A Real-Time Neural System for Color Constancy

Andrew Moore, *Student Member, IEEE*, John Allman, and Rodney M. Goodman, *Member, IEEE*

Abstract—This paper presents a neural network approach to the problem of color constancy. Various algorithms based on Land's retinex theory are discussed, with an eye on neurobiological parallels, computational efficiency, and suitability for VLSI implementation. The efficiency of one algorithm is improved by the application of resistive grids and is tested in computer simulations; the simulations make clear the strengths and weaknesses of the algorithm. A novel extension to the algorithm is developed to address its weaknesses. An electronic system based on the original algorithm was built, using subthreshold analog CMOS VLSI resistive grids, that operates at video rates. The system displays color constancy abilities and qualitatively mimics aspects of human color perception.

I. INTRODUCTION

ANYONE who has tried to take a picture of a friend or of a vase of flowers under different lighting conditions has realized that our present technology for capturing images is flawed. While the color of skin or of a rose may look the same to us at high noon or at sunset, a film or video camera just does not see it that way. Color constancy is the ability of the human visual system to judge, preattentively, the reflectance of objects in the visual world under a range of different illuminants. Color constancy is not perfect: if the illuminant is strongly saturated (lacking in white), we make errors. However, for natural variations, such as changing daylight conditions caused by varying cloud cover, we do rather well.

While the problem of color constancy has been recognized for some time (Helmholtz commented on it [1]), the computational essence of the problem has been grappled with only recently. In this paper, we present a system that addresses this problem for video images. The idea for the system originated in consideration of mammalian neurophysiology and human psychophysics; its validity was tested in computer simulations and it was implemented using analog VLSI. The electronic system is the first real-time instantiation of Land's retinex theory of color constancy for video imaging.

In the following, we first describe the neurobiological and computational aspects of the problem. Next we describe various manifestations of Land's retinex algorithm, improve on one of them by applying resistive grids, and propose a novel one. Results of computer simulations of the improved Land algorithm and the new algorithm are presented. Finally we describe an electronic system which performs the improved algorithm at video rates.

II. NEUROBIOLOGICAL AND COMPUTATIONAL ISSUES IN COLOR CONSTANCY

A. Neural Computation of Color

The only system which, at present, is capable of approximating color constancy in real time is the nervous system. As such, all algorithms for color constancy should be judged in comparison with the CNS. The following is a necessarily simplified sketch of the neurobiology of color vision, a subject of continuing extensive research.

We sense light with three classes of receptors, the cones (rod vision is not considered here). The three classes of cones have different spectral band-pass properties. They are called long, medium, and short (from the spectral bands that they are sensitive to), or colloquially, red, green, and blue. At the level of retinal ganglion cells, the output cells of the retina, the image has been transformed from three arrays of band-pass signals to three arrays of combinations of those signals. One set of outputs codes along the black–white axis of the color space, and the other two code along the red–green and yellow–blue axes.

Cortical visual area V4, many synapses "upstream" from the retina, receives inputs from lower visual areas that work with color difference signals. In early investigations V4 was dubbed the color area [2] because the cells could only be excited with color. (This view is now modified, as it is known that V4 cells can also respond to orientation and binocular disparity [3]. Here, only the spectral properties of V4 cells are considered.) Cortical neurons in this visual area are especially interesting since they seem to be responsive to perceived color, rather than wavelength; that is, they are "color constant" according to Zeki's informal study [4], [5]. An example of his work is as follows. With white illumination, he centered a cell's receptive field on one colored patch from a large field of many colored patches. A given cell responded only to a red patch, for example—yellow or green patches produced no cell firing under white light. Next he centered the cell receptive field on a yellow patch, turned off the white light, and carefully constructed a new illuminant such that *the spectrum of light coming from the yellow patch was the same for this illuminant as the spectrum from the red patch under the white illuminant.* To a human observer, the yellow patch still looked yellow, not red. Zeki found that the V4 cell did not fire when presented with the yellow patch which reflected red light, and so had discounted the illuminant. In contrast, he found that cells in the first visual area are sensitive to wavelength alone and so responded like a photometer, firing identically to a red patch under white light and a yellow patch under red light.

Desimone and his colleagues [6] obtained results from V4 cells that are in a sense supportive of Zeki's observations. In a study of the "nonclassical" receptive fields of extrastriate visual neurons, they found that V4 cells respond to white, and to many wavelengths, but have a maximal response at some wave-

length "analogous to a broad-band color filter, such as a piece of colored glass." They found that V4 cells are suppressed by stimuli in a large (30° or greater) "silent surround"; the suppression is maximum at the wavelength most effective in exciting the cell center, and falls off as the surround stimulus wavelength is moved away from the most effective center stimulus wavelength.

By comparing the color in the center of the receptive field with the color in a large area outside of the center, V4 cells judge *relative color*. This is the presumed basis for color constancy. If the illuminant is red, for example, a reddish cast is added to all parts of the scene. Though the cell center may see red, its surround does as well, and so it will not respond. Thus the cell discounts the illuminant and contributes to color constancy.

B. The Computational Essence of Color Constancy

Under normal variations (e.g., noon versus sunset or clear versus overcast sky), the spectrum of daylight varies somewhat. The variation is limited enough that is can be represented with three spectral basis functions [7]. A wide range of naturally occurring object reflectances can also be described with only three basis functions [8]. The light reaching a point on the retina, i.e., the set of three cone *quantum catches*, is just the product of the illuminant and the reflectance at a point in the world, to a first approximation (this is refined below). Thus, six unknowns determine the light impinging on each point of the retina, and only three data values, the quantum catches of the three cone classes, are available for further processing by the visual system. Yet we seem to be able to discount the illuminant and perceive the object reflectance [9]–[14]. *This is the computational problem of color constancy: How do we solve three equations in six unknowns?*

Various models of color constancy exist in the literature [15]–[23]. Several authors [9], [16]–[18] have shown that if the three basis functions for reflectance, illumination, and cone absorption are different, a color constancy algorithm must solve a tensor transformation from six unknowns to three knowns in order to find the reflectance from the cone signals with varying illumination. Each model makes assumptions to simplify the problem more or less. Here, we focus on Land's models since, through a powerful set of simplifying assumptions, he reduces the computational complexity of the color constancy problem tremendously [24], perhaps more than any other model.

III. LAND'S RETINEX THEORY

Land's assumptions and various versions of color constancy algorithms based on them are discussed in this section. For each algorithm, the biological basis, computational complexity, and suitability for VLSI implementation are noted.

