most limiting factor with respect to the human visual system, and this parameter was followed by the dark limit. Overall, the evaluation concludes that modern digital cameras do not rival the human eye. The functional gap ranges from 31 to 90 dB, or from 1.6 to 4.5 orders of magnitude.

Past trends in the area of electronic imaging were initially concerned with image quality of low dynamic range scenes. This was succeeded by image sensors more suitable for mobile devices. Image sensor capabilities can be improved by photodetector optimization and increased pixel-level data processing. But implementations in planar technologies result in degradation of the spatial resolution. Although more effort needs to be put into the development of reliable and economical fabrication methods, vertical stacking of pixel components is promising for boosting the performance of image sensors and digital cameras.

References

1. T. Suzuki, "Challenges of image-sensor development," *IEEE International Solid-State Circuits Conference (ISSCC) 2010*, Plenary Session, http://isscc.org/videos/2010_plenary.html#session3 (2010).
2. S. Franz, D. Willersinn, and K. Kroschel, "Assessment of image sensor performance with statistical perception performance analysis," *Proc. SPIE* **7539**, 75390I (2010).
3. B. Rodricks and K. Venkataraman, "First principles' imaging performance evaluation of CCD- and CMOS-based digital camera systems," *Proc. SPIE* **5678**, 59–74 (2005).
4. European Machine Vision Association, "1288 EMVA Standard Compliant," pp. 1–31 (2007).
5. A. Spivak, A. Belenky, A. Fish, and O. Yadid-Pecht, "Wide-Dynamic-Range CMOS Image Sensors—Comparative Performance Analysis," *IEEE Trans. Electron Devices* **56**, 2446–2461 (2009).
6. J. Janesick, "Lux transfer: Complementary metal oxide semiconductors versus charge-coupled devices," *Opti. Eng.* **41**, 1203–1215 (2002).
7. B. E. Rogowitz, T. N. Pappas, and J. P. Allebach, "Human vision and electronic imaging," *J. Electro. Imaging* **10**, 10–19 (2001).
8. R. Bremond, J. P. Tarel, E. Dumont, and N. Hautiere, "Vision models for image quality assessment: one is not enough," *J. Electron. Imaging* **19**, 043004 (2010).
9. Q. Ma, L. Zhang, and B. Wang, "New strategy for image and video quality assessment," *J. Electron. Imaging* **19**, 011019 (2010).
10. C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, USA (1989).
11. S. C. DeMarco, G. Lazzi, W. Liu, J. M. Weiland, and M. S. Humayun, "Computed SAR and thermal elevation in a 0.25-mm 2-D model of the human eye and head in response to an implanted retinal stimulator - Part I: models and methods," *IEEE Trans. Antennas Propag.* **51**, 2274–2285 (2003).
12. B. R. Straatsma, M. B. Landers, A. E. Kreiger, and L. Apt, "Topography of the adult human retina," in *The Retina: Morphology, Function and Clinical Characteristics*, edited by B. R. Straatsma, M. O. Hall, R. A. Allen, and F. Crescitelli, pp. 379–410 University of California Press, Los Angeles, CA, (1969).
13. A. H. Kashani, I. E. Zimmer-Galler, S. M. Shah, L. Dustin, D. V. Do, D. Eliott, J. A. Haller, and Q. D. Nguyen, "Retinal thickness analysis by race, gender, and age using stratus OCT," *Am. J. Ophthalmol.* **149**, 496–502 (2010).
14. H. Y. Cho, D. H. Lee, S. E. Chung, and S. W. Kang, "Diabetic retinopathy and peripapillary retinal thickness," *Korean J. Ophthalmol.* **24**, 16–22 (2010).
15. D. B. Henson, *Visual Fields*, 2nd ed., Butterworth-Heinemann, Great Britain (2000).
16. J. J. Kanski and J. A. McAllister, *Glaucoma: A Colour Manual of Diagnosis and Treatment*, 1st ed., Butterworths, Great Britain (1989).
