Spatial Filtering in Tone Reproduction and Vision

Thesis by

Andrew J. Moore

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1992

(Defended May 11, 1992

Acknowledgments

To:

- Ralph Evans, *scientist and author*,
- Deane Judd, *scientist and public servant*,
- Edwin Land, *scientist and entrepreneur*, and
- James and Rose Moore, *parents*,

for facing the unknown with courage and persistence.

Many thanks to my colleagues at Caltech and elsewhere for helpful discussions relating to this thesis: J. Allman, C. Andersen, M. Avalos, K. Boahen, P. DuBois, M. D'Zmura, K. Fleischer, G. Fox, J. Gallant, R. Goodman, B. Gupta, W. Hill, J. Harris, E. Heinemann, T. Horiuchi, E. Kaplan, C. Koch, J. Lazzaro, R. Lyon, M. Mahowald, L. Mathies, B. Mathur, J. McCann, C. Mead, M. O'Dell, F. Perez, J. Pine, T. Poggio, S. Ryckebush, M. Sereno, M. Sivlotti, S. Schein, D. van Essen, U. Wehemeier, I. Wong and S. Zucker.

Many thanks to all the folks involved with the Computation and Neural Systems Program at Caltech. The program has rigor, freshness, and integrity - it has been a great five years !

Thanks to the following organizations for support: the Parsons Foundation and the Pew Charitable Trust for fellowship support, the Office of Naval Research, the Joint Tactical Fusion Program, the Center for Research in Parallel Computation, and Pacific Bell for

assistantship support, DARPA for MOSIS fabrication services, and Hewlett Packard for computing support in the Mead lab. Much of the chip work was done in Carver Mead's lab - I am very grateful for his generous support.

Last but not least, thanks to Stacey Chascsa for being a patient model and a model of patience.

Abstract

This thesis is concerned with spatial filtering. What is its utility in tone reproduction ? Does it exist in vision, and if so, what constraints does it impose on the nervous system ?

Tone reproduction is just the art and science of taking a picture and then displaying it. The sensors available to capture an image have a greater dynamic range than the media that may be used to display it. Conventionally, spatial filtering is used to boost contrast; it ameliorates the loss of contrast that results when the sensor signal range is scaled down to fit the display range. In this thesis, a type of nonlinear spatial filtering is discussed that results in direct range reduction without range scaling. This filtering process is instantiated in a real-time image processor built using analog CMOS VLSI.

Spatial filtering must be applied with care in both artificial and natural vision systems. It is argued that the nervous system *does not* simply filter linearly across an image. Rather, the way that we see things implies that the nervous system filters nonlinearly. Further, many models for color vision include a high-pass filtering step in which the DC information is lost. A real-time study of filtering in color space leads to the conclusion that the nervous system is not that simple, and that it maintains DC information by referencing to white.

Contents

Chapter 1

The Range of Scenes, Sensors, and Displays

In this chapter, the central problem of tone reproduction is described: the problem of range mismatch. Studies of the range of luminances in a variety of scenes are reviewed, the challenge that this natural range presents to photographic, electronic, and biological vision systems is analyzed.

1.1 The Photometric Range of Typical Scenes

The goal of any tone reproduction system is the perfect copy: the precise light levels of any scene should be preserved for presentation. The light S reaching a sensor from an object in a scene is the product of the reflectance R of the object, the intensity I of the object illumination:

$$S = I * R. \tag{1.1}$$

The range of reflectances is limited, in practical terms, but illumination varies over orders of magnitude.

1.1.1 The range of reflectances

The reflectance or albedo of a surface is the degree to which it reflects light impinging upon it. For a perfectly black object the reflectance is zero; in practice certain fabrics such as felt are available with reflectances on the order of 1-2 %. A perfect mirror that reflects all light has a reflectance of one; bright, naturally occurring reflectances such as paper and fresh snow have a reflectance around 80 % [Hurvich, 1981]. Typical reflectances in a scene span about a factor of twenty [Walraven et al, 1990; Fink, 1955], with an average reflectance of 15 % [Stimson and Fee, 1953]. This latter study also found that the reflectance of (presumably caucasian) flesh is 38 %. A medium brown suit has a reflectance of 4 % [Moon and Spencer, 1954], and dark carpets may have a reflectance as low as 1.5 % [Moon and Cettei, 1938].

1.1.2 The range of scenes

Typical outdoor scenes exceed two orders of magnitude in range. The most widely accepted study of outdoor lighting levels is that of Jones and Condit [Jones and Condit, 1948]. These authors measured the light in several areas from 126 outdoor scenes. The maximum and minimum brightness in a given scene corresponded to the brightest and darkest portions of the image for which the viewer could see detail. Figure 1.1, taken from [Jones and Condit, 1948], shows a histogram of the photometric ranges of 126 scenes. The smallest range is 28, the largest is 760, and the average is 160.

Since the range of reflectances is small, these data suggest that the range of a scene is determined largely by lighting variations. Taking the reflectance range as a factor of 20, the lighting conditions in outdoor scene range from flat (overcast day, with diffuse lighting) to stark (clear, sunlit day, with direct light 38 times greater than reflected light in the shadows).

Moon and Spencer [1954] point out that the luminance extremes used by Jones and Condit do not correspond to the physical extremes of luminance in a typical scene. Should one use a specular highlight - for example, the sun reflecting from the polished chrome of

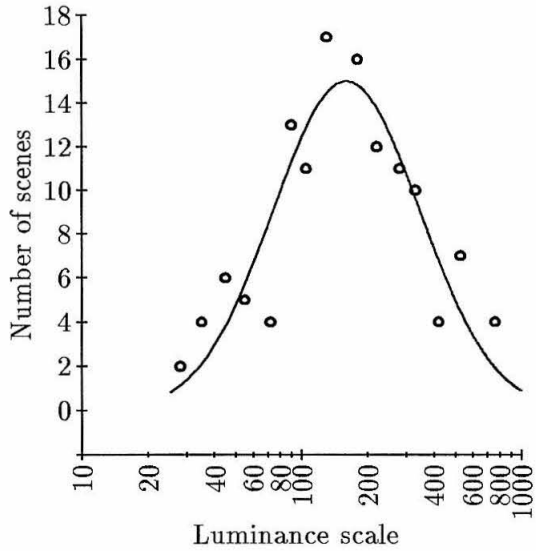


Figure 1.1: The photometric range of outdoor scenes, from Jones and Condit [1948].

a car - as the maximum, and the light beneath bushes, for example, as the minimum ? In that case, the ratio of the extremes may approach 10^9. Moon and Spencer are pressing the point that the significant range of a scene depends on the goal of the photographer. In this thesis, the Jones and Condit definition of photometric range is adhered to, since the goal is tone reproduction. That is, the reproduced image should resemble the original scene as closely as possible for a human observer.

1.2 Range Mismatch in Photography

In the early days of photography, nearly a century ago, pioneers of the technology faced a very difficult problem. While the photographic negative is sensitive to exposures ranging over three orders of magnitude (sufficient to capture the full range of most natural images), photographic papers can reproduce only a fraction of that range. Typically, photographic papers have a density range of a factor of thirty [Hunt, 1967]. Figure 1.2 shows the response curves of commonly available photographic papers [Jones, 1944]; the range or latitude of the papers varies from 0.7 log units to 1.7 log units. The response curve of a representative

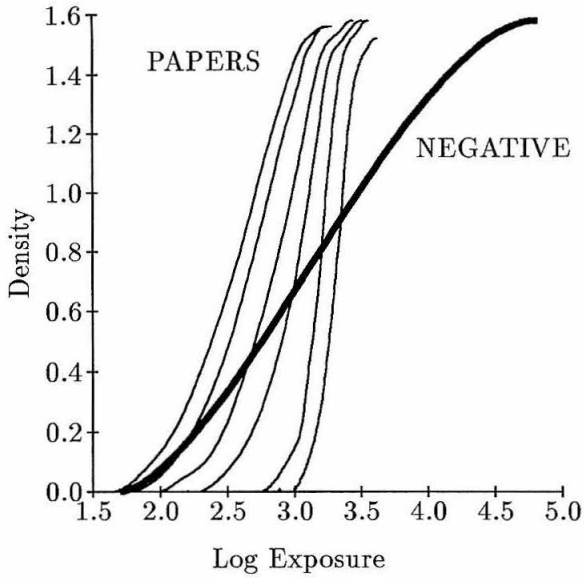


Figure 1.2: The response range of photographic media. A typical negative material can capture a thousand-fold range of luminance, but print papers can reproduce only over a range of about 30.

photographic negative, with far greater latitude than the papers, is superimposed on the paper curves.

1.3 Range Mismatch in Electronic Imaging

Figure 1.3 shows the range of the most commonly used electronic display device, the cathode ray tube or CRT. The screen luminance is plotted from an input voltage of 10 mV (the noise level) to the maximum allowed input voltage, 1 V. At the lower end of the input scale, the screen luminance is 0.9 foot-Lamberts; this is equal to the average ambient light level measured in a set of home television viewing rooms [Lisk, 1970]. Modern CRTs have a maximum screen luminance of about 50 fL; this gives an output range of 55, or 1.7 log units. Like photographic paper, the reproduction range of a CRT is below two orders of magnitude [Fink, 1955; Kallmann, 1940; Maloff, 1939].

Modern image sensors have a range of sensitivity comparable to that of photographic negative materials. Figure 1.3 shows the transfer function of a typical CCD imager, with a

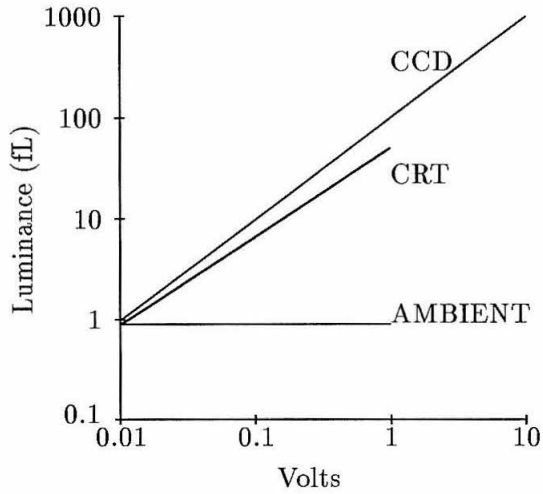


Figure 1.3: The response ranges of the sensor and the display commonly used in electronic tone reproduction systems. The CCD sensor can capture light over a range ten times larger than the range the CRT display can reproduce.

signal-to-noise ratio of 60 dB (range of three log units - see, for example, [Seige and Ress, 1987]). The noise level is set at 10 mV; this gives a maximum sensor signal of 10 V. Of course, most of this range must be thrown away since the maximum CRT input voltage is 1 V.

1.4 The Range of Sensitivity of the Visual System

1.4.1 Determinants of the visual response

The nervous system response to light on the retina is not as simple as the response of man-made media, photographic or electronic. Our perception of brightness depends on several factors: pupil diameter, mean light level, history of the mean light level, and spatial arrangement of the scene.

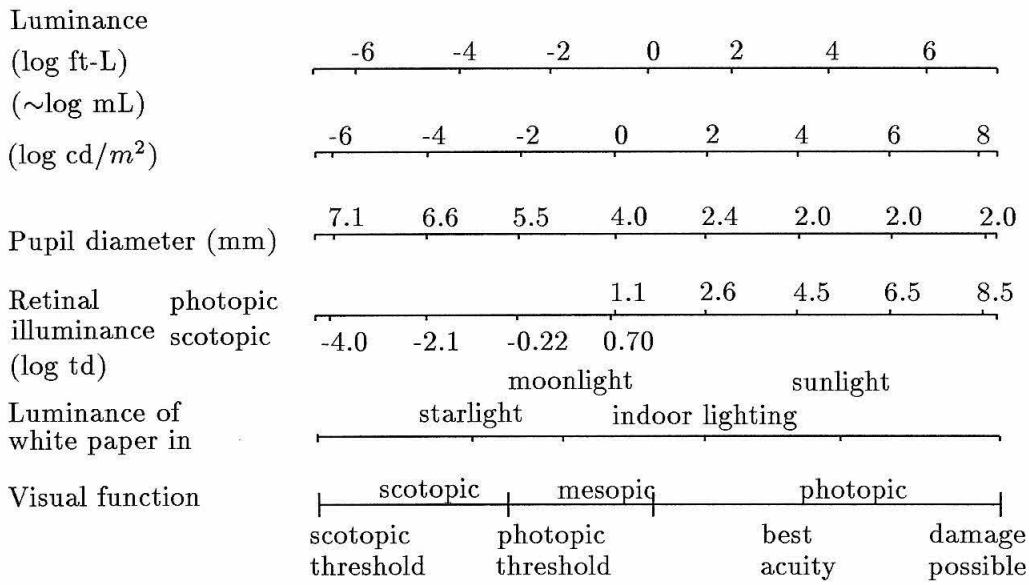


Figure 1.4: Human visual sensitivity and related physical quantities. Adapted from Walraven et al., 1990.

Pupil diameter

Pupil diameter corresponds to iris diameter in cameras. As the pupil is restricted, the light reaching the retina is scaled down. In cameras, practical iris diameters range from a millimeter to 5 centimeters, which represents a factor of 2,500 in area. In some species such as cats, the pupil ranges in diameter by a similar or larger factor [Barlow, 1972]. But in humans, pupil diameter range is rather more narrow - from 2 to 7 mm [Walraven et al., 1990], spanning a factor of ten in area [Barlow, 1972]. Therefore, by pupil constriction alone, the eye may adapt to an increase in light level over only one log unit.

Adaptation to the mean level

The visual system operates well over a range of light levels spanning more than 12 log units (Figure 1.4). Changes in pupil diameter, then, can extend the range of operation of the visual system over only a fraction of that tremendous range of sensitivity. Neural mechanisms are far more important in adapting the brightness response as the mean level of light reaching the eye varies.

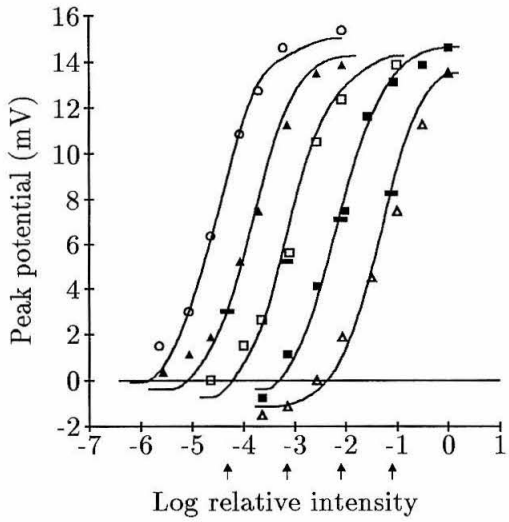


Figure 1.5: Adaptation in turtle cones. As the background intensity (arrows) is increased, the response shifts so that stimuli above and below the background intensity produce greater or lesser potentials than the resting potential (short horizontal bars). Since the shift is smaller than the change in background intensity, the resting potential increases with adaptation level.

Photoreceptor response

Figure 1.5 shows the response of turtle cones to spots of light flashed on backgrounds of various intensities [Normann and Pearlman, 1979]. The basic characteristic is an S-shaped curve; the curve shape changes little as it shifts left and right with the mean light level. While the adaptation levels are spaced by about one log unit, the shifts from level to level are less than one log unit. This means that as the adaptation level increases, the resting potential creeps upward. (The short horizontal bars indicate the resting potential of the turtle cone at each adaptation level.) Thus turtle cones show a compressive as well as an adaptive response to changes in light level. (In some species, photoreceptor responses are purely compressive, in others, they are purely adaptive, and in many, such as turtles, they are both compressive and adaptive [Shapely and Enroth-Cugell, 1984].)

Similar response curves have been measured from the photoreceptors of our close evolutionary relative, the macaque monkey. The measurement is quite difficult; the cells of higher mammals are small and fragile compared to the cells of simpler animals such as

turtles. To measure the cone response, Schnapf et al. [1990] isolated the cells in a dish, so that the normal neural connections were removed. In this preparation, they found that the monkey cone response is purely adaptive, with a shift that exactly corresponded with change in adaptation level - unlike turtles.

Due to the nature of the preparation, however, the conclusions of Schnapf et al. are more suggestive than definitive. Inputs from neural elements either directly or via diffusing neuromodulators are absent in this isolated condition, and so there may be compression *in situ* that could not be observed by Schnapf et al. Further, they measured the responses in an unnatural range of illumination, from daylight levels (4 log trolands) up near the level of pain and damage (9 log trolands). In tone reproduction, we are concerned with the neural response at far lower (viewing) levels, the mesopic range (1-3 log trolands). So, we still cannot say whether primate receptors are compressive as well as adaptive at normal light levels [Shapely and Enroth-Cugell, 1984].

Thalamic responses

The retinal output is routed to the lateral geniculate nucleus (LGN) of the thalamus. Two types of brightness response curves are found in the LGN: narrow-range responses and wide-range responses [Marrocco, 1975]. Virsu and Lee [1983] studied narrow-range cells in both cats and monkeys. The narrow-range response curve follows the photoreceptor response curve, i.e., is linear in the log of intensity over 1.5-2 log units, and slowly diminishes to zero response below this range. (It is difficult to observe the saturation of LGN cells *above* the log-linear response range because they adapt quickly. Stimuli more than two log units above threshold actually depress the response, and so the curve is not monotonic [Virsu and Lee, 1983].) Figure 1.6 shows data from a magnocellular narrow-range LGN cell measured in the mesopic and low photopic range [Virsu and Lee, 1983]. Both compression and adaptation are evident, as in turtle cones (Figure 1.5).

Cat responses showed no compression - the shift in the response curve corresponds to

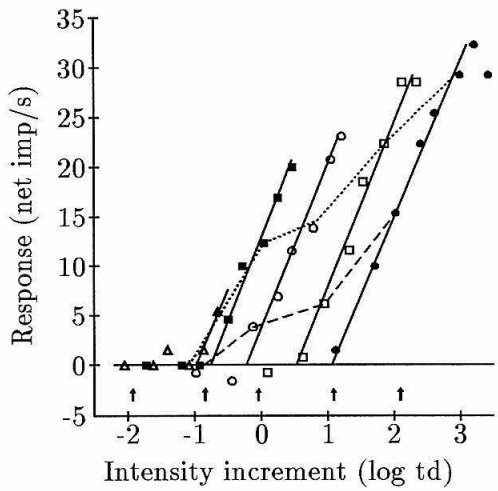


Figure 1.6: Adaptation in a macaque magnocellular LGN cell. The response is plotted as a function of the increment over the background intensity. Arrows indicate the absolute intensity of the background (adapting) stimulus. The dashed line connects responses to a contrast of 0.82 over the background; the dotted line connects responses to a contrast of 7.68 over the background. From Virsu and Lee, 1983.

the change in adaptation level [see also Sakmann and Creutzfeldt, 1969]. As a result, cat cells display contrast constancy. That is, stimuli with luminance that is a fixed multiple of the background (adapting) luminance produce the same cell response at all adaptation levels. However, the response to a fixed contrast (dashed and dotted curves in Figure 1.6) increases with adaptation level in monkeys. (Note that the *relative* increase in response as contrast is increased is fairly constant as the adaptation level changes.)

These investigators suggest that the reason for the difference between cats and monkeys is that cat visual responses originate mostly from rod inputs, while in monkeys the signals originate from both rods and cones. The dominant input is determined by the adaptation level: in the low mesopic range, cones are not active; in the photopic range, rods are saturated; in between these levels both rods and cones are active. If the gain of these two receptor signals is not equal, the transition from rods to cones as the adaptation level increases should be observable. The compressive character of Figure 1.6, they conclude, is evidence for gain mismatch between rod and cone signals [Virsu and Lee, 1983]. Since the

gain of signals of rod origin is smaller than those of cone origin, adaptation (the amount of curve shifting) is less than perfect.

The wide-range cells in the primate LGN respond over 5 log units of stimulation [Barlow et al., 1978; Kayama et al., 1979], and are found in both the parvocellular and magnocellular layers of the nucleus [Kayama et al., 1979]. Presumably, several cones with different thresholds form the input to this stream, since a single cone responds over only 2-3 orders of magnitude.

Cortical responses

As in the LGN, both wide-range and narrow-range cells are found in the striate cortex (V1) of primates. Kayama et al. [1979] found that 24 % of cells in V1 fire monotonically as stimulus intensity over several orders of magnitude. They call these cells "luxotonic" and point out that they are not found in cats.

Maguire and Bazier [1975] studied the flash response of orientation-selective cells and cells with no orientation preference in monkey striate cortex, and found different response ranges for the two classes. Orientation-selective cells responded to flashes over a range of about 1.5 log units when adapted to light in the mesopic range. Cells with a wide dynamic range, however, always lacked orientation selectivity. For parvocellular cells in striate cortex, a lack of orientation selectivity is associated with the blobs (anatomically distinct "islands" of cells scattered across the striate cortex) and cells with orientation specificity are not found in these "islands" [DeYoe and Van Essen, 1988]. Dynamic range thus appears to be one basis for separation of parvocellular cells at the level of striate cortex [Allman and Zucker, 1990]. Anatomical separation of magnocellular cells on the basis of dynamic range has not been found in striate or other regions of cortex.

1.4.2 The (psychophysical) response of the visual system

Human observers are able to estimate the magnitude of a flash of light over several orders of magnitude. Data from the pupillary reflex [Barlow et al., 1978] and from brightness

estimation experiments [Stevens and Stevens, 1963] indicate a photometer-like visual sensitivity that spans 8 orders of magnitude. The wide-range cells described above provide this capability.

This enormous range is available perceptually only because the response is so compressive; this type of response is not good for judging small differences, such as small reflectance variations. High contrast sensitivity vision, which is more normally used day-to-day, presumably arises from the narrow-range neural units described above. This second, more behaviorally important type of vision is the subject of the remainder of this discussion.

Logarithmic models of brightness sensation

The oldest model for the human visual response, the Weber-Fechner model, arises from just-noticeable-difference experiments. In this type of experiment, a subject looks at a small spot of light in a large field, and finds the lighting level of the small spot that makes it just barely discernible from the background. As the background light luminance is increased, the just-noticeable-difference luminance of the spot increases proportionally. Weber's law describes this result explicitly:

$$dL/L = constant. \tag{1.2}$$

The threshold increment of the spot, is 1-2 % of the background intensity [Kallmann, 1940]. If the constant or *Weber fraction* obtained is taken as the quantum of perception, then the brightness B is found from the the integral form of this equation:

$$B = m \log L + C. \tag{1.3}$$

This relation, known as Fechner's law, describes sensation above the just-noticeable-difference threshold. The slope m is inverse of the Weber fraction. Taking the brightness at the adapting luminance L_{av} as an arbitrary zero,

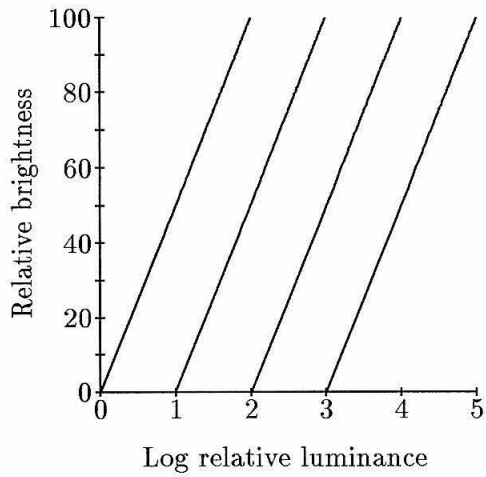


Figure 1.7: Visual response as predicted by the Weber-Fechner model. The response curve shifts to completely compensate for a change in adaptation level, and the slope of the curve does not change as it shifts.

$$C = -m \log L_{av}, \tag{1.4}$$

$$B = m \log L - m \log L_{av}, \tag{1.5}$$

$$B = m \log(L/L_{av}). \tag{1.6}$$

The family of curves predicted by the Weber-Fechner model is shown in Figure 1.7.

Power law models of brightness sensation

The Weber-Fechner model of brightness sensation relies on the assumption that the just-noticeable-difference is a quantum of sensation. A measurement made at threshold is used to extrapolate above threshold. Actual suprathreshold measurements do not bear out this assumption and extrapolation.

The Weber-Fechner model was for a long time the premier model of sensation because it used an objective measure, the Weber fraction, to establish a scale of sensation. Suprathreshold measurements rely on *subjective* judgements. As a result, subjective psy-

chophysics is controversial. Still, many studies have been done using a subjective scale, and the results are sufficiently consistent that they are taken seriously by tone reproduction scientists and engineers [Nelson, 1966, 1977].

In the simplest sort of subjective brightness measurement, the eye is dark adapted and the subject is presented with flashes of light and asked to assign numbers to the sensation. The experiment is repeated at other adaptation levels. As a check, the two eyes are adapted separately and the observer matches stimuli presented to the dark adapted eye to the other eye, which is adapted at an arbitrary level [Stevens and Stevens, 1963]. The estimation and matching experiments yield a consistent result: the brightness sensation is a power-law function of the stimulus.

$$B = (L - L_o)^n \tag{1.7}$$

The exponent n is about 1/3, and the threshold L_o increases with adaptation level. Since the test and adapting stimuli are presented separately, one after the other, this result is a measure of *successive contrast.* As the eye moves about a scene, for example from a region of low luminance to a region of high luminance, this relation describes the immediate sensation of brightness change.

Of course images are viewed not as a succession of flashed uniform fields, but a succession of complex arrays. A second psychophysical approach is taken to measure the brightness sensation of complex fields and the effects of *simultaneous contrast*, i.e., the way in which different regions in an image effect the region of fixation. In this second kind of subjective brightness measurement, a simple display is presented - such as a spot of light on an evenly illuminated background - and the observer is asked to adjust the spot luminance so that it looks half as bright. This second luminance is then used as a standard, and the observer is asked to adjust the spot luminance to produce a sensation that is half again as bright, and so on. This fractionation method yields a calibration curve that is used in the next stage of the experiment.

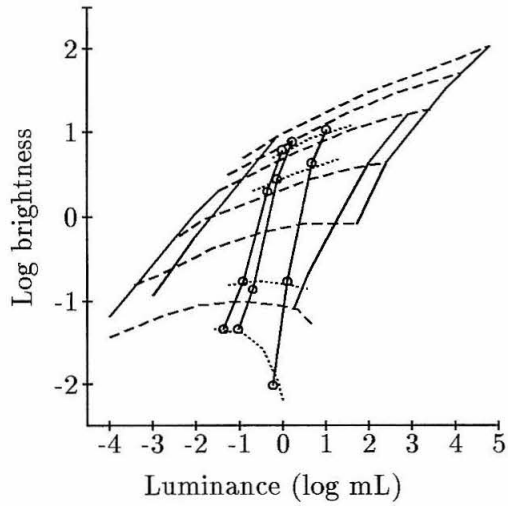


Figure 1.8: Subjective brightness response to complex stimuli. The pairs of curves at the far left and the far right were obtained by Bartleson and Breneman [1967a], and the three middle curves were obtained by Jameson and Hurvich [1961].

Now the observer is shown a complex image such as an array of patches that range from black to white, or an actual photograph of a scene. The observer views the complex display with one eye, and the standard simple display with the other. The observer adjusts the spot luminance in the simple display to match the subjective brightness of a point in the complex image [Jameson and Hurvich, 1961; Bartleson and Breneman, 1967a].

Jameson and Hurvich [1961], using this paradigm, made the important discovery that for complex displays, the slope of the brightness response curve (on a log-log plot) increases with adaptation level. It is expected from the results of successive contrast experiments such as the Stevens and Stevens [1963] results discussed above that a white piece of paper will look brighter at higher levels of illumination. Jameson and Hurvich also found that blacks look blacker at high levels of adaptation [see also Marimont, 1962]. Simultaneous contrast increases the "gain" of our sensitivity, as the slope of the Jameson and Hurvich curves is much greater than 1/3. Moreover, sensitivity seems to be referenced to an adaptation level that corresponds to a dark grey - thus whites appear brighter and blacks appear darker at higher light levels.

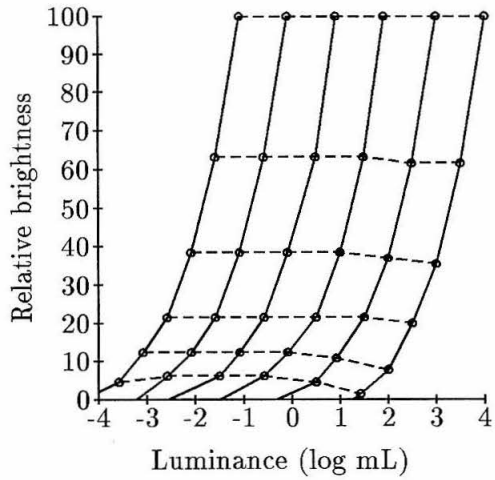


Figure 1.9: Relative subjective brightness response to complex stimuli, from Bartleson and Breneman [1967a]. A response of 100 is assigned to the sensation produced by a white patch in the scene.

Bartleson and Breneman [1967a], two Kodak researchers, extended the work of Jameson and Hurvich in two important ways. First, they studied subjective brightness of complex images (photographs) at adaptation levels that spanned 6 orders of magnitude, from moonlight (low mesopic) levels to sunlight (middle photopic) levels. Jameson and Hurvich worked only at dim light levels (the upper mesopic range), over 1.1 log units of adaptation range. Secondly, Bartleson and Breneman determined *relative* brightness from the absolute brightness data, i.e., the brightness of points in a scene relative to a reference white in the scene.

Four brightness response curves measured by Bartleson and Breneman are shown in Figure 1.8. A transparency and a print shown at high levels of adaptation yielded the pair of curves at the right of the figure, and the same stimuli shown at low light levels produced the pair of curves at the far left. The response curves for the prints are always lower than those for the transparencies, because of a white border around the prints. Several other pairs of curves (not shown) were obtained at various other adaptation levels. Each dashed line corresponds to a fixed luminance contrast above or below the adaptation level.

The data of Jameson and Hurvich are also shown in Figure 1.8. Note that the constant

contrast lines of Bartleson and Breneman are consistent with those found by Jameson and Hurvich (dotted lines). The slopes of the brightness response curves found in the two investigations are not consistent (for unknown reasons) but the general increase in response slope with adaptation level is found in both sets of data.

Using this data set, Bartleson and Breneman plotted on an arithmetic scale the brightness response *relative to a scene white* (Figure 1.9). Dashed lines in this figure trace the response to a particular point in a scene at different adaptation levels. The top dashed line is the response to a white point in the scene. Since these curves are normalized to the scene white, the top line is horizontal. The bottom line traces the response to a dark portion of a scene shown at different light levels. The black gets blacker with increasing adaptation level, but this effect, which seems so significant on a log-log plot, is revealed as quite small on an arithmetic brightness plot. The shape of the relative brightness response curve is practically constant with adaptation level.

It is interesting to compare this brightness response data with the response data of narrow-range LGN cells shown in Figure 1.6. On an absolute scale, the response to a fixed contrast above the adaptation level increases with adaptation level in Figure 1.6; this result is also seen in the psychophysical data (the top few dashed lines in Figure 1.8). However, the *change* in response from the lower constant-contrast line to the upper constant-contrast line in Figure 1.6 is about the same at all adaptation levels. The constant-contrast lines in the psychophysical relative brightness response curves (Figure 1.8) are also fairly parallel. This means that the shape of the *relative* brightness response curve does not change much with adaptation level.