A. Three Separate Lightnesses

One of Land's basic premises is that color constancy can be achieved by the computation of three separate *designators* or lightness values at each point, in three separate systems called retinexes. (Since he was not sure at first whether the computation took place in the retina or in the cortex, he coined the term *retinex*.) Further, he emphasizes the ability of the nervous system to perceive reflectance even though the illuminant is varying (albeit slowly) in space.

The three lightness signals are assumed to be independent;

Land does not state how this may occur, but Hurlbert and Poggio offer a derivation [24]. With this assumption, the color constancy tensor relation collapses to three independent equations. Within each channel i, the lightness l_i is the product of the illuminant m_i and reflectance p_i:

$$l_i(x) = m_i(x)\, p_i(x). \tag{1}$$

The log is taken to form a sum:

$$l_i'(x) = \log l_i(x) = \log m_i(x) + \log p_i(x)$$
$$= m_i'(x) + p_i'(x). \tag{2}$$

Once the two variables that make up the color signal are separated, the second of the three main assumptions of retinex theory is applied, namely, that the illumination is slowly varying in space but the reflectance signal varies mostly at sharp edges. Implementations of retinex theory work with this assumption by removing the slowly varying component $m'(x)$ to produce an image that depends only on reflectance $p'(x)$. Homomorphic filtering algorithms also use the logarithm to separate the components of the color signal in this way [25], [26]. The slowing varying component (the illuminant) is then separated by low-pass filtering via Fourier techniques. Retinex algorithms, in contrast, perform all operations in the spatial domain. Three implementations are presented.

B. Early Implementations

Land's original scheme [19], [20] considers the color signal at a point in one color plane of an image relative to a spatial average signal computed along a set of paths from other points in the image to the point in question. The starting point of a particular path is chosen randomly, and the logarithms of the ratio of color signals at transitions encountered along the path are accumulated if the transition represents a reflectance change versus a change in shading. A threshold operation is used to make this distinction. This procedure is repeated for many paths and the resulting values are averaged. The resulting average of logs is the log of the lightness of the point divided by a measure of the spatially averaged lightness. For an infinite number of paths and no thresholding, this measure of the spatial average is the geometric mean [27]. Land claims that he gets good results with 200 paths. Finally, the reported lightness is normalized to the lightest point in the image for this color plane.

The division by the average lightness and subsequent normalization imply the third main assumption of Land's theory, that the spatial average reflectance in each lightness channel is constant for all images. The retinex algorithm, then, operates under a *gray world assumption*. It is possible to defeat the algorithm by placing a strongly colored patch in a very simple scene so that the average reflectance is not gray [27].

This algorithm and variants of it can produce nice results (see, for example McCann's images in [28]). However, the procedures are cumbersome and it is difficult to see how the nervous system could carry them out. Further, the computational complexity, though reduced by assuming three separate lightness channels, is still daunting. For each point in the image, much of the rest of the image must be traversed by one of the paths to obtain the correct lightness. In other words, for an $N \times N$ image, on the order of N^4 calculations are required. In a VLSI implementation, each pixel would have to be connected with many other pixels.

Horn [15], [22] utilizes the Laplacian operator to compare lightness across edges. The Laplacian of the image is then

thresholded, to remove the slowly varying illuminant. Finally, the inverse Laplacian is performed. Analytically, this is done by convolving with the Green's function for the Laplacian, $(1/2\pi) \log (r)$. In a resistive grid framework, the Laplacian is inverted via a feedback network. In a digital implementation, the Poisson equation is solved iteratively via Gauss–Siedel elimination. Several variants of this implementation exist in the literature (e.g., [23] and [29]).

Marr [30] has proposed a scheme in which the nervous system may carry out this implementation, and the resistive grid framework is suitable for analog VLSI implementation since only nearest-neighbor connections between pixels are required. The undesirable spatial connectivity of the previous implementation is converted to the time domain. That is, the time required for the feedback network to settle is sufficient for information to cross the entire image space through the nearest-neighbor connections. To our knowledge, no one has attempted to build chips based on this algorithm.

C. Recent Implementations

In 1986, Land published an alternative to the algorithm described above [21]. This implementation involves computing an average weighted by distance from the point in question, and subtracting the log of this average from the log of the lightness of the point in question. This idea came from Land after his collaboration with Livingstone and Hubel [31] and Zeki [2] and has a strongly biological flavor to it. That is, the operator he uses looks like a cortical "nonclassical" receptive field, with a narrow center and a huge surround [6], [32].

In practical terms, the algorithm corresponds to subtracting from an image a blurred version of itself. The distance weighting (type of blurring) Land proposes varies as $1/r^2$, so the operation is a center minus surround operation, where the surround is the center convolved with a $1/r^2$ kernel:

$$l'_{\text{out},i}(x, y) - \log \left(l_i(x, y) \otimes \frac{1}{r^2} \right), \qquad r \neq 0. \qquad (3)$$

Hurlbert arrived at the same sort of operation analytically [24] with a Gaussian kernel:

$$l'_{\text{out},i}(x, y) = l'_i(x, y) - l'_i(x, y) \otimes e^{-r^2/\sigma} \qquad (4)$$

where σ is large enough that the kernel extends across most of the image. Except for the different kernels, the only difference between the two procedures is that Hurlbert's involves taking the log of the lightness of the surrounding points before rather than after averaging. She claims that in practice there is little difference between the two procedures [33].

This type of retinex algorithm, then, has a biological basis and sound computational underpinnings. But the complexity is too great. Since the required surround is so large, such a convolution across an $N \times N$ pixel image entails on the order of N^4 operations. On a chip, this corresponds to explicit connections from each pixel to most if not all other pixels.

A similar operation can be carried out much more efficiently

by switching from a convolution to a resistive grid calculation. The operations are similar since the weighting of neighboring points (Green's function) in a resistive grid decreases in the limit as the exponential of the distance from a given location on a resistive grid. With this type of kernel, the operation in each retinex (color channel) is

$$l'_{\text{out},i}(x, y) = l'_i(x, y) - l'_i(x, y) \otimes e^{-|r|/\lambda} \qquad (5)$$

where λ is the length constant or extent of weighting in the grid [34]. Since the calculation is purely local, the complexity is reduced dramatically from $O(N^4)$ to $O(N^2)$. On a chip, a local computation corresponds to connections only between nearest-neighbor pixels. So, in this novel retinex implementation, since a resistive grid is used to form the spatial average (i.e., to blur the image for subtraction from the original), the complexity is reduced to tractable levels, and the algorithm is appropriate for implementation in analog VLSI.