17. L. A. Zenteno, "Design of an optical fiber device for with a laser diode bar," *Opt. Rev.* **2**, 52–54 (1995).
18. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, U.S. Dept. of Commerce, Washington, D.C. (1972).
19. R. H. Simons and A. R. Bean, *Lighting Engineering: Applied Calculations*, 1st ed. Architectural Press, United Kingdom (2001).
20. A. Grzybowski, "Harry Moss Traquair (1875–1954), Scottish ophthalmologist and perimetrist," *Acta Ophthalmol.* **87**, 455–459 (2009).
21. A. S. Patel, "Spatial resolution by the human visual system. The effect of mean retinal illuminance," *J. Opt. Soci. Ame.* **56**, 689–694 (1966).
22. D. H. Kelly, "Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements," *J. Opt. Soci. Ame.* **51**, 422–429 (1961).
23. S. G. De Groot and J. W. Gebhard, "Pupil Size as Determined by Adapting Luminance," *J. Opt. Soci. of Ame.* **42**, 492–495 (1952).
24. J. W. Goodman, *Introduction to Fourier Optics*, McGraww-Hill, San Francisco, CA (1968).
25. O. Yadid-Pecht, "Geometrical modulation transfer function for different pixel active area shapes," *Opt. Eng.* **39**, 859–865 (2000).
26. A. El Gamal and H. Eltoukhy, "CMOS image sensors," *IEEE Circuits and Devices Magazine*, 6–20 (2005).
27. L. E. Arend, "Spatial differential and integral operations in human vision: Implications of stabilized retinal image fading," *Psychol. Rev.* **80**, 374–395 (1973).
28. W. F. Schreiber, *Fundamentals of Electronic Imaging Systems*, 3rd ed., Springer-Verlag, Germany (1993).
29. S. Hecht, "The visual discrimination of intensity and the Weber-Fechner law," *J. Gen. Physiol.* **7**, 235–267 (1924).

30. M. G. Helander, T. K. Landauer, and P. V. Prabhu, *Handbook of Human-Computer Interaction*, 2nd ed., Elsevier Science B.V., Amsterdam, The Netherlands (1997).
31. P. G. J. Barten, "Formula for the contrast sensitivity of the human eye," *Proc. SPIE* **5294**, 231–238 (2004).
32. J. G. Rogers and W. L. Carel, *Development of Design Criteria for Sensor Displays*, Hughes Airport Company, Office of Naval Research Contract No. N00014 72-C-0451, NR213-107, Culver City, CA (1973).
33. W. J. Smith, *Modern Optical Engineering*, 4th ed., McGraw-Hill, New York, NY (2008).
34. D. A. Atchison and G. Smith, *Optics of the Human Eye*, Butterworth-Heinemann, Edinburgh, United Kingdom (2003).
35. J. Gulbins, *Digital Photography from Ground Up*, 1st ed. Rocky Nook, Canada (2008).
36. S. Lim, J. Lee, D. Kim, and G. Han, "A high-speed CMOS image sensor with column-parallel two-step single-slope ADCs," *IEEE Trans. Electron Devices* **56**, 393–398 (2009).
37. J. Dubois, D. Ginhac, M. Paindavoine, and B. Heyrman, "A 10 000 fps CMOS sensor with massively parallel image processing," *IEEE J. Solid-State Circuits* **43**, 706–717 (2008).
38. S. Matsuo, T. J. Bales, M. Shoda, S. Osawa, K. Kawamura, A. Andersson, M. Haque, H. Honda, B. Almond, Y. Mo, J. Gleason, T. Chow, and I. Takayanagi, "8.9-megapixel video image sensor with 14-b column-parallel SA-ADC," *IEEE Trans. Electron Devices* **56**, 2380–2389 (2009).
39. A. Kitchen, A. Bermark, and A. Bouzerdoum, "A digital pixel sensor array with programmable dynamic range," *IEEE Trans. on Electron Devices* **52**, 2591–2601 (2005).
