Common Principles of Image Acquisition Systems and Biological Vision

BRIAN A. WANDELL, ABBAS EL GAMAL, FELLOW, IEEE, AND BERND GIROD, FELLOW, IEEE

Invited Paper

In this paper, we argue that biological vision and electronic image acquisition share common principles despite their vastly different implementations. These shared principles are based on the need to acquire a common set of input stimuli as well as the need to generalize from the acquired images. Two related principles are discussed in detail, namely, multiple parallel image representations and the use of dedicated local memory in various stages of acquisition and processing. We review relevant literature in visual neuroscience and image systems engineering to support our argument. Particularly, the paper discusses multiple capture image acquisition, with applications such as dynamic range, field-of-view, or depth-of-field extension. Finally, as an example, a novel multiple-capture-single-image complementary metal–oxide–semiconductor sensor is presented. This sensor has been developed at Stanford University and it illustrates the principles that are shared among biological vision and image acquisition.

Keywords—CMOS sensor, digital camera, dynamic range, human vision, image acquisition, image mosaicing, image registration, image sensor, memory architecture, multiple capture, multiple representations, visual pathways.

I. INTRODUCTION

Certain relationships between image systems engineering and visual neuroscience are well established. For example, there is much common ground between biological vision and the technical systems for image analysis or image quality evaluation. However, it is less obvious that biological vision shares common principles with image acquisition systems used for image reproduction. In the first part of the paper, we suggest that these two types of systems are conceptually cou-

pled and we describe principles that are basic to both fields. We argue that multiple image representations, a fundamental principle of visual neuroscience, can be applied to image acquisition systems. Then, we argue that local memory, a critical component in the design of electronic imaging systems with multiple representations, should be considered as an essential architectural feature when analyzing the function of neurons in the visual pathways.

In the second part of the paper, we point out the connection between the ideas presented here and various applications that use multiple image captures to render a single image. The success of multiple capture image acquisition applications provides further support for the generality of the ideas and the conceptual relationship between the biological and engineering systems.

In the third part of the paper, we describe an electronic imaging system that operates using these two principles. Specifically, we describe a multiple capture single image (MCSI) imaging architecture [1]. At the core of this architecture is a digital pixel sensor (DPS), a new image sensor architecture that performs multiple nondestructive image readouts within a normal exposure time. Processing circuits that combine computational and memory functions are fundamental to this architecture. We hope that lessons we have learned from engineering a multiple image representations system will benefit visual neuroscientists who seek to understand neural function.

II. IMAGING PRINCIPLES

Image reproduction is the largest application of electronic image acquisition systems. Biological systems, however, do not reproduce images: they interpret images. At first, then, it may appear that biological systems are not good models for the design of image acquisition and reproduction systems.

We believe, however, that there is a strong relationship between biological vision and electronic systems for image reproduction. There are two areas of commonality. First, bio-

logical vision and engineered image systems generally work with common input data: natural images. Hence, the two types of systems are linked by the need to work across the range of constraints imposed by natural images [2]. Natural image data span a particular dynamic range, have certain characteristic space-time correlations, and contain various typical types of motion. Encoding the image data accurately is a necessary step for both image analysis and image reproduction. The properties of natural images set common constraints on the encoding and representation of images in biological and engineering systems.

There is a second equally important source of commonality. Vision scientists since Helmholtz have argued that human visual perception is best understood as an explanation of the physical causes of the retinal image [3]–[5]. By interpreting the retinal images and not just storing them, biological systems are prepared to generalize from a particular viewpoint, ambient lighting, size, or distance of a specific image.

The ability to generalize from acquired data is an important capability for image reproduction systems as well. Image reproduction systems rarely render a precise replica of the original. The reproduction is usually smaller, seen from a different perspective and viewed in a different context than the original. Reproduction systems create images that are generalizations of the original image data; they do not replicate the image data.

Once we recognize that the acquired image data are used to reproduce a generalized view of the original, the connection between image reproduction and image interpretation becomes clear: the image reproduction process will be more successful if we interpret the original and apply appropriate transformations. For example, interpreting the shape of an original object can improve subsequent renderings from different perspectives: illumination estimation can improve color rendering, measuring motion can remove motion blur. In general, imaging systems that can interpret the original image data will have better image reproduction capabilities.

Our emphasis on the importance of image interpretation is an extension of current practice; modern electronic imaging systems already include control systems that make certain inferences about the scene. Exposure value systems analyze image intensity, white balance systems analyze the color distribution in the image, and focus systems measure the distance to a principal object. We believe that electronic imaging systems of the future will derive and encode much more information about the physical characteristics of the scene. Elsewhere, we have argued that these capabilities can be implemented and that doing so might prove to be much more important than simply adding to the spatial resolution of the digital sensor or simply integrating parts onto the sensor [6]. The ability to estimate the three-dimensional (3-D) spatial geometry of the objects, illumination direction, specularity, occlusion, and the properties of moving objects will provide information that can radically change the capabilities of the image reproduction system.

We begin, then, with the premise that biological and image reproduction share constraints based on the input data and the added value to both for incorporating image interpretation. However, what specific computational elements might link the processing in biological and electronic systems? We describe two principles, multiple representations and local memory, which we believe are shared by the visual pathways and electronic reproduction system design. We do not mean to imply that this is a complete list; indeed, we are certain that over time many other ideas will emerge.

A. Multiple Image Representations

On the one hand, it is obvious that the visual system makes multiple captures of a scene and renders, but a single perceptual experience: we have two eyes. It is surprising to learn that this is but one of many examples in which the visual pathways create multiple representations. Each retina includes multiple mosaics of neurons that separately represent the visual field. For example, image transduction uses two systems of photoreceptors: the rods and cones. Each system comprises a separate sampling mosaic of the retinal image. The rods encode the data for a system with low spatial resolution but high quantum efficiency. The cones encode the image data at much higher spatial resolution and lower quantum efficiency.

Rods and cones generally operate under different viewing conditions, but there are also many cases in which multiple representations of the image are obtained under a single viewing condition. For example, the cones can be subdivided into three sampling mosaics that expand the spectral encoding. The three cone mosaics also differ in their spatial sampling properties.

The use of multiple field representations does not end with the photoreceptors. In fact, the number of visual field representations increases at the next layer of neurons. Each foveal cone is contacted at roughly 250 synaptic sites by postreceptoral neurons. These contacts arise from eight to ten different types of postsynaptic neurons; each type of neuron forms part of a mosaic of similar neurons that represent the entire visual field [10], [11]. Hence, the three cone mosaics are transformed into a larger set of neural mosaics within the retina.

Among the best known mosaics is the one initiated by the midget cells of the retina that then projects to the parvocellular layers of the lateral geniculate nucleus and finally to layer 4Cb of the first cortical area (V1). The midget cell mosaic samples the entire visual field at very high resolution. In the central region of the visual field, a single cone drives the signal from a midget bipolar, which in turn drives the center of the midget ganglion cell. Because of their fine spatial sampling resolution, this mosaic is present in very high density. In addition to the midget cells, there are about a dozen other known retinal ganglion cell mosaics and several of these also send their outputs to the lateral geniculate nucleus and then to V1. By using multiple mosaics of photoreceptors and neurons, the visual system expands its operating range, perhaps including additional measurements along a variety of image dimensions.

In the last 30 years, we have learned that the visual cortex, too, contains multiple distinct representations of visual
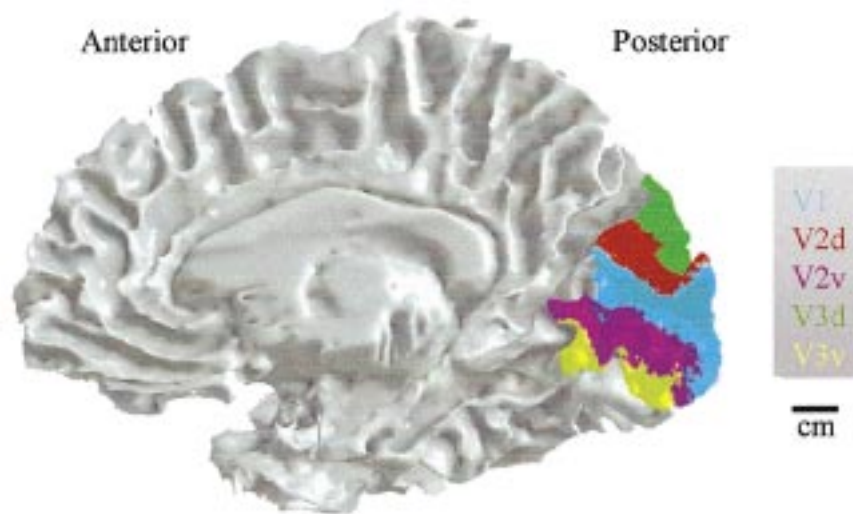
Fig. 1. Right hemisphere of a human brain seen in sagittal view. Rendered surface is the boundary between the white matter and gray matter surface. Boundary between the white and gray matter was measured using custom software.[1] Color overlays in the posterior portion of the brain (occipital lobe) show the positions of several different visual areas. These areas are determined from measurements of activity in the human brain using fMRI [7]. Signals from the retina and lateral geniculate are sent to primary visual cortex (V1). Areas V2 and V3 surround this region. Several other visual areas (not shown) can be measured on the dorsal and ventral surface of the occipital lobe [8], [9].

information. These representations take several forms, including multiple spatial maps of the visual field. These distinct maps have been observed using direct measurements of neural responses in animal brains [12], [13] and more recently using functional magnetic resonance imaging (fMRI) of activity in the human brain [7], [14]. Based on these maps and other experimental criteria, including single-unit receptive field properties, cell morphology, and anatomical connections, the visual cortex can be subdivided in many distinct *visual areas* that separately encode the visual field appear to serve different computational functions [15].