Brightness (in)constancy

The term "brightness constancy" refers to the fact that the constant-contrast lines in Figure 1.8 do not have a slope of one. That is, the ultimate response of the visual system, measured psychophysically, is not strictly correlated with the physical stimulus like a light

meter is. On the other hand, we do not respond in strict accordance with contrast either, as a "reflectance meter" would (this would correspond to constant contrast responses with zero slope). This result has been known for a long time. Thouless refers to it as "phenomenal regression to the real" and notes that a similar phenomenon is evident in many perceptual constancies such as size constancy and the constancy of shapes seen in perspective [Thouless, 1931a; Thouless, 1931b].

Given a reference, we can judge the reflectance of an object remarkably well, as relative increments are preserved (Figure 1.9). At the same time, we can judge the absolute physical light levels well. Thus arguments about whether constancy exists still rage in the literature [Gilchrist, 1988; Jacobsen and Gilchrist, 1988; Heinemann, 1989]. Perhaps when researchers agree on methods and terminology this debate will finally cease.

1.4.3 Range mismatch revisited

There is a fair match between the psychophysics and the neurophysiology of visual sensitivity. Subjective black is about two orders of magnitude below the adaptation level [Lowry and Jarvis, 1956; Maloff, 1939]; the narrow range off-cells in the visual system can respond over that range. Subjective white may be many orders of magnitude above the adaptation level, but fixation upon a bright stimulus changes adaptation quickly; adaptation will settle about two orders of magnitude below this white, which is again within the response range of narrow-range neurons (on-cells).

Jones and Condit [1948] (discussed above) found that the range of luminances within which detail can be seen is at most 760. (Condit later found a scene for which the dynamic range of the visual system spanned a factor of 1000 [McCann, personal communication].) Given both on- and off-cells within the visual system with response ranges of 1.5-2 orders of magnitude, an image range of 760 can be accommodated.

There is no range mismatch in the visual system analogous to the mismatch that exists in electronic and photographic tone reproduction systems. However, since we can sense

luminance over 3-4 orders of magnitude at any instant, to reproduce the richness of a scene, artificial systems must have this output range. As discussed above, they do not. Artificial systems must be designed to compress the range of an image so that it is rendered well within this constraint. This is the subject of the next chapter.

Chapter 2

Solutions to the Dynamic Range Mismatch in Tone Reproduction

Several methods are in use to provide automatic, mass-market image range compression. The simplest and crudest methods consist of scaling the image (zero-dimensional processing); sophisticated methods involve spatial filtering operations (two-dimensional processing). This chapter reviews the various range compression techniques in use today.

2.1 Scaling

The simplest way to compress an image is to distort its scale, by changing the slope and/or the shape of the input-output characteristic.

In photographic printing, a simple slope change is achieved by inserting a neutral density filter with the desired attenuation factor between the negative and the print paper in an enlarger. The slope, on a linear scale, is decreased by the filter factor. On a log scale, the output is decremented by the filter attenuation factor. To increase the slope, on a linear scale, a longer exposure time or a more intense exposing light is used. On a log scale, a faster (more sensitive) paper is used to increase the slope. A change in the slope on a linear scale corresponds to a gain change in electronic imaging; on a log scale, a slope change

corresponds to a change of the exponent in a power-law transformation.

In electronic imaging systems, scaling is performed immediately after the sensor [Neuhauser, 1957] or after sensor nonlinearity compensation circuits [Lappalainen and Hameenaho, 1976], in the camera or transmission system. Automated scaling is seldom performed in the receiver, though the brightness and contrast controls on the front of a television monitor allow for manual scale adjustments.

Lastly, scaling can be accomplished by careful lighting of a scene. In lighting for motion pictures, the ratio of the greatest amount of light falling on a subject must be 4:1 or less for good results. For television production, the subject lighting contrast must be even smaller - a ratio of 2:1 in lighting contrast combined with the 20:1 reflectance range serves to fill the dynamic range of television reproduction [Percy and Veal, 1954].

2.2 Saturation

Scaling is not the best solution to the range mismatch problem in tone reproduction. An image which has been greatly scaled down to reduce its range is not pleasing to viewers, who use terms such as "flat" and "low-contrast" to describe the result. Photographers have realized this for some time, and have determined that the viewers of photographic prints, for example, prefer a slope near one (i.e., no scaling) in the middle tones of an image [Nelson, 1967, 1977]. The only way to maintain a reasonable scale in the middle tones of a wide-range image is to compress the scale of the extreme tones.

Photographic media naturally compress the tone scale at the top and bottom of the range of sensitivity. That is, negatives, positive transparencies, and print papers compress both the shadow and highlight extremes of the image luminance range (Figure 2.1). The S-shaped transfer characteristic of photographic media, is also called the H and D curve (after the British photographic pioneers Hurter and Driffield), the D-log E curve (since it is a plot of the density of a negative or print paper as a function of the logarithm of the exposure [Farnell, 1966]), and simply the characteristic curve. The density of the print

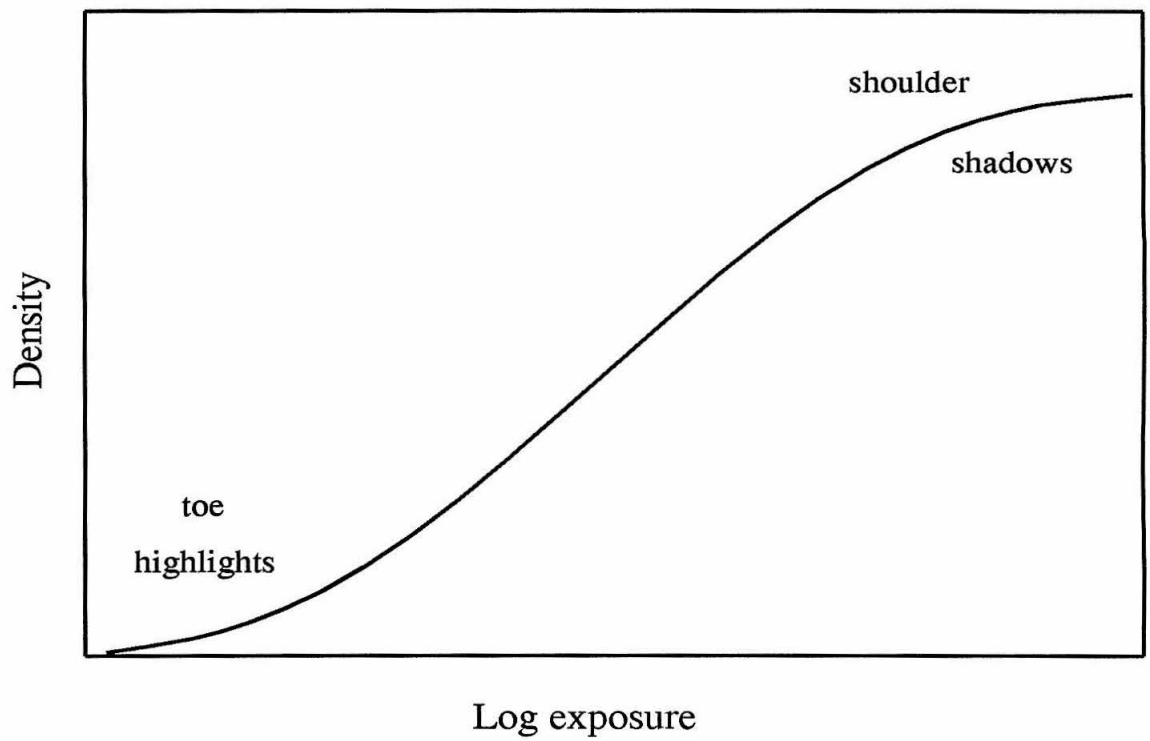


Figure 2.1: The characteristic curve of photographic print papers. The tone scale in the highlight ("toe") and shadow ("shoulder") regions of sensitivity is compressed, but in the middle tones there is little compression.

is the log of its reflectance; a change in slope of a D-log E curve is a power-law change in the mapping of the scene light to the reflectance of the reproduction. The low-density compressive region is called the "toe," and the high-density saturating region is called the "shoulder."

As noted in the last chapter, the range of photographic negatives is often sufficient to record the entire luminance range of a scene, but the range of print positives is relatively limited. The tone reproduction analysis method of Jones [Jones, 1944] outlines in graphical form the compromises that must be made given the range mismatch between scene and print ranges, with the added constraint of the compressive, saturating characteristic curves of negatives and print papers. Figure 2.2 is a typical Jones diagram [Mees, 1960]. In photographic practice, the lens aperture is adjusted so that the scene range falls either entirely on the linear portion of the negative D-log E curve, or somewhat lower, with the shadows of the scene falling on the toe of the negative curve. The latter practice, which results in some compression of the dark parts of an image (block I of Figure 2.2), is actually the standard. The positive material (block II) has a smaller range than the negative, and, in standard practice, the shadows are placed on the straight-line portion of the positive curve, so that no more compression of the shadows takes place. Instead, highlights are compressed by the shoulder of the positive characteristic curve. If the print is viewed in good conditions, without glare, the print density is carried through (Block III) to the reproduction characteristic (block IV). The final reproduction is a relative copy of the initial scene in the middle tones, and is compressed equally in the highlights and the shadows.

In electronic systems, the compromises made to cope with the range mismatch of imagers vs. displays are similar (Figure 2.3). An S-shaped curve on a log-log scale corresponds to a largely compressive characteristic on a linear-linear scale (curve A). This transfer function, identical to the photographic transfer function is used in electronic imaging [Fink, 1941], but not as commonly as the "knee" curve (curve B). On a log-log plot, most of transfer curve B is linear, with saturation mainly evident in the highlights. This saturation is produced

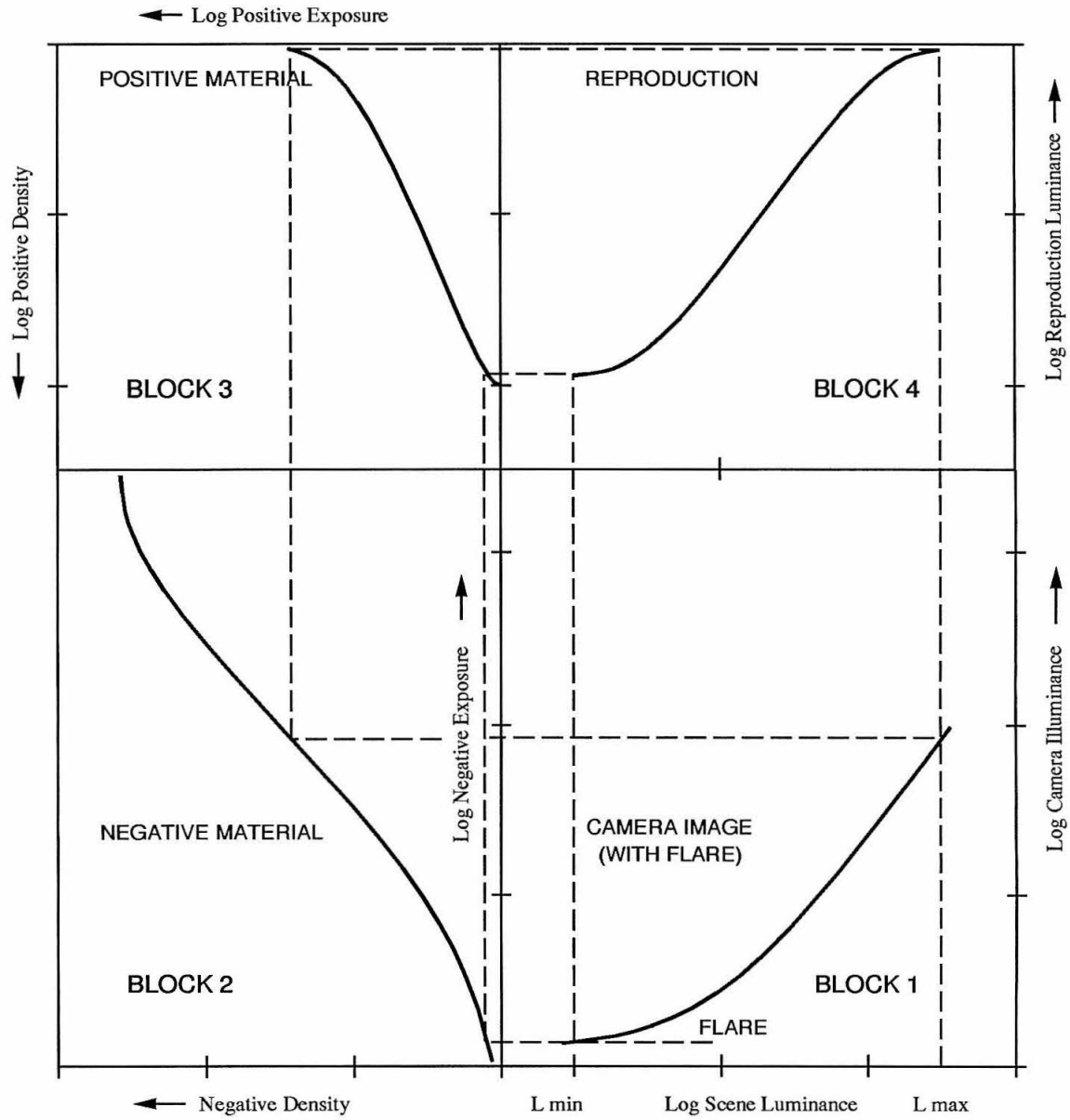


Figure 2.2: Typical photographic tone reproduction diagram. The shadows are compressed by the negative transfer function, and the highlights are compressed by the positive (print paper) transfer function. Adapted from Mees [1960].

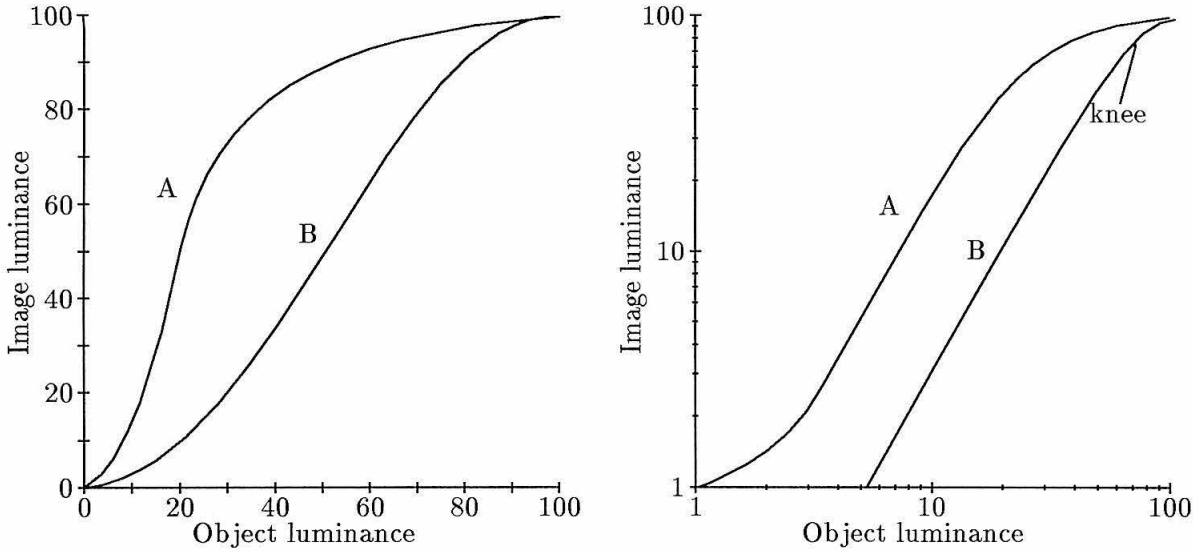


Figure 2.3: Transfer curves used in electronic tone reproduction. Curve A is identical to the photographic characteristic curve, but curve B, the "knee" characteristic, is more commonly used. From Fink [1941].

electronically; the point on the curve at which saturation begins (on a log-log plot) is called the "knee."

2.3 Scale and Filter Linearly

Consider a scene in which there is a great disparity in lighting amongst important subjects. Imagine, for example, a group portrait in which half of the people are standing in sunlight, and half are in shadow. Photographic manuals, of course, warn the photographer to not even try to capture such a scene. But photojournalists, for example, often have no choice, and amateur photographers often do not read the manuals. Such a scene, with important elements at very different lighting levels, can usually be captured by negative film, with its wide dynamic range, but cannot be reproduced as a print.

Mere saturation of the bright and dark portions of the image is the worst compromise in this case, as some people's faces will be bright with little contrast, and others will be dark with little contrast. A change in the exposure to shift one part of the group to the middle of the tone range inevitably shifts the other half of the group into complete darkness or

featureless white.

The traditional solution to problem images of this sort is to *mask* the negative, that is, to subtract something from the light passing between the negative and positive in the print making process. Masking with a slightly blurred version of the image reduces the range, as it is a form of high-pass filtering. By printing the masking image on a high-slope paper, a wide range can be reproduced with reasonable contrast. Figure 2.4 illustrates this [Nelson, 1966] for a step pattern with detail on each step. Without the masking, the detail is reproduced at low contrast (bottom right), but with masking to reduce the range of the steps, a higher gain paper can be used and the detail contrast is boosted (upper right).

This procedure represented the acme of the science of tone reproduction for the middle part of the century [Yule, 1967; Hunt 1967], and was rediscovered by the early image processing community [Oppenheim, Schafer, and Stockham, 1968] and by scientists modelling simultaneous contrast in the visual system. The latter group were especially interested in the spatial filtering aspect, because of visual effects called the *Mach bands* [Mach, 1897; Ratliff, 1965; Marimont, 1962; Marimont, 1963; Bekesy, 1968a; Bekesy, 1968b; Tolhurst, 1972; Heinemann, 1972; Fiorentini, 1972; Marr, 1982]. When humans view a staircase pattern of brightness like that diagrammed in Figure 2.4, they see overshoots on either side of the step, just like the overshoots that result from masking [Hurvich, 1981; Fiorentini et al., 1990]. The overshoot results because the mask smooths through the stairstep, and the mask value is a poor estimate of the intensity right next to the step - rather, it is a better estimate of the average of the two areas on either side of the step. The difference between an image point and the point in the blurred mask is added to the original to give the final result.

This is fine for staircase patterns, but what about the problem image discussed above, with a sharp division of the image by a deep shadow ? Photographic investigators know full well that the procedure can break down in this situation. If the illumination step is too great, the overshoot that results is quite noticeable. In fact, the overshoot itself may

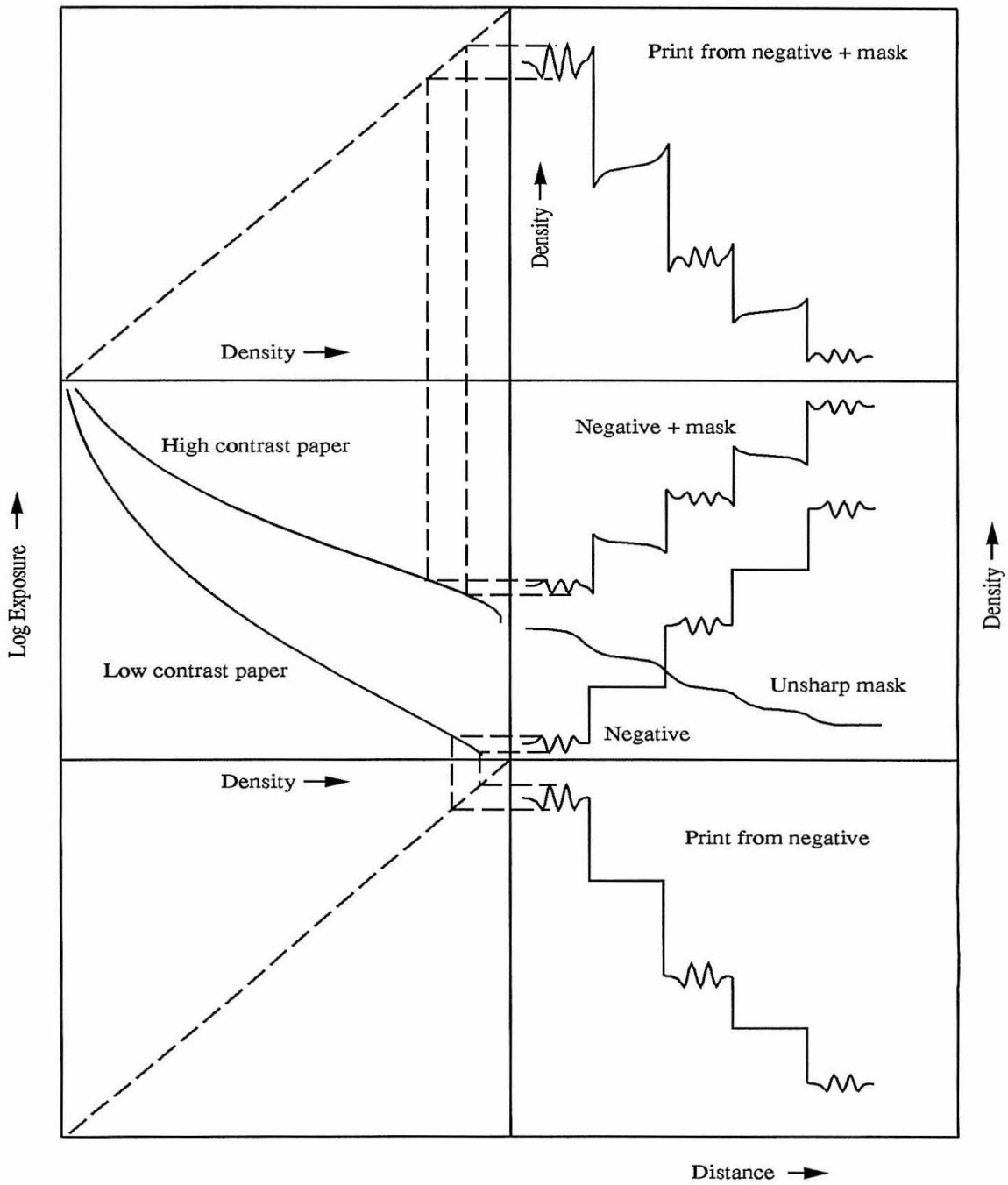


Figure 2.4: Unsharp masking. While a low contrast paper must be used with the negative to compress its range (from middle right, counterclockwise to lower right), an unsharp mask in combination with the negative compresses the range enough before printing that a high contrast paper can be used (from middle right, clockwise to upper right). From Nelson [1966].

exceed the range of the print medium, leaving no detail at all adjacent to the step; in this case, masking is not worth the trouble.

Masking to solve this problem, i.e., to pull up a dark area and pull down a dark area, is called *area masking* in photography [Yule, 1967] and in electronic imaging [Herbst et al., 1951; Herbst et al., 1952; Leistner, 1960]. The terms *unsharp masking* and *aperture equalization* refer to masking for improving sharpness, for aesthetic reasons, or to compensate for loss of sharpness in some stage of the photographic [Hunt, 1967; Yule, 1967] or electronic [Herbst et al., 1951; Herbst et al., 1952; McMann and Goldberg, 1968; Bayer and Powell, 1986; Brooks and Cosgrove, 1970; Nikoh et al., 1991] tone reproduction process. An area mask is usually more blurred than an unsharp mask. Unsharp and area masking are in common use in the graphic arts and the resultant overshoots are easy to spot in magazines and newspapers.

It is interesting to note that some image processing investigators completely ignored the problem of overshoots, and filtered to compress the range regardless of artifacts introduced by the filtering process [Oppenheim, Schafer, and Stockham, 1968; Stockham, 1982; see Schreiber, 1978, for a lucid criticism of this oversight]. Visual scientists continue to ignore it, for the most part, though there has been some discussion in the literature [Fiorentini, 1972; Ratliff, 1984; Fiorentini et al., 1990; Ross et al., 1989].

2.3.1 Histogram equalization

The digital image processing method known as histogram equalization is a form of scaling. In this method, the maximum and minimum pixel values are found, and a lookup table is generated that maps the minimum value to the output minimum and the maximum value to the output maximum. In the case of a linear mapping, this corresponds to scaling and adjusting the y-intercept to pass through the origin.

In the technique known as local area histogram equalization, a separate lookup table is found for each of many areas in an image. This corresponds to masking, as the average

value of the image is computed and removed from each block; the block size is akin to the kernel size of a low-pass filter. Not surprisingly, images processed this way look unnatural. Overshoots are present in areas around high contrast edges, and most of the background gradation of shadow is removed.

2.4 Scale and Filter Adaptively

As discussed above, high-pass filtering of an over-range image improves it in some cases, but actually does damage in others. Schreiber and coworkers [Schreiber, 1978; Curlander, 1974] carefully studied the limitations of the technique, and developed algorithms on digital computers that compensate for some of those limitations. Schreiber and co-workers used digital computers to obtain a "derivative image" from pointwise spatial derivatives, and to blur the derivative image. They used the value of the blurred spatial derivative to modulate the filtering at each point in the image. (This method was rediscovered by a later set of workers, who called the blurred derivative image the "edginess" image [Moore et al., 1991].) A block diagram of the procedure is shown in Figure 2.5.

In filtering, these workers first subtract the low-pass version of the image from the original, to derive a high-pass filtered image. This is because they use a scanning drum to digitize the image; on such scanners, a transparency or negative is scanned both with a sharp beam and with a blurred beam. Both signals are available for digitization, and it is a simple matter to subtract the two for fast high-pass filtering. The "highs" image and the original are digitized, and the latter is differentiated, rectified, and smoothed digitally to yield the derivative image (Figure 2.5). Next, the derivative image is transformed with a look-up table and then multiplied with the highs.

The look-up table, shown at top right in Figure 2.5, serves two purposes. First, in regions of low-contrast, the multiplier of the highs signal is zeroed. This prevents the accentuation of noise in smooth regions, a practice called *coring* in television engineering. Second, and more important for the purposes of this discussion, the lookup table weights high contrasts

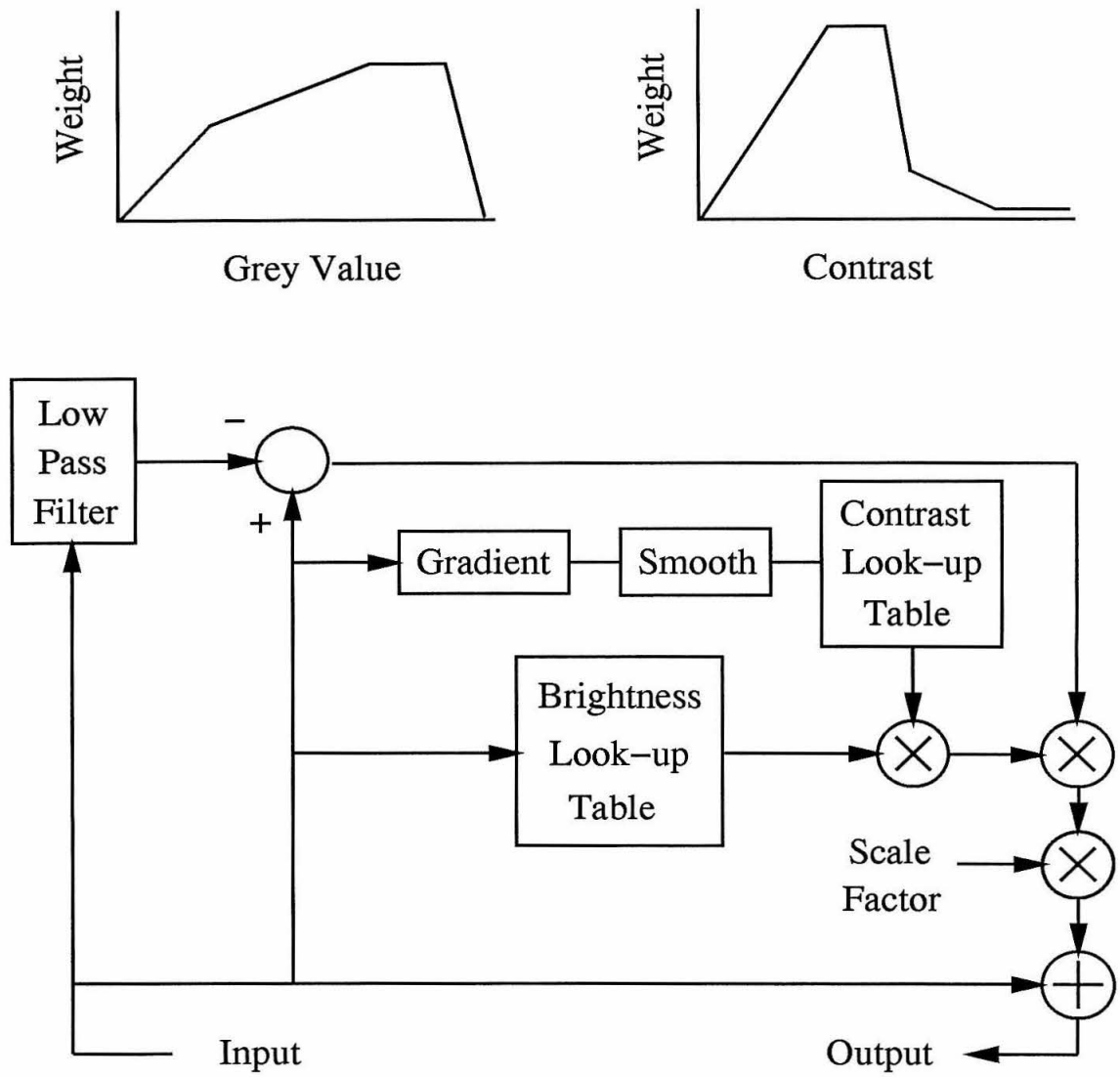


Figure 2.5: Adaptive filtering. A high-pass filtered version of the image is added to the original, but not at large steps in intensity, where filtering overshoots are noticeable. The weights of the contrast look-up table (top right) dictate the degree of filtering at various contrasts, and the brightness look-up table (top left) distorts the tone scale of the added edges from a linear to a cube-root scale to match the scale distortion of the visual system. From Curlander, 1974.

far less than middle contrasts. This prevents overshoots at high-contrast edges where they are most noticeable, by blocking the filtering process. In areas of medium contrast, where the addition of a highs signal enhances image sharpness and perceived contrast without adding overshoots that are *not* noticeable and annoying, filtering is performed normally. (Note also the brightness look-up table at top left of Figure 2.5. This table scaled the high-pass filtered image according to a cube-root scale, to maximize the psychophysical "fit" of the added edges.)

Masking for range compression has been well known in photographic circles for many decades. But it is not easy to form an edginess image optically, since there is no easy way to form the magnitude of the spatial derivative. The advent of the digital computer enabled the advance of the practice of tone reproduction, especially with regard to the problem of range mismatch. As such, Schreiber and his contemporaries introduced a new era in image processing.

Sophisticated as it is, this method has its drawbacks. First, one may remark that it embodies a compensation three times over. An image is scaled down to compensate for the range mismatch between the image and the reproduction medium - but this results in a loss of contrast. Filtering is introduced to boost the contrast, in compensation for the flatness brought on by scaling - but this results in overshoots. Adaptive filtering compensates for the overshoots at high-contrast edges brought on by filtering.