IV. SIMULATION RESULTS

A. Simulations of the Retinex Algorithm

Tools for simulating Land's most recent algorithm were developed and used to process both black and white images and color images. First, Hurlbert's results for one-dimensional black and white images were confirmed. A large spatial sample was obtained around each pixel by convolving with a filter whose weights drop off exponentially as the distance from the center pixel. This surround value was subtracted from the center pixel value. As Hurlbert and Poggio report [33], this scheme handily removes illumination gradients. Next, the simulation was extended to two dimensions with similar results and tremendous increase in run time, owing to the $O(N^4)$ complexity of the Gaussian convolution needed to form the spatial average for subtraction. For 128 × 128 pixel black and white images, the simulation took over an hour on a Sun 4 workstation.

Next, the same results were arrived at much more efficiently by switching from a convolution to a resistive grid calculation. The resistive grid simulation runs in a minute rather than an hour, since the calculation is purely local. With resistive grid code, color images were simulated next (Fig. 1(a) and (b)).

Specifically, in color simulations of the Land algorithm, 512 × 512 pixel images are subsampled to 128 × 128 resolution. Our frame grabber captures 8 bits each of R, G, and B. Within a color plane, the 8 bit pixel values are converted to floating point numbers and the log is taken. These values are then treated as input currents to a resistive grid; Kirchhoff's current law is used in local calculations to simulate the spread of the input across the grid. Several iterations are usually required for the voltages to settle down. (We stop the simulation when the difference in the node voltages across the grid between two iterations is less than one tenth of 1% of the maximum pixel value. About one hundred iterations are usually sufficient to meet this criterion.) Next the settled net values are subtracted from the log of the input values. Finally, the minimum of the corrected values in the three planes is found and subtracted from all values in the three planes, and all values are scaled up so that the maximum of all values in the three planes is set to the maximum value of our frame buffer, 255:

$$R_{\text{out}}(x, y) = 255 \cdot \frac{R_0(x, y) - \min\big(\min(R), \min(G), \min(B)\big)}{\max\big(\max(R), \max(G), \max(B)\big) - \min\big(\min(R), \min(G), \min(B)\big)}. \qquad (6)$$
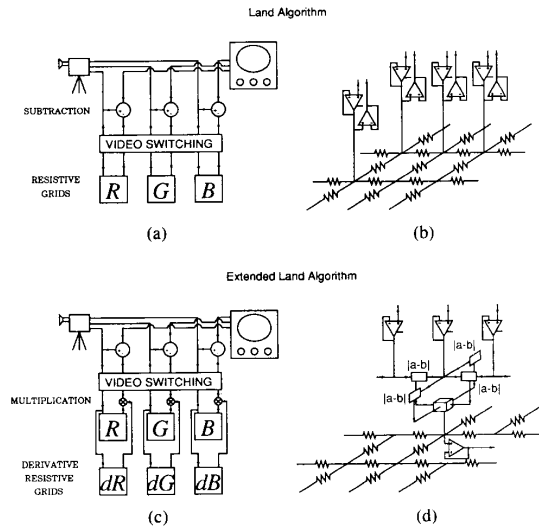
Fig. 1. (a) The Land algorithm. The three color camera outputs are smoothed on three separate resistive grids, labeled R, G, and B. The smoothed signal is subtracted from the camera output. (b) Resistive grid for smoothing images. (c) Extended Land algorithm. The magnitude of the local spatial derivative is smoothed for each color channel on the resistive grids labeled dR, dG, and dB and used to modulate the strength of the smoothed image before subtraction from the original. (d) The scheme for computing edginess. The average of the magnitudes of the local derivatives serves as the input to a resistive grid.

This last step is a form of gain control and is crucial; without it, all colors would tend to gray since the subtraction of the blurred image is a compressive operation. Note that it is the only step that requires operations across the three color planes. All prior steps proceed independently within each color plane. We go to this trouble since there are some images in which there is little or no information in a given channel. For example, in forest scenes there is not much signal in the blue channel. Normalizing independently in each channel for such a scene would artificially expand the pixel values in the blue channel, causing noise to be accentuated and generally changing the image color globally in the wrong way.

Fig. 2 shows the results of simulations of the Land algorithm. At top are three images obtained directly from the video camera. For images (a) and (b) the color output controls of the video camera were adjusted to match skin color and a color card fairly well under ordinary fluorescent illumination. One image (top left, (a)) was captured under this illuminant; it will be called the (camera) corrected or fluorescent image. The fluorescent lights were then turned off, and the same subject was illuminated with incandescent light. A second image (middle top, (b)) was captured without correcting the camera color settings under this new illuminant; it will be called the uncorrected or incandescent image. While colors in the scene looked a bit shifted to the red to us in the room when the second image was taken, they were not as bad as those captured by the camera— the second image is unacceptable. The skin color is too red, the background is lost in darkness, and the shadows are very deep. Unfortunately, in these respects it resembles many amateur video images taken indoors.

The middle row of images in Fig. 2 show the result of applying the retinex operation to the original images. The corrected image (middle left, (d), corresponding to (a) above it) is somewhat improved in terms of contrast enhancement. Note, for example, the highlights in the hair that are not visible in the original. The color is less saturated (i.e. less pure, more washed out, more gray) but hue is well preserved. The uncorrected image is strikingly improved (center image, (e), corresponding to (b)). Skin color is more muted, the shadows across the face are softened, and detail is visible in the background. While color correction is not perfect, it is significant. The contrast enhancement inherent in this algorithm is at least as significant.

One drawback of this algorithm, however, is apparent in these images, namely, color induction across edges. Close examination of image (e) of Fig. 2 reveals that the (black) border of the CIE diagram poster has been tinged with red above and to the right of the horseshoe-shaped diagram and tinged green below the diagram. The discoloration decreases with distance from the edge of the diagram. The unwanted color, overlaid on the black border, is the complementary color of the area on the other side of the edge: induced red abuts the green region on the right and induced green abuts the red region on the bottom. Red and green are complementary colors. From these facts one may conclude that color induction across abrupt edges is inherent in the algorithm. Consider, for example, a point in the black border area just adjacent to the CIE diagram on the right side of the poster. Its surround is strongly weighted green by the nearby region of the color diagram. This (mainly green) surround is subtracted from the black center to yield black plus green's complement, red. (Along with the red value, the blue value is raised over the green channel in this region. So in this sense, it could be said that green's complement is red plus blue. What we perceive, however, is mostly the complement to green, which is more red than blue.) A black border point further from the color edge is less induced to red since the green area is further away, and thus weighted less in forming the surround. Image (d) is similarly distorted but the distortion is less noticeable by inspection. This effect is quantified below.