Fig. 1 illustrates the locations of several distinct visual areas in the right hemisphere of a human. Each of the colored regions contains a map of the left half of the visual field; a corresponding map of the right visual field is present in the analogous region in the left hemisphere. These visual field maps represent one of the clearest examples, at a large spatial scale, of multiple spatial representations of the image. The surface area of these visual areas varies from 4 to 30 cm^2. Visual cortex in human ranges from 2 to 5 mm in thickness and there are approximately 10^5 neurons/mm^3 [16]. Most remarkable, it is estimated that each cubic millimeter of cortex contains a length of 3 km of fibers that interconnect these neurons [17], [18]. Like the retina, each cortical area further contains a multiplicity of cell types.

Why should the visual pathways contain so many mosaics that encode the visual field? One hypothesis is that each area represents some perceptual feature. Some visual areas are said to represent color, another motion and depth, another form, and so forth [9], [19]–[24]. Animal and human data lend modest support to this principle. However, the association between visual areas and specific visual functions is far from proven. In fact, even the general form of this hy-

pothesis, that visual areas should be organized by their role in representing consciously available perceptual features, is unproven. We suspect that over time many other hypotheses about the purpose of computations performed within different visual areas will emerge.

A second hypothesis explaining the need for multiple representations is that each representation is designed to *compute* a specialized quantity. For example, roughly 30 years ago, behavioral studies suggested that the visual pathways contain maps that code different spatial and temporal scales [25]–[28]. This has been an important theme in the human experimental literature ranging from relatively complex theories of brightness perception [29] to theories of pattern detection and discrimination [30]. Some data suggest that orderly clusters of neurons within individual visual areas represent the multiple spatial resolutions [31]–[34]. The multiple spatial resolution architecture has proven to be very useful in digital imaging applications (e.g., [35]). For example, multiple scale representations are used in search algorithms; initial solutions are found at a coarse scale and these solutions initiate the search at a finer scale [36], [37]. Hence, both behavioral data and algorithmic efficacy support the value of multiple representations of different spatial and temporal scale.

Behavioral data also suggest that the visual pathways contain several different temporal representations [38], [39]. One interesting demonstration of the relatively small number of temporal representations is this: human observers cannot discriminate between lights flickering at different rates. In fact, lights flickering at temporal frequencies beyond 10 Hz all appear to be flickering at the same rate [40], [41]. While the flicker itself is very visible, the flicker rate is impossible to assess. This observation suggests that the representation of temporal frequency is based on only a few sample measurements, just as the representation of wavelength is based on a few samples [42]. Compared to the vast literature on spatial mecha-

---

[1]Available at http://white.stanford.edu/mri/unfoldSegment.htm.

nisms [30], there is a modest amount of work on the neural basis of temporal mechanisms. This asymmetry is surprising: the evidence for several temporal resolution maps is equal to that of multiple spatial resolutions and the problem should be simpler because there appear to be only two or three different temporal representations. The problem deserves further study because, as we explain below, sensor dynamic range can benefit greatly from multiple temporal representations.

The visual pathways contain a large number of neural spatial representations of the visual field or spatial sampling mosaics. These neural mosaics appear to be specialized for representing visual features or for localized computations within a portion of the visual field. These multiple representations appear to be a fundamental architectural feature and one that might well be exploited by image reproduction systems.

B. Local Memory

In this section, we suggest a principle related to multiple representations: the use of local memory to assist computations and representations at many points within the imaging pathways. In engineering applications, local memory is found in many forms. Registers, buffers that coordinate timing when input and output between different devices, and local memory, to contain the results of intermediate calculations, are all widespread. In what follows, we consider the hypothesis that such local memory circuits will be found distributed in the visual pathways, even at the earliest stages.

The visual neuroscience literature contains many experiments that document the significance of a general-purpose short-term visual memory, sometimes called a *visual scratchpad*. In the behavioral literature, for example, short-term visual store can be demonstrated using a very simple experiment [43]–[45]. Suppose a 3×3 array of letters is flashed for a few milliseconds and the observer is asked to remember as many of the letters as possible. If no further indication is given, the observer will remember four to six of the nine letters in the array, suggesting that about half of the letters were encoded.

Now, suppose that a short time after presentation of the array, a tone is sounded. The pitch of the tone (high, middle, or low) indicates which row of letters (top, middle, or bottom) should be recalled. Typically, the observer will name all three letters in that row correctly. In this experiment, then, it appears as if the observer has encoded all the letters, not half. This phenomenon can be explained by positing that all the letters are stored in a short-term visual memory; if the tone is presented while the letters are in this store, all of the letters can be retrieved. However, if the observer tries to retrieve all of the letters, as in the first experiment, the memory trace fades during retrieval. Hence, only half the letters can be recalled. In addition to these behavioral studies of short-term memory, there is evidence for the presence of short-term memory specific to some visual tasks [46], but not others [47]. A widespread assumption is that the visual scratchpad is located outside of primary sensory areas [48]–[50].

In addition to the visual scratchpad, we urge consideration of the hypothesis that visual memory capabilities are distributed locally throughout visual cortex. At this point, we urge consideration of this point mainly based on the practical design considerations that motivated us to design systems with memory located near the image sensor. In that work, we found various uses for local memory. For example, the availability of a small amount of per-pixel memory makes it possible to separate the timing demands imposed by image acquisition and image communication. We will discuss the value of local memory in the engineering design more fully later in this paper. We pause here to consider what might be known about neurons that store local computational results within the visual pathways.

To understand how local memory would fit with current thinking in visual neuroscience, it is worth considering the current view of computation and communication. Perhaps a consensus description would be something like this: neural measurements of the image are continuously updated by peripheral neurons. These measurements are communicated to output cells by variations in the firing rate. For example, simple retinal signals derive information about image contrast features that are highly localized space and time; this information is continuously sent to cortical circuits that compute information spread over larger regions of space and time (local orientation, local motion). These results, in turn, are continuously sent to brain regions that analyze data spread across large amounts of space and time (object features).

Within this framework, local storage of information memory functions can be implemented by a variety of simple neural signals. For example, storing neurons whose signals represent multiple moving averages, each reflecting averaging over a different temporal periods, would serve as one type of memory. If such signals are present, they can be compared with relatively recent signals, say, comparing a signal averaged over the last 50 ms with one measured over the most recent 100 ms. The ability to make such comparisons is essential to motion analysis and other visual functions.

For example, a persistent signal of a few hundred milliseconds can be of value in registering information from just before and after an eye movement. Also, longer integration times usually provide more accurate estimates of intensity or contrast, but only if the scene is static. By storing a sequence of measurements in local memory, one can decide whether the long duration measurement reflects a high-quality measurement of a static scene or merely measures the time average of a set of moving objects. In this example, we use the array of signals to decide which is the most precise measurement of the image. It is likely that an array of memory cells, each providing a measure of the scene over the last few hundred milliseconds, can be used to create various types of context-sensitive algorithms that aid visual performance.

On this view, multiple representations and memory are closely linked. One copy representation measures a value close to the instantaneous stimulus-driven activity. A second represents more about the recent time history of the response, either by having a longer integration period or other specific computations. The value of a computed signal derived from these separate representations is the output buffer, another type of system memory that caches the result. By separating

this output from values driven continuously by the signal over time, the early sensory pathways can communicate a more reliable signal to other visual areas. This was among the reasons we use near-pixel memory in our imaging hardware.

The properties of memory circuits for other cortical functions, such as motor control or decision mechanisms, have been analyzed and various biophysical mechanisms have been proposed and studied (see, e.g., [51]–[53]). It is also useful to consider the system-level characteristics that visual neural random access memory (NRAM) circuits might possess. Memory circuits should represent information accumulated over various time scales, not just the (near) instantaneous response to the signal itself. The contents of the memory might reflect the result of some intelligent processing based on several measurements. In the event of a substantial change in the imaging environment (eye movements, blinks), it should be possible to reset the memory and begin afresh.