40. B. Hoefflinger, *High-Dynamic-Range (HDR) Vision*, Springer, Berlin Heidelberg (2007).
41. G. Storm, R. Henderson, J. E. D. Hurwitz, D. Renshaw, K. Findlater, and M. Purcell, "Extended dynamic range from a combined linear-logarithmic CMOS image sensor," *IEEE J. Solid-State Circuits* **41**, 2095–2106 (2006).
42. M. Mase, S. Kawahito, M. Sasaki, Y. Wakamori, and M. Furuta, "A wide dynamic range CMOS image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic A/D converters," *IEEE J. Solid-State Circuits* **40**, 2787–2795 (2005).
43. T. Lule, M. Wagner, H. Keller, and M. Bohm, "100 000-Pixel, 120-dB imager in TFA technology," *IEEE J. Solid-State Circuits* **35**, 732–739 (2000).
44. Texas Instruments Incorporated, *TC237B 680 x 500-pixel CCD Image Sensor*, pp. 1–15 (2001).
45. Eastman Kodak Company, *Kodak KAF-50100 Image Sensor*, Device Performance Specification Revision 4.0, pp. 1–40 (2010).
46. Dalsa Professional Imaging, *FTF2010M 4M Full-Frame CCD Image Sensor*, Preliminary Product Specification, pp. 1–22 (2009).
47. Eastman Kodak Company, *Kodak KAI-02050 Image Sensor*, Device Performance Specification Revision 1.0, pp. 1–38 (2008).
48. Atmel Corporation, *AT71299M 8M-pixel Color Image Sensor*, Rev. 2133A, pp. 1–19 (2003).
49. Sony Corporation, *ICX285AL Diagonal 11 mm (Type 2/3) Progressive Scan CCD Image Sensor with Square Pixel for B/W Cameras*, E00Y42-TE, pp. 1–21 (2003).
50. Philips Semiconductors, *FXA 1012 Frame Transfer CCD Image Sensor*, Objective Specification, pp. 1–17 (2000).
51. Hynix Semiconductor Inc., *YACD5B1S 2M CMOS Image Sensor*, Q3 2010 Databook Revision 0.0, pp. 3 (2010).
52. Samsung Electronics, *S5K4E1GA(EVT3) 1/4-Inch QSXGA CMOS Image Sensor*, Data Sheet Revision 0.23, pp. 1–94 (2010).
53. Cypress Semiconductor Corporation, *CYIL1SM4000AA LUPA 4000: 4 MegaPixel CMOS Image Sensor*, Revision D, pp. 1–31 (2009).
54. Micron, *MT9P031: 1/2.5-Inch 5-Mp Digital Image Sensor*, Preliminary, pp. 1–12 (2006).
55. Aptina Imaging, *MT9M001: 1/2-Inch Megapixel CMOS Digital Image Sensor*, MT9M001 DS Rev.1, pp. 1–27 (2004).
56. Samsung Electronics, *S5K3A1EA (1/3-Inch SXGA CMOS Image Sensor)*, Preliminary Specification, pp. 1–34 (2004).
57. STMicroelectronics, *ST VS6502 VGA Color CMOS Image Sensor Module*, pp. 1–48 (2004).
58. National Semiconductor Corporation, *LM9638 Monochrome CMOS Image Sensor SXGA 18 FPS*, Advance Information, pp. 1–2 (2002).
59. Hynix Semiconductor Inc., *HV7141D CMOS Image Sensor*, DA41010108R 1.0, pp. 1–23 (2001).
60. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. John Wiley and Sons, New York (1982).
61. J. M. Palmer, "Radiometry and photometry: Units and conversions," in *Handbook of Optics, Volume II - Design, Fabrication and Testing, Sources and Detectors, Radiometry and Photometry*, edited by M. Bass, V. N. Mahajan, and E. Van Stryland, McGraw Hill, USA (2010).
62. D. Joseph and S. Collins, "Modeling, calibration, and correction of nonlinear illumination- dependent fixed pattern noise in logarithmic CMOS image sensors," *IEEE Trans. Instrum. Meas.* **51**, 996–1001 (2002).