Color induction is not mentioned in any of the studies of retinex theory except the most recent one by Land [21]. In this paper, he notes induction in terms of *Mach bands*, a well-known phenomenon in psychophysics. Fig. 5 of that paper shows how "spill-over" of the surround is responsible for a relatively dark region in the light region adjacent to a dark–light edge and a complementary, relatively light region in the dark area near the edge. Normally one hears only of Mach bands along the achromatic (black–white) lightness axis. Whether color Mach bands are visible is controversial. However, it suffices to say that we do not perceive effects as strong as the effects produced by the Land algorithm with video camera inputs; we do not see, for example, a green halo surrounding a red ball placed against a gray background.

Another limitation of the Land algorithm is revealed by the images in the right column of Fig. 2. At top (image (c)) is the output of a video camera shot of a still life in which a large portion of the scene is composed of just one color. This is a common situation; often half of an image is filled with sky or foliage. The scene was deliberately captured under dim illumination, to study the contrast enhancement capabilities of this algorithm. At middle right (image (f)) is the result of retinex processing. Although the shadows were softened considerably, much of the image is gray, not green. This illustrates how the gray world assumption can go wrong. Since we are subtracting a blurred version of the image from the original image, in this case we are subtracting green from green, leaving gray.
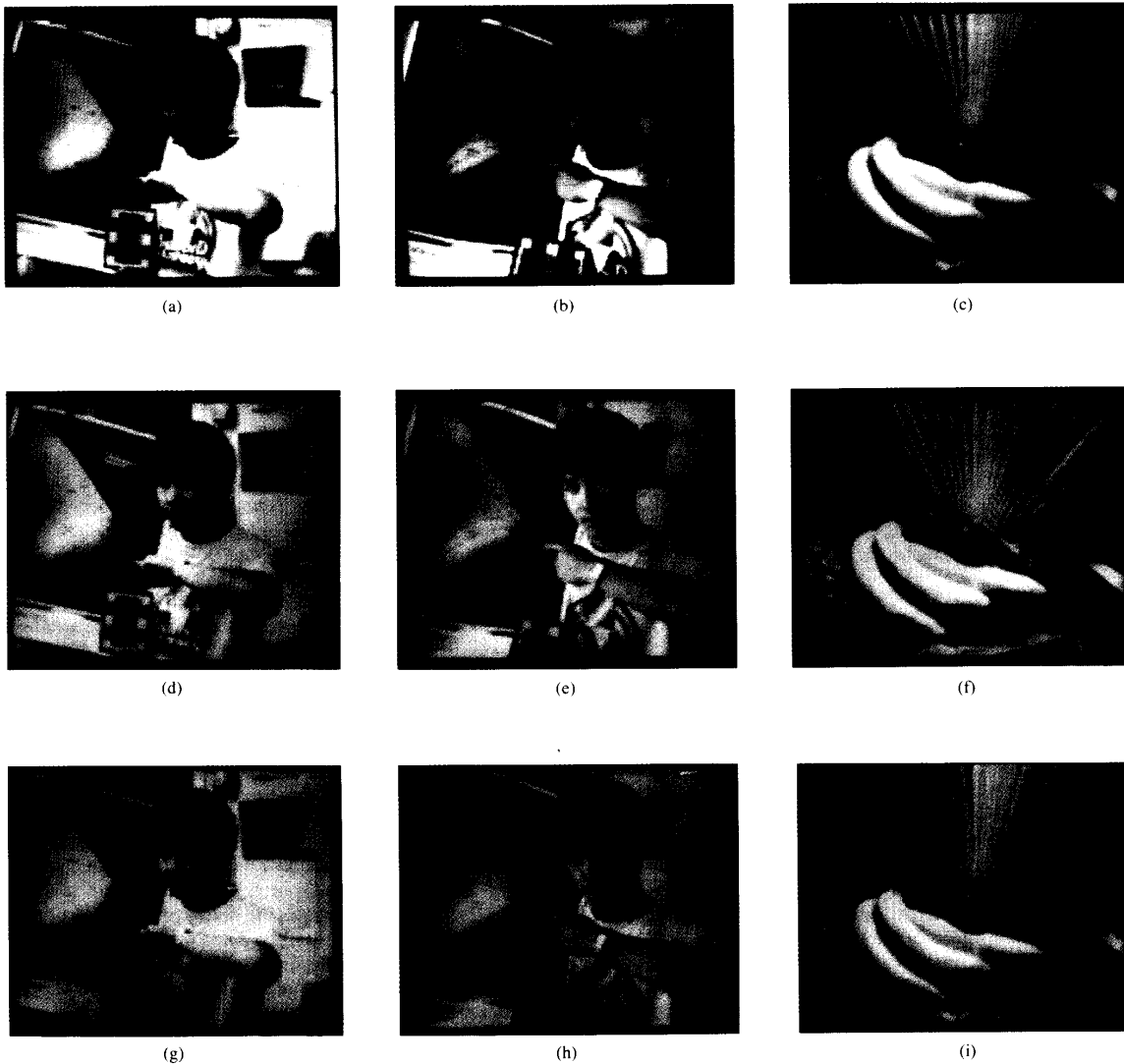
Fig. 2. Results of simulations. At top are the original images. In the middle row are the results of retinex processing. At bottom are the results after applying the extension to the retinex algorithm.

The upshot of all of this is that the Land model is too simple in at least two ways. First, it embodies but a simplification of a static aspect of visual processing that psychophysicists call *simultaneous contrast* [14], [35]-[37]. (In this sense it could also be called the Cornsweet model, the Jameson and Hurvich model, or even the Mach model as all of these researchers have pointed to simultaneous contrast as a mechanism for color constancy.) Land's model of simultaneous contrast is insufficient in that it ignores edge information and thus suffers from induction across borders. While retinex proponents point to cortical visual area V4 as being a site of surround suppression in color processing, they do not cope with the fact that V4 cells respond well to edges [3], [6]. Second, the model suffers from overreliance on the gray world assumption. As we shall see below, edge information can also help with this problem.

B. An Extension to the Retinex Algorithm

A modification of the retinex algorithm was applied next to the same color images (bottom row of Fig. 2). The magnitude of the spatial derivative is smoothed on a second resistive grid, to yield a measure of "edginess"; this measure is used to weight the surround before subtraction from the center (parts (c) and (d) of Fig. 1). In other words, while for a retinex simulation we have

$$\text{output} = \text{center} - \text{surround} \tag{7}$$

$$I'_{\text{out},i}(x, y) = I'_i(x, y) - I'_i(x, y) \otimes e^{-r/\lambda} \tag{8}$$

to ameliorate induction effects and lessen reliance on the gray world assumption, we need to modify the surround weight from point to point. In particular, if edginess is given a value close

to 0 in homogeneous regions such as the black border of the poster in the left images, and is given a value close to 1 in more detailed regions such as the colored shirt, we have a better formulation as follows:

$$\text{output} = \text{center} - \text{surround} \cdot \text{edginess}. \qquad (9)$$

In this relation, the surround is effectively zeroed in smooth areas before it is subtracted, so that induction is diminished—more of the original color is retained.