While there is not yet compelling overwhelming evidence for local memory neurons within visual or other sensory areas, such as V1, there are some recent intriguing reports. Neurons that appear to satisfy certain memory requirements have been reported widely across motor areas, including peripheral areas such as the spinal cord [54]. We note with particular interest that some neurons within area V1 mimic neighboring responses, as if they are representing the recent activity [55]. Local field potential measurements suggest that some neurons within V1 appear to have the characteristics needed for a short-term visual store [56]. Differential sensitivity to neighboring responses is one of the markers for local memory circuits and we hope that the relation between this neuronal response property and memory will be explored further.

III. MULTIPLE CAPTURE IMAGING SYSTEMS

An increasing number of image systems use multiple capture methods. These range from standard applications, such as the use of color filter arrays to acquire wavelength samples, to sophisticated 3-D graphics applications. Here, we describe some innovative applications that use multiple capture. We review methods that might integrate well with simple image sensor hardware.

A. Dynamic Range Extension

The most challenging scenes for image capture are those that include a wide range of ambient lighting. For example, natural viewing conditions often contain shadow and direct illumination regions. Intensity differences between such regions can exceed three orders of magnitude; neither photoreceptors nor typical image sensors can capture this intensity without significant loss of contrast information. Algorithms for selecting a best exposure value (combination of integration time and aperture) are important to electronic image acquisition.

The large intensity range is often caused by spatially varying illumination. Thus, single variable mechanisms,

such as controlling the lens aperture (pupil) or exposure time, are not a satisfactory control parameter for managing the large dynamic range. Instead, space-varying methods for extending dynamic range are needed.

One way to extend the dynamic range of an image acquisition system and still allow for spatial variation is to acquire multiple exposure durations of the image. For example, suppose one takes several images of a scene at different exposure durations. Then, one can assemble a single image from different locations within this image collection [57], [58]. In this way, each region of the final image has its own unique exposure duration.

Several groups have used methods for integrating data from multiple images captured at different temporal or spatial sensitivity. For example, Takahashi *et al.* [57] describe a temporal approach for extending image dynamic range. They combine multiple captures obtained using different exposure durations from a conventional charge-coupled device (CCD) sensor. Because the CCD readout process is quite slow and destroys the accumulated charges, multiple temporal captures combined with multiple sensor readouts, offchip processing is required.

Street [59] described a spatial approach for extending image dynamic range. He proposed creating a sensor array with photosensitive elements that each have a different light sensitivity. By combining measurements from sensors with high and low sensitivity, the overall dynamic range of the array could be extended.

Nayar and Mitsunaga also use this principle [60]. They propose to place neutral density filters that reduce the sensitivity of individual sensors. Placing a mosaic of neutral filters with several densities extends the dynamic range, much as using different temporal integration times will do so. This method trades spatial resolution for dynamic range and it fails to take advantage of the shorter integration time (better temporal resolution) that can be obtained for bright image regions. A complementary metal–oxide–semiconductor (CMOS) imager implemented by Bajovic, however, exploits this effect. Rather than fixing the exposure time and measuring the accumulated charge, this sensor measures the exposure time necessary to obtain a given charge level, thus exchanging the role of time and accumulated charge [61]. This design does not involve multiple captures, but it offers a space-varying approach to extending dynamic range.

B. Active Illumination

Kimachi and Ando use a combination of active illumination and multiple captures in a video sequence to permit rendering of the image as if illuminated by one of a variety of illuminant sources [62]. In their method, each illuminant flickers at a separate frequency. The image data can be demodulated to produce a separate still image that shows the scene illuminated by only one of the various sources.

To implement the method efficiently, Kimachi and Ando have designed and implemented a special purpose CMOS sensor. The sensor demodulates the time-varying charge at each pixel to measure the signal amplitude at a specific temporal frequency [62], [63]. Simpler forms of this method, say

differencing the images acquired using the ambient illuminant and sum of the ambient illuminant and an active illuminant, can be used for a variety of purposes, including reflectance estimation and illuminant estimation [63], [64].

C. Image Mosaicing, Stabilization, and Superresolution

For many applications, it is desirable to create high-resolution pictures with a wide field of view. Widening the image rendered on a fixed sensor trades field of view for spatial resolution. It is required, then, to develop methods for combining multiple captures to increase the field of view and retain the spatial resolution intrinsic to the optics and sensor [65], [66]. These are called *image mosaicing* methods.

Image-mosaicing applications range from aerial imaging to video retrieval to surveillance. The key computational step is the registration of overlapping regions in the individual images. If the camera only tilts and pans and there is, thus, no parallax between the images, the computation is relatively simple. Registration of planar patches, viewed from different angles such as the registration of aerial photographs of flat terrain, is also straightforward. Integrating views of a 3-D scene from different vantage points, however, requires the reliable and precise estimation of extrinsic and intrinsic camera parameters and 3-D scene structure [67], [68]. An example of such 3-D mosaicing is shown Fig. 2. The example shows that combining multiple images from different time instances can fill in occluded background, such that a complete map of the background results. Such a map is never visible in any individual image.

Image mosaicing is closely related to electronic image stabilization, where small involuntary pan and tilt motions are removed from a video sequence [69]. Such techniques are common today even in consumer camcorders.

Once image-stabilization algorithms yield multiple registered images, one can attempt to combine these images such that noise is reduced or even the spatial resolution is increased. Superresolution schemes have received considerable attention in the image processing research community over the last years and impressive results have been shown [70]–[73]. Superresolution techniques typically exploit imperfect antialiasing filtering before image sampling. In this case, frequency components beyond the Nyquist limit are still present in each image as aliasing. By combining multiple images with suitable linear shift-varying interpolation, these aliased components can be recovered and subpixel accuracy can be achieved. Interestingly, binocular human visual acuity is better than monocular acuity [74]. However, the binocular resolution gain is consistent with the gain expected by noise reduction and does not necessarily suggest superresolution involving aliasing cancellation from multiple images.

D. Multiple Apertures and Multiple Viewpoints

Information about object structure can be obtained by measuring multiple images using optics with different-sized concentric apertures or a variety of focal plane depths [69], [75]. By observing the relative blur between these images, one can learn something about the distance to the object and
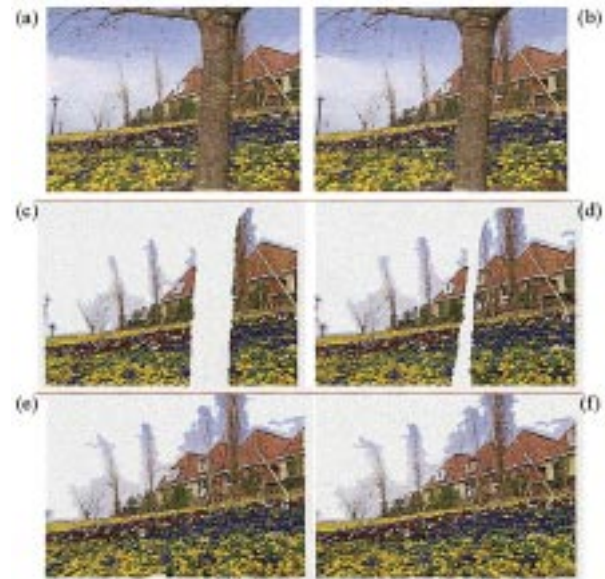

Fig. 2. Example of 3-D mosaicing. (a) and (b) Two frames from the video sequence *Flowergarden*. 3-D camera rotation and translation, as well as 3-D scene structure, are recovered from the sequence. Using this information, the tree can then be removed by using proximity as a cue. By combining multiple images from different time instances, the background occluded by the tree is filled in. (c) First frame, with the foreground removed, is shown. Constructed images after combining (d) 9, (e) 17, and (f) 25 images are shown. Background is never completely visible in a single input frame, but 3-D mosaicing nevertheless fills in the occluded background [67].

its 3-D structure. These algorithms, sometimes called depth from defocus, can be based on sophisticated image filtering and analyses [76]. Image content can be manipulated in a depth-dependent way by clever linear combinations of the images, for example, to create new images that represent a variety of depth of fields that are not directly measured [77].

Rather than capturing two or more images with different concentric apertures, images of the same object or scene are often captured with the same aperture, but from different vantage points. Classic stereo imaging is certainly among the oldest examples [78]. Stereo image pairs are viewed with an appropriate display system that directs each image to the corresponding eye. The analogy between stereo image acquisition and binocular biological vision is obvious.

It is desirable to build image applications in which a viewer can: 1) see a scene rendered an image from many vantage points; 2) fly through a rendered 3-D scene; and 3) change an object's 3-D position. Computer graphics algorithms accomplish this easily for a given 3-D geometry, surface reflectance and illumination. For natural scenes, where 3-D geometry, reflectance and illumination are not known *a priori* (and might even be ill-defined) multiple captures architectures can be used in systems designed to render a scene from several different vantage points.