Second, two smoothing operations are needed, one to form the low-pass image from the original, and one to form the edginess image from the derivative image. The adaptive filtering method championed by Schreiber and coworkers makes use of optical blurring for one of the smoothing operations; designers of all-electronic tone reproduction systems will not have this speed-of-light smoothing operation available.

Thirdly, adaptive filtering can distort an image, though the distortion is more subtle than the overshoot distortion that results from straight, non-adaptive filtering. In textured areas, contrast is often high. This method will prevent filtering in such areas, and so the loss

of contrast introduced by scaling carries through to the final reproduction. For example, a grove of saplings, backlit with a bright sky, will be reproduced with bark that is flatter than the original, but a close-up of a sapling will be reproduced with much more detailed bark. As another example, the clouds in the sky of an image filtered adaptively will be quite detailed - except in the neighborhood of a flock of birds in the sky. These are subtle effects, that are probably not noticeable in still images. In moving images, though, the effect may more noticeable. No real-time system of this sort has been built yet, so it remains to be seen whether this solution, which works so well in the graphic arts, transfers well to video imaging.

2.5 Dodging

Ironicly, the best solution to the problem of range mismatch in tone reproduction is also one of the oldest. This is the photographic darkroom technique known as *dodging.* Dodging consists of shaping, by hand, the tone scale of different regions of an image by interposing an opaque light blocker between the light from the negative and the photographic print paper. The opaque blocker, such as a wand with a circle at the end (Figure 2.6), is waved in the enlarging light in areas that are very dark, to diminish exposure in those areas. If the dodger is not artful, and the blocking perimeter does not exactly line up with the high-contrast step between a light and a dark area, an overshoot results.

This technique is very effective, because the segmentation of regions of different tone scales is performed by the most sophisticated image processor available - the human brain. But, the method is time-consuming, painstaking, and labor-intensive, and so it is seldom used except on a custom basis.

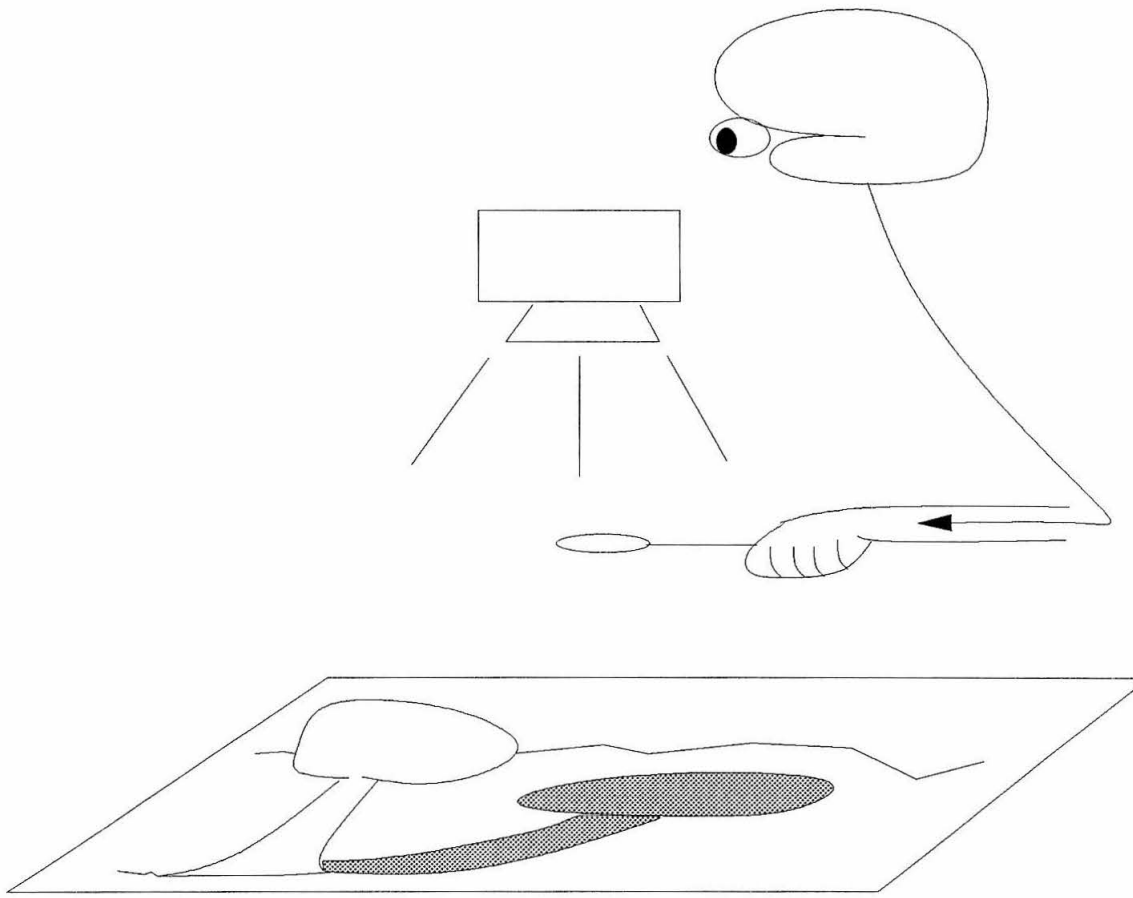


Figure 2.6: One of the oldest and most effective range compression methods: dodging. The cost of the image processor needed in this technique has prevented its application in mass-market tone reproduction.

2.6 Summary

All of the *automated* techniques described in this chapter - scaling, saturation, filtering, histogram equalization, and adaptive filtering - go some distance in solving the problem of range mismatch in tone reproduction, but even the most sophisticated of them distorts the image. The last method described, dodging, depends on a human operator to segment regions of an image that are out of range (presumably due to lighting effects).

In the next chapter, compact hardware for electronic image filtering is described, and is integrated into systems for masking and adaptive filtering. In chapter 4, the hardware is operated in a mode that performs both segmentation and filtering and an automated method similar to dodging is described.

Chapter 3

An Analog VLSI Resistive Grid for Video Image Spatial Filtering

3.1 The Chip

3.1.1 Introduction

Image smoothing is an important operation in artificial and natural vision systems. Commonly this is done in man-made systems by digitizing the output of a video camera, performing fast convolutions digitally, and, if display of the smoothed result is required, reconverting the image to analog voltages compatible with video monitors. Such systems are compact only if the degree of spatial smoothing is limited to a few pixels (e.g., 7 by 7 pixels).

Analog methods of smoothing images have been previously described by Mead [Mead, 1989]. In his analog retina, the image is formed by on-chip photoreceptors; the smoothing is carried out in real time with a resistive grid. In contrast to conventional digital techniques, smoothing with a resistive grid built from subthreshold analog VLSI is performed with very low power consumption, and with a compactness that does not vary with the degree of smoothing; the equivalent convolution kernel can be varied in size from a small fraction of the image (one pixel) to the whole image with no loss of compactness and little increase in

power consumption. Other investigators [Luo, Koch, and Mead, 1989] have accomplished smoothing with resistive grids, but only "off-line" - the image is input to the chip by an external digital system, at rates much lower than frame rate. In this chapter a chip is described that can receive images directly from a video camera and perform smoothing via a resistive grid at the video frame rate.

3.1.2 System description

Physical description of the chip

The chip was fabricated through MOSIS with 2 micron CMOS design rules. It is approximately 5 mm by 7 mm in size, and contains a 46 by 47 node hexagonal resistive grid, sample-and-hold circuitry, and switching circuitry. A four by four representation of the chip is shown in Figure 3.1. Across the bottom is a row of cells which perform a sample-and-hold operation at the line rate, and select columns of the resistive grid for analog signal input and output. On the right is a column of cells which select the row of the grid which is written to and read from. In the middle is a two dimension array of cells which perform a second stage sample-and-hold, smooth the input from this sample-and-hold by way of resistive interconnections, and output a current proportional to the smoothed image.

Input sample-and-hold strategy

A two stage sample-and-hold design is the key to the video rate performance of the chip. In the first stage, the video signal is switched onto one of 46 capacitors through pass transistors. At one microsecond per pixel, 46 of the approximately 52 microseconds of video is utilized per line of the image. (With appropriate video switching, all 52 us could be utilized, however. The one microsecond per pixel sample time was convenient given the crystals at hand.) Followers fed by these capacitors pass the signal to a second sample and hold circuit within each cell of the two-dimensional array *during the horizontal blanking interval.*

This strategy is essential to the task at hand, since sampling must proceed at high rates

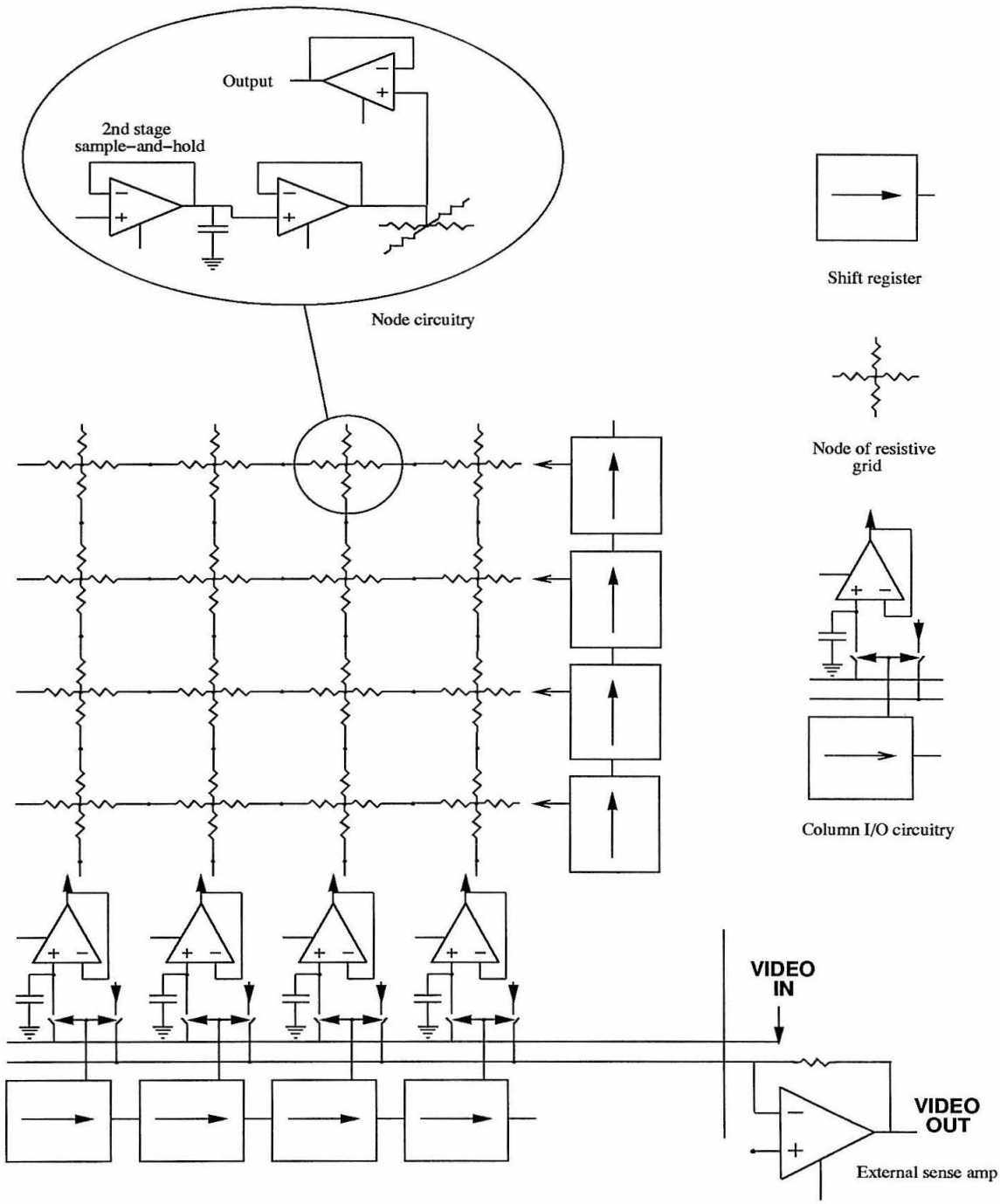


Figure 3.1: Schematic of the resistive grid chip. Incoming video is sampled on capacitors at the line rate. The row of sampled values is fed to a row inside the main array where is sampled again (inset).

(1 us), but holding must last for at least one video field (16 ms). This is a difference of four orders of magnitude; the two stage design has circuits that operate independently at each of the two extremes of this range. The settling time of a circuit comparable to our second stage is on the order of 10 us. The horizontal blanking interval lasts for about 9 us in NTSC video; this system makes use of this "dead time" to charge up the second stage of sample-and-hold circuits inside the two dimensional array.

Since a standard video signal has a resolution of roughly 500 by 500 pixels, split into two fields of roughly 250 lines each, this system samples a video image at a very low resolution. At 1 us per pixel horizontal sampling time, about ten video pixels are averaged (integrated in time by the first stage sample-and-hold circuit) for input to one resistive grid cell. Similarly, with a 47 row resistive grid, several lines must be averaged to form a single row of resistive grid input. In particular, five rows per field are fed into a given row of second stage sample-and-holds for each field; at two fields per frame, 10 lines are averaged (integrated in time by the second stage sample-and-hold circuit) to form the input to a single line of the resistive grid. Thus each pixel input to the grid for smoothing represents an average of 10 by 10 video pixels.

Resistive net

The circuit details within each pixel are similar to those of the analog retina [Mead, 1989]. 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. (This circuitry is discussed in detail in the next chapter.) The output of each node is a follower-connected transconductance amplifier, run above threshold, which produces a current proportional to the node voltage. This current

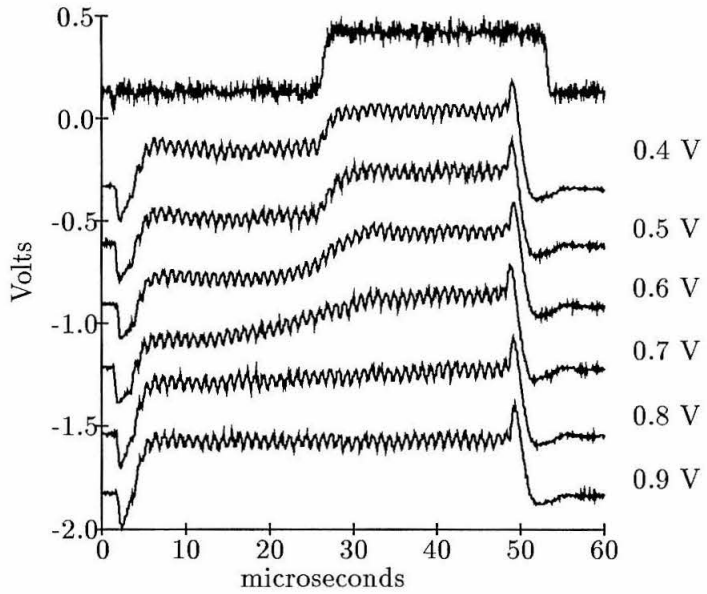


Figure 3.2: Chip input (top trace) and output (bottom traces) at different degrees of image smoothing at the video line rate. Each trace is a horizontal line of the chip output. Bumps of one microsecond duration in the output traces are the result of current switching transients inherent in the current-sense output scheme.

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.

Output circuitry

Signals are read out from the resistive grid as currents. They are sensed by an off-chip high gain current sense amplifier built from a conventional op amp. At any given moment, a single row of the resistive net puts current onto the set of columnar output lines, and a single column is selected to pass its current off-chip through pass transistors. The pixel being charged up in the first stage sample-and-hold at the moment corresponds to the selected column, so that a given pixel is written to and read from simultaneously.

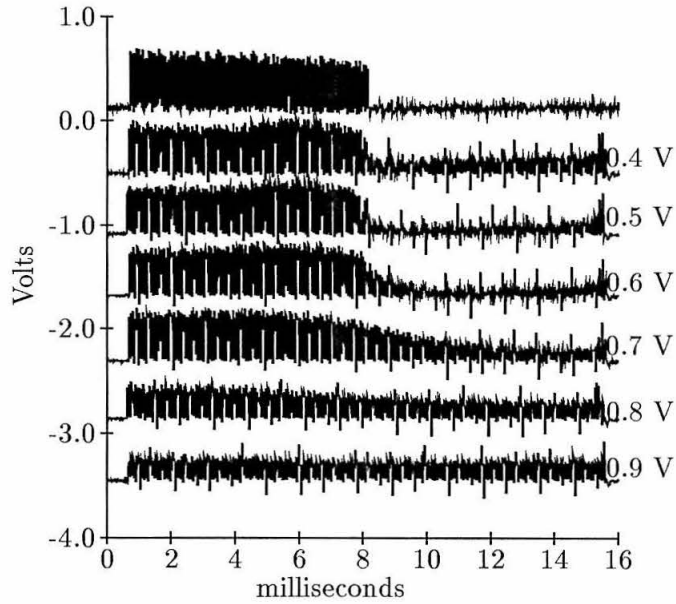


Figure 3.3: Chip input (top trace) and output (bottom traces) at different degrees of image smoothing at the video field rate.

3.1.3 System performance

Horizontal and vertical smoothing are depicted in Figure 3.2 and Figure 3.3. As described in Mead [1989], the space constant of smoothing is proportional to the square root of ratio of the transverse and input conductances, for a one dimensional resistive grid. This relationship also holds in the approximation for a two dimensional grid. In Figure 3.2 and Figure 3.3, the bias of the input transconductance amplifier is held at 0.55 Volts and the bias of the transverse resistor circuit is varied. The value of the transverse resistor bias is shown to the right of each trace. The output essentially replicates the input at low transverse conductance values, spatially averages the input at high transverse conductance values, and spatially smooths the input at values in between.

3.2 Video Switching Circuitry

A custom digital circuit that derives switching signals for the resistive grid processor from a camera sync signal is described in the following.

3.2.1 System clocks

The system clocks for the video switching circuit are developed as shown in the top left block of Figure 3.4. A 16 MHz crystal is placed between pins XTAL1 and XTAL2. With an inverter between them, the signals on these two pins must be of opposite polarity; the signal CCLK at the output of the inverter is a robust 16 MHz digital square wave. Used as the clock of a pair of D flip-flops, one of which is inverting, CCLK is divided by four, to yield an 4 MHz system clock, SCLK. (SCLK is routed off chip and back onto a special pin that distributes it to all parts of the chip).

Since the shift registers of the resistive grid chip are of the two-phase type, a pair of non-overlapping shift pulses are needed to increment X and Y shift registers. The first shift pulse, PHI1, occurs in the first half of the negative phase of SCLK, and the second, PHI2 occurs in the first half of the positive phase of SCLK. The PHI signals are developed first as the two intermediate signals INCA and INCB; both are high for 62.5 ns of the 250 ns period of SCLK, separated by 125 ns. To drive the vertical and horizontal video shift registers, the PHIs must be gated by a signal that is high for one full cycle of SCLK. Those gating signals, XINC00 and YINC0, are generated by the other 8 blocks of the switching circuit.

3.2.2 Sync shaping and vertical-horizontal countdown

Vertical and horizontal sync signals from a camera are fed into D flip-flops for capture, and processed in the temporal context of the digital system clock in the bottom left block. A pair of D flip-flops triggered at opposite phases of SCLK, along with an AND-NOT gate produce the shaped VERT and HORZ1 signals, which are high for half of a cycle of SCLK once per vertical and horizontal sync.

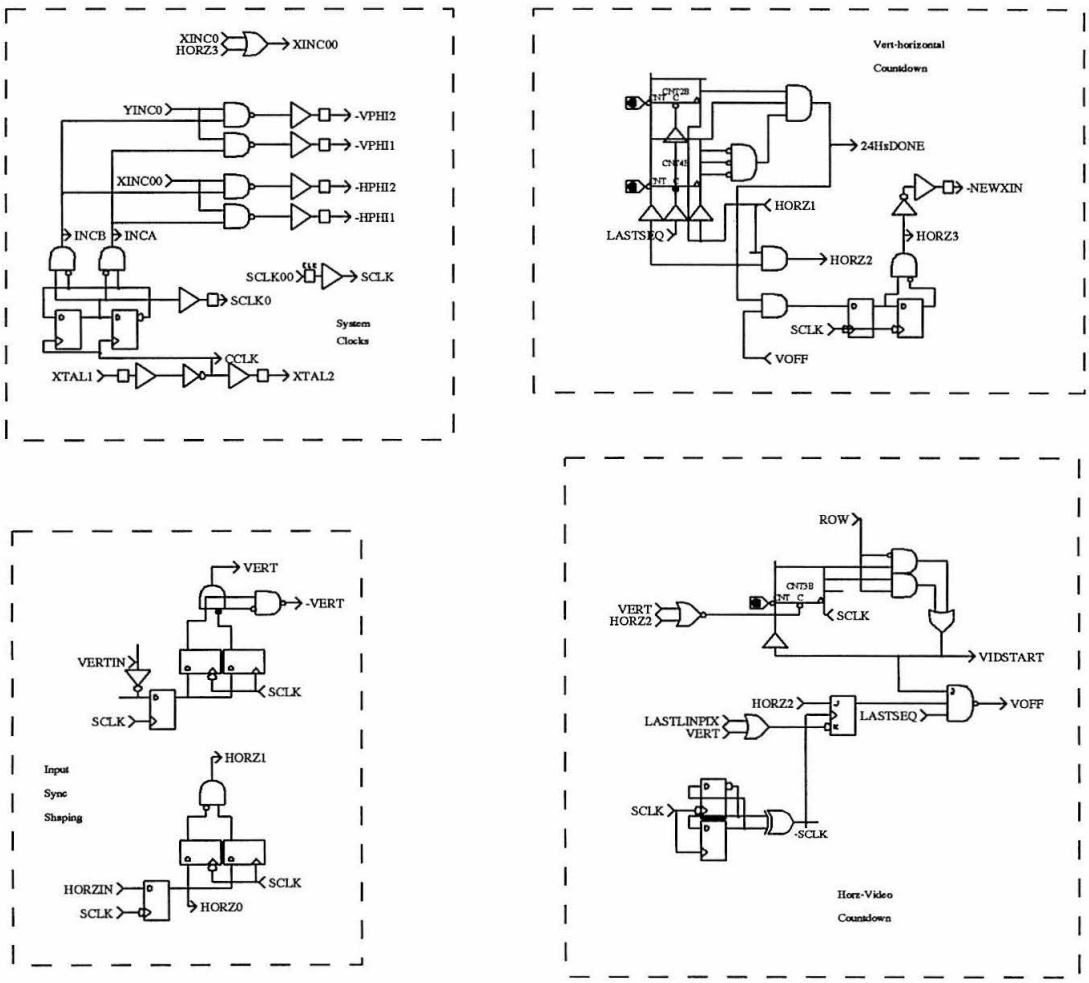


Figure 3.4: A portion of the video switching circuitry. See text for description.

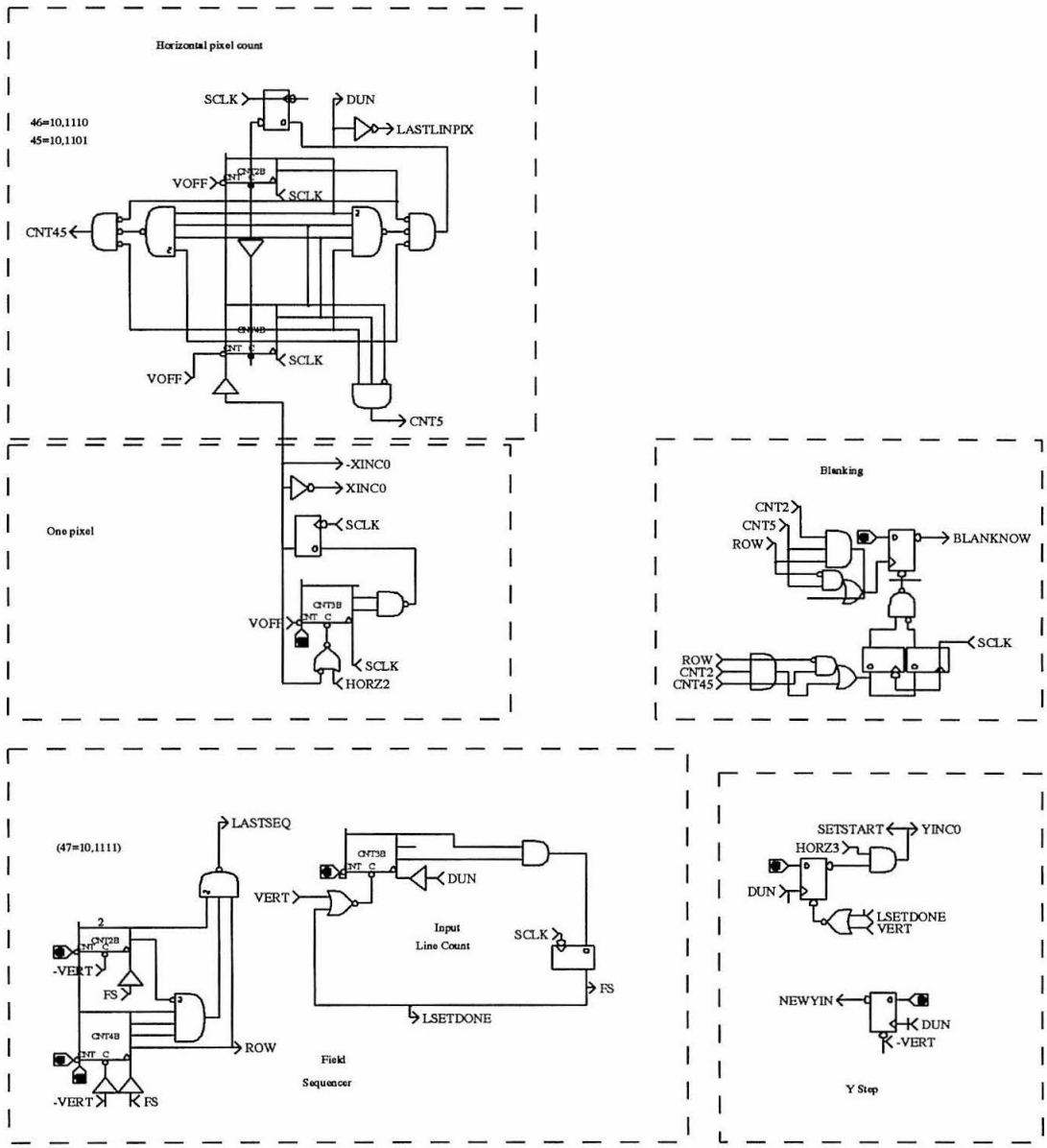


Figure 3.5: The remainder of the video switching circuitry. See text for description.

In NTSC video, there is no video information transmitted in the first 24 horizontal intervals following a vertical sync. The countdown from the vertical sync is performed in the upper right block by a five bit counter. The counter, which is implemented as a cascade of a 4 bit and a 2 bit counter, is cleared after all of the lines of a video field pass. The signal LASTSEQ is this clearing signal. LASTSEQ goes high immediately following a vertical sync, and so releases the counter circuit from a clear state. Subsequent horizontal syncs increment this counter, and after twenty four horizontal intervals, the counter stops and the signal 24HsDONE goes high and stays high until cleared by LASTSEQ after many more horizontal intervals, at the end of the video field.

Two more signals are derived from this "countdown complete" signal. The first, HORZ2, results from the gating of the HORZ1 shaped sync signal by the 24Hs countdown signal. This signal indicates the start of horizontal intervals that contain video information. The second signal, HORZ3, results from the gating of the VOFF signal by the 24Hs countdown signal. VOFF is high during the horizontal blanking interval. Thus a signal is available that follows the end of every horizontal line that contains video information. With shaping by a flip-flop delay line and an AND-NOT gate, a signal is obtained from it that is high for a short interval before video starts in the horizontal interval. That short interval is synchronized with HPHI1 and HPHI2, the pair of horizontal shift register driving signals, so that it forms the "new bit" that is inserted into the shift registers at the beginning of the horizontal interval. The "new bit" signal, -NEWXIN, is the inverse of HORZ3. After this bit is inserted, subsequent HPHIs push it down the shift register.

3.2.3 Horizontal-to-video countdown

The nodes of the resistive grid are arranged in a hexagonal array, with even rows shifted right one-half pixel with respect to odd rows. Video should be fed to and read from the first pixel of even rows a little later after the horizontal sync than the time after sync that odd rows receive and output video. Each pixel of the resistive grid receives video for 1

microsecond; this means that the delay between even and odd rows should be 0.5 us, or two cycles of SCLK. Also, there is a set delay of 0.25 us from the time of the derived sync signal, HORZ2, and the time that video is available in the line interval. These delays are produced with a counter in the block labeled HORZ-VIDEO COUNTDOWN of Figure 3.4. This counter is cleared upon a horizontal sync. After two clock ticks for even rows (ROW high) or four clock ticks for odd rows (ROW low), the counter stops and the signal VIDSTART goes high and remains high.

What goes up must come down, so a signal must be derived from the VIDSTART signal - which indicates the start of video after a horizontal sync - that also indicates when video ends about 50 us after a horizontal sync. This derived signal is VOFF, which is low when there is video during the line interval, and is derived by gating VIDSTART with two signals and inverting the result. The first gating signal is LASTSEQ, which is low during the vertical blanking interval - this keeps VOFF high during that interval. The second gating signal is an inverted copy of the signal LASTLINPIX; the copy goes low after the proper number of pixels in a row have been counted, and so it pulls VOFF high during the horizontal blanking interval.

3.2.4 Horizontal pixel count

The two blocks at the top left of Figure 3.5 accomplish the pixel counting during the horizontal video interval. The bottommost counter is cleared upon a horizontal sync, and counting is enabled after the proper delay from the sync when VOFF goes low. After four ticks of SCLK, which corresponds to 1 microsecond and one pixel, the counter resets and the signal XINC0 goes high briefly. XINC0 is synchronized with the PHI1 and PHI2 signals so that it is a suitable gating signal for a PHI1-PHI2 pair. This gating occurs in the top left block of Figure 3.4, and is routed to the resistive grid chip.

The top block at left contains circuitry for counting horizontal pixels. A stack of a 4 bit and a 2 bit counter is clocked with the inverse is XINC0, and is enabled by VOFF. This 6

bit counter increments until a count of 46 is reached, and then it resets itself. Three signals are derived from the 6 bit counter outputs. The first, CNT5, goes high after 5 pixels. The second, CNT45, goes high after 45 pixels. The function of these two signals is described below. The third signal derived from the pixel counter, DUN, goes high after 46 pixels. An inverted copy of DUN, LASTLINPIX, pulls VOFF high after the last pixel of the resistive grid is written to and read from. DUN is also used to count lines, as described next.

3.2.5 Line count and Y increment

The resistive grid chip has 47 rows, and there are 262.5 lines in each field of a video frame. To utilize as much as possible of the video field, each row of the resistive grid is fed five lines of video - this gives 47 * 5 or 235 utilized lines. In the left block at the bottom of Figure 3.5, video lines are counted and signals are developed to increment the vertical (row) shift register of the resistive grid chip.

A 3 bit counter is used to count five video lines. This counter is cleared by a vertical sync signal, and is incremented by the end-of-line signal DUN. After a count of five is reached, the counter is reset. The signal LSETDONE goes high at the count of five (this signal is also called FS, which stands for "field sequence"). LSETDONE is synchronized with the two-phase shift register increment signals PHI1 and PHI2 by combination with the DUN signal from the pixel counter to produce the signal YINC0. This logic is shown in the block labeled Y STEP at the bottom right of Figure 3.5. YINC0 gates PHI1 and PHI2 to give the resistive grid chip vertical shift register increment signals VPHI1 and VPHI2. These signals advance a bit along a shift register.