Parts (c) and (d) of Fig. 1 show how edginess is computed and used. The 512 × 512 image is again sampled at a low resolution. The magnitude of the first spatial derivative, labeled $|a - b|$, is computed between points; the average of the absolute value of the four local spatial derivatives are fed as a current into each node of the grid. The output voltage of this resistive grid is multiplied with the surround value read out from the first resistive grid. This modified surround is then subtracted from the camera output, to yield a color-corrected signal. Signifying the averaged magnitude of local spatial derivatives as $|\partial l_i'(x, y)|$, the mathematical expression for the resistive grid smoothing of that quantity is the convolution of it with an exponential distance weighting function, so the complete expression for the extended algorithm is

$$l_{\text{out},i}'(x, y) = l_i'(x, y) - l_i'(x, y) \otimes e^{-r/\lambda}$$

$$\cdot |\partial l_i'(x, y)| \otimes e^{-r/\lambda}. \qquad (10)$$

The bottom figures of Fig. 2 show images processed with this extended retinex algorithm. The color induction is much less noticeable upon inspection in the middle and left images, and color is returned to the palm frond at bottom right. The extended algorithm effectively varies, point by point, the degree of subtraction of the blurred version of the image from the original. In detailed areas, edginess is high, so the subtraction is carried out as for the original algorithm. In smooth areas, however, the degree of correction (weight of surround subtracted) varies as the distance from the nearest edgy area. In smooth areas, more of the original image "passes through," and so there is less color correction. Color constancy will be worse for such areas. For example, in Fig. 2(h) the skin tone is redder than in Fig. 2(e). The extended algorithm, then, is a working compromise between color constancy via strict application of the gray world assumption and no color constancy at all.

Some of these results are quantified in Fig. 3. A horizontal and a vertical line through the images in places that show induction artifacts were selected (Fig. 3(a) and (b)). The green intensity at each pixel in each line was subtracted from the red intensity at the pixel to show the value of the red–green axis of color at the pixel in the original images, in the images processed with the retinex algorithm, and in the images processed with the extended retinex algorithm. Concentrating on the black border area of the poster, note that for the original images (thick lines) the pixel value is zero in these regions—red and green are balanced in the achromatic, black region. A shift from zero here results from induction. At the top of the poster border, red is strongly induced in the retinex-processed image (dashed line at pixels 10–30 of plots (e) and (f) of Fig. 3). It is induced by the neighboring green area (pixels 30–50). The extended retinex algorithm produces less induction (thin line). Similarly, green is induced in a black region next to the reddish face area after retinex processing in a region crossed by the horizontal line (pixels 50–70 of plot (d)). The extended algorithm (thin line)

is not much better than the original retinex algorithm (dashed line) in this instance.

Other resistive grid methods for color correction have been explored in simulation. If at each point of input to a grid computing the surround for subtraction, the input resistance is modulated by the local spatial derivative, a surround is formed that consists of areas "filled in" or interpolated between edgy regions. Here the local spatial derivatives form an input confidence [38], [39]. Mach bands are lessened in this algorithm in comparison with the Land algorithm, but the degree of smoothing required to form good surrounds varies from image to image, so the algorithm is not as robust as the extension detailed above. We have also tried varying the lateral resistances according to local spatial derivatives, with disastrous results; variation of the lateral resistances strongly disturbs current flow in the grid, segmenting the image into discrete areas [40], [41]. As a result, subtraction of the grid outputs leads to patches of gray in most smooth areas of the input image. In other words, variation of the lateral resistance by local values is more appropriate for segmentation than for normalization. We have not tried to vary the input or lateral resistances according to smoothed edginess, though it may be more comparable to the extended method discussed above.

These results are anecdotal and limited in nature, but they show the strengths and weaknesses of Land's algorithm and allow us to see ways to improve the algorithm. The extension explored, modulation of the surround by a measure of edginess calculated by smoothing the magnitude of the spatial derivative on a second resistive grid, is easy to implement in VLSI.

V. VLSI IMPLEMENTATION OF THE RETINEX ALGORITHM

From the simulation results, it appears that the Land algorithm and simple extensions to it may be effective in color correction. We have implemented the Land algorithm in analog CMOS VLSI. Fig. 1(a) shows the outline of a system of video camera color correction based on Land's algorithm. The three color outputs of a video camera (labeled red, green, and blue here) are fed onto three separate resistive grids built from subthreshold analog CMOS VLSI. Each 48 by 47 node resistive grid was built using 2 μm design rules and contains about 60 000 transistors.