The design of such systems is based upon the idea that the rays scattered from a surface contains a multiplicity of views (the so-called light field) and that by appropriate measurement these views can be obtained by a small array of cameras: Adelson and Wang trace this notion back to Leonardo's

notebooks [79]. Methods for acquiring multiple images and then reordering and interpolating the data to provide a variety of viewpoints of the scene are now used extensively in computer graphics. These techniques are part of the general trend toward *image-based rendering* [80], which complements the classic geometry-based computer graphics methods.

Adelson and Wang describe an integrated sensor design that acquires multiple images of the scene from slightly different points of view [79]. Their design uses a single camera lens coupled with a lenticular array that separates the image onto an interleaved mosaic of sensors. While it is elegant, the Adelson and Wang design only measures a narrow range of viewpoints. Systems that move a single camera (or the object, or both) achieve a larger range of viewpoints. Many groups have built these types of systems for static objects or scenes. To our knowledge, the largest light field that has been acquired in this fashion to date is that of the Statue of Night by Michelangelo. In March 1999, Levoy and a group of his students from Stanford University acquired 24 304 images, each with resolution of 1300×1000 red-green-blue pixels. One other noteworthy recent design is *Concentric Mosaics* where a camera is mounted at the end of a horizontal arm, looking outwards. As the arm swivels around, the camera acquires a full 360° panorama of the environment and simultaneously sweeps a range of vantage points [84]. For acquiring motion video from many simultaneous vantage points, systems with arrays of video cameras have been built, for example by a team at Carnegie Mellon University [85]–[87].

Biological vision systems infer a great deal about the 3-D structure of their environment by moving through it. Motion parallax is one of the strongest depth cues. Owls, when otherwise sitting still, can be observed to move their heads sideways to improve their depth perception by motion parallax and recognize prey. How 3-D scene structure is represented internally in biological vision systems is still an open question, but, we believe, multiple parallel representations will be found, analogous to efficient light-field representations in technical systems that combine both many views and coarse geometry to aid processing and compression [88]–[90].

IV. Example—The Stanford CMOS Sensor

In this section, we turn from general principles and a survey of techniques to specific systems we have developed for image acquisition and reproduction. By describing these applications, we hope to explain further why we have concluded that multiple image representations and local memory are of general importance to engineering and biological systems.

Earlier, we described how it is possible to acquire several images of a scene at different exposure durations and then to combine these distinct images to obtain a single high dynamic range image. A more efficient method of achieving the same purpose is to take one picture but make a series of sensor measurements at different times. From this series of measurements, we obtain a series of nested pictures with increasing exposure durations. Thus, from a single acquisition, we acquire multiple representations of the image at different exposures.

The CMOS sensors designed and implemented at Stanford University perform this process in hardware within a single sensor at high speed [91]–[93]. During a single exposure, the scene intensity is sampled at high speed using an analog-to-digital converter (ADC) placed within each pixel. The ADC measures the current stored charge nondestructively. Hence, the first time sample obtains a brief exposure; each subsequent time sample measures an increasingly long exposure. The length of the exposure duration is set at a level that permits the darkest image region to be estimated. We call the pixel design that incorporates the local ADC a DPS.

A. Multiple Capture DPS

The first implementation of the DPS design was a 640×512 CMOS sensor built using 0.35-μm technology. Each pixel spanned 10 μm on a side and contained a light sensor (photodiode), a bit-serial ADC, and one bit of dynamic random access memory (DRAM). In the first implementation, every ADC was shared among a block of four adjacent pixels (Fig. 3). A number of the sensor properties were under programmable control. For example, by manipulating external signals, the sensor's *transduction function* (the function that governs the relationship between light level and digital value) could be varied. In addition, under programmable control, measurements could be obtained from each pixel separately or from the sum of any combination of the four pixels within the block [1], [91]. Such spatial binning is well known, though perhaps less well known is a creative suggestion from Cornsweet and Yellott [94]. These authors suggest that the extent of the spatial pooling should depend on the local image intensity. The DPS imager we have developed uses this notion with respect to temporal integration.

Fig. 4 illustrates how this MCSI architecture extends dynamic range. Fig. 4(a)–(c) show measurements of a scene made by three temporal samples while the pixels were charging. The image in Fig. 4(a) is from the first sample. Portions of the image are under-exposed and contain considerable spatial noise. The image in Fig. 4(c) is from the last temporal sample, so that portions of this image are saturated. Because of the nonuniform illumination, no single exposure value captures both the light and dark parts of the image.

To create a high dynamic range image, we select pixels from the different sample times in Fig. 4(a)–(c), choosing an appropriate duration sample. The image in panel Fig. 4(d) is constructed from the three separate images in Fig. 4(a)–(c) by selecting the pixel reading from the last sample prior to pixel saturation (or the final sample if the pixel never saturates). The coloring in panel Fig. 4(e) indicates the image source for each pixel. Because the DPS design uses nondestructive reads, the total image acquisition time is no longer than the time needed to measure the dark image regions.

Fig. 5 illustrates graphically how a high dynamic range image is assembled from the multiple images. Fig. 5(a) illustrates how pixel digital value increases linearly with exposure duration (assuming a constant input intensity). Measurements from four pixels with different intensities are illustrated; the slope measures the pixel intensity. The bright pixel (highlight) would ordinarily reach saturation quickly. In
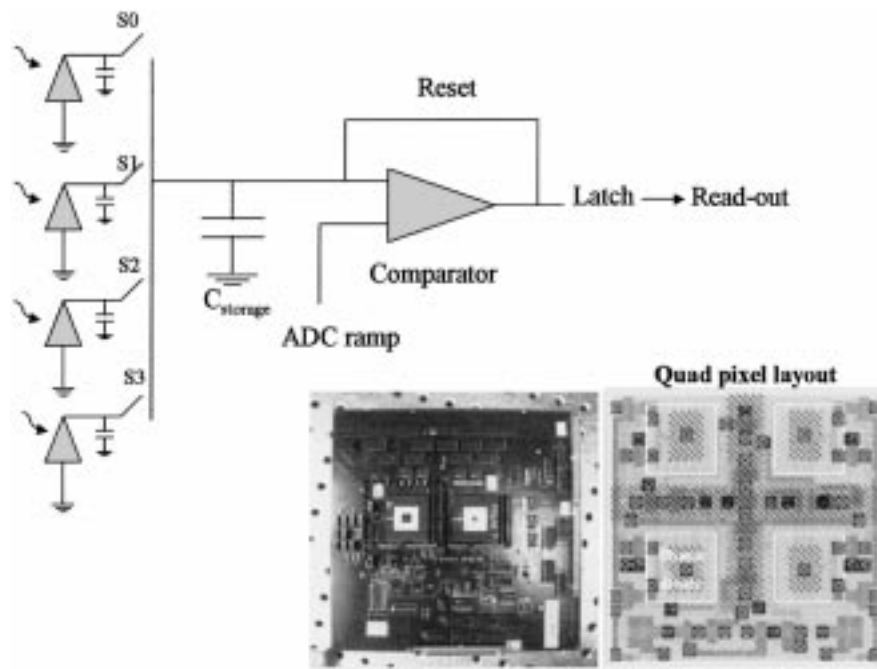
Fig. 3. Design of a CMOS image sensor with a programmable DPS. In this design, one ADC was shared amongst four pixels. Data were sampled several times, so that each pixel could have its own exposure duration. In addition, through programmable control of the switches, the pixels could be read out singly (high spatial resolution) or the charge from all of the pixels could be pooled and read (low spatial resolution). The sensor was implemented in 0.35-μm technology [91]. Pixel circuit is diagrammed on the upper left; an image of the sensor and the pixel layout are shown in the lower right.
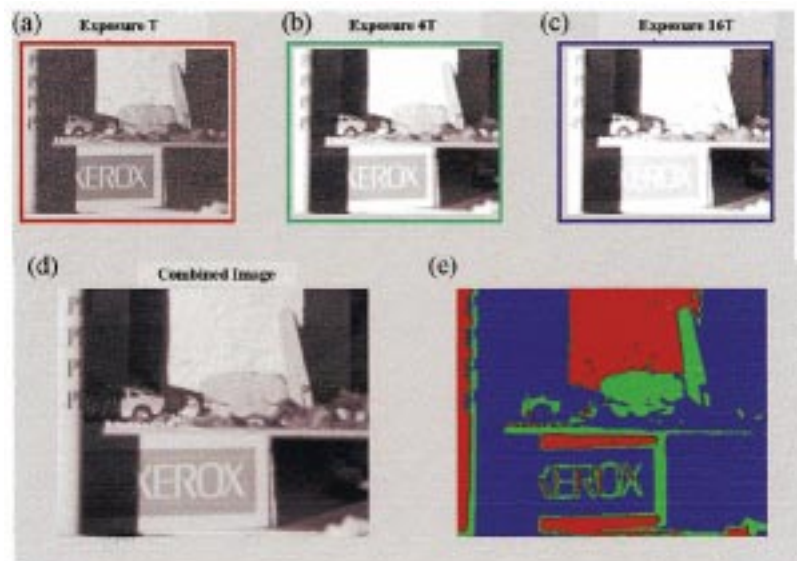

Fig. 4. Multiple capture architecture for dynamic range extension is illustrated. (a)–(c) Images were acquired during a single exposure by sampling the sensor repeatedly. Each of the exposure durations is problematic, containing either a noisy region or a saturated region. (d) Combining the last good sample and scaling the value appropriately, we can assemble the data into a single high dynamic range image. (e) Pixel color corresponds to the colored boxes surrounding the images in (a)–(c). Pixel color indicates the source image for each of pixels in the combined image (d).

a conventional camera, the exposure value algorithm might choose a brief duration, say, at the first time sample to avoid saturation, and pixels in the low lights would record low values that are not significantly different from the lowest quantization level or the system noise. Hence, using a single short capture, the dark regions are measured poorly.