The bit is initially placed in the first latch of the shift register upon vertical sync. The bit is present for latching in (the signal NEWYIN in the Y STEP block) after the vertical sync and up to the end of the first line of pixels. It is latched in by the vertical sync since YINC0 goes high in response to a vertical sync as well as a LSETDONE signal.

A stack of a 4 bit and a 2 bit counter in the bottom left block of Figure 3.5 forms a six

bit row counter. This counter is cleared by the vertical sync signal and is incremented by FS signal, and so counts rows of the resistive grid. Two signals are derived from the output of this counter. The first, ROW, is high for even rows of the resistive grid, and is used in the horizontal-to-video countdown as described above. The second, LASTSEQ, goes low after 47 rows are counted. This signal has two functions. First, it clears the vertical-to-video horizontal sync counter (top right block of Figure 3.4) as described above, so that this counter is reset before the vertical sync occurs. Second, LASTSEQ gates the VIDSTART signal in the horizontal-to-video countdown block, and thereby prevents VOFF from going low until after the next vertical sync.

3.2.6 Video blanking

The resistive grid chip does not process all of the video in a given line interval. Because it is a hexagonal grid, a direct display of the video processed by the resistive grid has side borders that are crenelated. This is annoying to many viewers, and so it is preferable to blank in such a way as to produce straight borders. A blanking signal is developed from the logic in the top right block of Figure 3.5. The blanking signal BLANKNOW, goes low at the beginning of the fifth pixel for odd rows and in the middle of the fifth pixel for even rows. The signal CNT2 marks the middle of a pixel and the signal CNT5 marks the fifth pixel. (The first few pixels are distorted by an artifact that arises in the resistive grid chip, and so they are blanked.) BLANKNOW goes high at the end of the 45th pixel for even rows and in the middle of the 45th pixel for odd rows. The signal CNT45 marks the 45th pixel.

3.2.7 Conclusion

In this chapter, a system is described that is capable of image smoothing at video rates. The degree of smoothing is continuously variable across a wide range.

This development is significant since previous analog smoothing chips have relied upon on-chip photoreceptors (which are noisy and have limited signal range) or upon an external

digital system followed by an digital-to-analog converter feeding a single on-chip sample and hold stage (which is slow). The system uses as its input a conventional video signal, and thus builds upon the many man-years of engineering that have gone into building reliable video cameras. Thus, analog VLSI has progressed to the point at which it can take advantage of the highly developed infrastructure that comprises commercial video systems.

Image smoothing is but one of the applications that can build upon the sample-and-hold strategy embodied in this system. Other applications that have been developed using analog VLSI, such as image segmentation [Harris et al., 1989] and motion detection [Tanner and Mead, 1987] could be furthered by incorporating this or a similar design.

Chapter 4

Non-linear Spatial Filtering and Image Tone Reproduction

As described in chapter 2, there are several solutions to the problem of range mismatch in imaging. All of the solutions that are used widely in industry are compromises: either contrast is lost in a range scaling step, or artifacts are introduced near high-contrast edges in a filtering step. Non-linear filtering is necessary to avoid artifacts, but in the past this has required human participation in the image reproduction process and therefore has not been used widely. In this chapter, an electronic imaging system is described in which non-linear filtering is used in image range compression automatically.

First, a means for nonlinear spatial filtering is described. Then a system is built up, step by step, that uses the nonlinear filter. Scale transformations and gain control are incorporated, and crucial system parameters are obtained through the analysis of the physics of images and image reproduction media. Lastly, examples of image range compression with a working model of the system are discussed.

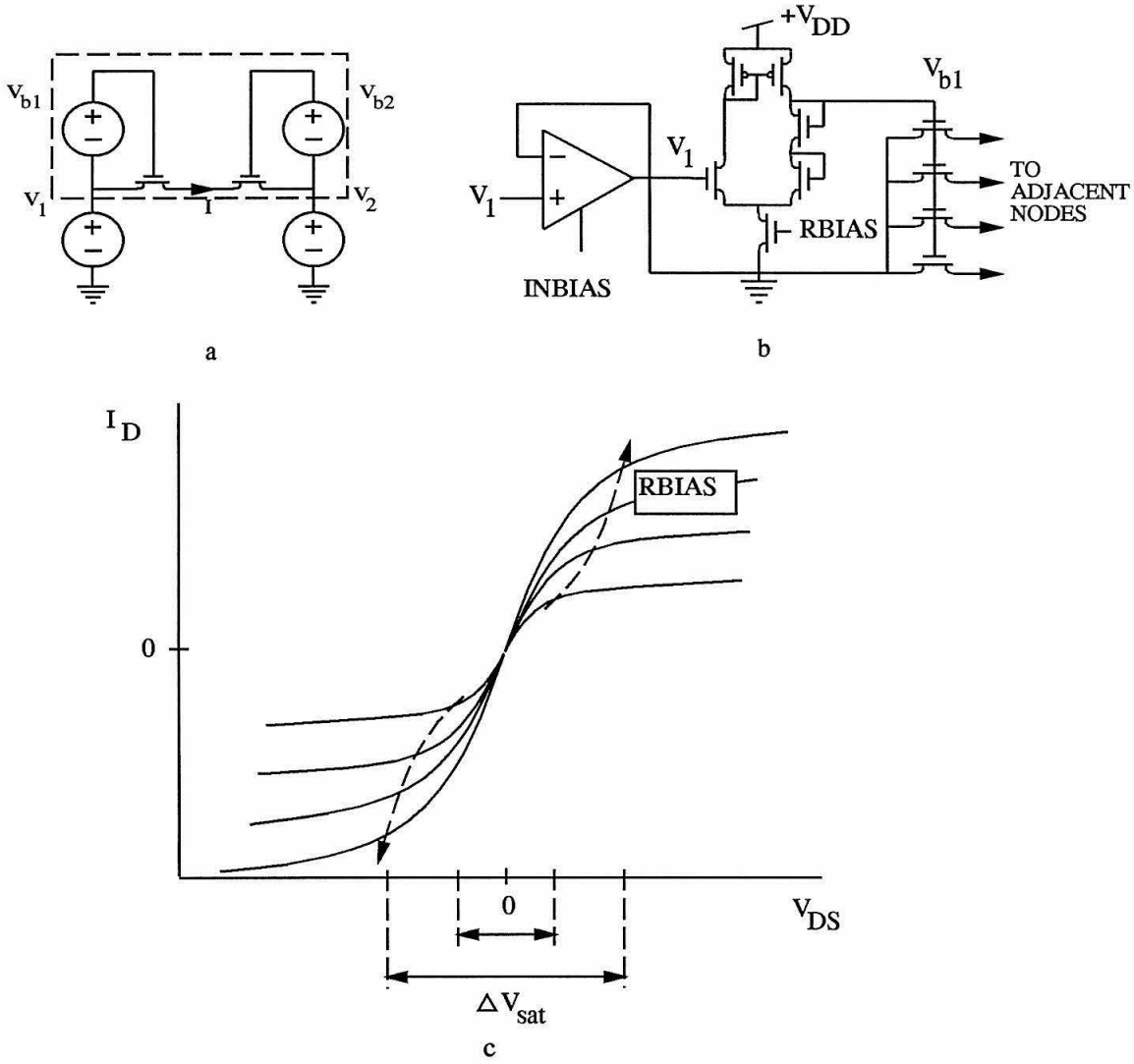


Figure 4.1: The saturating resistor. Resistors are formed from a pair of FETs in series (a). Current variation with FET source voltage changes (the "back-gate effect") are prevented by biasing each FET relative to its source voltage. The bias level V_b is set by the RBIAS control of the biasing circuit (b), a follower with an extra diode in its output arm. The I-V characteristic of the resistor (c) is linear in some region about the $I_D = 0$, $V_{DS} = 0$ point, as expected of a resistor, but the current saturates as V_{DS} increases. The saturation voltage is a function of RBIAS. Adapted from Mead [1989].

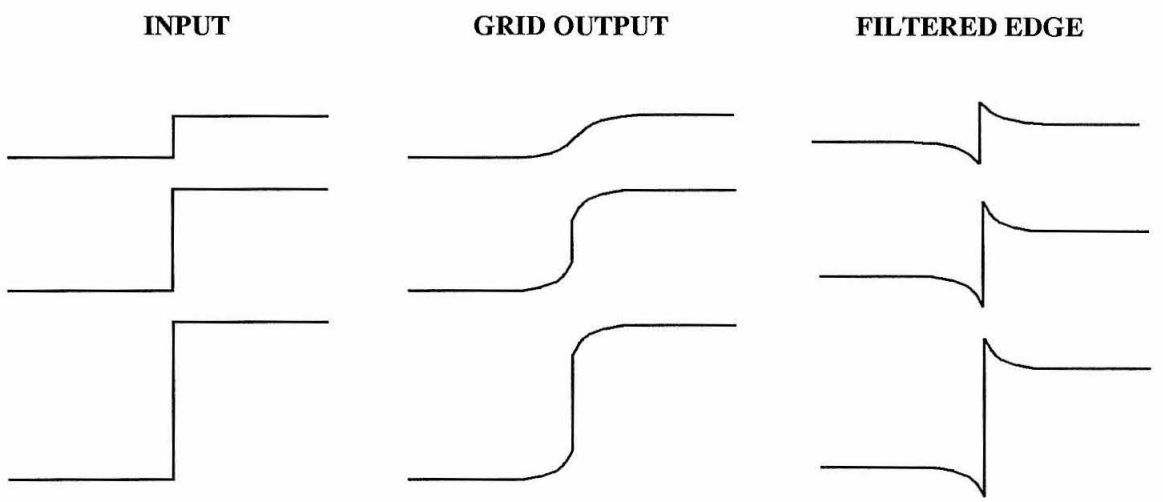


Figure 4.2: Edge enhancement and edge preservation by a saturating resistive grid. At top left is a small intensity step. The resistive grid smooths across this step (top middle). Suppose that the grid output is scaled down by a factor of one half and the result is subtracted from the input. This yields a high-pass filtered step (top right). At middle left is a step large enough to saturate the grid. The grid preserves about half of the step (middle) and the same scaling and subtraction of the grid output from the original gives a step with small overshoots on either side of the step. For an even larger step (bottom left) the grid preserves most of the step, and after scaling and subtraction, the size of the overshoots of the filtered step is even smaller relative to the filtered step size.

4.1 Non-linear spatial filtering with a saturable resistive grid

An electronic substrate for nonlinear spatial filtering was first described by Mead [1989]. The resistive grid built by Mead and coworkers has as its lateral, diffusive elements resistors built from transistors. The transistor gates are held at a constant voltage relative to the potential of the diffusing medium passing through them (the source voltage). Figure 4.1*a* shows the resistor circuit, Figure 4.1*b* details the biasing circuitry at each node, and Figure 4.1*c* shows the current-voltage characteristic of a FET with such a biasing arrangement. The abscissa of the plot in Figure 4.1*c* is the voltage across the transistor. As the current through the transistor increases, the voltage across it increases, up to a certain point. At some point, depending on the gate-to-source bias voltage, the current through the device is limited by channel saturation and the drain-to-source voltage across the device increases essentially independently of the current through it. The device becomes a current source, or, equivalently, a current limiter.

The saturating nature of the circuit is illustrated in Figure 4.2. For a small step input to the saturating resistive grid (top curves), the grid smooths across the step. For a larger step (middle curves), with a height that exceeds the saturation range of the lateral elements, the step is not smoothed across completely. The grid output dips down on the positive side of the step and up on the negative side; current from the resistor at the edge, which is saturated, pulls the step down to the right and up to the left. For a still larger step (bottom curves), the same dips on either side of the step are produced by the grid, but they represent a smaller percentage of the step. The dip size is fixed by the magnitude of the current through the saturated resistor, the resistance of the unsaturated resistors adjacent to the step, and the conductance of the input amplifiers of the resistive grid. Thus as a step input to the saturating grid is increased in height beyond the point of grid saturation, the dip size remains constant.

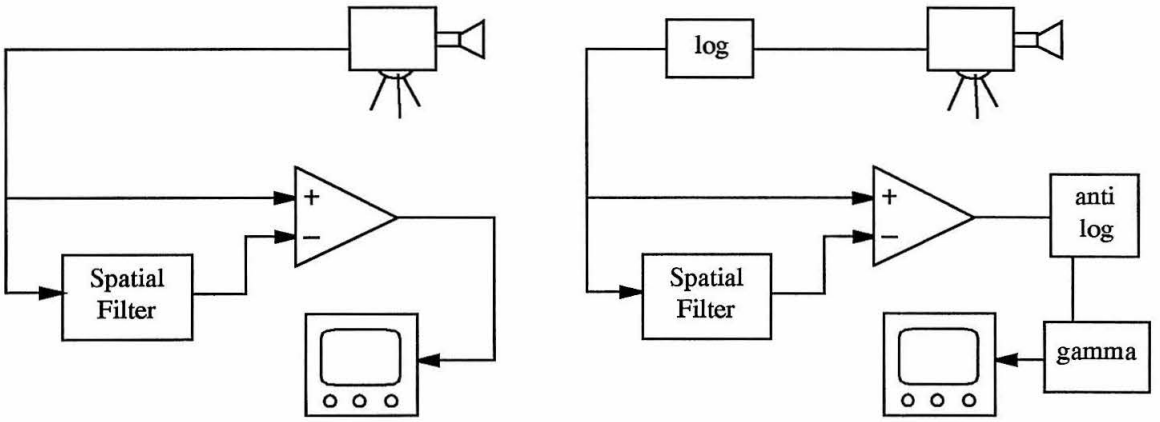
4.2 A Masking System with Nonlinear Spatial Filtering

A minimal system for range compression that incorporates a saturating resistive grid is shown in Figure 4.3*a*. The camera signal is fed to a saturating resistive grid. The grid may filter the image linearly or nonlinearly, depending on the amplitude of steps in the image. The filtered result is subtracted from the camera image for display.

4.2.1 Log, antilog, and gamma transformations

Intelligent masking requires a consideration of the physics of image formation, and its effect on image range. As discussed in chapter 1, the reflectance of naturally occurring surfaces vary over a limited range, but the lighting of a scene may vary over a much wider range. Any masking system should selectively diminish the range of reproduced illumination but preserve scene reflectances as much as possible. Since image intensity is the product of scene illumination and the reflectance of objects in a scene, the logarithm of the intensity is the sum of illumination and reflectance (neglecting geometric effects). The logarithm provides a way to separate illumination and reflectance in a masking system, and so a simple log block is added to the system immediately after the camera in the system (Figure 4.3*b*).

Direct display of the masked image is not desirable for two reasons. First, display media seldom have a linear input-output characteristic. For example, the cathode ray tubes used in the vast majority of electronic tone reproduction systems have a transfer characteristic (from input voltage to output luminance) that is approximately a square law. Predistortion of the image tone scale to account for the display characteristic is known as *gamma correction* in electronic imaging [Maloff, 1939; Kallman, 1940; Fink, 1941]. A system block for this predistortion is included in Figure 4.3*b*. Second, display of an image that has been transformed to a logarithmic scale is not pleasing visually. An antilog block is included before gamma correction in Figure 4.3*b*. These two blocks could be combined in practice.



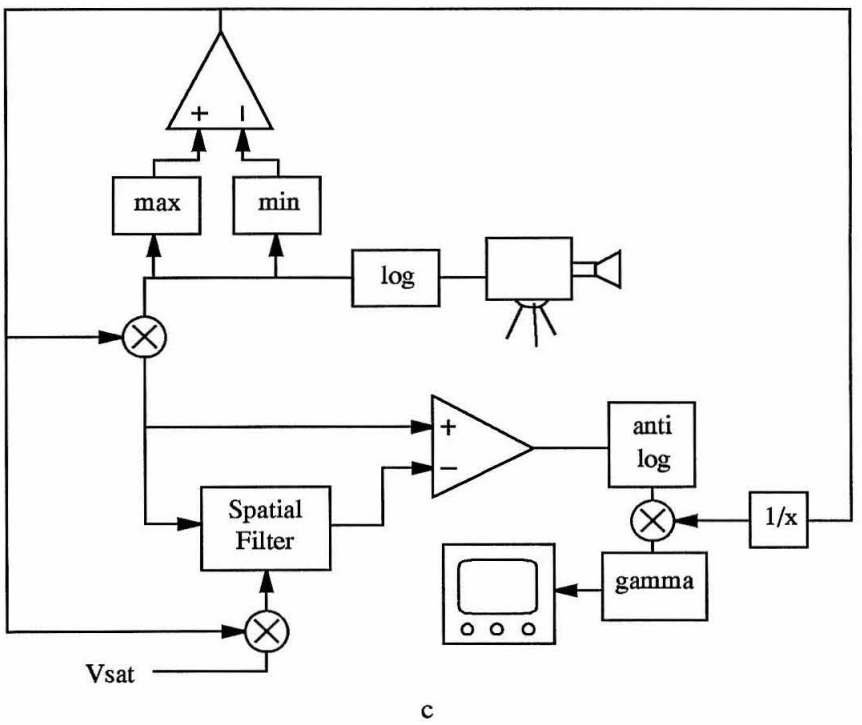


Figure 4.3: Progressively more sophisticated masking systems. In the most rudimentary system (*a*) the mask is simply subtracted from the input to form the output. Scale transformation to a log scale allows masking to work in concert with the physics of image formation, and antilog and gamma scale corrections optimize the image for display (*b*). In system (*c*), the input range is measured and used to scale the signal to match the mask input range and saturation level; this scaling is reversed before gamma correction and display.

4.2.2 Resistive grid parameters

The input signal to the masking chip must, of course, be within the range of the input amplifiers of the resistive grid. The logarithmically transformed camera signal must be scaled so that it never exceeds this range. This block, and a complementary block to rescale the filtered image to conform to the camera scale, are introduced into the system in Figure 4.3*c*.

More importantly, the saturation characteristics of the resistive grid must be set. The typical range of reflectances in an image determines this parameter; the range is taken as a factor of twenty. For example, if the noise level of the image signal input to the resistive grid is 5 mV, the lateral resistors should be set to saturate for steps greater than 100 mV. This parameter input to the masking chip is labeled V_{sat} in Figure 4.3*c*.

4.2.3 Gain control

In a tone reproduction system, the maximum possible ranges of the image capture medium and of the display medium are always known. For example, in a video imaging system the CRT input voltage must not exceed 1 volt. This sets the maximum level of the signal input to the gamma correction block of the system under consideration, and in turn sets the maximum signal level input to the antilog block of the system. Let us assume that the maximum allowable voltage input to the antilog block is also 1 volt (which is possible since the antilog is expansive but the gamma correction is compressive). This means that the filtered image signal which is subtracted from the camera signal should be scaled so that the masked signal (after subtraction) is one volt or less.

The system block diagram in Figure 4.4 includes a gain control scheme to keep the masked signal within the allowable range. The maximum of the camera signal is found with a peak detection block, and is compared with the known maximum allowable level (taken here as 1 V). The maximum and minimum of the masked signal (the output of the saturating resistive grid) are also found. The amplitude (max minus min) of the filtered

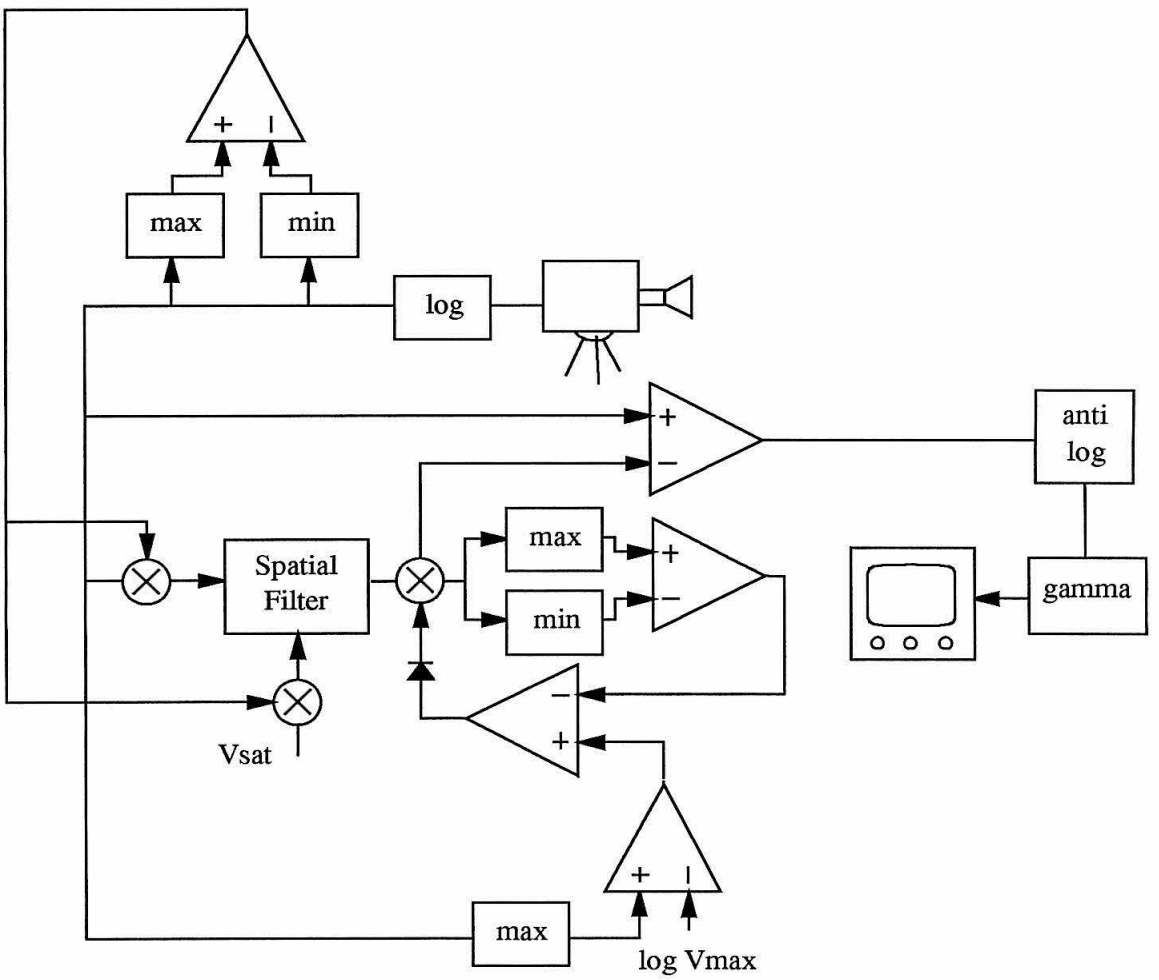


Figure 4.4: Gain control of the mask output. The range *max* - *min* of the mask output is scaled up to match the amount by which the input max exceeds a predetermined level V_{max}.

image is compared with the value by which the original image signal exceeds the allowed maximum, and the result of this comparison is used to servo the scaling of the image signal. With this feedback loop, gain control is accomplished that always assures that the masked signal range is within the range of the image display device.

It may seem at first that only the maximum of the resistive grid output is needed to characterize its range, but the minimum is needed too. As Harris [1991] noted, a saturating resistive grid always compresses a step input about its mean. The collective filtering of the grid not only "pulls down" the upper portion of the step - it also "pulls up" the lower portion. This compression about the mean is image dependent. An ideal one-dimensional step with equal image area on either side of the step is compressed equally toward the mean on both sides of the step. However, if the area of the upper portion of a one-dimensional step is much greater than the area of the lower portion, the compression toward the mean is greater for the lower portion. In general, the compression of an image region depends on its area and the perimeter around it for which the lateral resistors are saturated: a small area with a relatively large perimeter, such as a single noise or "outlier" point is maximally compressed [Harris, 1991]. Given this compression toward the mean, neither the maximum nor the minimum of the filtered image may be known in advance and so both must be measured.

4.2.4 DC restoration

Since the resistive grid filtering process produces compression toward the mean, the DC level of the filtered image, with reference to the original, is lost in the process. Restoration of the DC level of the filtered image must be performed because of this.

Figure 4.5 shows the correct DC restoration. The minimum of the image to be subtracted from the original is set to equal the black level of the original image. For black-and-white images, this is not strictly necessary. It may be desirable to lighten the dark regions of an image, for example - in this case the minimum of the filtered image should be set above the

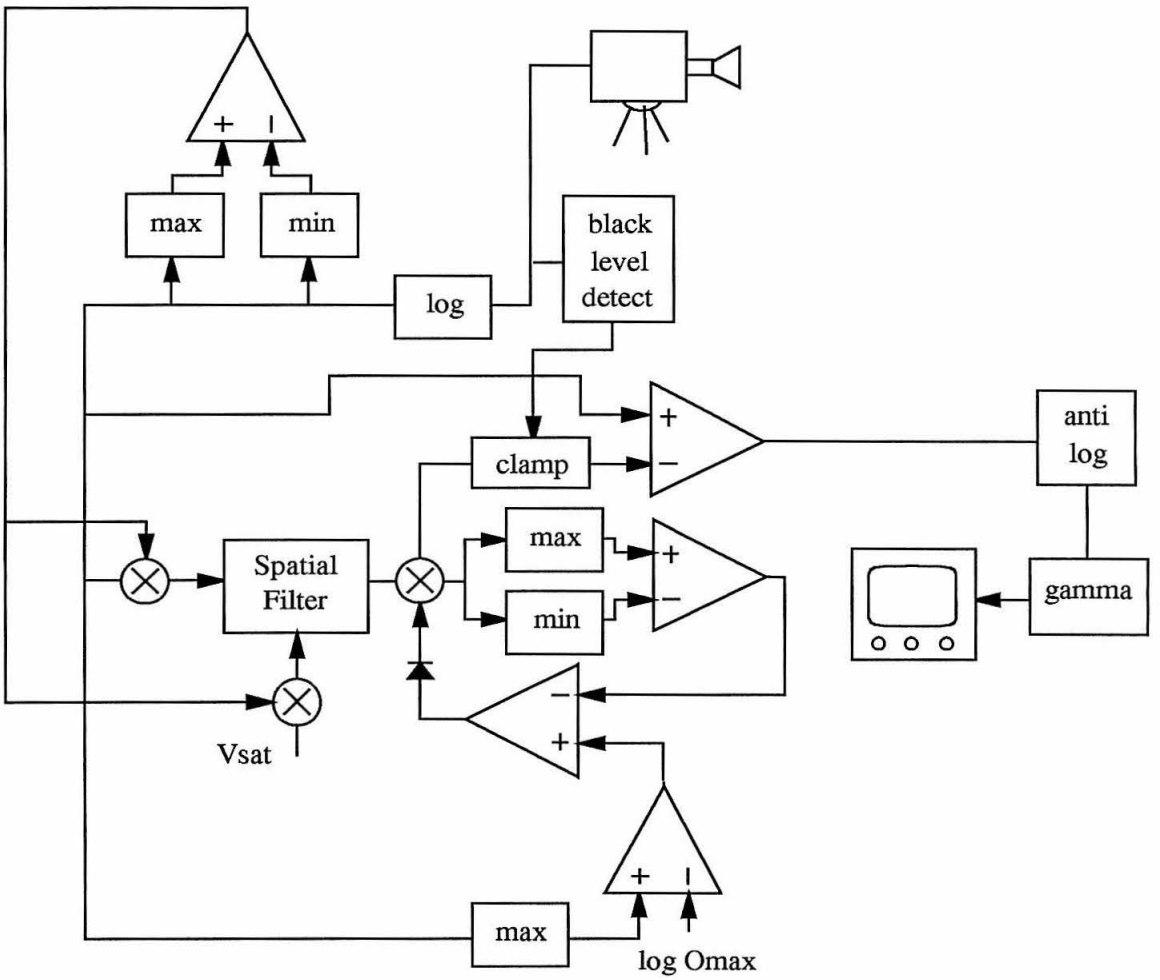


Figure 4.5: Optimal masking system. In addition to proper scale transformations and gain control, DC restoration is realized with a black level detector and a clamp circuit.

black level of the original. Setting the mean of the filtered image to the mean of the original would accomplish this rather simply. For colored images, however, the scheme shown in Figure 4.5 is necessary. Any other DC restoration procedure would introduce artifacts in the displayed image. For example, setting the means of the original and filtered images equal would produce a distortion called the *Helson-Judd effect* (see chapters 5 and 6). Consider a sequence in which a red truck drives into the view of the camera in such a system. As the red area increased over several frames, the rest of the scene would get greener and greener; this is because green is the complement of red. Setting the filtered image minimum to the original black level corresponds to normalization to white (or to the maximum in each color dimension).

The function of the final system shown in Figure 4.5 then may be described concisely as follows. The amount by which an image exceeds the range of the display is measured, and the range of the image is reduced by that amount.

4.3 Results of Nonlinear Spatial Filtering in Image Masking

The block diagram of a prototype electronic masking system for image range compression is shown in Figure 4.6. The log and antilog blocks of the ideal system were omitted, and the image signal was taken from a camera with built-in gamma correction circuitry. Only rudimentary gain control and DC restoration was used. Primitive as it is compared with the ideal system shown in Figure 4.5, the advantages of nonlinear masking are easy to demonstrate with this system.

Plate 1 illustrates the power of image masking with a saturable resistive grid. At top is a challenging image captured with a video camera at two settings of the camera iris. In the field of view is a window and a set of stripes and squares against an office wall below the window. With the camera iris open wide (Plate 1*a*), the indoor scene is visible, but the scene outside the window is awash in white. At a low iris setting, the house outside the window, lit with direct sunlight, is visible, but the indoor scene, lit with indirect light,

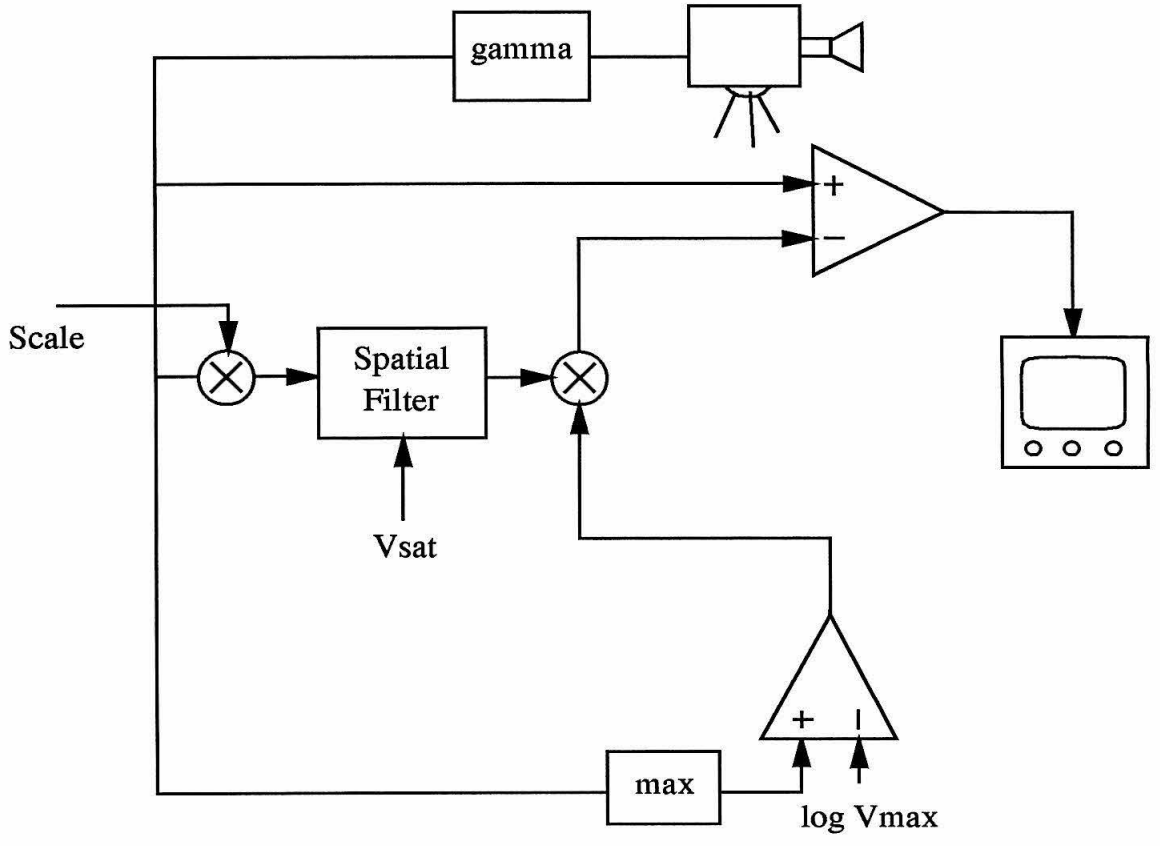


Figure 4.6: System used in this study. Mask input signal scaling and DC level correction (not shown) are available, but not automatic. A feedforward gain control circuit scales the mask output so that the final result is always within the range of the display.

is rendered black on a video monitor (Plate 1*b*). The light reaching the camera from the house is about 200 times greater than the light from the indoor scene (measured with a Tektronix J6512 photometer).