Since a single chip can contain only a small grid (roughly 50 by 50), the 525 × 525 video image must be sampled at a low resolution with appropriate video switching and sample-and-hold circuitry. Perhaps the most novel aspect of this design is in its sample-and-hold architecture. A horizontal line of NTSC video is about 50 μs in duration; 48 horizontal pixels must be fed with the video input averaged onto a capacitor over 1 μs. However, the data must be held for input to the resistive grid for the field duration, which is about 16 ms for NTSC video. Thus the sample time and the hold time differ by over four orders of magnitude. The crucial design feature of these chips is that a two-stage sample-and-hold scheme is used. At the bottom of the chip, 48 capacitors are charged up at the line rate. Followers broadcast these voltages into the array, where the currently selected row of nodes reads the 48 values and integrates them into a second sample-and-hold circuit. This second circuit is a follower-connected transconductance amplifier, set to run in the subthreshold range, feeding a capacitor. Five video lines
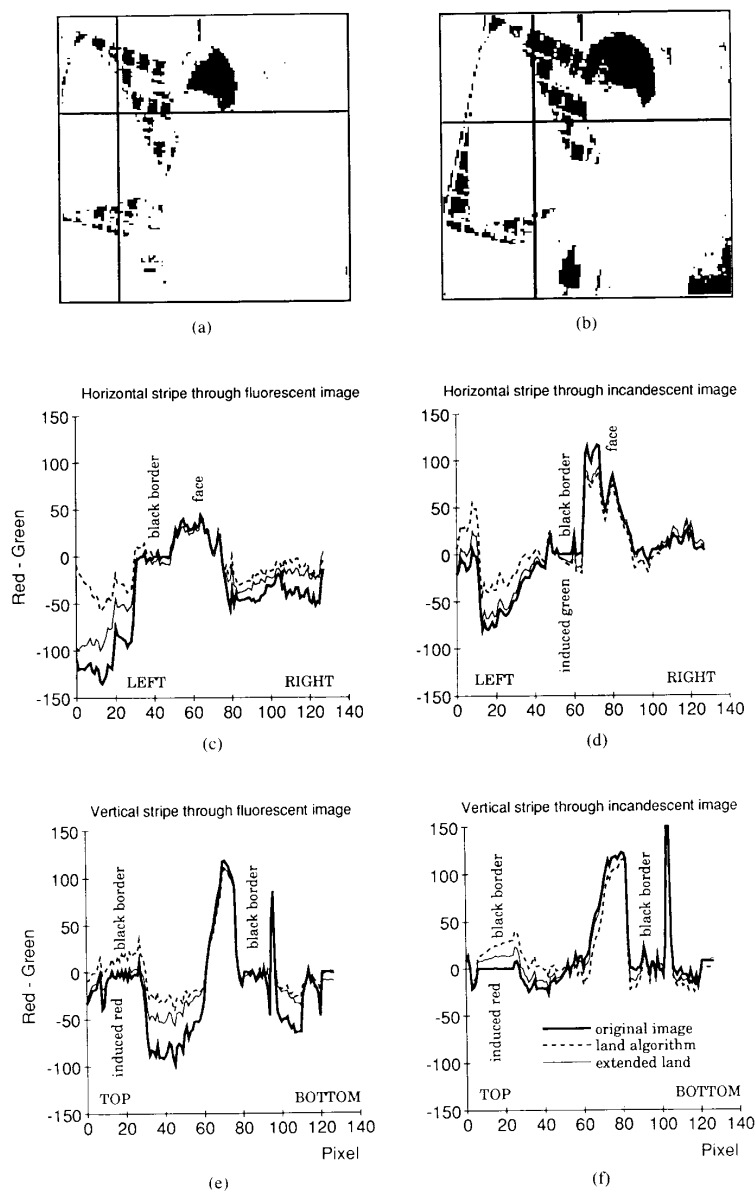
Fig. 3. Quantification of simulation results. At top ((a) and (b)) are binary representations of the images in the left two columns of Fig. 2. As in Fig. 2, the fluorescent-illuminated image is at left (a) and the incandescent-illuminated image is at right (b). A horizontal and vertical line of pixels was selected through each image, in areas where the color induction effects of the color correction algorithms are notable. At bottom ((c)-(f)) are plots of the red pixel value minus the green pixel value along the selected lines for the original image (Fig. 2(a) and (b), thick line), the retinex-processed images (Fig. 2 (d) and (e), dashed line), and the images processed with the extended Land algorithm (Fig. 2(g) and (h), thin line). See text for details.

are integrated by each of the 47 rows of the resistive grid in each field of the video frame [42].

The circuit details within each pixel are similar to those of the analog retina [34]. A current proportional to the node sample-and-hold capacitor voltage is injected into the grid with a follower-connected transconductance amplifier run in the subthreshold range. The grid consists of n-type transistors interconnecting the input nodes; a "horizontal resistor" bias circuit at each node sets the gate bias of the interconnect transistors so that the resistance is linear regardless of the transistor source voltage. These bias circuits are also set to run in the subthreshold range. The output of each node is a follower-con-

nected transconductance amplifier, run above threshold, which produces a current proportional to the node voltage. This current is sensed and transformed to a voltage by an off-chip current sensing amplifier. Current steering identical to that used in the analog retina directs one node output at a time to the sense amp.

Fig. 4 shows the ability of the system to correct skin color under a common variation in lighting. At top are the two original images. At left is an image under fluorescent lights, with the camera corrected for this illuminant; skin color looks normal. In the right image, the illuminant is incandescent light, but the camera is still set up for fluorescent light; the skin color is too red. At bottom are the outputs of the system after a smoothed version of the image is subtracted. Though the image at bottom right is more red than the one at bottom left, the color difference is less between the bottom images than between the top images. The system-corrected images are of a poorer quality than the camera images, because of switching noise, cross-talk, etc. The point is that the color is more constant for the processed images than for the camera.[1]

Conventional methods are capable of this level of correction. Many video cameras have an ambient light sensor attachment, which is used to sense the illuminant; a global subtraction of the global value corrects skin color as well as our system. In fact, simply averaging the red, green, and blue signal over a video frame and subtracting this average will work with richly colored scenes such as these (i.e., with scenes for which the gray world assumption is valid). The strength of this algorithm and its value as a model of the biology lie in its use of a spatially varying average for subtraction. This feature enables it to enhance contrast, soften shadows, and reproduce color shifts that are observed by humans. Our electronic system is not very good at shadow softening and contrast enhancement, for three reasons. First, we are not taking the log of the video signal before processing, so we are not taking advantage of as much of the signal as we are in the simulations. Second, the noise in the surrounds produced by nonidealities in the analog CMOS fabrication technology distorts the image in dark areas. Third, the resolution of the surround is much lower than the resolution of the original image (50 × 50 versus 512 × 512). However, we are able to reproduce one aspect of human color perception with this electronic system, an aspect that illustrates the spatially varying nature of the color normalization: color darkening in light regions of a scene and color lightening in dark regions, as shown in Fig. 5.

Fig. 5 shows three set of images of color bars. At top are the video camera outputs under fluorescent light, fluorescent and blue light, and fluorescent and green light. The middle row of images shows the corrected system output for identical lighting conditions as the top three images. The bottom row of images are the direct output of the resistive grids, with no smoothing. The color constancy among images in the middle row is impressive compared with the top and bottom rows. These images also show the spatial aspect of the color correction. The red bars in these images are cut from the same piece of paper. Note that in the top row the red bar next to the white bar looks darker

than the red bar next to the black bar. When the red bars are examined in isolation the color is identical. This is easiest to see by cutting two holes in a piece of paper so that one hole lies over the top red bar, and the other hole lies over the bottom red bar. In isolation the colors are identical, but the perception is influenced by nearby colors; this is simultaneous contrast [14], [35]-[37]. In the color-corrected images of the second row, the red bar in the light region appears darker than the red bar in the dark region, even when viewed in isolation. This demonstrates that the system is using local information to perform the correction; this is the first system to show these color effects with video images.

The bottom row of Fig. 5 shows images of the same color bars taken from the resistive grid outputs, under the same lighting conditions as the top two rows. (The smoothing is set to zero here, to show the bars clearly; for correction, the image is smoothed greatly, so that the resistive grid outputs are an uninteresting blur.) As expected, the color varies as the lighting is changed just as in the video camera images (top row). The two-hole test described above reveals that the top and bottom red bars in this row of images are identical in color.