The DPS sensor improves system sensitivity by making multiple image measurements. The bright regions are measured at the first time sample and the low lights are measured at later times. After enough time the signals from the dark image regions accumulate significant charge and become reliably different emerge from the lowest quantization level.
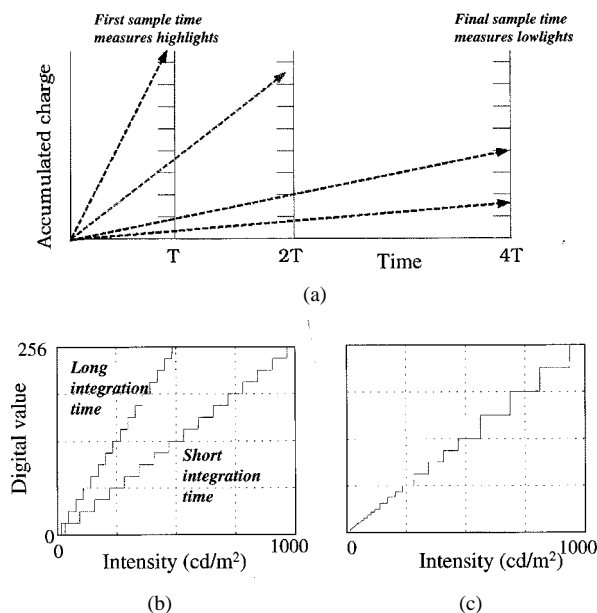
Fig. 5. Conversion of intensity to digital value (transduction function) is analyzed for a DPS. (a) For a constant image, charge accumulates linearly at each pixel. The four lines show charge accumulating in regions with four different intensities. In the DPS, the charge value is sampled at multiple times, indicated by the vertical dashed lines. Bright regions are measured by early time samples and dark regions are measured by later samples. Pixel response is coded both by the sampled level and the measurement time. (b) In a conventional pixel, the transduction function is linear and the size of the quantization bins is inversely related to the exposure duration. (c) DPS architecture combines data from short and long exposure durations. The transduction function is again linear, but the quantization bin increases with signal intensity.

With this design, the pixel digital value is estimated from the last sample of the image prior to saturation, when sensitivity is best. Each pixel is captured using a duration optimized for that pixel's intensity and the exposure duration is spatially varying.

Fig. 5(b) and (c) shows the transduction function that maps light intensity to digital value. The transduction function for a conventional single time sample imager is shown in Fig. 5(b). The stair patterns on the left and right of that graph show the functions for a long- and short-duration exposure, respectively. Changing the exposure duration alters the slope of the function and changes the size of the quantization bins. For each duration, the quantization bins are spaced evenly and the bin sizes are inversely related to exposure duration.

Fig. 5(c) shows the transduction function using the MCSI architecture. Again, the transduction function is linear. The quantization bin sizes vary systematically: low intensity levels are quantized more finely than high intensity levels. This occurs because the high intensity values are obtained from the short-exposure durations and the low intensity values are obtained from the long-exposure durations.

The uneven size of the quantization bins has an interesting visual consequence. The mean intensity difference needed to discriminate between high intensities is greater than that required to discriminate low intensities. This discrimination function corresponds well to human visual sensitivity and the observation that is generally called Weber's Law: threshold to a change in intensity increases proportionally to the signal level [18], [95]. Ordinarily, Weber's Law is modeled by the changing slope of a logarithmic transduction function. In the MCSI architecture, however, the transduction function is linear, which is beneficial for image-processing steps (e.g., demosaicing, color-balancing). The match to visual intensity discrimination is implemented by the varying quantization bin sizes inherent in the MCSI architecture.

Many types of rules can be used to derive a digital value from the multiple samples of the pixel. For example, it is possible to include a statistical analysis of the reliability of signals obtained from each of the samples. It is further possible to improve the quality of the measurements by analyzing the multiple captures and deciding whether or not the samples are consistent with a still image. Hence, in addition to expanding the dynamic range, we believe that multiple captures will also be helpful in interpreting image motion and, thus, image blur-free [6], [96].

Are there biological vision specializations that correspond to the multiple exposure durations? Measurements of human temporal integration show substantial variation with mean intensity level. At low mean intensity levels, temporal integration of the signal appears to extend over a period of more than 100 ms; at high mean intensity levels temporal integration is on the order of 10 ms [97]. Further, it appears that the visual system contains multiple temporal sensitivity mechanisms with peak sensitivities centered at different resonant frequencies [39], [40]. It is possible that these differences in measured temporal integration are mediated by signals from different mosaics and that a multiplicity of mosaics serves to increase the effective dynamic range of the photoreceptor signals.

Finally, we have emphasized that the DPS sensor differs from a CCD sensor because data are read nondestructively. In this regard, the DPS reading process parallels the way photoreceptor signals are read by postreceptoral neurons. Many different postreceptoral cells are driven by each photoreceptor and the photoreceptor signal follows the light continuously. In this way, postreceptoral neurons can be structured to measure photoreceptor signals in a variety of ways, averaging differentially over space and time.

One challenge that results from acquiring high dynamic range images is this: conventional displays do not have the capability to display such a dynamic range. To render the image shown in Fig. 4(d), we compressed the data using a logarithmic transformation. More sophisticated algorithms have been explored and more work on this topic is needed [98]–[100].

B. Local Memory

All electronic image sensors include temporary memory to support external transfer of image data from the pixels array. In CCD image sensors, the temporary memory is in the form of analog shift registers (vertical and horizontal CCDs). In CMOS image sensors with analog pixels [101], [102], the memory is located at the bottom of the sensor array and temporarily stores at least one line of image data.

In the first DPS implementation, each pixel has one bit of temporary memory to store the "bit-plane" before it is read
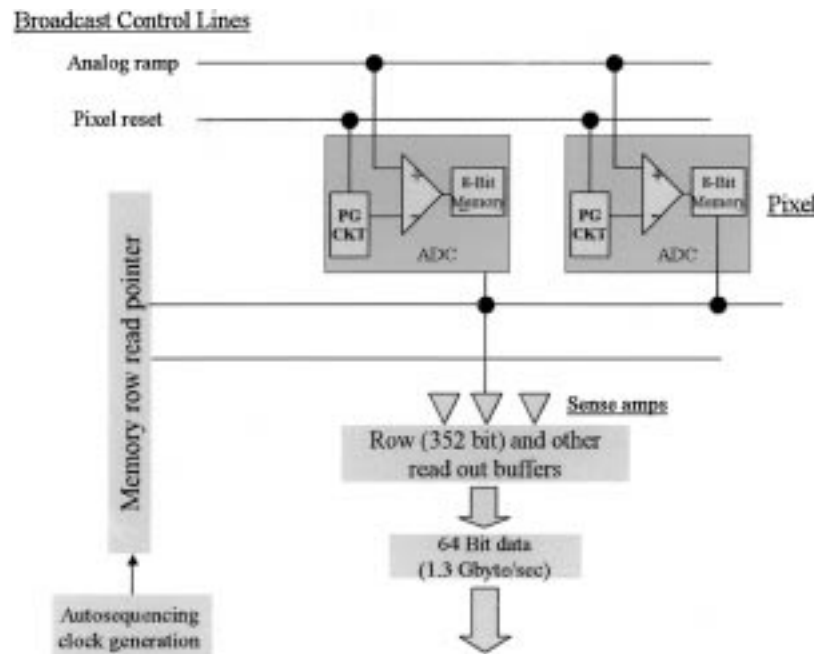
Fig. 6. Partial DPS circuit diagram.

out. With such a limited amount of local memory, intermediate results had to be transferred from the sensor to computer memory where the image was assembled. The next ADC step could not complete until the single-bit memory value was transmitted offchip; consequently, the ADC process was limited by the interdevice bandwidth.

In the second DPS chip, we added more memory to store the ADC results of each pixel. In this implementation, each pixel contained a bit-parallel ADC and 8 bits of DRAM (see the partial circuit diagram in Fig. 6) [93]. By placing more memory at the pixel, we could use a simpler ADC and communication mechanisms. This sensor, which has 352×288 pixel array, was built using 0.18-μm CMOS technology. The digital value measured at each sample time is stored temporarily in the 8-bit memory. Because the data is stored digitally on the sensor, the entire image array can be read at a very high rate (1 GB/s). Moreover, while memory readout takes place, the sensor continues to integrate charge (pipelining). The added memory improved the logical separation between the ADC process and the data communication.