With the camera iris set in between the two extremes shown in the top row of Plate 1, the output of the resistive grid chip is shown in the second row of Plate 1. At left (Plate 1*c*), the resistive grid is configured as a linear low-pass spatial filter, and its output is a smoothly varying transition from bright to dark; the chip is smoothing across the step of intensity. At a different parameter setting (Plate 1*d*), the resistive grid smooths at the top and bottom of the image, but not across the light-to-dark step.

Subtraction of the resistive grid output from the original image gives the results in the image shown at the bottom of Plate 1. The linearly masked image (Plate 1*e*), for the most part, does not exceed the camera range. Detail is visible on the video monitor at both light levels - the house outside and the objects indoors can be seen simultaneously at a single camera iris setting. However, near the windowsill, the familiar overshoot distortion is quite noticeable. A dark band extends into the indoor scene below the windowsill, and a light band extends into the outdoor scene above it. However, if a nonlinear mask is subtracted from the image, as shown in Plate 1*f*, the result is much more pleasing. Both the indoor and outdoor scenes are visible without the overshoots near the windowsill.

Since, in the linear mode of operation, the resistive grid performs a blurring operation, the result at bottom left corresponds to photographic and electronic unsharp masking. In the nonlinear mode of operation, the resistive grid not only blurs - it also segments or dodges. At such, the result at bottom right corresponds to electronic dodging, and represents the first time that automatic dodging has been achieved at video rates completely external to the video camera.

Plate 2 shows three examples of gain control. At left are images of three scenes with the iris set low. No part of these images is out of range, but important parts are lost in shadow (the books in *a* and *d* and the lab bench in *g*, for example). To pull these objects out of

shadow, the iris must be opened further. But this results in images that are out of range (middle column) because of the lights in b and e and the glare of sunlight off the whiteboard in h. With gain control (Figure 4.6), enough of the mask is subtracted to keep the images in range when the iris is opened. The right column of images are the output of the masking system, with the middle column of images as inputs. In Plate 2c, the books that are lost in shadow in a and the whiteboard that is lost in "glare" in b are all visible. The images in Plate 2f and i also reproduce a wider range of tones than is normally possible.

4.4 Analysis of Illuminant Reduction by Masking

With a nonlinear masking system for range compression, the effects of illumination are diminished while image reflectances are preserved. In this section the process is described mathematically.

For an image signal S that is linear in image luminance, the signal is the product of the reflectance R in a scene and the scene illumination I:

$$S = RI. \tag{4.1}$$

Forming the log of S, the product becomes a sum:

$$\log S = \log R + \log I. \tag{4.2}$$

This signal is fed to a saturating resistive grid; the saturation may be described by a threshold. This filtered image is scaled by a value G that is determined by the gain control subsystem, and the result is subtracted from the original to yield a signal $\log O$:

$$\begin{aligned} \log O &= \log S - T\left[\log S\right] \\ &= \log R + \log I - G \cdot T\left[\log R + \log I\right]. \end{aligned} \tag{4.3}$$

Since the threshold process preserves the illumination component only, this becomes

$$\begin{aligned}
\log O &= \log R + \log I - G \cdot T\left[\log I\right] \\
&= \log R + \log I - G \cdot \log I \\
&= \log R + (1 - G) \cdot \log I \\
&= \log R + \log \left[I^{1-G}\right] \\
&= \log \left[RI^{1-G}\right]. \qquad (4.4)
\end{aligned}$$

Taking the antilog, the output signal O is obtained,

$$O = RI^{1-G}, \qquad (4.5)$$

and so the illumination is reduced according to a power law relationship.

Range compression takes place only if the camera signal S exceeds the maximum output signal. This maximum, V_{max}, is compared with $\log S$ (Figure 4.5), and, neglecting signal compression by the saturating resistive grid, the multiplier G of the filtered signal is equal to the amount by which $\log S$ exceeds V_{max}.

$$G = \log S - V_{max}, \quad \log S - V_{max} > 0. \qquad (4.6)$$

V_{max} is the logarithm of the true output range maximum O_{max}, and so

$$G = \log S - \log O_{max}, \quad \log S - \log O_{max} > 0. \qquad (4.7)$$

The typical range of a CRT is a factor of 50, and in a typical scene the reflectance spans a factor of twenty in range. This means that the maximum allowable range of illumination

is about a factor of two.

$$\begin{aligned} \log R + \log I &\leq \log O_{max} \\ \log I &\leq \log O_{max} - \log R \\ \log I &\leq \log (50/20) \\ I &\leq 2.5. \end{aligned} \tag{4.8}$$

In Equation 4.7 $G = 0$ means that the range of I is small enough that the inequality in Equation 4.8 is satisfied, and $G = 1$ means that that range of I and R combined exactly match the range of the output device, O_{max}.

4.5 Filtering and Segmenting Independently with a Resistive Grid

Spatial filtering is used for two reasons in imaging systems. First, filtering is carried out to offset the adverse effects of image tone scaling. Second, filtering is carried out to compensate for resolution lost in camera optics and/or in the display step. This is called *aperture correction.* A resistive grid can perform spatial filtering operations to either end, independently of segmentation for nonlinear masking.

There are three parameters that determine how a resistive grid processes a signal: the input signal range and the input amplifier and lateral resistor i-v characteristics. It is always desirable to keep the input signal within the range of the input amplifiers. However, the range of the lateral elements relative to that of the input signal determines whether segmentation will occur at a large step, and the degree of smoothing is determined by the transconductance of the input amplifiers relative to the conductance of the lateral resistive elements. From the design standpoint, the input range of the resistive grid should be as wide as possible, and so the input characteristic is fixed in a practical sense. The lateral

resistance and the saturation level of the resistors vary together as the bias of the horizontal resistance circuit is changed (Figure 4.1).

Thus an arbitrary amount of filtering is available by varying the lateral resistance relative to the input resistance. Once the lateral resistance is set, the lateral range is known, and the input signal range can be varied to produce segmentation at any desired step size.

4.6 Conclusion

In this chapter, a nonlinear spatial filtering method was described and applied to the problem of image range compression. Scale transformations and gain control principles necessary for this application were discussed. In chapter 7, this system is refined further by applying the psychophysical principles discussed in chapter 1.

This thesis now turns from the topic of tone reproduction in black-and-white to a consideration of spatial filtering in color science and engineering.

Chapter 5

Color Constancy Theories and Algorithms

5.1 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 due to varying cloud cover, etc., we do rather well.

While the problem of color constancy, to be described below, has been recognized for some time (Helmholtz commented on it [Helmholtz, 1911]), the computational essence of the problem has been grappled with only recently. Land thrust the problem on the scientific world with startling demonstrations 30 years ago [Land, 1959], and has championed one school of thought on the topic since. In the last decade, two other schools of thought

have emerged. Buchsbaum [Buchsbaum, 1980] and Maloney and Wandell [Maloney and Wandell, 1986] have applied formal mathematical techniques to the problem; they offer two alternatives to the formulation promulgated by Land and others. These three approaches utilize different assumptions to simplify the problem and render it tractable.

In the following, the neurobiological and computational aspects of the problem are described. Next various theories of color constancy are described, and three classes of color constancy algorithms are compared. In the following chapters, Land's retinex theory and algorithm are explored through computer simulation and a hardware implementation, and the results are applied to the practice of tone reproduction.

5.2 A sketch of the neurobiology of color vision

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 to the CNS. In this section the main points of the neurobiology of color vision are reviewed.

5.2.1 Retina and LGN

We sense light with three classes of receptors, the cones (rod vision is not considered here). Turtle cones respond to light with a voltage that is, to a first approximation, proportional to the log of the intensity of light incident upon them, normalized to the time averaged intensity in the immediate past. In other words, turtle cones *adapt* to the mean light level (as discussed in chapter 1) [Normann and Pearlman, 1979]. Whether primate cones share this property exactly is as yet controversial; they certainly adapt, but whether the slope at each adaptation level is identical has not been determined decisively [Shapley and Enroth-Cugell, 1984].

The three classes of cones have different spectral bandpass properties. They are called long, medium, and short (due to the spectral bands that they are sensitive to) or red, green,

and blue colloquially. At the level of retinal ganglion cells, the image has been transformed from a single cone array into two arrays, called parvo and magno. The receptive fields of retinal ganglion cells are organized in a center-surround fashion, with a surround somewhat larger than the center; this receptive field organization resembles a difference of Gaussians. The parvo receptive fields are 3-5 times smaller in radius than the magno receptive fields, and are chromatically *opponent*. That is, these cells receive center and surround inputs from different cone classes; these cones can adapt independently, and thus the center and surround strengths are independent. The opponencies formed are commonly called red-green (long-medium) and yellow-blue (long+medium/short).

It is well known that an efficient method of representing images is to encode in terms of opponencies. Principle component analysis has shown that most of the information in an image is along the black-white axis, less is along the red-green axis, and only a few percent is along the yellow-blue axis [Lennie and D'Zmura, 1988]. The source of the black-white axis in the CNS is still controversial. The magno pathway is largely color-blind and low-resolution, yet it is known that achromatic acuity is better than chromatic acuity. The parvo stream still contains information about black and white, which can be recovered by appropriate linear combinations of opponents. So, it may be that opponency is a trick used in the parvo stream to multiplex color information and achromatic information in the same high-resolution stream.

5.2.2 Striate cortex

At the level of cortex, color is processed in two different streams [DeYoe and Van Essen, 1988; Livingstone and Hubel, 1984]. (I am discussing diurnal primates here; nocturnal primates have poor color vision but still have the streams discussed below.) The parvo-blob stream seems to be concerned with representing color and low spatial frequencies. The receptive fields of blob cells are in a center-surround configuration, as in the retina and LGN, but the receptive fields are larger, and the opponencies are more complicated. Rather

than having a red center and a green surround, for example, these cells can have centers and surrounds that are sensitive to both red and green. There is some recent evidence that the blobs are divided into two classes, one that processes red-green information and another that processes yellow-blue information, and that there is communication amongst blobs of the same type [Ts'o and Gilbert, 1988].

The parvo-interblob stream is jointly selective for spectral wavelength, binocular disparity and orientation. The receptive fields are elongated and are strongly excited by short bars [DeYoe and Van Essen, 1988].

5.2.3 Area V4

This extrastriate visual area receives projections from all three streams (parvo-blob, parvo-interblob, and magno). In early investigations it was dubbed the "color area" [Zeki, 1980] in that 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 disparity [DeYoe and Van Essen, 1988]. In this discussion, only the spectral properties of V4 cells are presented. Most interestingly, they seem to be responsive to perceived color, rather than wavelength; that is, they are "color constant" according Zeki's study [Zeki, 1983a, 1983b]. 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 [Desimone et al., 1985; Schein and Desimone, 1990] obtained results from V4 cells that are in a sense supportive of Zeki's observations. In a study of the "non-classical" receptive fields of extrastriate visual neurons, they found that V4 cells:

1. are rarely color opponent,

2. respond to white, and to many wavelengths, but have a maximal response at some wavelength, "analogous to a broad-band color filter, such as a piece of colored glass,"

3. are suppressed by stimuli in a large (30 deg 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.

This last property is quite like that expected from Land's retinex theory: the suppressive surround of the neuron corresponds to the large surround subtracted from each image pixel in the algorithm. By comparing the color in the center of the receptive field to the color in a large area outside of the center, V4 cells judge *relative color*. This is presumably one 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.

Finally, one group reports that monkeys who have had V4 removed surgically show little or no impairment in wavelength discrimination, but perform poorly on tasks that require color constancy [Wild et al., 1985].

5.2.4 Summary: spectral processing in the CNS

From this broad review of the neurobiology, it is evident that there are several stages of color processing in the CNS:

1. The retinal cones adapt to the mean level of the illuminant.

2. Color opponencies are formed in the retina; thus retinal ganglion cells (and LGN cells) are excited by a narrower band than cones.

3. A more complicated form of opponency is evident in striate cortex. Wavelength sensitivity is present in two different, parallel streams.

4. Opponency is rare in V4, but cells in this visual area are perturbed by stimuli in a huge "non-classical" surround.

A rather condensed list of what we do not know about color vision is:

1. The spatial and temporal characteristics of cone adaptation (i.e., how the mean illuminant is obtained) is not understood.

2. Whether adaptation or a "silent surround" suppression of spectral response exists for cells in the LGN, V1, or extrastriate areas leading to V4 is not known but would not be surprising.

The two known adaptive mechanisms (at the receptor level and in V4) and perhaps some unknown ones are presumably essential in solving the problem of color constancy, the subject of the next section.

5.3 Theories of Color Constancy

5.3.1 Psychophysical foundations

Color constancy theories are all based on a few basic visual effects: color adaptation, successive contrast, simultaneous contrast, and degenerate (aperture) color phenomena.

Color adaptation

The colors of objects do not change much even though the spectral content of natural illuminants varies widely. Objects look the same to us at noon, under a bluish illumination from the sky, and at sunset, under a rather more reddish illumination. Just as the eye

adapts to changes in the intensity of illumination, as discussed in chapter 1, it adapts to spectral variation. The main points of the black-and-white discussion in chapter 1 apply here: we are aware of the changes in illumination but can compensate for it rather well using relative judgement and cognitive cues.

Successive contrast

Color adaptation is a process that occurs over time, and persists over time. If you stare at a green card for a couple of minutes and then look at a white wall, you will see a red patch. This afterimage of the complementary color of a previous adapting stimulus will fade slowly over time.

Simultaneous contrast

A grey piece of paper on a blue background looks slightly yellower than the same grey paper on a neutral background. Yellow is the complementary color of blue, and so the added color in the former situation arises from contrast. In general, an object color is distorted in our perception by the color of its surroundings.

Aperture phenomena

Suppose that you are standing in a field and looking at a red barn. The barn looks red. If you walk up very close to the barn, it still looks red, even though the barn occupies the large proportion of your visual field. Now suppose that your eyes are completely covered with white, semi-transparent surfaces (such as split ping-pong balls) and red light is shone onto the surfaces. In a short time, the redness will diminish, and when adaptation is complete, the surface will not look colored at all.

How can we reconcile these two perceptions? In both cases, mostly red light reaches the eye - but in one case the eye completely adapts to the red light and in the other it does not.

Consider another pair of examples. Suppose that you are looking at an array of neutral papers, from dark grey to white, in a normal everyday viewing situation. If red light is shone

on the papers, but not on the rest of the room you are in, your perception of the papers will not change much. You might say, "Well, they are redder, but they are still neutral." Now suppose that you are in a completely light-tight room, and the only thing visible besides the blank walls of the room is that array of grey papers. If the array is viewed under red illumination, with middle-grey walls as a background, the papers with higher reflectance will look reddish (taking on the illuminant color) and the papers with reflectance lower than the background will look greenish (taking on the complementary color of the illuminant) [Helson, 1938]. This is called the *Helson-Judd effect.* In the light-tight room, the visual system seems to lose the ability to normalize the color sensation correctly.

Again, two different viewing conditions produce two different percepts for the same stimulus. In normal viewing, the color of a portion of the visual field is judged in the context of the entire visual experience. This judgement is largely preattentive - automatic and fast. (One study found that humans make reflectance judgements well even when a scene is viewed for only a fraction of a second [Land and Daw, 1962].) When a stimulus is viewed without a context, strange things happen. Psychophysicists commonly remove context by having a subject view a stimulus through a dark tube or other insulating aperture; these strange effects are called *aperture phenomena.* Simultaneous contrast effects are noticeable, but small. Aperture effects, on the other hand, can be quite strong.

5.3.2 Color constancy theories

Most theories of color constancy attempt to explain the major psychophysical effects discussed above (for reviews see [D'Zmura and Lennie, 1986; Maloney, 1984; Buchsbaum, 1980; Jameson and Hurvich, 1989; Helson, 1943; Evans, Hanson, and Brewer, 1953]. Jameson and Hurvich's theory is representative [Hurvich, 1981; Jameson and Hurvich, 1989]. First, they propose that the receptors adapt multiplicatively so that sensitivity is lowered at higher illumination levels. This is called *von Kries adaptation* after the first proponent of receptor sensitivity scaling. They assume that receptor adaptation results from spatial sampling and

temporal integration. As the eye roams from place to place in a scene, adaptation occurs over time, determined not only by the color of the most recent fixation point but also by the many colors of previous fixation points, over several seconds and minutes. Secondly, Jameson and Hurvich postulate a postreceptoral differencing stage in which color opponencies are formed. The differencing has both spatial and temporal characteristics; they assume that adaptative processes result from this differencing that operate in addition to von Kries adaptation.

This model accounts for the three main visual effects described above. Multiplicative, Von Kries gain scaling explains most color adaptation. Lateral inhibition in the second stage explains simultaneous contrast. Temporal integration and differencing in the second stage explains successive contrast.

All of the aperture effects mentioned above can also be explained with a theory this complete. In particular, aperture effects arise from normal visual mechanisms operating in an abnormal situation. Von Kries scaling eliminates the color of whole-field stimuli, in the ping-pong ball "experiment," but the barn stays red in a normal viewing situation because of eye movements (occasional glances at the sky, etc.) and influences of stimuli in the visual periphery (such as the grass at one's feet and even the bridge of one's nose). Small as these "non-barn" aspects of the visual experience seem, they are sufficient to maintain relatively normal adaptation so that the barn never appears grey. Of course, this means that the temporal aspect of adaptation is fairly long lived (one does not glance at the sky often) and that the spatial computation involved in the adaptive process extends across most of the visual field.

The difference between the appearance of a set of neutral patches illuminated by red light in normal and aperture (light-tight room) situations is explained in the Jameson and Hurvich framework by the second stage color-opponent lateral inhibition mechanism. The receptor adaptation is set mainly by the background, which appears neutral. Papers whiter than the background produce excitation relative to the background, and so they appear

red. For papers darker than the background, the excitation from the receptors *is less than* the inhibition from the background, and so the sensation is grey minus red, or green. In the light-tight room situation with a strongly colored illumination, the effect is quite strong [Helson, 1938; Hurvich, 1982] in normal viewing situations the effect is quite small and not very noticeable. The Helson-Judd effect is important because it demonstrates the opponent stage of color processing, and also because it shows that normalization can go awry. It shows that the "reference" used by the nervous system in the aperture condition is the scene average, equated to grey. Whether middle grey is the adaptation reference color in everyday viewing conditions is not clear from this Helson-Judd experiment. As we shall see in the following discussion of algorithms, this question of normalization is crucial.

Almost all color scientists eschew the term *color constancy* and instead use terms such as *approximate color constancy* [Hurvich, 1981], *color compensation* [Helson, 1943], and even *color inconstancy* [Buchsbaum, 1980] ! Adaptation to the illuminant is seldom complete, so that we can always tell when the illumination changes in normal viewing situations. Still, receptor adaptation and lateral inhibition are effective in providing a good deal of compensation. These compensatory mechanisms continue to fascinate visual scientists; in turn, visual scientists continue to propose mechanistic theories that incorporate the latest developments in neurobiology and psychophysics.

Perhaps the most mechanistic kind of theory is the algorithmic theory. In the final part of this chapter, the problem of color constancy is formally defined and several algorithmic models are described and compared.

5.4 The Problem of Color Constancy

5.4.1 Overview

Under normal variations (such as different degrees of cloud cover and different times of day), the spectrum of daylight varies somewhat. The variation is limited enough that it can be represented with three spectral basis functions [Judd et al., 1964]. A wide range of

naturally occurring object reflectances can also be described with only three basis functions [Maloney, 1986]. The light reaching a point on the retina, i.e., the set of 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 [D'Zmura and Lennie, 1986; Blackwell and Buchsbaum, 1988; Arend and Reeves, 1986; Worthey, 1985; Ingle, 1985; Cornsweet, 1970]. *This is the computational problem of color constancy: how do we solve three equations in six unknowns ?* This conundrum is stated mathematically in this section.

5.4.2 Definitions

λ	wavelength (in the visible range, 400 nm to 700 nm)
$Q_k(\lambda)$	sensitivity of kth cone class
$q_k(\mathbf{x})$	quantum catch by kth cone type at location $\mathbf{x}$
$S(\mathbf{x}, \lambda)$	surface reflectance or albedo
$S_i(\lambda)$	reflectance basis function i
b_1	number of reflectance basis functions
$p_i(\mathbf{x})$	weight of ith reflectance basis function at location $\mathbf{x}$
$I(\mathbf{x}, \lambda)$	illuminant
$I_i(\lambda)$	illuminant basis function i
b_2	number of illuminant basis functions
$m_i(\mathbf{x})$	weight of ith illuminant basis function at location $\mathbf{x}$
$L(\mathbf{x}, \lambda)$	intensity of light reaching sensors

5.4.3 The intensity equation and the color signal

In the 3D world, the intensity of light reaching a point on a 2D array of sensors is:

$$L(\mathbf{x}, \lambda) = I(\mathbf{x}, \lambda) F(\mathbf{v}, \mathbf{s}, \mathbf{n}) S(\mathbf{x}, \lambda), \tag{5.1}$$

where $I(\mathbf{x}, \lambda)$ is the illuminant, $S(\mathbf{x}, \lambda)$ is the surface reflectance (albedo) of the object illuminated, and $F(\mathbf{v}, \mathbf{s}, \mathbf{n})$ is a viewing geometry factor [Horn, 1985]. The geometry factor takes into account the viewer direction $\mathbf{v}$, the source direction $\mathbf{s}$, and the surface normal $\mathbf{n}$. For simplicity we assume that $F(\mathbf{v}, \mathbf{s}, \mathbf{n}) \equiv 1$, i.e., the light source and viewer are perpendicular to the surface. With this simplification, the intensity equation becomes:

$$L(\mathbf{x}, \lambda) = I(\mathbf{x}, \lambda) S(\mathbf{x}, \lambda). \tag{5.2}$$

At the retina, this impinging light is captured by retinal cones, which we will assume come in three varieties. The *quantum catch* (i.e., the light flux captured) by a cone of the kth class is

$$q_k(\mathbf{x}) = \int Q_k(\lambda) L(\mathbf{x}, \lambda)\, d\lambda = \int Q_k(\lambda) I(\mathbf{x}, \lambda) S(\mathbf{x}, \lambda)\, d\lambda \tag{5.3}$$

where $Q_k(\lambda)$ is the spectral sensitivity of the kth cone type.

Both illuminants and reflectances can be described with a limited set of basis functions:

$$I(\lambda) = \sum_{i=1}^{b_1} m_i I_i(\lambda) \Rightarrow I(\mathbf{x}, \lambda) = \sum_{i=1}^{b_1} m_i(\mathbf{x}) I_i(\lambda) \tag{5.4}$$

$$S(\lambda) = \sum_{i=1}^{b_2} n_i S_i(\lambda) \Rightarrow S(\mathbf{x}, \lambda) = \sum_{i=1}^{b_2} p_i(\mathbf{x}) S_i(\lambda). \tag{5.5}$$

There is no *a priori* value for either b_1 or b_2. However, for naturally occurring variations in daylight $b_1 = 3$ is sufficient [Judd et al., 1964], and coincidentally three basis functions ($b_2 = 3$) also describe a large set of naturally occurring reflectances [Maloney, 1986]. Thus

we can rewrite Equation 5.3 as

$$q_k(\mathbf{x}) = \int Q_k(\lambda) \left(\sum_{i=1}^{3} m_i(\mathbf{x}) I_i(\lambda) \right) \left(\sum_{i=1}^{3} p_i(\mathbf{x}) S_i(\lambda) \right) d\lambda, \tag{5.6}$$

or,

$$q_k(\mathbf{x}) = \mathbf{m}^{\mathbf{T}}(\mathbf{x}) \mathbf{G}_k \mathbf{p}(\mathbf{x}), \tag{5.7}$$

where $\mathbf{G}_k$ is a $b_1 \times b_2$ (3×3) tensor,

$$\mathbf{G}_k = \begin{bmatrix} \int Q_k(\lambda) I_1(\lambda) S_1(\lambda)\, d\lambda & \int Q_k(\lambda) I_1(\lambda) S_2(\lambda)\, d\lambda & \int Q_k(\lambda) I_1(\lambda) S_3(\lambda)\, d\lambda \\ \int Q_k(\lambda) I_2(\lambda) S_1(\lambda)\, d\lambda & \int Q_k(\lambda) I_2(\lambda) S_2(\lambda)\, d\lambda & \int Q_k(\lambda) I_2(\lambda) S_3(\lambda)\, d\lambda \\ \int Q_k(\lambda) I_3(\lambda) S_1(\lambda)\, d\lambda & \int Q_k(\lambda) I_3(\lambda) S_2(\lambda)\, d\lambda & \int Q_k(\lambda) I_3(\lambda) S_3(\lambda)\, d\lambda \end{bmatrix}, \tag{5.8}$$

and $\mathbf{m}^{\mathbf{T}}(\mathbf{x}) = [m_1(\mathbf{x})\ m_2(\mathbf{x})\ m_3(\mathbf{x})]$ and $\mathbf{p}^{\mathbf{T}}(\mathbf{x}) = [p_1(\mathbf{x})\ p_2(\mathbf{x})\ p_3(\mathbf{x})]$. For three cone signals $\mathbf{q}^{\mathbf{T}}(\mathbf{x}) = [q_1(\mathbf{x})\ q_2(\mathbf{x})\ q_3(\mathbf{x})]$, so

$$\mathbf{q}(\mathbf{x}) = \mathbf{m}^{\mathbf{T}}(\mathbf{x}) \mathbf{G} \mathbf{p}(\mathbf{x}). \tag{5.9}$$

Here $\mathbf{G}$ is a $b_1 \times b_2 \times 3$ $(3 \times 3 \times 3)$ tensor, one component of which (for one cone class) is defined in Equation 5.8.

The aim of any color constancy algorithm is to recover the reflectance information at each location, $\mathbf{p}(\mathbf{x})$, from the cone quantum catches. Thus we have a tensor transformation from 6 unknowns to 3 knowns. Now that the computational problem of color constancy is stated we turn to ways of rendering Equation 5.9 tractable.

5.5 Three Approaches to the Problem

5.5.1 Overview: three sets of assumptions

Assumptions must be made to simplify the underdetermined system of Equation 5.9. The three schools of thought make different assumptions about the illuminant, about the mathematical properties of the illuminant and reflectance basis functions, and about the typical spatially averaged reflectance in a scene. Table 1 shows the various assumptions of the three approaches. For brevity, they are referred to as the Land approach, the Buchsbaum approach, and the Maloney approach, after the first person historically to champion a particular set of assumptions.

Table 1. Assumptions of three approaches to color constancy

	spatially constant illuminant	*separable* $I(x)$ *and* $S(x)$	*constant* $\overline{S(x)}$	*separate lightnesses channels*	$2D$ *reflectance*
Land		√	√	√	
Buchsbaum	√		√		
Maloney	√				√

These three approaches are discussed in turn below.

5.5.2 Land's retinex theory

Three separate lightnesses

What I am calling the Land school here includes Land and his colleagues at Polaroid and the Rowland Institute [Land, 1959; Land and McCann, 1971; McCann, 1987; McCann et al., 1976; Land, 1983; Land, 1986], Horn [Horn, 1974; Horn, 1985], and Hurlbert [Hurlbert, 1986; Hurlbert and Poggio, 1987]. 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 three *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.

First, it is assumed that the retinal cones output a lightness signal l that is a logarithm of the incoming flux.

$$l_k(\mathbf{x}) = \log q_k(\mathbf{x}) = \log \int Q_k(\lambda) L(\mathbf{x}, \lambda)\, d\lambda = \log \int Q_k(\lambda) I(\mathbf{x}, \lambda) S(\mathbf{x}, \lambda)\, d\lambda \tag{5.10}$$

Second, the three lightness signals are assumed to be independent; Land did not state how this may occur, but Hurlbert and Poggio derive it as follows [Hurlbert and Poggio, 1987]. A set of illuminant and reflectance basis functions are assumed (Equation 5.4 and Equation 5.5). Further, the illuminant and reflectance basis functions are assumed to be identical ($S_i = I_i$) and orthogonal with respect to the cone sensitivity functions. With these assumptions the tensor of Equation 5.9 collapses to a diagonal matrix:

$$l_k(\mathbf{x}) = \log q_k(\mathbf{x}) = \log \mathbf{m}^{\mathbf{T}}(\mathbf{x}) \mathbf{G}_k \mathbf{p}(\mathbf{x}) \tag{5.11}$$

$$\int Q_k(\lambda) I_i(\lambda) I_i(\lambda)\, d\lambda = \begin{cases} \kappa_{ijk} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{5.12}$$

$$\mathbf{q}_k(\mathbf{x}) = \mathbf{m}^T(\mathbf{x}) \begin{bmatrix} \kappa_{k11} & 0 & 0 \\ 0 & \kappa_{k22} & 0 \\ 0 & 0 & \kappa_{k33} \end{bmatrix} \mathbf{p}(\mathbf{x}). \tag{5.13}$$

This gives us three equations in six unknowns. The antilog is taken and the diagonal matrix $\mathbf{G}_k$ is inverted to form three new lightness channels $l'_i(\mathbf{x})$:

$$\mathbf{G}_k^{-1} \exp q_k(\mathbf{x}) = l'_i(\mathbf{x}) = m_i(\mathbf{x}) p_i(\mathbf{x}). \tag{5.14}$$

The log is taken once more to get the relation that Land and Horn originally worked with:

$$l_i''(\mathbf{x}) = \log q_i'(\mathbf{x}) = \log m_i(\mathbf{x}) + \log p_i(\mathbf{x}) = m_i'(\mathbf{x}) + p_i'(\mathbf{x}). \tag{5.15}$$

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 illumination component $\mathbf{m}'(\mathbf{x})$ to produce an image that depends only on reflectance $\mathbf{p}'(\mathbf{x})$. Homomorphic filtering algorithms also use the logarithm to separate the components of the color signal in this way [Oppenheim, Schafer, and Stockham, 1968; Faugeras, 1979]. 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.