In summary, our real-time system, which forms a blurred version of the image on resistive grids for subtraction from the original, demonstrates color constancy and simultaneous contrast effects. Other effects produced by Land's retinex algorithm, such as color Mach bands, have been observed with the electronic system, but are not shown here.

VI. CONCLUSION

Land's retinex theory is a model for our natural ability to see color as roughly constant as the lighting varies widely. The neurobiology and psychophysics of color constancy support the plausibility of his model; computational analysis of the problem shows that his is an elegant solution. We have applied resistive grid processing to his model, greatly reducing its complexity. Through computer simulations we have explored the strengths and weaknesses of the retinex theory; we have developed an extension of it that lessens its weakness. Impressed with its strengths, we have implemented the retinex algorithm using analog VLSI. The system, based on three resistive grids, is capable of color correction and displays color shifts that qualitatively mimic those of human perception. The system operates at video rates, and as such is the first of its kind. With further development, systems such as this, designed to implement the retinex algorithm and simple extensions to it, would be useful in a variety of video applications.

Is this system a neural network? Even though there are no weights, thresholds, energy surfaces, or the like in its architecture and operation, we feel that it is. We were led to this problem after exploring the nonclassical receptive fields of cortical cells that process visual motion [32]. We turned to psychophysics to understand the problem and to computational theory to understand the models proposed to solve it. Finally, after computer simulation, we had the confidence to build the system. It is neural in the sense that it is a realization of the premier model of how the brain accomplishes color constancy. In the introduction, we pointed out that our present technology for capturing images is flawed—it is too simple. By studying the brain we have been able to build a system that does it better.

---

[1]In preparing the final images for Fig. 4, the *scene dependence* of color constancy was discovered. In order to obtain the level of constancy of Fig. 4(d), the subject had to be placed between a dark region and a bright region. By accident rather than design, this is how the subject was arranged in parts (b), (e), and (h) of Fig. 2. This further weakness of the Land algorithm may not have been discovered without a real-time system. For further details, see [43].
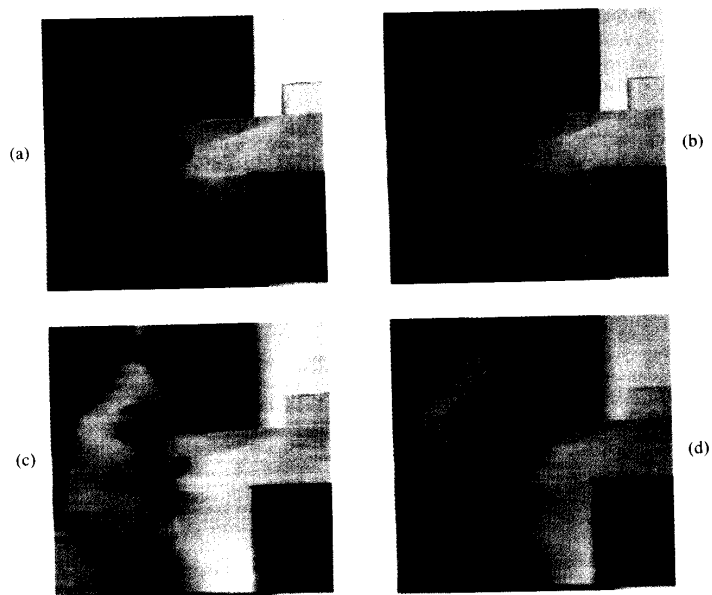
Fig. 4. Skin color correction with an electronic implementation of the Land algorithm. At top are the camera outputs under (a) fluorescent and (b) incandescent light. The camera was adjusted to report colors well under fluorescent light. The bottom images show the output of the color correction system for (c) fluorescent and (d) incandescent illuminants. The skin tone in the bottom images changes less for the two conditions than the camera images.
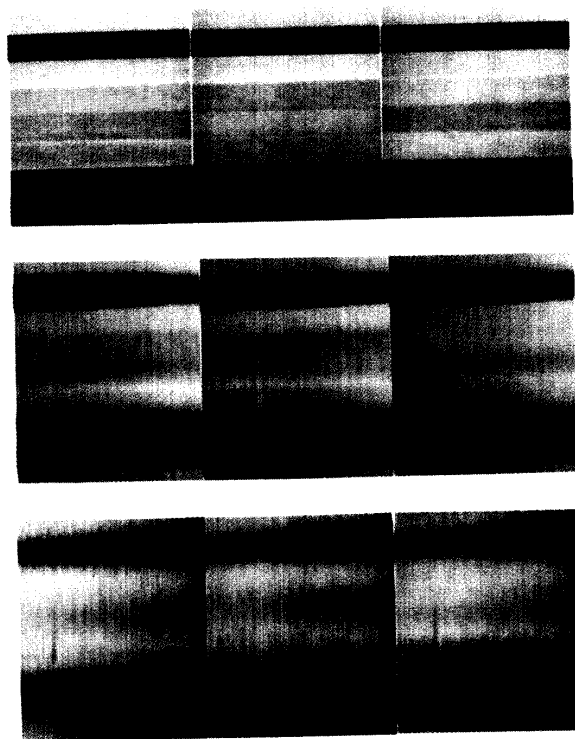

Fig. 5. Color constancy results from the electronic system. The top images are the video camera outputs, the middle images are the color-corrected outputs, and the bottom images are the outputs from the three resistive grids with no smoothing. The color bars are lit with fluorescent light in the left column. Narrow-band blue light is added in the middle column, and narrow-band green light is added in the right column.

REFERENCES

[1] H. von Helmholtz, *Treatise on Physiological Optics*, vol. 2, 3rd ed., J.P.C. Southall, Ed. New York: Dover, 1962.

[2] S. M. Zeki, "The representation of colours in the cerebral cortex," *Nature*, vol. 284, p. 412, 1980.

[3] E. A. DeYoe and D. C. Van Essen, "Concurrent processing streams in monkey visual cortex," *Trends Neurosci.*, vol. 11, p. 219, 1988.

[4] S. M. Zeki, "Colour coding in the cerebral cortex: The reaction of cells in the monkey visual cortex to wavelengths and colours," *Neurosci.* vol. 9, 1983.

[5] S. M. Zeki, "Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition," *Neurosci.*, vol. 9, 1983.

[6] R. Desimone, S. J. Schein, J. Moran, and L. G. Ungerleider, "Contour, color, and shape analysis beyond the striate cortex," *Vision Res.*, vol. 25, p. 441, 1985.