Fig. 7 shows the image of a propeller rotating at 220 r/s. The images were acquired sequentially with an integration time of 100 μs per frame and no interframe delay (10 000 frame/s).

The additional memory affects a space-time tradeoff. We provide more space to store the ADC result, more data communication lines to address the memory, and a memory controller to read the results. In exchange, we simplify the communication timing for transmitting the results along the imaging pipeline: with added memory available to each pixel, the ADC process overlaps less with data communication. The initial reason for adding memory, then, was to improve the timing between the ADC measurements and communication channel timing.
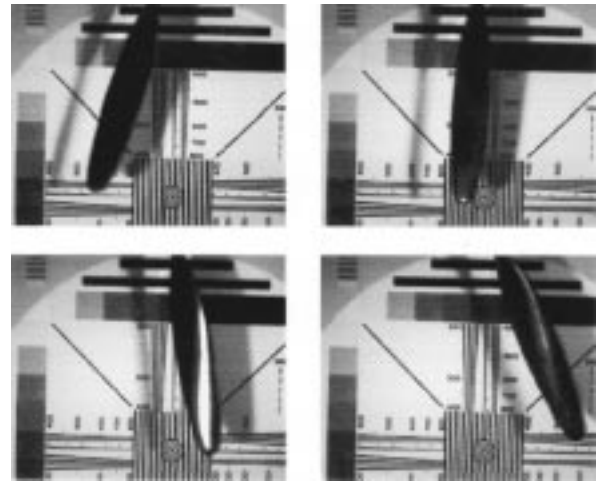

Fig. 7. Four images obtained from a high-speed DPS with 8 bits of local memory [93].

The next generation implementations of the our CMOS sensor will take further advantage of local memory and processing. By expanding the processing capabilities, it should prove possible to create a better quality sensor data. As one example, Liu and El Gamal [96], [103] have shown that by adding a small amount of extra processing, one can use near-optimal statistical methods to combine multiple samples of the pixel data on the sensor itself. These statistical methods account for the different levels of reliability of the data acquired at different times. Further developments in this area will include the detection of nonlinear growth in the time measurements that can account for image or camera motion. Lim and El Gamal have developed methods to combine multiple image samples to measure motion flow [6], [104]. These methods may lead to algorithms that will reduce or eliminate

motion blur. The Stanford University group has also developed algorithms using multiple samples in order to perform color balancing and rendering [63], [64].

Our experience with this hardware design suggests that algorithms based on local spatial and temporal computation are well suited to being carried out on the sensor. The value of local memory to enable this processing has impressed us and for this reason we suspect that even the peripheral visual pathways are served by local memory circuitry.

V. DISCUSSION—ELECTRONIC IMAGING AND VISUAL NEUROSCIENCE

There have been many attempts to characterize the connections between biological vision and electronic imaging systems. While it is beyond the scope of this paper to review this literature, we do note that the ideas developed here differ from the best known efforts to share ideas between the disciplines of biological and electronic imaging systems [4], [105], [106]. Both Marr and Mead argued that we should look to the biological design to inspire electronic design. In certain instances, we agree that the insights will necessarily flow from biological vision to engineering design. Image quality metrics for evaluating image systems must be based on models of human visual performance (e.g., CIELAB) and not the common metrics used in engineering design (e.g., RMSE) [107]–[111].

The approach here is based on the premise that ideas should flow in both directions. There is much value in examining the necessary elements of electronic implementation and asking whether these functions might be present in the neural implementation. Our focus here is on the value of local memory; it is also possible that other basic circuit functions, such as timing circuitry, are present in the biological circuitry (e.g., [112], [113]). Hence, efforts to think through clean designs of electronic imaging systems will raise novel questions and insights about neural circuitry.

We have identified principles that span the fields of electronic imaging and biological vision. The principle of multiple representations of the image is well established in the neuroscience literature and we believe that this idea is also relevant to the process of capturing and transforming images for reproduction. Based on our experience with system design, we believe that temporary storage of intermediate output is a valuable tool at all stages within the imaging pipeline and we suggest that such functionality is likely to be found distributed within the visual pathways. We are certain that many other common design principles can be found.

Despite the enormous differences in implementation detail between electronic circuits and biological vision, we believe that understanding these broad principles may be a natural bridge for the exchange of ideas. We hope this article helps create a bridge to carry ideas between these two disciplines.

REFERENCES

[1] B. Wandell, P. Catrysse, J. DiCarlo, D. Yang, and A. El Gamal, "Multiple capture single image with a CMOS sensor," in *Proc. Chiba Conf. Multispectral Imaging*, Chiba, 1999.

[2] R. N. Shepard, "Psychophysical complementarity," in *Perceptual Organization*, M. Kubovy and J. R. Pomerantz, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1981, pp. 279–341.

[3] H. von Helmholtz, "Helmholtz's Treatise on Physiological Optics," Thoemmes Press, Bristol, U.K., 2000.

[4] D. Marr, *Vision*. San Francisco, CA: Freeman, 1982.

[5] J. J. Gibson, *The Perception of the Visual World*. Boston, MA: Houghton Mifflin, 1950.

[6] S. H. Lim and A. E. Gamal, "Integrating image capture and processing—beyond single chip digital camera," in *Proc. SPIE Electronic Imaging Conf.*, San Jose, CA, 2001, pp. 219–226.

[7] B. A. Wandell, "Computational neuroimaging of human visual cortex," *Annu. Rev. Neurosci.*, vol. 10, no. 22, pp. 145–173, 1999.

[8] W. T. Press, A. A. Brewer, R. F. Dougherty, A. R. Wade, and B. A. Wandell, "Visual areas and spatial summation in human visual cortex," *Vis. Res.*, vol. 41, no. 10–11, pp. 1321–1332, 2001.

[9] N. Hadjikhani, A. K. Liu, A. M. Dale, P. Cavanagh, and R. B. H. Tootell, "Retinotopy and color sensitivity in human visual cortical area V8," *Nature Neurosci.*, vol. 1, no. 3, pp. 235–241, 1998.

[10] D. J. Calkins, "Representation of cone signals in the primate retina," *J. Opt. Soc. Amer. A*, vol. 17, no. 3, pp. 597–606, 2000.

[11] R. W. Rodieck, *The First Steps in Seeing*. Sunderland, MA: Sinauer, 1998.

[12] D. J. Felleman and D. C. V. Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[13] S. Zeki, *A Vision of the Brain*. London, U.K.: Blackwell, 1993.

[14] R. B. Tootell, A. M. Dale, M. I. Sereno, and R. Malach, "New images from human visual cortex," *Trends Neurosci.*, vol. 19, no. 11, pp. 481–489, 1996.

[15] S. Zeki, "Parallelism and functional specialization in human visual cortex," in *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor, NY: Cold Spring Harbor Lab., 1990, vol. 55, pp. 651–661.

[16] V. Braitenberg and A. Schüz, *Anatomy of the Cortex: Statistics and Geometry*. Berlin, Germany: Springer-Verlag, 1991.

[17] C. Stevens, "What form should a cortical theory take," in *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. Davis, Eds. Cambridge, MA: MIT Press, 1994, pp. 239–255.

[18] B. A. Wandell, *Foundations of Vision*. Sunderland, MA: Sinauer, 1995.

[19] J. Meadows, "Disturbed perception of colors associated with localized cerebral lesions," *Brain*, vol. 97, no. 4, pp. 615–632, 1974.

[20] S. M. Zeki, "The representation of colors in the cerbral cortex," *Nature Lond.*, vol. 284, no. 5755, pp. 412–418, 1980.

[21] S. Zeki, "A century of cerebral achromatopsia," *Brain*, vol. 113, no. 6, pp. 1721–1777, 1990.

[22] C. D. Salzman, C. M. Murasugi, K. H. Britten, and W. T. Newsome, "Microstimulation in visual area MT: Effects on direction discrimination performance," *J. Neurosci.*, vol. 12, no. 6, pp. 2331–2355, 1992.

[23] W. T. Newsome and C. D. Salzman, "The neural basis of motion perception," *Ciba Found. Symp.*, vol. 174, pp. 217–30, 1993.

[24] K. Tanaka, "Inferotemporal cortex and object vision," *Annu. Rev. Neurosci.*, vol. 19, pp. 109–139, 1996.

[25] F. W. Campbell and J. G. Robson, "Application of fourier analysis to the visibility of gratings," *J. Physiol. Lond.*, vol. 197, no. 3, pp. 551–566, 1968.

[26] J. G. Robson, "Spatial and temporal contrast sensitivity functions of the visual system," *J. Opt. Soc. Amer.*, vol. 56, no. 8, pp. 1141–1142, 1966.

[27] ——, "Neural images: The physiological basis of spatial vision," in *Visual Coding and Adaptability*, C. S. Harris, Ed. Hillsdale, NJ: Lawrence Erlbaum, 1980.