Land's original implementation

At each point $\mathbf{x}_0$ in an image, in each lightness channel, Land's original scheme [Land and McCann, 1971] computes the cone signal relative to a spatial average signal. To set the reference for the relative computation to grey the operations are simple; to set the reference to white the operations are a bit more complicated. First the grey normalization is described.

A second point $\mathbf{x}_{P,0}$ in the image is chosen randomly, and along a path P from this far point to the point of interest $\mathbf{x}_0$, small steps are taken. The lightness transition encountered in each step is considered. *If it exceeds a certain threshold* , the difference of the two lightnesses is summed into an accumulator. (Note that this is the difference of two logs, and thus a log of a ratio.) The threshold operation removes any slow gradients in the image that would result from uneven illumination. Thus at the end of the path, back at the point

$\mathbf{x}_0$ whose relative brightness is being computed, the accumulator just holds the log of the ratio of the starting point lightness and the end point lightness:

$$\begin{aligned}\log \frac{l_i''(\mathbf{x}_0)}{l_i''(\mathbf{x}_{P,0})} &= \log \frac{l_i''(\mathbf{x}_0)}{l_i''(\mathbf{x}_{P,n})} + \log \frac{l_i''(\mathbf{x}_{P,n})}{l_i''(\mathbf{x}_{P,n-1})} + \ldots + \log \frac{l_i''(\mathbf{x}_{P,1})}{l_i''(\mathbf{x}_{P,0})} \\ &= \log \left[\frac{l_i''(\mathbf{x}_0)}{l_i''(\mathbf{x}_{P,n})} \frac{l_i''(\mathbf{x}_{P,n})}{l_i''(\mathbf{x}_{P,n-1})} \cdots \frac{l_i''(\mathbf{x}_{P,1})}{l_i''(\mathbf{x}_{P,0})} \right]. \end{aligned} \tag{5.16}$$

This procedure is repeated for many points (paths $P1$ through PN) at random distances and directions from the point of interest, and the resulting values are averaged:

$$l_i'''(\mathbf{x}_0) = \frac{1}{N} \log \left(\frac{l_i''(\mathbf{x}_0)}{l_i''(\mathbf{x}_{P1,0})} \right) + \ldots + \log \left(\frac{l_i''(\mathbf{x}_0)}{l_i''(\mathbf{x}_{PN,0})} \right) \tag{5.17}$$

$$= \log \left[\frac{l_i''(\mathbf{x}_0)}{(l_i''(\mathbf{x}_{P1,0}) \ldots l_i''(\mathbf{x}_{PN,0}))^{1/N}} \right]. \tag{5.18}$$

The resulting average of logs is the log of the lightness of point $\mathbf{x}_0$ divided by a measure of the spatially averaged lightness. For an infinite number of paths, this measure of the spatial average is the geometric mean. 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 retinex.

The division by the average lightness and subsequent normalization imply the third main assumption of Land's theory, that the spatial average reflectance $\overline{S(x)}$ in each lightness channel is constant. This simple retinex algorithm, then, operates under a *grey world assumption.* Normalization to the lightness of a few hundred random points corresponds to high-pass spatial filtering. The lightness of the point of interest is set relative to that of its neighbors, with near neighbors more influential than more distant neighbors (assuming distance and angle are selected from uniform probability distributions). It is possible to mess up this simple version of the algorithm by placing a strongly colored patch in a very

simple scene so that the average reflectance is not grey [Brainard and Wandell, 1986].

To achieve normalization to white rather than grey in this scheme, Land and McCann add another non-linear operation at each step along a path about a point of interest. The ratio is set to one at the first, distant, point in the path. With this initial condition it is assumed, as it were, that the distant point is the lightest point in the image. As the path is traversed, the ratio decreases as long as points are encountered that are darker than the original point in the path and increases above one when a pont is encountered that is lighter than the original. If the ratio is never allowed to exceed one, the path averaging is *reset* every time a lightness is encountered that exceeds that maximum lightness encountered so far. This simple rectifying nonlinearity changes the algorithm from a grey referencing method to a white referencing method. The lightest patch in an image is always divided by one when the path values about it are averaged, but all other patches are divided by some value greater than one, and so are assigned values darker than white.

McCann in particular has advanced the notion of normalization to white as opposed to normalization to grey. The nonlinear retinex algorithm predicts human perception well [McCann, McKee, and Taylor, 1976; Worthey, 1985] even when the scene average is not grey, but preponderately red, green, blue, or yellow [McCann, 1987]. Brainard and Wandell report that the both simple grey-world algorithm and the nonlinear algorithm which normalizes to white are strongly influenced by the background color, and that both produce effects like the Helson-Judd effect [Brainard and Wandell, 1986]. McCann argues that the nonlinear model is not so scene dependent and these investigators made technical errors in their formulation of the model that overemphasized local contrast effects relative to more global normalization effects [McCann, 1989]. In his hands, the nonlinear algorithm, which normalizes to the maximum in each color plane by way of nonlinear spatial filtering, always reproduces white fairly well. A white patch in a scene illuminated with a colored light will take on the illuminant somewhat, in accordance with the Helson-Judd grey normalization effect, but the distortion from white is small compared to the physical color shift produced by the

colored light. This agrees with our everyday perception of reflectances lit with non-white illumination; the adaptation is not perfect, but is more robust than a strict grey-world adaptation. In the aperture condition that produces a strong Helson-Judd effect, "the extent of the surround is large enough to influence the global normalizing mechanisms," according to McCann [1989]. If the white that would normally serve as the reference for the algorithm is dwarfed in area by an adapting stimulus, the reference floats to grey, with the adapting background as the reference grey.

A subtle but important point about this nonlinearity of path summation is that it not only affects the normalization but also affects the spatial filtering process. The lightest patch is not filtered at all spatially, since all paths into it are reset at is boundaries, but patches darker than the maximum are normalized with respect to the maximum averaged with neighbor lightness values, weighted with distance. It was pointed out in chapter 2 that high-pass filtering produces overshoots at step edges in black and white images. In colored images, the overshoots are manifest along an arbitrary axis of the color space, and can be quite noticeable. In the retinex algorithm with a rectifying reset nonlinearity, overshoots are not produced around the lightest patch in a given color plane, but may be produced around any other patch in the image.

This algorithm and variants of it can produce nice results (see McCann's images in [La Brecque, 1988] for example). 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 NxN 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. An iterative, multiscale approach reduces the complexity, but requires complicated digital hardware [Frankle and McCann, 1983].

Horn's implementation

Horn [Horn, 1974; Horn, 1985] 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, to "spread out" the edge image and fill in smooth ares of the image. 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 nearest-neighbor feedback circuit. In a digital implementation, the Poisson equation is solved iteratively via Gauss-Siedel elimination. In all implementations, after the system is converged, the maximum value is found and is assigned to white. Several variants of this scheme have been described (e.g., [Blake, 1985; Funt and Drew, 1988]).

Horn's resistive grid idea is especially interesting since it requires only nearest-neighbor connections. The long spatial connections of the Land-McCann path architecture are eliminated, as the long-range spatial weighting is carried out in the temporal domain rather than the spatial domain. That is, it takes time for the feedback network to settle, and in that time information crosses the entire image space through the nearest-neighbor connections.

Marr [Marr, 1974] argues that the nervous system is equipped to carry out this operation. To our knowledge, no one has attempted to build chips based on this algorithm.

Land's latest implementation, and Hurlbert's implementation

In 1986, Land published an alternative to the algorithm described above [Land, 1986]. 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 [Livingstone and Hubel, 1984] and Zeki [Zeki, 1980], and has a strongly biological flavor to it. That is, the operator he uses looks like a cortical "non-classical" receptive field, with a narrow center and a huge surround [Allman et al., 1985; Desimone et al., 1985; Schein and Desimone, 1990].

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:

$$l'_{out,i}(x,y) = l'_i(x,y) - \log\left[l_i(x,y) \otimes \frac{1}{r^2}\right] \qquad r \neq 0. \tag{5.19}$$

Hurlbert arrived at the same sort of operation analytically [Hurlbert, 1986] with a Gaussian kernel:

$$l'_{out,i}(x,y) = l'_i(x,y) - l'_i(x,y) \otimes e^{-\frac{r^2}{\sigma}}, \tag{5.20}$$

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 [Hurlbert and Poggio, 1987].

This type of retinex algorithm, then, has a biological basis and sound computational underpinnings. But from an algorithmic standpoint, the complexity can be great. Since the required surround is so large, such a convolution across an NxN pixel image entails on the order of N^4 operations. (Note that this is not true if the kernel is separable. Both Briant Funt at Simon Fraser and Steve Omohundro at Berkeley have pointed out that this complexity figure is too general. With a separable kernel, such as Hurlbert's Gaussian, a 2D convolution can be reduced to a pair of 1D convolutions - one in the x direction and one in the y direction. Since a Gaussian kernel is separable, the operation in Equation 5.20 can be reduced to $2N$ calculations at N^2 pixels for a total of $O(N^3)$ calculations.)

On an analog chip, a convolution model with a $2D$ (non-separable) kernel corresponds to explicit connections from each pixel to most if not all other pixels. A similar operation

can be carried out 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

$$l'_{out,i}(x,y) = l'_i(x,y) - l'_i(x,y) \otimes e^{-\frac{|r|}{\lambda}}, \tag{5.21}$$

where λ is the length constant or extent of weighting in the grid [Mead, 1989]. Since the calculation is purely local, the complexity is small, $O(N^2)$. As in the Horn algorithm, complexity is moved from the spatial domain to the time domain. The image diffuses through the nearest-neighbor local connections in the grid and is smoothed over time. (Note that we are not getting something for nothing here. Convolution with a separable kernel requires N^3 operations; a resistive grid has $O(N^2)$ complexity in space but a factor N complexity in time. The diffusion in the grid with a large length constant is not complete until the weight of a given point is felt on the other side of the grid, N nodes or so distant. Any chip built to filter through diffusion must diffuse quickly to outperform a digital convolver.)

As described in chapter 3, a resistive grid can be built on a chip that diffuses at the video rate. In the next chapter this chip is used in a novel implementation of this retinex algorithm.

5.5.3 Buchsbaum's approach

Another approach to the problem originates in Buchsbaum's analysis [Buchsbaum, 1980], and has been extended by D'Zmura and Lennie [D'Zmura and Lennie, 1986]. This approach depends on two stages; first the illuminant is estimated, then the reflectance at each point is estimated. The estimations are based on two main assumptions: 1) that the spatially averaged reflectance of any scene is a "standard grey" and 2) that the illuminant is spatially uniform and can be estimated via application of the "standard grey" reflectance assumption to a sample of the incoming light averaged across the entire visual field.

Buchsbaum also assumes that the illuminant and all reflectances can be represented in terms of three basis functions; he assumes that the basis functions of each are fixed functions, not necessarily orthogonal, but sufficient to span the same space as the cones. With these specifications the tensor of Equation 5.9 becomes a 3×3 matrix; he works with its transpose. We thus rewrite the expression for the quantum catch by a cone of the kth class as

$$\mathbf{q}_k(\mathbf{x}) = \mathbf{p}^T(\mathbf{x})\mathbf{G}_k^T\mathbf{m}(\mathbf{x}) = \mathbf{p}^T(\mathbf{x}) \begin{bmatrix} \kappa_{k11} & \kappa_{k12} & \kappa_{k13} \\ \kappa_{k21} & \kappa_{k22} & \kappa_{k23} \\ \kappa_{k31} & \kappa_{k32} & \kappa_{k33} \end{bmatrix} \mathbf{m}(\mathbf{x}). \tag{5.22}$$

Buchsbaum's Algorithm

The first stage of Buchsbaum's procedure [Buchsbaum, 1980] entails comparing the incoming light across the visual field to a "standard grey" visual field and using this comparison to estimate the illuminant. The spatial average of the incoming light, as processed by the cones, is

$$\overline{\mathbf{q}} = \int_{vs} b(\mathbf{x})\mathbf{q}(\mathbf{x})\, d\mathbf{x}, \tag{5.23}$$

where the subscript vs indicates that the integral is taken over the entire visual space, the overbar indicates the spatial average, $\mathbf{x}$ is understood to correspond to a coordinate in a 2D space (for notational simplicity), and $b(\mathbf{x})$ is a weighting factor that Buchsbaum says may vary with the number of points in the field and/or distance from the center of gaze. Again for simplicity here we will set $b(\mathbf{x}) = 1/N$ where N is the number of points in the retinal image. The estimated average retinal signal for the kth cone class is

$$\overline{q}_{k,est} = \frac{1}{N} \int_{vs} q_{k,est}(\mathbf{x})\, d\mathbf{x} \tag{5.24}$$

$$= \frac{1}{N} \sum_{i=1}^{3} \sum_{j=1}^{3} \kappa_{kij} m_{j,est} \int_{vs} p_{i,grey}\, d\mathbf{x} \tag{5.25}$$

$$= \sum_{i=1}^{3} \sum_{j=1}^{3} \kappa_{kij} m_{j,est} p_{i,grey} \tag{5.26}$$

$$= \mathbf{p}_{grey} \mathbf{G}_k^T \mathbf{m}_{est}, \tag{5.27}$$

where the subscript *est* indicates an internally estimated value, and $p_{i,grey}$ is the weight of the reflectance basis function $S_i(\lambda)$ that corresponds to the "standard grey." For the set of cones, we have

$$\overline{\mathbf{q}}_{est} = \mathbf{p}_{grey} \mathbf{G}^T \mathbf{m}_{est}. \tag{5.28}$$

Now the actual and estimated spatial averages are equated and solved for the estimated illuminant basis vector $\mathbf{m}_{est}$:

$$\mathbf{m}_{est} = [\mathbf{p}_{grey} \mathbf{G}^T]^{-1} \overline{\mathbf{q}}. \tag{5.29}$$

The second step is to estimate the local reflectance $\mathbf{p}(\mathbf{x})$ by using the estimated illuminant just obtained. The actual data (cone quantum catches $\mathbf{q}(\mathbf{x})$) are equated to the estimated data $\mathbf{q}_{est}(\mathbf{x})$:

$$\mathbf{q}(\mathbf{x}) = \mathbf{p}_{est}(\mathbf{x}) \mathbf{G}^T \mathbf{m}_{est} \tag{5.30}$$

$$\mathbf{p}_{est}(\mathbf{x}) = \mathbf{q}(\mathbf{x}) [\mathbf{G}^T \mathbf{m}_{est}]^{-1}. \tag{5.31}$$

Substituting for $\mathbf{m}_{est}$,

$$\mathbf{p}_{est}(\mathbf{x}) = \mathbf{q}(\mathbf{x}) [\mathbf{G}^T [\mathbf{p}_{grey} \mathbf{G}^T]^{-1} \overline{\mathbf{q}}]^{-1} \tag{5.32}$$

and so the reflectance can be estimated from the cone data assuming that the spatially

averaged reflectance is grey.

D'Zmura and Lennie's Treatment

In an attempt to reconcile Buchsbaum's approach with the neurobiological fact of adaptation, D'Zmura and Lennie [D'Zmura and Lennie, 1986] change the algorithm in two ways. First, they use reflectance and illuminant basis functions that resemble the opponencies formed at the output of the retina. Second, they modify the second step to include two adaptive stages: one is explicitly stated, and models cone adaptation; the other is a vaguely described "second stage" adaptation.

5.5.4 Maloney's approach

Maloney's solution to the problem [Maloney, 1984; Maloney and Wandell, 1986] is derived from a parameter counting argument. He assumes only that the illuminant is spatially uniform, and that the illuminant and reflectance can be represented by a limited number of basis functions (b_1 and b_2, respectively). With no further assumptions he maintains that, with three cone signals as data at each point, we cannot solve the problem with $b_1 = b_2 = 3$. This is because, with light from one point in space, we obtain three equations in six unknowns; with light at two points in space, we have six equations in nine unknowns, etc. There are always more equations than unknowns. Thus he is forced to assume that $b_2 = 2$, that is, that the reflectances found in any scene are spanned by a 2D basis. This is contrary to our intuition, and indeed, to a study of a common set of reflectances by Maloney himself [Maloney, 1986]! Nonetheless, this is the procedure that he proposes.

Since his analysis leaves us in a color-impoverished world, and is based entirely on decidedly non-neural procedures, it will not be detailed here.

5.6 Conclusion

A variety of color visual effects can be explained with simple descriptive models of visual function. Algorithmic models go beyond description to actual specification of architectures and processing steps. In the next chapter, one of the retinex algorithms is evaluated.

Chapter 6

Real-time Retinex

Humans have the remarkable ability to perceive object colors as roughly constant even if the color of the illumination is varied widely. Land models the computation that results in this ability as three identical center-surround operations performed independently in three color planes, such as red, green, and blue [Land, 1986].

As discussed in the last chapter, a resistive grid efficiently computes the surround and is well-suited for implementation in VLSI. In this chapter, this simple algorithm is evaluated using computer simulations and a real-time system based on resistive grids built using analog VLSI. The evaluation leads to two conclusions. First, normalization according to a grey world assumption is not at all desirable since it results in strongly scene dependent color rendering. Rather, normalization to the maximum value in each color plane is a better rule of thumb. Second, a center-surround operation leads to color Mach bands which can be quite noticeable. Two modifications are suggested to minimize this kind of distortion: modulation of the surround weight by a measure of scene structure, and nonlinear spatial smoothing in surround formation.

6.1 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 [Hurlbert and Poggio, 1987], this scheme handily removes illumination gradients. Next, the simulation was extended to two dimensions with similar results and a tremendous increase in run time, due to the $O(N^4)$ complexity of the $2D$ Gaussian convolution needed to form the spatial average for subtraction. For 128x128 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 (Figure 6.1).

Specifically, in color simulations of the Land algorithm, 512x512 pixel images are subsampled to 128x128 resolution. The frame grabber used captures eight bits each of R, G, and B. Within a color plane, the eight 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; Kirchoff'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. (The simulation is stopped when the difference in the node voltages across the grid between two iterations is less than one-tenth of one percent of the maximum pixel value. About one hundred iterations is usually sufficient to meet this criterion. As Brian Funt has suggested, one hundred steps are needed in an iterative diffusion simulation to propagate voltages from one side of the grid to another for images of this size. The speedup from one hour for a $2D$ convolution to one minute for a resistive grid simulation is about a factor of 100, consistent

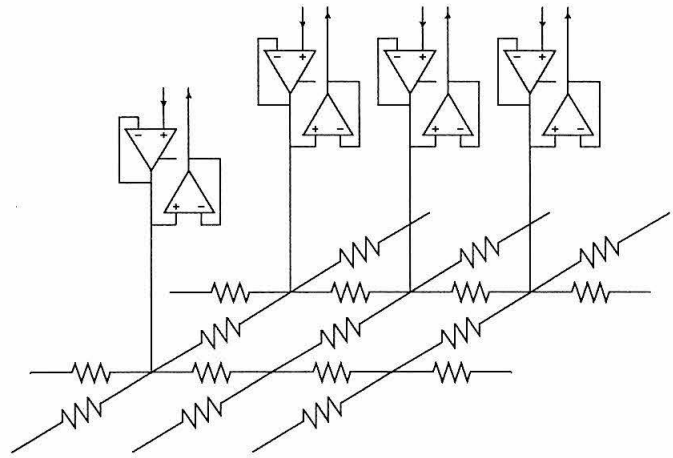
with a reduction in complexity from $O(N^4)$ to $O(N^3)$.) Next the settled net values are subtracted from the log of the input values. Finally, the minimum of the corrected values in the three planes is found and subtracted from all values in the three planes, and all values are scaled up so that the maximum of all values in the three planes is set to the maximum value of our frame buffer, 255 (Figure 6.2). For the red channel, this operation is

$$R_{out}(x,y) = 255 \cdot \frac{R_0(x,y) - min(min(R), min(G), min(B))}{max(max(R), max(G), max(B)) - min(min(R), min(G), min(B))}. \tag{6.1}$$

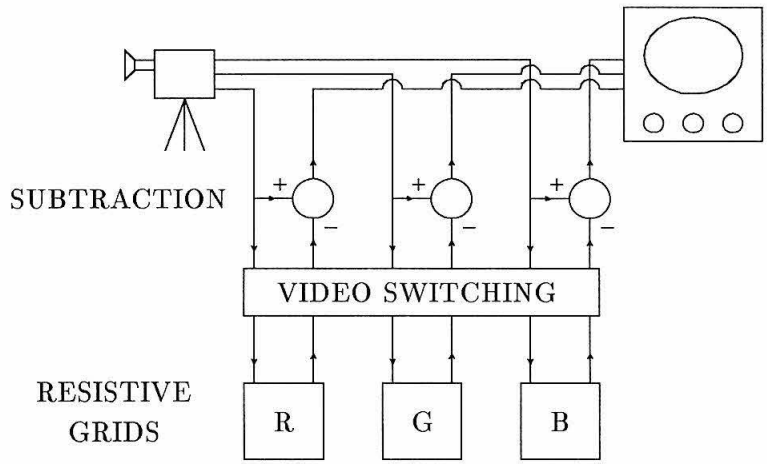
This last step restores the DC level (which is lost in high-pass filtering) and normalizes the result to the same range as the input image range (Figure 6.2). This sort of normalization can produce distortions in featureless images, but for the natural images we studied, with good blacks and whites present, Equation 6.1 just fits the output image range to the digital range of the frame buffer. Equation 6.1 is the only step that requires operations across the three color planes; all prior steps proceed independently within each color plane. Since it preserves the relative scale between color planes, Equation 6.1 does not affect the more important range-setting and normalization provided by the algorithm.

Since the images are acquired with a video camera with a variable iris, normalization by the algorithm builds on the "gain control" effects of the camera auto-iris system. This system sets the CCD sensor exposure for all three color channels identically. For scenes without perfect color balance, there is a need for additional gain control since the image range is, in general, different for the three color channels.

Plate 3 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



a



b

Figure 6.1: The Land algorithm. *a* Resistive grid for smoothing images. *b* 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.

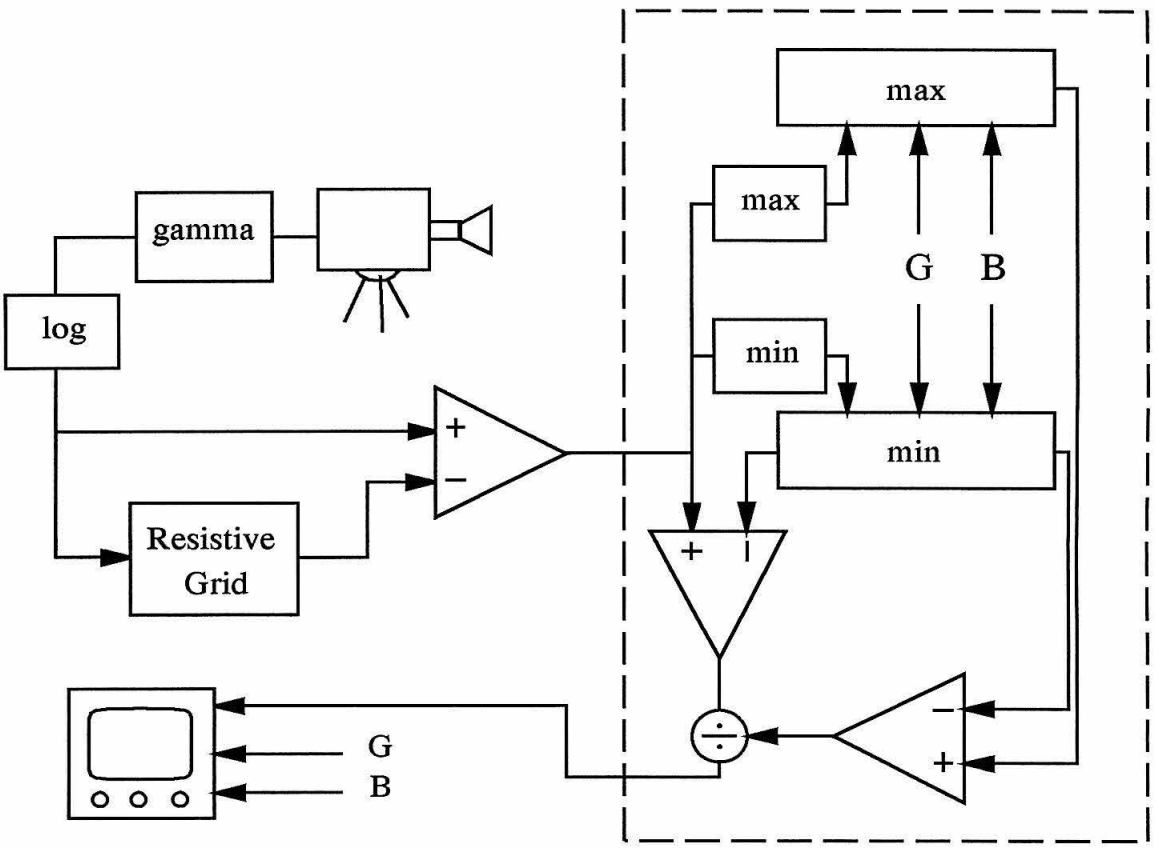


Figure 6.2: Gain control in the simulation of the retinex algorithm (only the red channel is shown). The circuitry in the dashed box at right performs DC restoration and range normalization on the result for display, but preserves the relative range and DC value of each color channel.

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 looks horrible. 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. This is a poor image for two reasons Figure 6.3. First, the scale of the red component of the image is greater than the scale of the green and blue components, due to the reddish illumination. (Its range is shifted upwards relative to the range of the green and blue components on a logarithmic plot, Figure 6.3b). The auto-iris, like our pupils, cannot correct this imbalance. Second, since the lighting in the incandescent image is less diffuse than the lighting in the fluorescent image, the range of the image exceeds the range of the computer monitor (see chapter 2) and so details in the highlights and shadows are compressed by the "knee circuit" of the camera.

The middle row of images in Plate 3 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 grey) 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 highlight and shadow contrast enhancement provided by this algorithm is at least as significant. The algorithm shifts the red channel down along the logarithmic axis to correct for the shift in the original incandescent image and compresses the range of both the fluorescent and incandescent images (Figure 6.3c).

One drawback of this algorithm, however, is apparent in these images - namely, color induction across edges. Close examination of image *e* of Plate 3 reveals that the (black)

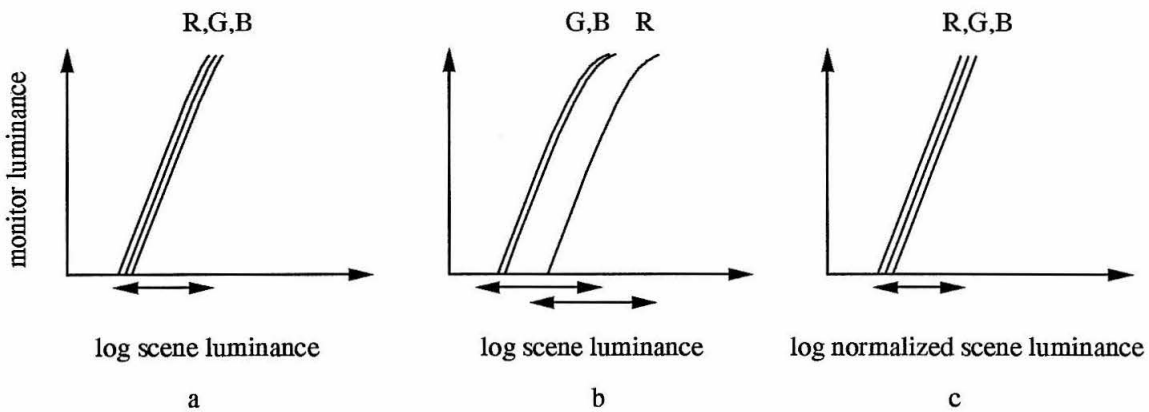


Figure 6.3: Range shifting and range compression by the retinex algorithm. The fluorescent image (*a*) has good color balance and is lit well so that there is only slight highlight compression. The scale of the red component is greater than that of the blue and green components in the incandescent image (*b*), and the relative range of each component exceeds the range of the CRT monitor. The retinex operation balances the color and compresses the range of each component so that the reproduction of either images approaches the ideal (*c*).

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. 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 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 [Land, 1986]. In this paper, he notes induction in terms of *Mach bands*, a well-known phenomenon in psychophysics. Figure 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 grey background.

Another limitation of the Land algorithm is revealed by the images in the right column of Plate 3. 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 comprised 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 range compression 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 grey, not green. This illustrates how the grey 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 grey.

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* [Cornsweet, 1970; Hurvich, 1981; Creutzfeldt et al., 1987; Lennie and D'Zmura, 1988]. 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 [DeYoe and Van Essen, 1988; Desimone et al, 1985; Schein and Desimone, 1990]. Second, the model suffers from over-reliance on the grey world assumption. As we shall see below, edge information can also help with this problem.

6.1.1 An extension to the retinex algorithm

A modification of the retinex algorithm was applied next to the same color images (bottom row of Plate 3). 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 (Figure 6.5). In other words, while for a retinex simulation, we have

$$output = center - surround, \tag{6.2}$$

$$l'_{out,i}(x,y) = l'_i(x,y) - l'_i(x,y) \otimes e^{-\frac{r}{\lambda}}, \tag{6.3}$$

to ameliorate induction effects and lessen reliance on the grey world assumption, we

need to modify the surround weight from point to point. In particular, if edginess is given a value close to zero in homogeneous regions like the black border of the poster in the left images, and is given a value close to one in more detailed regions such as the colored shirt, we have a better formulation as follows:

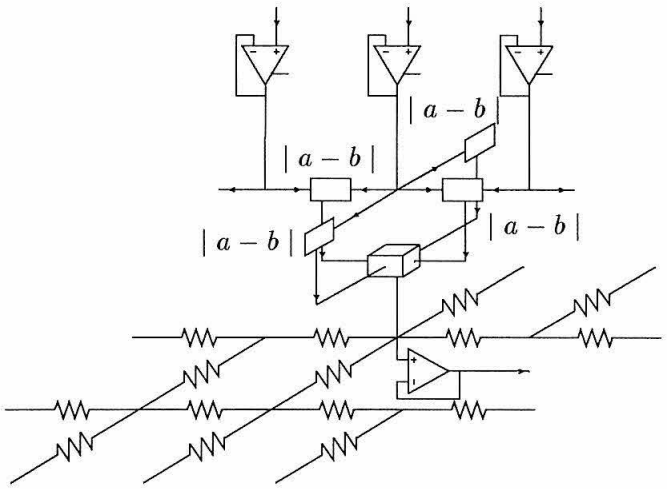
$$output = center - surround \cdot edginess. \tag{6.4}$$

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.