[7] D. B. Judd, D. L. MacAdam, and G. Wysecki, "Spectral distribution of typical daylight as a function of color temperature," *J. Opt. Soc. Amer.*, 1031, 1964.

[8] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Amer.*, A3, p. 1673, 1986.

[9] M. D'Zmura and P. Lennie, "Mechanisms of color constancy," *J. Opt. Soc. Amer.*, vol. A3, p. 1662, 1986.

[10] K. T. Blackwell and G. Buchsbaum, "Quantitative studies of color constancy," *J. Opt. Soc. Amer.* vol. A5, p. 1772, 1988.

[11] L. Arend and A. Reeves, "Simultaneous color constancy," *J. Opt. Soc. Amer.*, vol. A3, p. 1743, 1986.

[12] J. A. Worthey, "Limitations of color constancy," *J. Opt. Soc. Amer.*, vol. A2, p. 1014, 1985.

[13] D. Ingle, "The goldfish as a retinex animal," *Science*, vol. 227, p. 651, 1985.

[14] T. N. Cornsweet, *Visual Perception*. New York: Academic Press, 1970.

[15] B. K. P. Horn, *Robot Vision*. New York: McGraw-Hill, 1985.

[16] G. Buchsbaum, "A spatial processor model for colour perception," *J. Franklin Inst.*, vol. 310, no. 1, 1980.

[17] L. T. Maloney, "Computational approaches to color constancy," Ph.D. thesis, Stanford University, 1984.

[18] L. T. Maloney and B. A. Wandell, "Color constancy: A method for recovering surface spectral reflectance," *J. Opt. Soc. Amer.* vol. A3, p. 29, 1986.

[19] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, p. 1, 1971.

[20] E. H. Land, "Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image," *Proc. Nat. Acad. Sci. U.S.*, vol. 80, p. 5163, 1983.

[21] E. H. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci. U.S.* vol. 83, p. 3078, 1986.

[22] B. K. P. Horn, "Determining lightness from an image," *Comput. Graph. Image Proc.*, vol. 3, p. 277, 1974.

[23] B. Funt and M. Drew, "Color constancy computations in near-Mondrian scenes using a finite-dimensional linear model," *Proc. IEEE Conf. Comp. Vis. Patt. Rec.*, June 1988.

[24] A. Hurlbert, "Formal connections between lightness algorithms," *J. Opt. Soc. Amer.*, vol. A3, p. 1684, 1986.

[25] A. V. Oppenheim, R. W. Shafer, and T. G. Stockham, Jr., "Non-

[26] linear filtering of multiplied and convolved signals," *Proc. IEEE*, p. 1264, 1968.

[26] O. D. Faugeras, "Digital color image processing within the framework of a human visual model," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, p. 380, 1979.

[27] D. H. Brainard and B. A. Wandell, "Analysis of the retinex theory of color vision," *J. Opt. Soc. Amer.*, vol. 3A, p. 1611, 1986.

[28] M. La Brecque, "Retinex: Physics and the theory of color vision," *Computers in Physics*, Nov/Dec 1988.

[29] A. Blake, "Boundary conditions for lightness computation in mondrian world," *Comp. Vis. Graph. Image Proc.*, vol. 32, 1985.

[30] D. Marr, "The computation of lightness by the primate retina," *Vision Res.*, vol. 14, p. 1377, 1974.

[31] M. S. Livingstone and D. H. Hubel, "Anatomy and physiology of a color system in primate primary visual cortex," *J. Neurosci.*, vol. 4, p. 309, 1984.

[32] J. Allman, F. Miezin, and E. McGuinness, "Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT)," *Perception*, vol. 14, p. 105, 1985.

[33] A. Hurlbert and T. Poggio, "Learning a color algorithm from examples," MIT AI Memo 909, 1987.

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

[35] L. M. Hurvich, *Color Vision*. Sunderland, MA: Sindauer Associates, 1981.

[36] O. Creutzfeldt, B. Lange-Malecki, and K. Wortmann, "Darkness induction, retinex, and cooperative mechanisms in vision," *Exp. Brain Res.*, vol. 67, p. 270, 1987.

[37] P. Lennie and M. D'Zmura. "Mechanisms of color vision." *CRC Crit. Rev. Neurobiol.*, vol. 3, p. 333, 1988.

[38] S. Grossberg and E. Mingolla, "Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading," *Psych. Rev.*, vol. 92, 1985.

[39] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *IEEE Computer*, vol. 21, 1988.

[40] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, 1990.

[41] J. Harris, C. Koch, and J. Luo, "A two-dimensional analog VLSI circuit for detecting discontinuities in early vision," *Science*, vol. 248, 8 June 1990.

[42] A. Moore and R. Goodman. "Image smoothing at video rates with analog VLSI," in *Proc. IEEE Conf. Syst., Man, Cybern.*, Nov. 1990.

[43] A. Moore, G. Fox, J. Allman, and R. Goodman, "A VLSI neural network for color constancy," in *Advances in Neural Information Processing 3*, D. S. Touretzky and R. Lippman, Eds. San Mateo, CA: Morgan Kauffmann, 1991 (in press).

Andrew Moore (S'90) was born in Springfield, IL, and attended the University of Illinois at Urbana. He received his B.S. in electrical engineering, with honors, in 1983. He worked in computer graphics, bioelectromagnetics, and medical imaging before entering the doctoral program in Computation and Neural Systems at the California Institute of Technology, in 1987. At Caltech, he has realized a long-held goal of combining engineering with neuroscience.

John Allman received the Ph.D. degree from the University of Chicago in 1970 and then did postdoctoral research in the Laboratory of Neurophysiology at the University of Wisconsin. He joined the Division of Biology at the California Institute of Technology in 1974, where he is currently the Hixon Professor of Psychobiology. Together with his colleagues, he has discovered, through microelectrode recordings, a series of topographically organized maps of the visual field in the cerebral cortex of primates that collectively comprise about half of the cerebral cortex. In these maps, information from small regions of the visual field is constantly compared with the entire visual field—color constancy mechanisms are a special example of this phenomenon.

Rodney M. Goodman (M'85) was born in London, England, on February 22, 1947. He received the B.Sc. degree in electrical engineering from Leeds University, Yorkshire, U.K. in 1968 and the Ph.D. in electronics from the University of Kent at Canterbury, U.K., in 1975.

In 1985 he joined the faculty of the Department of Electrical Engineering at the California Institute of Technology as Associate Professor. His research interests are in error control coding, cryptography, neural networks, and expert systems. He is currently a consultant on these topics for the Jet Propulsion Laboratory and Pacific Bell.

Dr. Goodman is a Chartered Electrical Engineer of the I.E.E. in the U.K.