[28] C. Blakemore and F. W. Campbell, "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images," *J. Physiol. Lond.*, vol. 203, no. 1, pp. 237–260, 1969.

[29] J. M. H. d. Buf and S. Fischer, "Modeling brightness perception and syntactical image coding," *Opt. Eng.*, vol. 34, no. 7, pp. 1900–1911, 1995.

[30] N. Graham, *Visual Pattern Analyzers*. Oxford, U.K.: Oxford Univ. Press, 1989.

[31] R. M. Everson, A. K. Prashanth, M. Gabbay, B. W. Knight, L. Sirovich, and E. Kaplan, "Representation of spatial frequency and orientation in the visual cortex," *Natl. Acad. Sci.*, vol. 95, no. 14, pp. 8334–8338, 1998.

[32] N. P. Issa, C. Trepel, and M. P. Stryker, "Spatial frequency maps in cat visual cortex," *J. Neurosci.* , vol. 20, no. 22, pp. 8504–8514, 2000.

[33] R. T. Born and R. B. Tootell, "Spatial frequency tuning of single units in macaque supragranular striate cortex," *Natl. Acad. Sci.*, vol. 88, no. 16, pp. 7066–7070, 1991.

[34] R. B. Tootell, M. S. Silverman, and R. L. De Valois, "Spatial frequency columns in primary visual cortex," *Science*, vol. 214, no. 4522, pp. 813–815, 1981.

[35] "JPEG," Joint Photographic Experts Group, vol. 2001, 2001.

[36] P. J. Burt, "Smart sensing within a pyramid vision machine," *Proc. IEEE*, vol. 76, pp. 1006–1015, Aug. 1988.

[37] M. S. Lew and T. S. Huang, "Optimal multi-scale matching," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 1, 1999.

[38] J. A. J. Roufs and F. J. J. Blommaert, "Temporal impulse and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique," *Vis. Res.*, vol. 21, no. 8, pp. 1203–1221, 1981.

[39] A. B. Watson, "Temporal sensitivity," in *Handbook of Perception and Human Perfomance*, K. R. Bof, L. Kaufman, and J. P. Thomas, Eds. New York: Wiley, 1986, ch. 6.

[40] M. B. Mandler and W. Makous, "A three channel model of temporal frequency perception," *Vis. Res.*, vol. 24, no. 12, pp. 1881–1887, 1984.

[41] C. Yo and H. R. Wilson, "Peripheral temporal frequency channels code frequency and speed inaccurately but allow accurate discrimination," *Vis. Res.*, vol. 33, no. 1, pp. 33–45, 1993.

[42] R. J. Snowden, R. F. Hess, and S. J. Waugh, "The processing of temporal modulation at different levels of retinal illuminance," *Vis. Res.*, vol. 35, no. 6, pp. 775–789, 1995.

[43] G. Sperling, "The information available in brief visual presentation," *Psychol. Monographs*, vol. 74, no. 498, 1960.

[44] E. Averbach and A. S. Coriell, "Short-term memory in vision," *Bell Syst. Tech. J.*, vol. 40, no. 2, pp. 309–328, 1961.

[45] V. Di Lollo and P. Dixon, "Two forms of persistence in visual information processing," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 14, no. 4, pp. 671–681, 1988.

[46] S. Magnussen and M. W. Greenlee, "The psychophysics of perceptual memory," *Psychol. Res.*, vol. 62, no. 2–3, pp. 81–92, 1999.

[47] T. S. Horowitz and J. M. Wolfe, "Visual search has no memory," *Nature*, vol. 357, no. 6693, pp. 575–577, 1998.

[48] G. Krieger, I. Rentschler, G. Hauske, K. Schill, and C. Zetzsche, "Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics," *Spat. Vis.*, vol. 13, no. 2–3, pp. 201–214, 2000.

[49] K. Schill and C. Zetzsche, "A model of visual spatio-temporal memory: The icon revisited," *Psychol. Res.*, vol. 57, no. 2, pp. 88–102, 1995.

[50] M. W. Becker, H. Pashler, and S. M. Anstis, "The role of iconic memory in change-detection tasks," *Perception*, vol. 29, no. 3, pp. 273–286, 2000.

[51] J. M. Fuster, *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. Cambridge, MA: MIT Press, 1995.

[52] H. S. Seung, D. D. Lee, B. Y. Reis, and D. W. Tank, "The autapse: A simple illustration of short-term analog memory storage by tuned synaptic feedback," *J. Comput. Neurosci.*, vol. 9, no. 2, pp. 171–185, 2000.

[53] X. J. Wang, "Synaptic basis of cortical persistent activity: The importance of nmda receptors to working memory," *J. Neurosci.*, vol. 19, no. 21, pp. 9587–9603, 1999.

[54] Y. Prut and E. E. Fetz, "Primate spinal interneurons show pre-movement instructed delay activity," *Nature*, vol. 401, no. 6753, pp. 590–594, 1999.

[55] V. Dragoi, C. Rivadulla, and M. Sur, "Foci of orientation plasticity in visual cortex," *Nature*, vol. 411, no. 6833, pp. 80–85, 2001.

[56] H. Super, H. Spekreijse, and V. A. Lamme, "A neural correlate of working memory in the monkey primary visual cortex," *Science*, vol. 293, no. 5527, pp. 120–124, 2001.

[57] K. Takahashi, T. Hieda, C. Satoh, T. Masui, T. Kobayashi, and K. Yoshimura, "Image sensing device with diverse storage times used in picture composition," U.S. Patent 5 638 118, 1995.

[58] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, Aug. 1997, pp. 130–135.

[59] R. Street, "High dynamic range segmented pixel sensor array," U.S. Patent 5 789 737, 1998.

[60] S. K. Nayar and T. Misunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 472–479.

[61] V. M. Brajovic, R. Miyagawa, and T. Kanade, "Temporal photoreception for adaptive dynamic range image sensing and encoding," *Neural Netw.*, vol. 11, no. 7–8, pp. 1149–1158, 1998.

[62] A. Kimachi and S. Ando, "Time-domain correlation image sensor: CMOS design and integration of demodulator pixels," in *Proc. SPIE*, vol. 3950, 2000, pp. 220–227.

[63] F. Xiao, J. DiCarlo, P. Catrysse, and B. Wandell, "Image analysis using modulated light sources," in *Proc. SPIE Electronic Imaging Conf.*, San Jose, CA, 2001, pp. 22–30.

[64] J. DiCarlo, P. Catrysse, F. Xiao, and B. Wandell, "System and method for estimating physical properties of objects and illuminants in a scene using temporally modulated light emission," Proc. 9th Color Imaging Conf., pp. 27–34, 2001, submitted for publication.

[65] R. Szeliski, "Video mosaics for virtual environments," *IEEE Comput. Graph. Applicat.*, vol. 16, pp. 22–30, Mar. 1996.

[66] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1144–1154, Oct. 2000.

[67] E. Steinbach, P. Eisert, and B. Girod, "Motion-based analysis and segmentation of image sequences using 3-D scene models," *Signal Processing*, vol. 66, no. 2, pp. 233–247, 1998.

[68] S. Srinivasan and R. Chellappa, "Fast structure from motion recovery applied to 3D image stabilization," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 6, 1999, pp. 3357–3360.

[69] A. Pentland, S. Sherock, T. Darrell, and B. Girod, "Simple range cameras based on focal errors," *J. Opt. Soc. Amer. A*, vol. 11, no. 11, pp. 2925–2934, 1994.

[70] S. Srinivasan and R. Chellapa, "Image sequence stabilization, mosaicking and superresolution," in *Handbook of Image and Video Processing*, A. Bovik, Ed. New York: Academic, 2000, pp. 259–268.

[71] A. Zomet and S. Peleg, "Efficient super-resolution and applications to mosaics," in *Proc. 15th IEEE Int. Conf. Pattern Recognition*, vol. 1, 2000, pp. 579–583.

[72] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.

[73] A. J. Patti, M. I. Sezan, and M. A. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Processing*, vol. 6, pp. 1064–1076, Aug. 1997.

[74] R. B. Cagenello, A. Arditi, and D. L. Halpern, "Binocular enhancement of visual acuity," *J. Opt. Soc. Amer. A*, vol. 10, no. 8, pp. 1841–1848, 1993.

[75] M. Subbarao and N. Gurumoorthy, "Depth from defocus: A spatial domain approach," *Int. J. Comput. Vision*, vol. 13, no. 3, pp. 271–294, 1994.

[76] M. Watanabe and S. K. Nayar, "Rational filters for passive depth from defocus," *Int. J. Comput. Vision*, vol. 27, pp. 203–225, 1997.

[77] K. Aizawa and K. Kodama, "Signal processing based approach to image content manipulation," in *Image and Multidimensional Signal Processing '98*, H. Niemann, H. Seidel, and B. Girod, Eds. Alpback, Austria: IEEE Signal Processing Society, 1998, pp. 267–270.