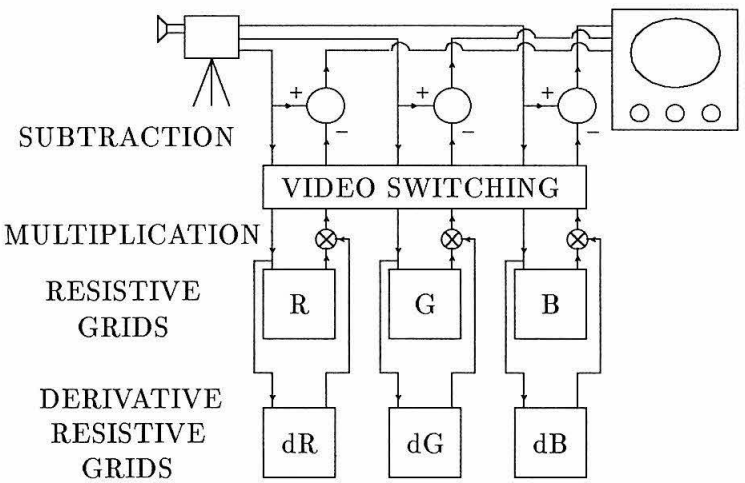
Figure 6.4 shows how edginess is computed and used. The 512 x 512 image is again sampled at a low resolution. The magnitude of the spatial derivative is computed between points; the average of the absolute value of the four local spatial derivatives are fed as a current into each node of the grid. The output voltage of this resistive grid, normalized to span the range 0-1 (Figure 6.5) 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

$$l'_{out,i}(x,y) = l'_i(x,y) - l'_i(x,y) \otimes e^{-\frac{r}{\lambda}} \cdot \left|\partial l'_i(x,y)\right| \otimes e^{-\frac{r}{\lambda}}. \tag{6.5}$$

The bottom figures of Plate 3 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



a



b

Figure 6.4: Extended Land algorithm. *a* The scheme for computing edginess. The average of the magnitudes of the local derivatives serves as the input to a resistive grid. *b* 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.

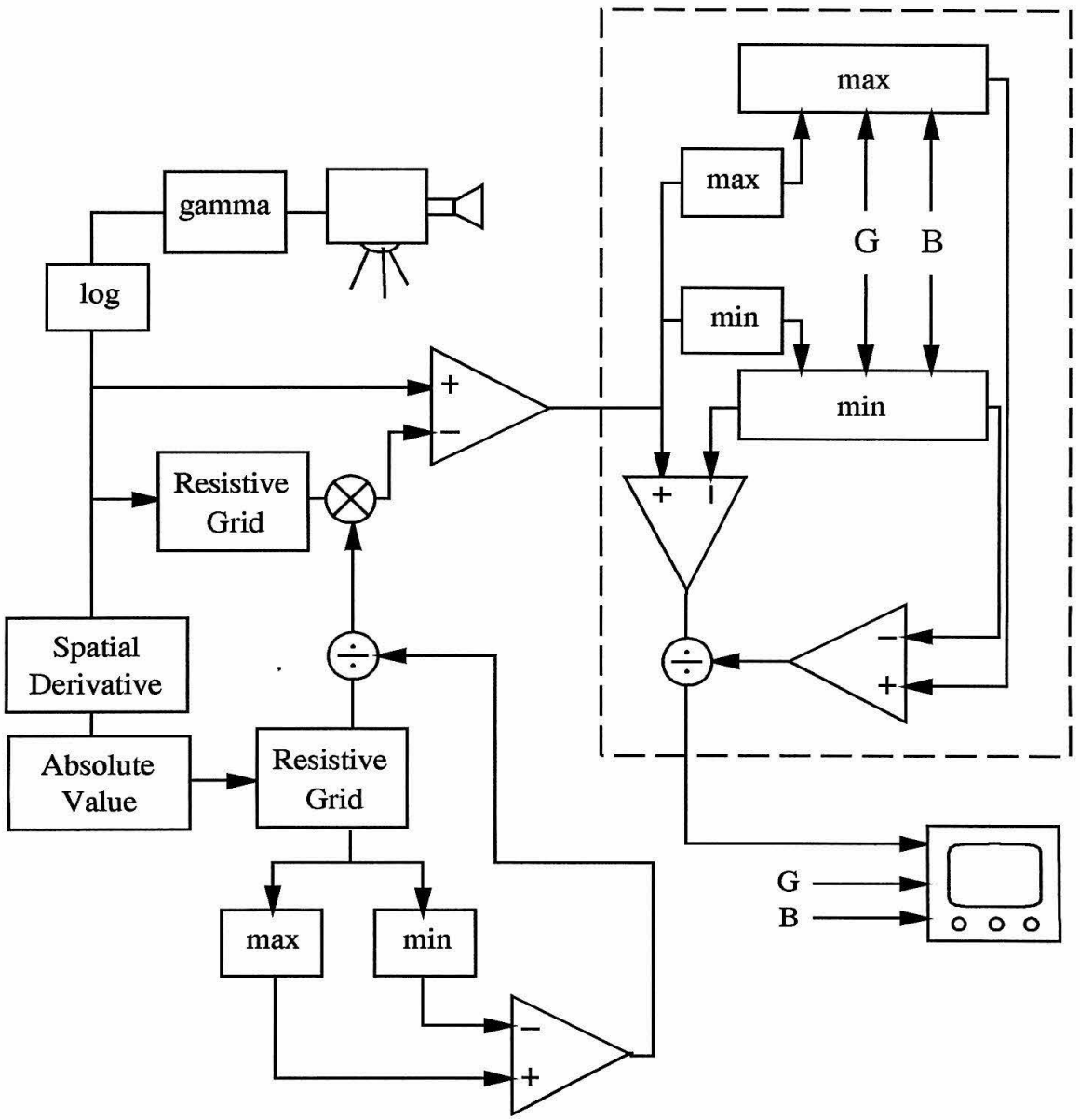


Figure 6.5: Gain control scheme used in the extended Land algorithm. The smoothed edge image is normalized to a maximum of unity and is used to modulate the strength of the smoothed image subtracted from the original image at each point.

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 Plate 3*h* the skin tone is redder than in Plate 3*e*. The extended algorithm, then, is a working compromise between color constancy via strict application of the grey world assumption and no color constancy at all.

Some of these results are quantified in Figure 6.6. A horizontal and a vertical line through the images in places that show induction artifacts were selected (Figure 6.6 *a,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 Figure 6.6). 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 [Grossberg and Mingolla, 1985; Hutchinson et al., 1988]. Mach bands are lessened in this algorithm as compared to 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

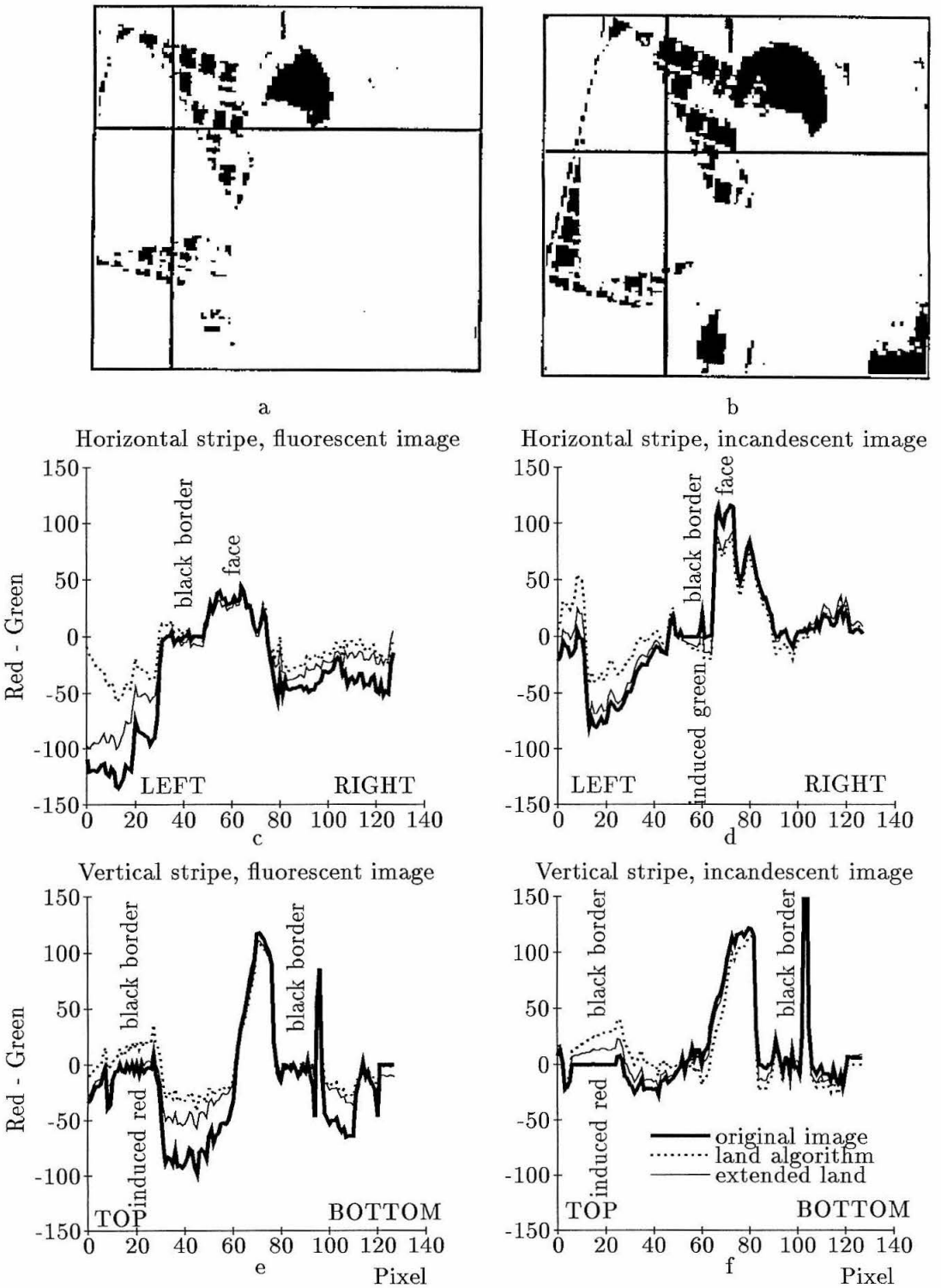


Figure 6.6: Quantification of simulation results. *a* and *b* are binary versions of Plate 3*a* and 3*b*. At bottom are plots of the R - G values along the lines through *a* and *b*.

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 [Perona and Malik, 1990; Harris et al., 1990]. As a result, subtraction of the grid outputs leads to patches of grey 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.

6.2 VLSI implementation of the retinex algorithm

6.2.1 Calibration

From the simulation results, it appears that the Land algorithm and simple extensions to it may be effective in color correction. Figure 6.1 *a* shows the outline of a color correction system, based on Land's algorithm, with video camera input. The three color outputs of a video camera - red, green, and blue - are fed onto three separate resistive grids built from subthreshold analog CMOS VLSI. For details of the resistive grid chips, see chapter 3.

A real-time implementation allows great flexibility in experimental verification of an algorithm. Why not point the camera at the lights, zoom in on a featureless area in a scene, stick your hand in front of the lens? In a computer simulation study, an image is a file, usually large, that takes time to store, retrieve, and process. An image is treated as an entity unto itself, outside of the context in which it was captured. With a real-time system, an image is not so burdensome and not so sacred. An image comes with

no experimental overhead in a real-time system (after the system is built and tweaked). System performance on an image is considered good only if the system also renders well images from other scenes in the surrounding environment. It is easy to become shackled to a "toy world" idealization in computer studies of an algorithm, but delusions of this sort are dissipated quickly in a real-time study. An investigator is naturally encouraged to compare a new color compensation technique to the tried-and-true techniques that are already implemented in the camera. The result is a perspective of the algorithm that is a bit hard-nosed, which, some may argue, is good for science.

In the following description of the evaluation of the Land (1986) algorithm the real-time system, gain control and iris settings were as follows, unless noted otherwise. The resistive grid space constant was set at a large value, so that the image to be subtracted from the original image was smoothed greatly; the kernel size in a comparable convolution smoothing operation would extend across most of the image, dropping in strength to half its maximum strength at a distance from the kernel center of about one fifth of the image width. In color comparisons, the camera autoiris was set only once, at the beginning of the experimental session, to place the exposure of the camera to reasonable levels given the lighting conditions (e.g., day *vs.* night). The color compensation circuits were set with a neutral reference and then frozen (turned off). The resistive grid DC level was set at the center of the camera peak-to-peak range, and the gain of the processed image (after subtraction of the smoothed image) was set so that white signal levels matched before and after subtraction, for a calibration image consisting of simple black and white bars. As a check, with a resistive grid length constant (degree of smoothing) decreased from a large value to zero, the system was set to diminish the excursion of blacks and whites toward grey, leaving only a grey field for smoothing set to zero. The system was thus balanced to grey, with a large degree of smoothing in normal operation.

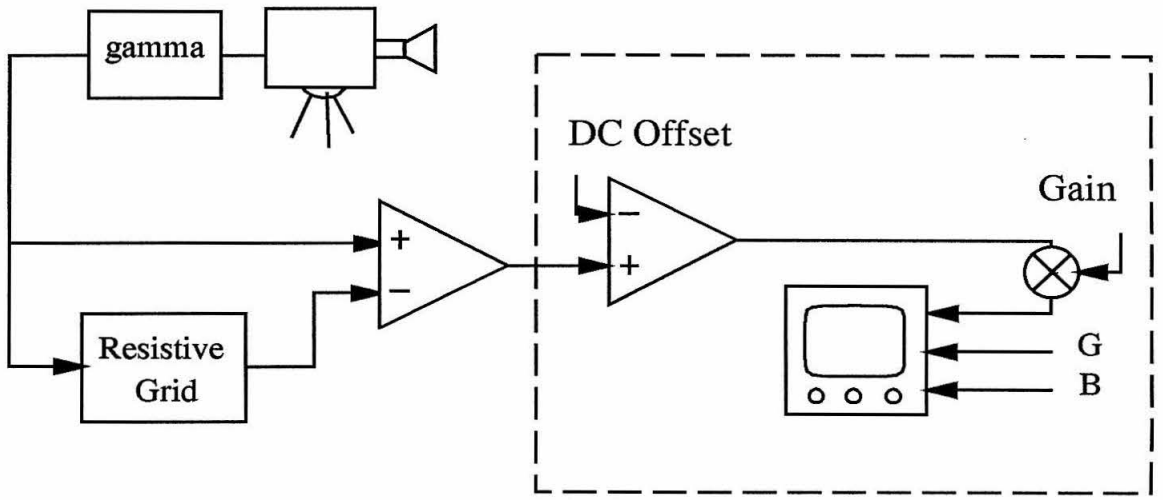


Figure 6.7: Gain control in the simulation of the retinex algorithm (only the red channel is shown). The circuitry in the dashed box at right performs DC restoration and range normalization on the result for display, but preserves the relative range and DC value of each color channel.

6.2.2 Real-time verification of color constancy

Color Plate 6 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, due to switching noise, cross-talk, etc. The point is that the color is more constant for the processed images than for the camera.

Conventional methods may be 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 grey world assumption is valid). The strength of this algorithm and its value as a model of the biology lies 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.

For a richly colored scene, the Land [1986] algorithm can remove strongly colored illumination. As a second example, a color Mondrian with many differently colored patches of paper was constructed and illuminated with ordinary fluorescent light plus various colored lights. Under a wide range of conditions, the color of the Mondrian as viewed on a video monitor changes with the illumination while it looks fairly stable to an observer. After passing the images through the electronic color compensation system, the image is also fairly stable for a wide variety of illumination conditions (Plate 4).

6.2.3 Color Mach bands and greying of large regions

To our surprise, the color Mach band effect, explained in the simulation work above, is less pronounced in the real-time system than we expected; for many scenes the induction effects are not noticeable. It is possible to the see the Mach bands clearly by placing colored cards on a grey background - the complementary color of the card surrounds the card as a halo that diminishes with distance from the card. This artifact is quite beautiful (Plate 5), and may be useful as a video special effect. In the practical, everyday work with video imaging, however, this effect would be worrisome. Would the yellow halo around the image of a man in a blue suit be noticeable to viewers? Clearly, efforts to diminish this color distortion are needed. Modification of spatial filter characteristics by a measure of spatial structure are useful (see above and chapter 2), and a nonlinear spatial filtering method (see [Harris, 1991] and chapter 4) is promising.

Since this version of the retinex approach relies on the grey world assumption, this algorithm fails where this assumption fails to hold. With the real-time system, we have

demonstrated this in many ways. For example, if the video camera is pointed at the color Mondrian and the hand of a Caucasian investigator (with a reddish skin tone) is slowly moved in front of the camera lens, the Mondrian in the background slowly grows more green (Plate 4 *m*). Green is the complementary color of red, and so it is added to the background in regions lacking in red by the extended spatial influence of the large red stimulus. Another example of practical importance is revealed by zooming in on a particular patch of the Mondrian. As more and more of the image is filled with this patch, the patch grows greyer and greyer, because the correction system subtracts the patch color from itself as shown in Plate 4 *n* and *o*.

6.3 Conclusion

Is the grey world assumption valid in explaining human vision ? The answer is yes, for extraordinary, controlled conditions. If a human observer really saw only what a video camera sees, he/she may process color as this system does (excluding color Mach band effects) [Helson, 1934]. But in normal visual experience, the answer is no, since our field of view is many times larger than our region of scrutiny, and the rendering of the region is greatly affected by points in the field outside the particular region of fixation. This means that a sample of the color signal, even though it extends across much of the region of scrutiny, is not an adequate *normalizing reference* for the visual system. Evidently, we use every available color signal, even from the light coming from the bridge of the nose, to set the normalization of the incoming color signal. Perhaps this is the proper interpretation of the neurophysiological data on color vision in the cortex (discussed in chapter 5) - normalization, relatively speaking, is a more or less global process, but the spatial scale of scrutiny is set at the whim of the observer.

Is the Mach model valid in color vision? In ordinary conditions, these effects are not noticeable, so in practical terms, the answer is no. We do not see a yellow halo around a man wearing a blue suit or a green halo around a red ball on a grey background. The

spatial filtering processes that underlie color normalization in the nervous system are not as influential in vision as they are in this algorithm.

Both conclusions - that this algorithm models both the normalization and spatial filtering processes of the visual system poorly - have implications in tone reproduction engineering and in visual science, as discussed in the next chapter.

Chapter 7

Tone Reproduction and Vision

So far, this thesis has separately asked applied (tone reproduction) and theoretical (vision) questions. In this last chapter, both areas are discussed together. The science of tone reproduction has historically borrowed much from visual science, and in this chapter this tradition is continued. Further, ideas are transferred in the other direction in this chapter - from the practical world of tone reproduction to the more abstract, theoretical world of vision.

7.1 From vision to tone reproduction

7.1.1 Reproducing subjective brightness effects

At the end of chapter 1, psychophysical brightness response curves were described. The relative brightness at a given adaptation level is monotonic, but not linear, in luminance. Figure 7.1 is a plot of subjective brightness relative to the white reference in a given scene [Bartleson and Breneman, 1967a]. Bartleson and Breneman found that the relative brightness response curve is strongly affected by the viewing conditions, due to contrast. Thus the response curve is different when viewing a print (with a bright surround) and viewing a transparency or motion picture (with dark surround). The former curve resembles the latter displaced downward, presumably as a result of contrast with the bright surround [Bartleson

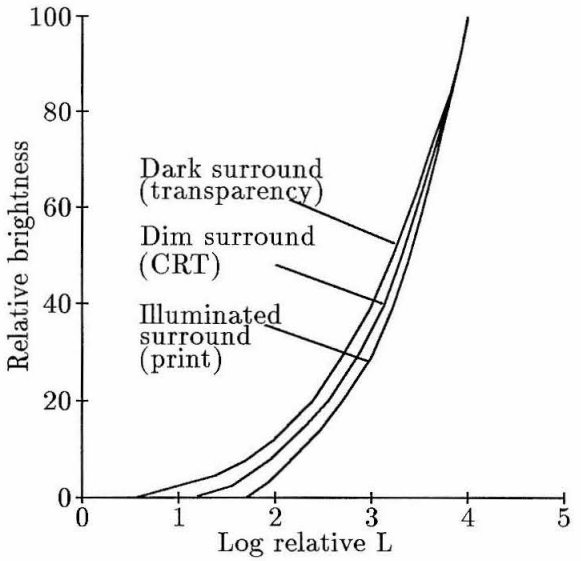


Figure 7.1: Human subjective brightness response, relative to scene white. With a dark surround, the response extends over more than 3 orders of magnitude. Illuminated surrounds displace the curve downward. The response is most shallow with dark surroundings, and steepens for dim and bright surrounds. From Bartleson and Breneman [1967b].

and Breneman, 1967b; Nelson, 1966; Nelson, 1977]. Television is normally viewed with a dim surround, and so the brightness response curve is in between the curves for prints and transparencies.

Bartleson and Breneman used these curves to explain why TV viewers turn up the brightness and contrast knobs on their sets, and why motion picture film is printed with a slope greater than 1 (on a log-log plot). Figure 7.2 shows their analysis for television viewing. The goal is to reproduce the brightness response evoked by the original scene even though the viewing conditions are different. At top right is the response curve for the original scene, with three reflectances (white, skin, and dark grey) shown. At bottom left is the response curve in the viewing conditions. It is shallower than the curve for the original scene, and so the television transfer curve from scene luminance to CRT luminance must have a slope greater than 1 to reproduce relative brightness. At bottom right, the unity

gain transfer characteristic (dashed diagonal line) is plotted with the preferred *transfer* curve (thick line) for comparison. The effect is small, but real, and reproducible.

Incremental contrast is highest near white and is rather low near black. With the limited range of CRTs and photographic prints, it has in the past been impossible to reproduce with these media a scene with a range that matches the range of the visual response at a given adaptation level. With nonlinear masking, as described in chapter 4, it is now possible to do so. What does the relative brightness response characteristic tell us about using this procedure in tone reproduction?

Consider the window scene of chapter 4 (Plate 1). The portion of the image seen through the window is bright, and the indoor area is dim. In the past, only one of these two areas, distinct in their respective intensity ranges, could be reproduced. Contrast would be rendered uniformly in most of the range, centered on the middle tones, but would be decreased to zero at the extremes of the display range. This S-shaped rendering curve is the classic compromise solution to the problem of range mismatch between the sensor and the display (chapter 2). One of the two distinct ranges of the window scene would be chosen, and rendered as faithfully as possible (i.e., relative brightness would be reproduced as closely as possible). Given the limited range of the display medium, the visual system would be excited only in a portion of its response range - the upper 1.5 orders of magnitude of the response curve. This portion of the response curve is steep compared to the lower portion; contrast is greatest near white.

With nonlinear spatial masking it is possible to render the whole window scene on a limited range device, through intelligent range compression. How should it be rendered? Again, the display still excites the eye in the upper, high-slope region of the response curve, but the various elements of the original scene excite the eye over its entire response range. In the original scene, the eye would "render" the bright portion of the image with greater contrast (white squares in Figure 7.3, top right) and the dim, indoor portion with lower slope (white circles in Figure 7.3, top right). The aim is to reproduce a scene *as the eye*

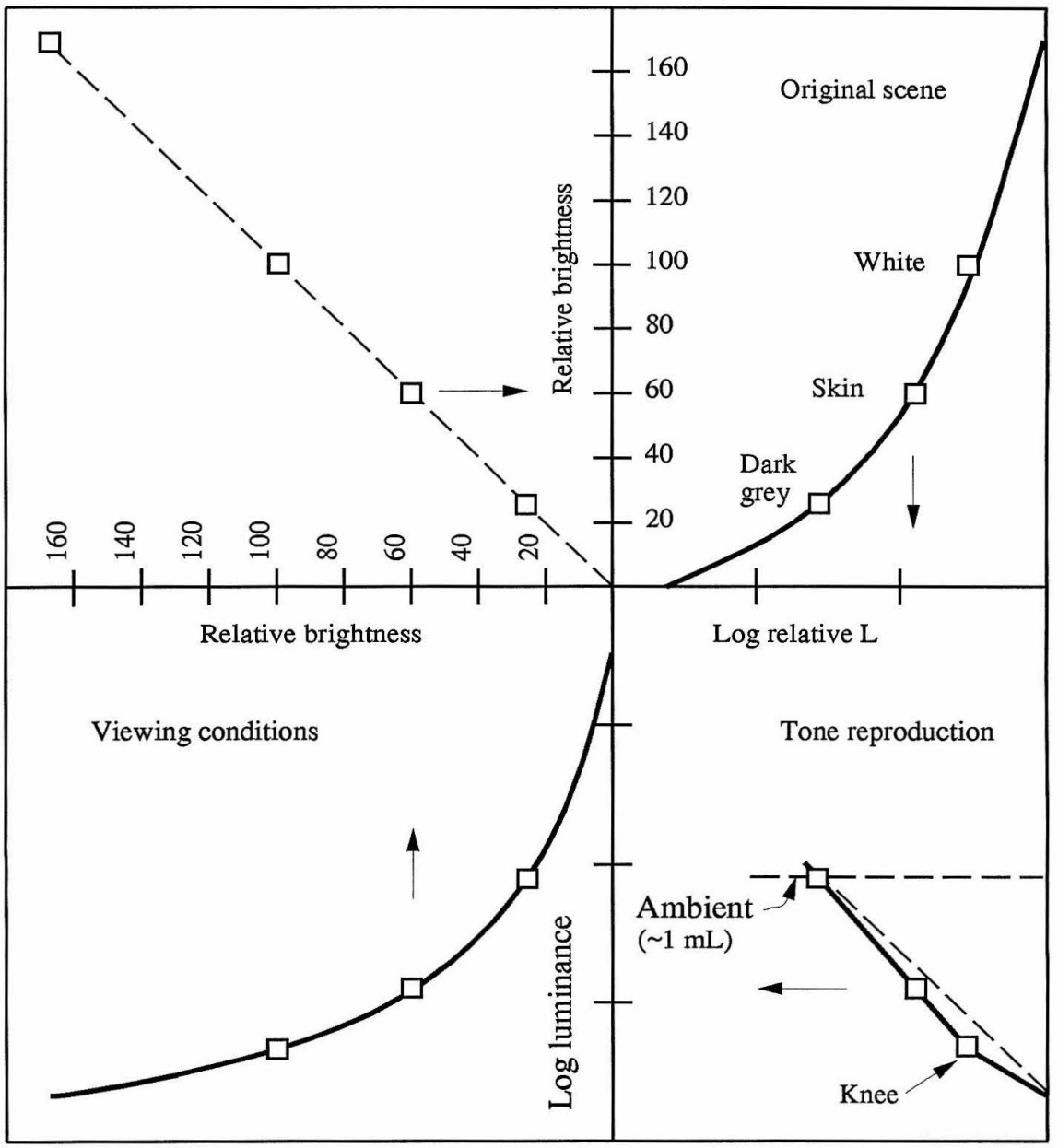


Figure 7.2: Analysis of subjective television tone reproduction. A transfer curve with unity slope (dashed diagonal line, bottom right block) would reproduce a scene darker than the actual visual response to the scene, since the response curve in the dim viewing conditions (bottom left) is shallower than the response curve in the original environment (upper right). Viewers prefer a transfer curve that compensates for the viewing conditions - a curve with a slope greater than one (thick line, bottom right). Adapted from Bartleson and Breneman [1967b].

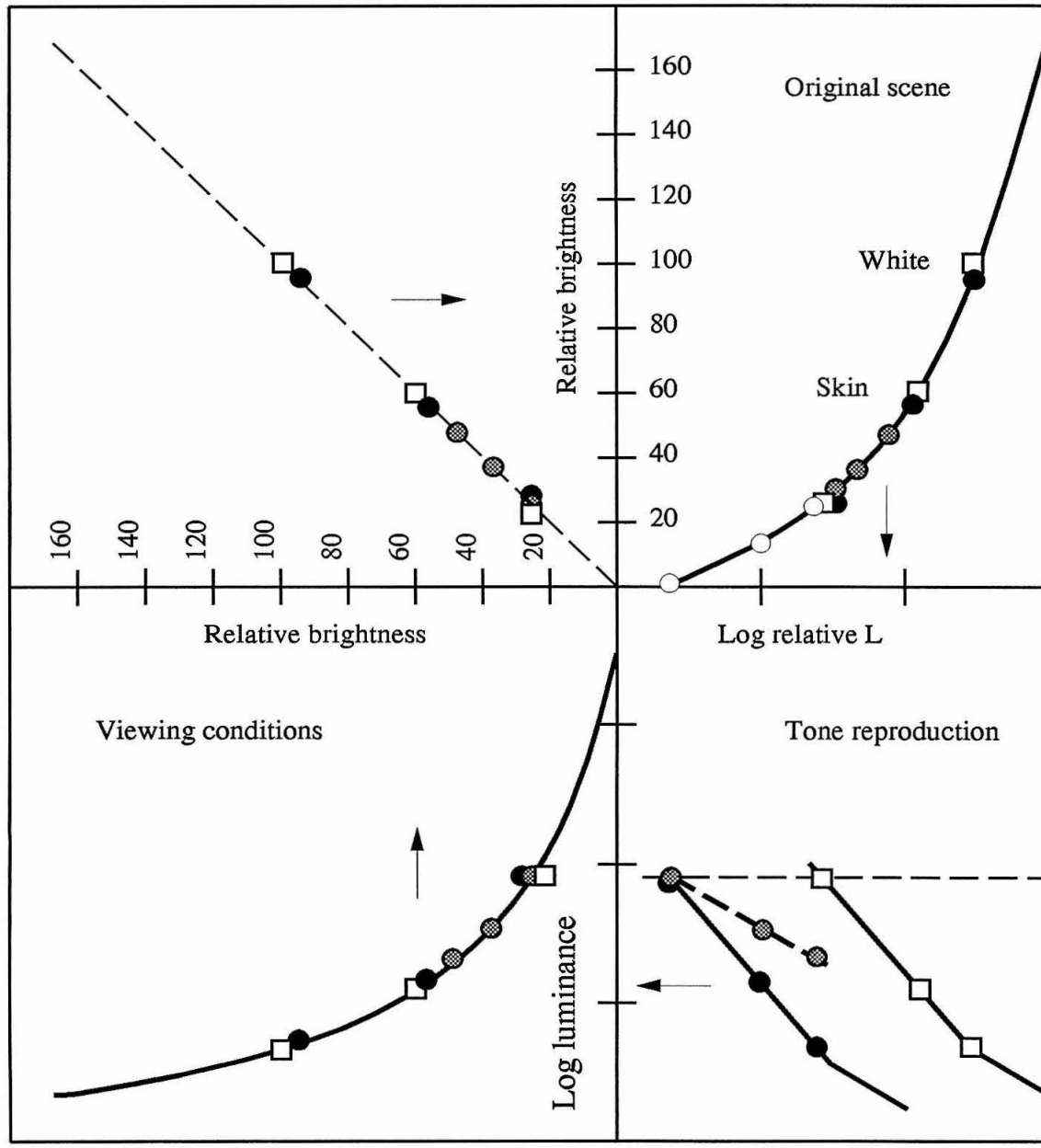


Figure 7.3: Subjective television tone reproduction with nonlinear masking. If the same transfer curve is used for shadow and well-lit regions of an image, the reflectances in shadow (white circles, top right block) are reproduced with too much contrast (black circles). A transfer curve with a lower slope (thick dashed line, bottom right block), is needed to match the reproduced contrast with the original contrast (grey circles).

sees it. If the dim portion of the image is reproduced using the same transfer characteristic as that used for the bright portion, it will not look right - the indoor area will look too "contrasty" (black circles in Figure 7.3). This can be understood by following the white circles from top right to bottom right of Figure 7.3. They are reproduced (black circles) with the standard, high-slope transfer characteristic. As seen in the bottom left block, the dim portion of the image (black circles) falls on the same portion of the viewer response range as the well-lit portion of the image (white squares). The two portions of an image which to a viewer of the original scene have different incremental contrasts and tone ranges are reproduced with the same contrast and tone range.