63. C. Steger, M. Ulrich, and C. Wiedemann, *Machine Vision Algorithms and Applications*, Wiley-VCH, Berlin, Germany (2008).
64. E. R. Fossum, "CMOS image sensors: Electronic camera-on-a-chip," *IEEE Trans. Electron Devices* **44**, 1689–1698 (1997).
65. S. Mendis, S. E. Kemeny, and E. R. Fossum, "CMOS active pixel image sensor," *IEEE Trans. Electron Devices* **41**, 452–453 (1994).
66. J. V. Forrester, A. D. Dick, P. McMenamin, and W. R. Lee, *The Eye — Basic Sciences in Practice*, WB Saunders Compant LTD, United Kingdom (1996).
67. D. X. D. Yang, B. Fowler, and A. El Gamal, "A Nyquist-rate pixel-level ADC for CMOS image sensors," *IEEE J. Solid-State Circuits* **34**, 348–356 (1999).
68. J. Rhee and Y. Joo, "Wide dynamic range CMOS image sensor with pixel level ADC," *Electron. H. Lett.* **39**, 360–361 (2003).
69. Y. M. Chi, U. Mallik, M. A. Clapp, E. Choi, G. Cauwenberghs, and R. Etienne-Cummings, "CMOS camera with in-pixel temporal change detection and ADC," *IEEE J. Solid-State Circuits* **42**, 2187–2196 (2007).
70. T. Lule, S. Benthien, H. Keller, F. Mutze, P. Rieve, K. Seibel, M. Sommer, and M. Bohm, "Sen- sitivity of CMOS based imagers and scaling perspectives," *IEEE Trans. Electron Devices* **47**, 2110–2122 (2000).
71. C. Niclass, C. Favi, T. Kluter, F. Monnier, and E. Charbon, "Single-photon synchronous detection," *IEEE J. Solid-State Circuits* **44**, 1977–1989 (2009).
72. N. Massari, M. Gottardi, L. Gonzo, D. Stoppa, and A. Simoni, "A CMOS image sensor with programmable pixel-level analog processing," *IEEE Trans. Neural Networks* **16**, 1673–1684 (2005).
73. W. R. Davis, J. Wilson, S. Mick, J. Xu, H. Hua, C. Mineo, A. M. Sule, M. Steer, and P. D. Franzon, "Demystifying 3D ICs: The pros and cons of going vertical," *IEEE Design Test Computers* **22**, 498–510 (2005).
74. D. M. Kwai and C. W. Wu, "3D integration opportunities, issues, and solutions: A designer perspective," *Proc. SPIE* **7520**, 752003 (2009).
75. J. A. Burns, B. F. Aull, C. K. Chen, C. Chen, C. L. Keast, J. M. Knecht, V. Suntharalingam, K. Warner, P. W. Wyatt, and D. W. Yost, "A wafer-scale 3-D circuit integration technology," *IEEE Trans. Electron Devices* **53**, 2507–2516 (2006).
76. X. Dong and Y. Xie, "System-level cost analysis and design exploration for three-dimensional integrated circuits (3D ICs)," *Design Automation Conference 2009. ASP-DAC 2009, Asia and South Pacific*, IEEE, Yokohama, Japan, pp. 234–241 (2009).

Orit Skorka received a BS degree (summa cum laude) in electrical and computer engineering from the Ben Gurion University of the Negev, Israel, in 2001, a MS degree in electrical and computer engineering from the Technion-Israel Institute of Technology, Israel, in 2004, and a PhD degree in electrical and computer engineering from the University of Alberta, Canada, in 2011. Her current research concerns sensors and systems for electronic imaging.

Dileepan Joseph received a BS degree in computer engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1997 and a PhD degree in engineering science from the University of Oxford, Oxford, United Kingdom, in 2003. In 2004, he joined the Faculty of Engineering at the University of Alberta, Edmonton, AB, Canada, where he developed a research program in electronic imaging that ranges from image sensors to image understanding.