[78] C. Wheatstone, "Contributions to the physiology of vision. Part I: On some remarkable and hitherto unobserved, phenomena of binocular vision," *Philos. Trans. R. Soc. Lond.*, vol. 128, 1838.

[79] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 99–106, Feb. 1992.

[80] S. B. Kang, "Survey of image-based rendering techniques," Digital Equipment Corp. Cambridge Res. Lab., Cambridge, MA, Tech. Rep. 97/4, 1997.

[81] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. ACM SIGGRAPH*, Aug. 1996, pp. 31–42.

[82] P. E. Debevec, "Image-based modeling and lighting," *Computer Graphics*, vol. 33, no. 4, pp. 44–50, 1999.

[83] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proc. ACM SIGGRAPH*, Aug. 1996, pp. 43–54.

[84] H.-Y. Shum and L.-W. He, "Rendering with concentric mosaics," in *Proc. ACM SIGGRAPH*, 1999, pp. 299–306.

[85] P. J. Narayanan and T. Kanade, "Virtual worlds using computer vision," in *Proc. IEEE ATR Workshop on Computer Vision for Virtual Reality Based Human Communications*, 1998, pp. 2–13.

[86] S. Vedula, P. Rander, H. Saito, and T. Kanade, "Modeling, combining and rendering dynamic real-world events from image sequences," in *Proc. 4th Int. Conf. Virtual Systems and Multimedia*, 1998.

[87] L. Guernsey, "Turning the super bowl into a game of pixel," *The New York Times*, Jan. 25, 2001.

[88] M. Magnor, P. Eisert, and B. Girod, "Multi-view image coding with depth maps and 3-D geometry for prediction," in *Proc. Visual Communication and Image Processing*, 2001.

[89] B. Girod and M. Magnor, "Two approaches to incorporate approximate geometry into multi-view image coding," in *Proc. IEEE Int. Conf. Image Processing*, 2000.

[90] M. Magnor, P. Eisert, and B. Girod, "Model-aided coding of multi-viewpoint image data," in *Proc. IEEE Int. Conf. Image Processing*, 2000.

[91] D. X. D. Yang, A. El Gamal, B. Fowler, and H. Tian, "A 640*512 CMOS image sensor with ultra wide dynamic range floating-point pixel-level ADC," in *Proc. 1999 IEEE Int. Solid-State Circuits Conf.*, San Francisco, CA, 1999, pp. 1821–1834.

[92] B. Fowler, D. Yang, and A. El Gamal, "A CMOS area image sensor with pixel-level A/D conversion," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 1994, pp. 226–227.

[93] S. Kleinfelder, S. H. Lim, X. Q. Liu, and A. El Gamal, "A 10 000 frames/s 0.18 μm CMOS digital pixel sensor with pixel-level memory," in *Proc. 2001 IEEE Int. Solid-State Circuits Conf.*, 2001, pp. 88–89.

[94] T. N. Cornsweet and J. I. Y. Jr, "Intensity-dependent spatial summation," *J. Opt. Soc. Amer. A*, vol. 2, no. 10, pp. 1769–1786, 1985.

[95] H. B. Barlow, "Optic nerve impulses and weber's law," in *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor, NY: Cold Spring Harbor Lab., 1965, vol. 30, pp. 539–546.

[96] X. Q. Liu and A. El Gamal, "Photocurrent estimation from multiple nondestructive samples in a CMOS image sensor," in *Proc. SPIE Electronic Imaging Conf.*, vol. 4306, 2001, pp. 450–458.

[97] H. B. Barlow, "Temporal and spatial summation in human vision at different background intensities," *J. Physiol.*, vol. 141, pp. 337–350, 1958.

[98] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Trans. Visual. Comput. Graphics*, vol. 3, pp. 291–306, Oct.–Dec. 1997.

[99] J. Tumblin and G. Turk, "LCIS: A boundary hierarchy for detail-preserving contrast reduction," in *Proc. ACM SIGGRAPH*, 1999, pp. 83–99.

[100] J. DiCarlo and B. Wandell, "Rendering high dynamic range images," in *Proc. SPIE*, vol. 3965, 2000, pp. 392–401.

[101] E. R. Fossum, "Active pixel sensors: Are CCD's dinosaurs," in *Proc. SPIE*, vol. 1900, 1993, pp. 2–14.

[102] ——, "CMOS image sensors: Electronic camera on a chip," in *Proc. Int. Electron Devices Meeting*, 1995, pp. 17–25.

[103] X. Liu and A. El Gamal, "Simultaneous image formation and motion blur restoration via multiple capture," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, 2001, p. IMDSP-SF3.4.

[104] S. Lim and A. E. Gamal, "Optical flow estimation using high frame rate sequences," in *Proc. Int. Conf. Image Processing*, vol. 2, 2001, pp. 925–928.

[105] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.

[106] ——, "Neuromorphic electronic systems," *Proc. IEEE*, vol. 78, pp. 1629–1636, Oct. 1990.

[107] M. Fairchild, *Color Appearance Models*. Reading, MA: Addison-Wesley, 1998.

[108] R. W. G. Hunt, "The Reproduction of Colour," Fountain Press, Kingston-upon-Thames, U.K., 1987.

[109] B. Wandell and L. Silverstein, "Digital color reproduction," in The Science of Color, 2nd ed, S. Shevell, Ed. Washington, DC: Opt. Soc. Amer., to be published.

[110] B. Girod, "Psychovisual aspects of image communication," *Signal Processing*, vol. 28, no. 3, pp. 239–251, 1992.

[111] ——, "What's wrong with mean squared error?," in *Visual Factors of Electronic Image Communications*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

[112] W. Singer and C. Gray, "Visual feature integration and the temporal correlation hypothesis," *Annu. Rev. Neurosci.*, vol. 18, pp. 555–586, 1995.

[113] W. Singer, "Neuronal synchrony: A versatile code for the definition of relations?," *Neuron*, vol. 24, no. 1, pp. 49–65, 1999.

Brian A. Wandell was born in New York, NY, on October 6, 1951. He received the B.S. degree in mathematics and psychology from the University of Michigan, Ann Arbor, in 1973 and the Ph.D. degree from the University of California, Irvine, in 1977.

He joined the Faculty of Stanford University in 1979. In engineering, he founded the Image Systems Engineering Program at Stanford. He is Co-Principal Investigator of the Programmable Digital Camera program, an industry-sponsored effort to develop programmable CMOS sensors. He authored *Foundations of Vision* (Sunderland, MA: Sinauer, 1995), a textbook on vision science. His research includes the image systems engineering and vision science. His work in vision science uses both functional MRI and psychophysics and includes the computation and representation of color and measurements of reorganization of brain function during development and following brain injury.

Dr. Wandell is a Fellow of the Optical Society of America. He received the 1986 Troland Research Award from the U.S. National Academy of Sciences for his work in color vision, the McKnight Senior Investigator Award in 1997, and the Macbeth Prize from the Inter-Society Color Council in 2000.

Abbas El Gamal (Fellow, IEEE) received the B.S. degree in electrical engineering from Cairo University, Cairo, Egypt, in 1972, and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1977 and 1978, respectively.

From 1978 to 1980, he was an Assistant Professor of Electrical Engineering with the University of Southern Caliornia. He joined the Stanford University faculty in 1981, where he is currently a Professor of Electrical Engineering. From 1984 to 1988, while on leave from Stanford University, he was Director of LSI Logic Research Laboratory and then a Cofounder and Chief Scientist of the Actel Corporation. From 1990 to 1995, he was a Cofounder and Chief Technical Officer of Silicon Architects, which was acquired by Synopys. He is currently a Principal Investigator on the Stanford Programmable Digital Camera project. He serves on the board of directors and advisory boards of severel integrated circuit and computer-aided design companies. He has authored or coauthored over 100 papers and 25 patents in these areas. His current research intersts include CMOS image sensors and digital camera design, image processing, network information theory, and electrically configurable VLSI design and computer-aided design.

Dr. El Gamal is a Member of the IEEE International Solid-State Circuits Conference Technical Program Committee.

Bernd Girod (Fellow, IEEE) received the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1980 and the Doctoral degree from the University of Hannover, Germany, in 1987.

He is currently a Professor of Electrical Engineering and (by courtesy) Computer Science with the Information Systems Laboratory, Stanford University, Stanford, CA. Previously, he was a Chaired Professor of Telecommunications with the Electrical Engineering Department, University of Erlangen-Nuremberg, from 1993 to 1999. Prior visiting or faculty positions include the Massachusetts Institute of Technology, the Georgia Institute of Technology, and Stanford University. He has also been involved with several startup ventures as Founder, Director, Investor, or Advisor, among them Vivo Software, 8 × 8, and RealNetworks. He authored or coauthored one book and more than 200 technical papers in his field. His current research intersts are in the areas of networked multimedia systems, video signal compression, and three-dimenstional image analysis and synthesis.