Thus the correct reproduction strategy is to lower the incremental contrast of the indoor portion of the reproduction (grey circles in Figure 7.3, bottom right) so that, even though it is sensed within the high-slope region of the response curve upon reproduction (grey circles in Figure 7.3, bottom left), it is rendered at relatively low contrast. By using a less steep transfer curve for the dim portion of the image, that portion is rendered with a contrast that is similar to that which a viewer of the original scene would experience.

A simple addition to the masking system described above can simulate this effect, and thus can produce images that are more consistent with the way that we see. Figure 7.4 shows a camera signal with three distinct tone ranges. The lowest range (thin line) extends from black to about the mean of the signal range, the middle range (dashed line) straddles the signal mean, and the upper range extends from the signal mean to the signal maximum (thick line). In this example, the segments are shown as overlapping, and it is assumed that the amount of overlap is sufficient to produce segmentation in the saturating mask. Also shown in Figure 7.4 *a* is the masking signal, which is scaled linearly on this log-log representation by the gain control system described in chapter 4; the mask signal is scaled so that the difference between its maximum and the camera signal maximum is V_{max}, the maximum output range for the display device. The signal that results from the subtraction of the mask signal from the camera signal is represented in Figure 7.4 *b*. All three segments

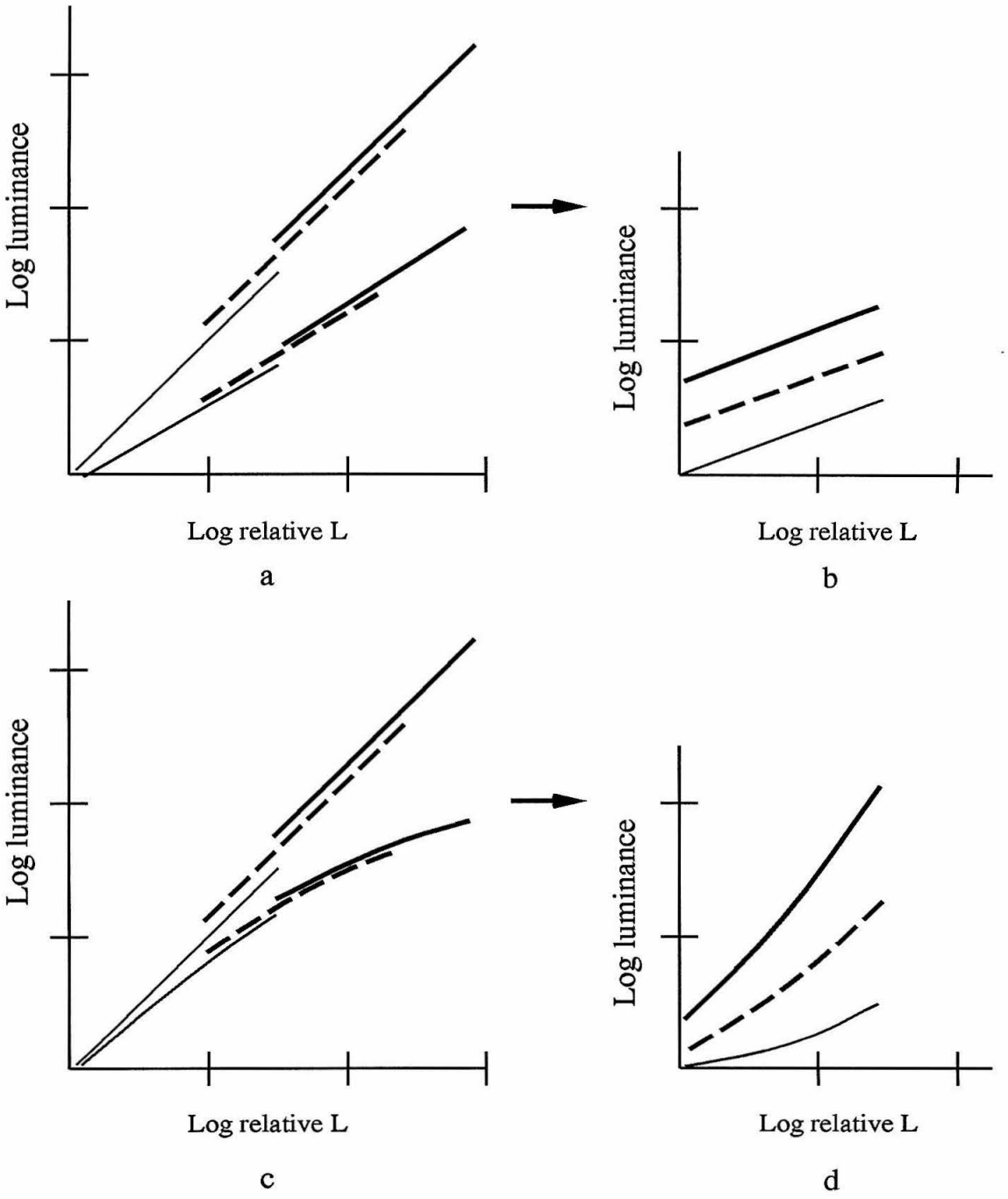


Figure 7.4: Mask compression to achieve optimal subjective tone reproduction of a wide-range image. A linearly scaled mask (bottom 3 lines in *a*) results in only a vertical shift after subtraction (*b*). If a compressive scale is applied to the mask signal (bottom 3 lines in *c*) A slope change as well as a vertical shift results after subtraction (*d*). Thick lines: upper tone range; dashed lines: middle tones; thin lines: lower tone range.

are reproduced with the same slope - there is no difference in contrast for any *tone range*.

Differential contrast can be produced as illustrated in Figure 7.4 *c*. Before scaling by the gain control system, the mask signal is compressed. After subtraction (Figure 7.4 *d*), the most luminous portion of the image (thick line) is reproduced with a greater slope than the least luminous (thin line).

These examples are rather qualitative. The exact nonlinear mask compression characteristic that produces the most pleasing results is unknown and should be investigated psychophysically. The procedure advocated comes from analysis (Figure 7.3) and from anecdotal remarks from people who have looked at the window image. One student said that it lacked depth: although it was clear that outdoor portion was more distant than the indoor portion, they seemed too "flat." Relative shading between the two image regions is preserved in this reproduction, and shading is a strong cue in our perception of depth [Evans, 1943], and so that visual "need" is fulfilled. What is lacking is an incremental contrast difference between the two portions of the image. This topic is worthy of further study.

7.1.2 Color normalization

The retinex investigation detailed in the last chapter showed that a grey-world assumption is not generally useful in modeling vision, and that a system built around this assumption produces artifacts when the assumption fails.

Color film is commonly corrected with a grey-world normalization. Global estimates of the signal strength in each of the three color planes are used to produce a grey balanced final result. This removes many lighting variations, but produces artifacts in non-grey images. These effects are said to be due to *subject failure* [Evans, 1943; Bartleson and Huboi, 1956]. Film consumers are *frustrated* by prints with color distortion, but since it takes time to develop and print negative film, they are not *outraged.* Video users will not be so forgiving, as they will be able to see the results instantly.

A better way to normalize color in video imaging is to use the maximum rather than the mean as the reference, and to mask the video signal so that the maximum value in each color channel never exceeds the maximum allowable signal level. Since the video range is sufficient to span the range of common reflectances, this procedure will reduce illumination effects while preserving reflectance variations. Masking to keep the signal within the allowed range will never turn a red barn into a grey barn, as a grey-world normalization will, but it will remove much of the strong coloration that results from reddish light.

An example of this procedure is shown in Plate 7. The color camera was color balanced for fluorescent light, and the camera was pointed at a lovely young lady. The entire video signal was within the allowable video range, and so no masking occurred using the scheme described in the last paragraph. Next incandescent illumination was added to the fluorescent, with the iris unchanged. The skin tone of the subject was noticeably redder, but still, the red video signal did not exceed the maximum allowable signal level, and so no masking was performed. Next the incandescent intensity was increased still more; on the video monitor, the skin tone of the subject was a blazing red. This is shown in Plate 7*a*. The red signal did exceed the maximum allowable level in this instance, and so masking "kicked in." The result of masking is shown in Plate 7*b*. The skin tone is not nearly as red in *b* as it is in *a*.

(Note that in this experiment, masking was carried out on the red channel only; the blue and green components in the incandescent illumination are not attenuated and result in an unnatural skin tone. The point is that the red signal is diminished effectively in extreme cases.)

This scheme is a compromise. It is not possible to determine whether the skin tone of the subject is reddish when the signal is in range because of lighting, on the one hand, or because of a sunburn or excessive application of rouge on the other hand. So be it - this scheme will at least prevent the catastrophic imbalance that results from intense, strongly colored illumination.

7.2 From tone reproduction to vision

7.2.1 Nonlinear spatial filtering and vision

Mach bands only go so far

Many biological and psychophysical investigators assert that linear spatial filtering is used to remove illumination gradients and normalize the image range. Linear filtering is called many things in visual science: center-surround, difference of Gaussians, lateral inhibition, etc., [Mach, 1897; Ratliff, 1965; Marimont 1963; Bekesy, 1968a; Bekesy, 1968b; Tolhurst, 1972; Heinemann, 1972; Fiorentini, 1972; Marr, 1982; Hurvich, 1981; Land, 1986; Cornsweet, 1970].

This sort of filtering mechanism is proposed to explain Mach bands and related phenomena. While it is true that lateral inhibition effects are observable in certain cases, effects that are predicted by the same mechanism are not observed in other circumstances. For example, Mach bands are seen at the top and bottom of a spatially ramped stimulus, and a similar percept arises in the Chevreul or staircase illusion [Hurvich, 1981]. (Mach bands extend only about 6 minutes of arc from the top or bottom of a ramp, but the spatial variation produced in the staircase illusion extends much further, to more than one degree of arc. Bekesy [Bekesy, 1968b] differentiates between the two effects, and calls them Mach and Hering inhibition effects, respectively.) However, Mach bands are not visible at the edges of a large step in intensity, and the strong overshoots that one would expect from a linear spatial filtering model at the top of a staircase of intensity are not observed.

This inconsistency has not escaped the notice of some researchers [Fiorentini et al., 1990]. Two models exist which depend on filters with odd-symmetric line spread functions to explain the absence of Mach bands at a step edge [Tolhurst, 1972, and Ratliff, 1984; Ross et al., 1989]. Neurons with odd-symmetric line spread functions are found in the visual cortex [Hubel and Weisel, 1968] and are often called edge detectors [Tolhurst, 1972]. However, such filters are not found in the retina; even symmetric filters (center-surround

cells) are. Since there is no feedback from the cortex to the retina, it is hard to see how the mechanisms suggested by these models might be manifest at the level of the retina.

The present study suggests another possible mechanism: nonlinear spatial filtering at the retinal level. It is known that the horizontal cells in the retina are coupled with gap junctions. If the gap junctions have current limiting characterististics, nonlinear filtering of the sort described in chapter 4 results.

There is one psychophysical study which supports this hypothesis (Figure 7.5). In this experiment, a thin line is superimposed on a step intensity display at various locations. The subject is asked to adjust the intensity of the line until it disappears, and this null intensity is plotted as a function of distance for three different amplitudes of the step. For a small step, Mach bands are uncovered by this procedure, even though the subject does not see them. Somehow they are integrated later by the visual system so that they are subliminal and not noticeable. For a step that is 2-3 times greater in amplitude, the subliminal overshoots revealed by this method are 2-3 times greater, as predicted by a linear model. For a step that is about 8 times greater in amplitude than the smallest step studied, however, the subliminal overshoots are roughly equal in size to the overshoots produced by the middle-sized step. The overshoots have not grown linearly with the stimulus, but remain constant. This saturation in the overshoot size with step size is consistent with the nonlinear spatial filtering process described in chapter 4.

This single study does not make a theory out of this hypothesis. The experiment illustrated in Figure 7.5 has not been repeated, to our knowledge, nor has it been published in a peer-review journal - the results of the experiment must be regarded as tentative. Further, the Heinemann effect of Figure 7.5 is consistent with the Ratliff and Tolhurst and the Ross et al. conjectures as well as the nonlinear diffusive inhibition hypothesis.

A simple neurophysiological measurement could decide this mechanistic question. Enroth-Cugell and Robson [Enroth-Cugell and Robson, 1966; Fiorentini, 1972] measured and have published retinal responses to a spatial step of intensity in the cat. They used step sizes

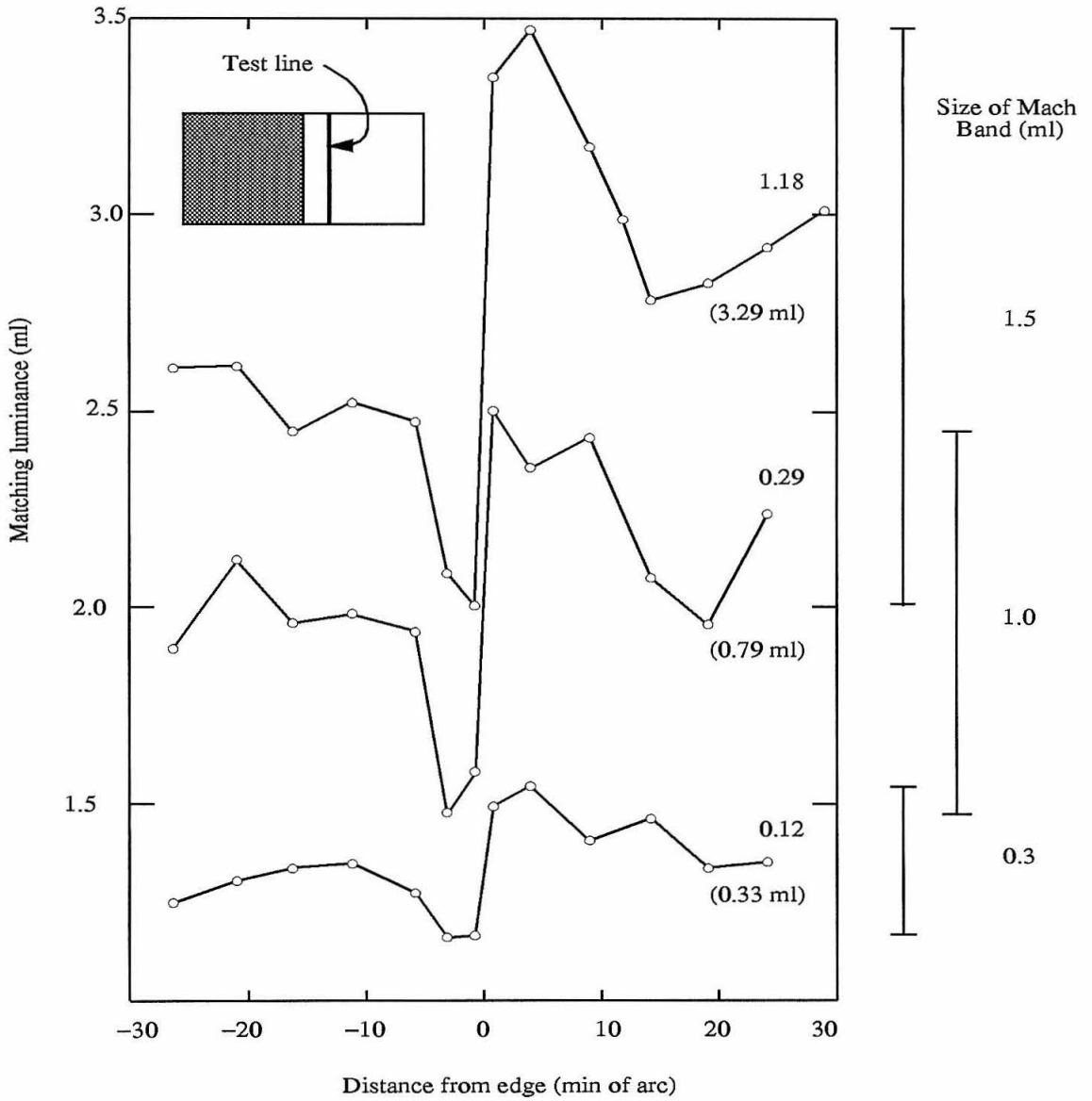


Figure 7.5: Saturation of Mach band size with stimulus step size. The curves trace the matching luminance of a thin line imposed on a step stimulus (inset) with distance from the step edge. The luminance of the dark side of the step is 2.72 mL; responses to three different luminances on the lighter side are shown. For a step with a luminance ratio of 0.12, which is a luminance step of 0.33 mL over the baseline (parameters at right of bottom curve), the overshoot size is 0.3 mL. For a step which is about 2.5 times higher, the overshoot size increases by about a factor of 3. This is a roughly linear increase, and is expected from linear filtering theory. For a step which is about 10 times higher, however, the Mach bands are only about 5 times greater in size. The top two curves are shifted up by 0.7 mL for clarity. Adapted from Heinemann [1972].

that are close to the smaller two step sizes studied by Heinemann (Figure 7.5) and show that the overshoot increases linearly with step size. Unfortunately, if they or other neurophysiologists have made such a measurement in the retina with larger step sizes, the results have not been published. If the overshoot amplitude stays constant after the step size is increased beyond a certain level, this would provide evidence for saturation in the lateral inhibitory mechanisms in the retina, and would eliminate the possibility that edge-detecting neurons are involved in the effect.

Linear filtering can cause trouble

No matter what the mechanism is, it is clear that nonlinear spatial filtering is important in survival. Consider the images of the house in Plate 1 of chapter 4. With linear spatial filtering (Plate 4*c* and *e*), overshoots near the window sill are so large that the imaging system "hits the rails." If the nervous system filtered linearly, we might *see* these distortions, and so would be functionally blind near large step edges (such as light-shadow boundaries). Hungry predators would learn to wait at the edge of a shadow for us to wander by and then eat us. Perhaps at some stage of evolution some species did suffer from the ill effects of linear spatial filtering. If so, they did not wander the face of the earth for long.

7.2.2 Spatial filtering and color vision

A couple of influential color vision models [Land, 1986; Hurvich, 1982] are linear spatial filtering models. In this real-time study, it was easy to explore the possibility that a simple center-surround mechanism accounts for color perception. Plate 5 shows the output of an instantiation of a simple center-surround model in color space. Such a model predicts that we will see a halo of color around a colored object on a grey background. According to this sort of model we would see a blue halo around the sun and a green halo around a person wearing a green sweater standing against a white wall. This is contrary to experience. Perhaps spatial integration after center-surround processing [Arend et al., 1971; Horn, 1974; Wallace and Lockhead, 1987; Grossberg and Mingolla, 1985; Burr, 1987] removes such

artifacts; perhaps saturating diffusive inhibitory processing prevents the occurrence of such aberrations. This question can be best resolved with focussed neurophysiological study.

7.3 Conclusions

In this chapter, analytic methods and psychophysical data from tone reproduction science were used to optimize the appearance of an image reproduced with the aid of a nonlinear mask.

Further, insights gained from the real-time study led to two theoretical questions. First, is there nonlinear lateral inhibition in the retina ? A simple experiment can answer this question. Second, if lateral inhibition exists in the color visual pathways, why don't we see color Mach bands ? Extensive study is required to answer this question, but perhaps it is time to start.

Chapter 8

Conclusions

8.1 Tone reproduction

The dynamic range of the visual system at a given adaptation level is quite large - about four orders of magnitude. Electronic image sensors and photographic negative materials can capture the range for which we can see detail at a given adaptation level, about three orders of magnitude. The dynamic range of display media is rather smaller - 1.5 to 2 orders of magnitude.

There are several solutions to the problem of range mismatch between the sensors and displays in tone reproduction systems. Nonlinear spatial filtering produces fewer distortions in the reproduction of wide dynamic range images with large step edges than traditional solutions. A saturating resistive grid is a substrate for nonlinear spatial filtering that can be integrated into high-speed video imaging systems in the near future.

Range compression presents an interesting scaling question: how should a wide-range image be reproduced with a limited range medium so that the reproduction is believable? The psychophysics of the human brightness response provides an answer to this question: tones near white should be reproduced with greater contrast than tones near subjective black.

Many psychophysical models suggest that normalization to the mean in a color image

is the optimal procedure in color tone reproduction, and this procedure is followed in mass-market photofinishing. A real-time video study of this procedure reveals that normalization to grey is not desirable. Rather, normalization to white is preferable, even though it does not correct for colored illumination except in extreme circumstances. This compromise is necessary to provide continuity in video imaging. The dynamics of video imaging (zooming, panning, etc.) constrain the normalization method - an abrupt change in the scene must not produce an abrupt shift in color normalization. Normalization to grey results in abrupt shifts, but normalization to white does not.

8.2 Vision

The vast majority of spatial vision models include a linear filtering stage. Linear spatial filtering produces severe artifacts at steps in intensity and color. Such a stage must be followed by an integrative stage to remove the distortion. Alternatively, a nonlinear spatial filtering stage may be substituted. A saturating resistive grid is an attractive substrate for this nonlinear filtering stage since it

- requires only local interconnections,
- does not depend on annealing or reset processes, and
- maps directly onto known neural hardware, the horizontal cell plexus of the retina.

Simple neurophysiological experiments can determine whether nonlinear spatial filtering is indeed manifest at the level of the retina, and so can decide between this model and other nonlinear models which invoke cortical processing primitives. Of course, if such experiments reveal that nonlinear spatial filtering is present at the level of the retina, the filtering process is not necessarily a saturating diffusive process; other models that explain this effect will arise, and new experiments must be designed to elucidate the true mechanism. For now, in the 1990's, 130 years after Mach's landmark treatment of spatial vision, it is important to see whether his theory describes retinal processing generally, or whether it describes retinal processing only in limited, specific circumstances.

Many models of color vision propose that the visual system normalizes to grey. *Aperture* experiments support this proposal, but everyday experience does not. Many models of vision propose that visual signals mix and interact only locally, over some small portion of the visual field. Everyday experience is better explained by normalization to white (as suggested by Land and McCann) with the normalizing reference propagated over large portions of the visual field. Cues from the visual periphery, such as the bridge of the nose, affect our perception in the center of gaze. This set of facts presents a challenge to neural scientists: are there anatomical substrates and physiological processes that can distribute a normalizing reference widely over the visual field?

8.3 Real-time Study

It is easy to be fooled by paper models and computer simulations. Counterexamples that may disprove a model do not always come to mind in an armchair experiment. A real-time study, however, provides the opportunity to readily test a model with many, many examples. Crucial counterexamples arise through directed inquiry, and, perhaps more importantly, through serendipity when one is "playing" with a real time system. In this age, when with a bit of sweat it is possible to build a real time system to test models, perhaps scientists should be wary of intellectual constructs which have not been tested in this ultimate, objective crucible.

References

[Allman and Zucker, 1990] Allman, J. and Zucker, S. (1990). Cytochrome oxidase and functional coding in primate striate cortex: a hypothesis. *Cold Spring Harbor Symp. Quant. Biol.*, 55:979–982.

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

[Arend and Reeves, 1986] Arend, L. and Reeves, A. (1986). Simultaneous color constancy. *J. Opt. Soc. Am.*, A3:1743.

[Arend et al., 1971] Arend, L.E., Buehler, J.N., and Lockhead, G.R. (1971). Difference information in brightness perception. *Perception and Psychophysics*, 9:367-370.

[Barlow, 1972] Barlow, H.B. (1972). Dark and light adaptation: psychophysics In Jameson, D. and Hurvich, L.M., editors, *Handbook of Sensory Psychology VII/4: Visual Psychophysics*, pages 3-28. Springer Verlag, Berlin.

[Barlow et al., 1978] Barlow, R.B., Jr., Snodderly, D.M., and Swadlow, H.A. (1978). Intensity coding in primate visual system. *Exp. Brain Res.*, 31:163-167.

[Bartleson and Breneman, 1967a] Bartleson, C.J. and Breneman, E.J. (1967). Brightness perception in complex fields. *J. Opt. Soc. Am.*, 57:953-957.

[Bartleson and Breneman, 1967b] Bartleson, C.J. and Breneman, E.J. (1967). Brightness reproduction in the photographic process. *Phot. Sci. Eng.*, 11:254-262.

[Bartleson and Huboi, 1956] Bartleson, C.J., and Huboi, R.W. (1956). Exposure determination methods for color printing: the concept of optimal correction level. *J. SMPTE*, 65:205-215.

[Bayer and Powell, 1986] Bayer, B.E. and Powell, P.G. (1986). A method for the digital enhancement of unsharp, grainy photographic images. *Adv. Computer Vision Image Processing*, 2:31-88.

[Bekesy, 1968a] Bekesy, G. von (1968). Brightness distribution across the Mach bands measured with flicker photometry, and the linearity of sensory nervous interaction. *J. Opt. Soc. Am.*, 58:1-8.

[Bekesy, 1968b] Bekesy, G. von (1968). Mach- and Hering-type lateral inhibition in vision. *Vision Res.*, 8:1483-1499.

[Blackwell and Buchsbaum, 1988] Blackwell, K.T. and Buchsbaum, G. (1988). Quantitative studies of color constancy. *J. Opt. Soc. Am.*, A5:1772.

[Blake, 1985] Blake, A. (1985). Boundary conditions for lightness computation in mondrian world. *Comp. Vis. Graph. Image Proc.*, 32:32.

[Brainard and Wandell, 1986] Brainard, D.H. and Wandell, B.A. (1986). Analysis of the retinex theory of color vision. *J. Opt. Soc. Am.*, A3:1611.

[Brooks and Cosgrove, 1970] Brooks, R.R. and Cosgrove, W.J. (1970). Combed aperture equalization for color television cameras. *J. SMPTE*, 79:21-25.

[Buchsbaum, 1980] Buchsbaum, G. (1980). A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1.

[Burr, 1987] Burr, D.C. (1987). Implications of the Craik-O'Brien illusion for brightness perception. *Vision Res.*, 27:1903-1913.

[Cornsweet, 1970] Cornsweet, T.N. (1970). *Visual Perception*, . Academic Press, New York.

[Creutzfeldt et al., 1987] Creutzfeldt, O., Lange-Malecki, B., and Wortmann, K. (1987). Darkness induction, retinex, and cooperative mechanisms in vision. *Exp. Brain Res.*, 67:270.

[Curlander, 1974] Curlander, P.J. (1974). *Image enhancement using digital adaptive filtering.* Master's Thesis, Massechusetts Institute of Technology.

[Desimone et al., 1985] Desimone, R., Schein, S.J., Moran, J., and Ungerleider, L.G. (1985). Contour, color, and shape analysis beyond the striate cortex. *Vision Res.*, 25:441.

[DeYoe and Van Essen, 1988] DeYoe, E.A. and Van Essen, D.C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.*, 11:219.

[D'Zmura and Lennie, 1986] D'Zmura, M. and Lennie, P. (1986). Mechanisms of color constancy. *J. Opt. Soc. Am.*, A3:1662.

[Enroth-Cugell and Robson, 1966] Enroth-Cugell, C., and Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells in the cat. *J. Physiol.*, 187:517-522.

[Evans, 1943] Evans, R.M. (1943). Visual processes and color photography. *J. Opt. Soc. Am.*, 33:579-614.

[Evans, Hanson, and Brewer, 1953] Evans, R.M., Hanson, W.T., and Brewer, W.L., Jr. (1953). *Principles of Color Photography.* Wiley, New York.

[Farnell, 1966] Farnell, G.C. (1966). The relationship between density and exposure. In Mees, C.E.K, and James, T.H., editors, *The theory of the photographic process, Second Edition.* MacMillan, New York.

[Faugeras, 1979] Faugeras, O.D. (1979). Digital color image processing within the framework of a human visual model. *IEEE Trans. Acous. Speech Sig. Proc.*, 27:380.

[Fink, 1941] Fink, D.G. (1941). Brightness distortion in television. *Proc. IRE*, 29:310-321.

[Fink, 1955] Fink, D.O. (1955). Color television vs. color motion pictures. *J. SMPTE*, 64:281-290.

[Fiorentini, 1972] Fiorentini, A. (1972). Mach band phenomena. In Jameson, D. and Hurvich, L.M., editors, *Handbook of Sensory Psychology VII/4: Visual Psychophysics*, pages 3-28. Springer Verlag, Berlin.

[Fiorentini et al., 1990] Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P., and Thomas, J.P. (1990). The perception brightness and darkness: relations to neuronal receptive fields. In Spillman, L. and Werner, J.S., editors, *Visual Perception: The Neurophysiological Foundations*, pages 129-160, Academic Press, New York.

[Frankle and McCann, 1983] Frankle, J.A. and McCann, J.J. (1983). Method and apparatus for lightness imaging. *US Patent 4,384,336.*

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

[Gilchrist, 1988] Gilchrist, A.L. (1988). Lightness contrast and failures of constancy: A common explanation. *Percep. Psychophys.*, 43:415-424.

[Grossberg and Mingolla, 1985] Grossberg, S., and Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psych. Rev.*, 92:92.

[Harris, 1991] Harris, J.G. (1991). *Analog models for early vision.* Ph.D. Thesis, California Institute of Technology.

[Harris et al., 1989] Harris, J., Koch, C., Luo, J. and Wyatt, J. (1989). Resistive fuses: analog hardware for detecting discontinuities in early vision. In Mead, C.and Ismail, M., editor, *Analog VLSI Implementations of Neural Systems.* Kluwer, Norwell, MA.

[Harris et al., 1990] Harris, J. G., Koch, C., and Luo, J. (1990). A two-dimensional analog VLSI circuit for detecting discontinuities in early vision. *Science*, 248:1209–1211.

[Heinemann, 1972] Heinemann, E.H. (1972). Simultaneous brightness induction. In Jameson, D. and Hurvich, L.M., editors, *Handbook of Sensory Psychology VII/4: Visual Psychophysics.* Springer Verlag, Berlin.

[Heinemann, 1989] Heinemann, E.G. (1989). Brightness contrast, brightness constancy, and the ratio principle. *Percep. Psychophys.*, 45:89-91.

[Helmholtz, 1911] Helmholtz, H. von (1911/1962). *Treatise on Physiological Optics, Volume 2, Third Edition.* Southall, J.P.C., Editor, Dover, New York.

[Helson, 1938] Helson, H. (1938). Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. Exp. Psychol.*, 23:439-476.

[Helson, 1943] Helson, H. (1943). Some factors and implications of color constancy. *J. Opt. Soc. Am.*, 33:555-567.

[Herbst et al., 1951] Herbst, P.J., Drew, R.O., and Johnson, S.W. (1951). Electrical and photographic compensation in television film reproduction. *J. SMPTE*, 57:289-307.

[Herbst et al., 1952] Herbst, P.J., Drew, R.O., and Brumbaugh, J.M. (1952). Factors affecting the quality of kinerecording. *J. SMPTE*, 58:85-104.

[Horn, 1974] Horn, B.K.P. (1974). Determining lightness from an image. *Comput. Graph. Image Proc.*, 3:277.

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

[Hubel and Weisel, 1968] Hubel, D.H. and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, 195: 215-243.

[Hunt, 1967] Hunt, R.W.G. (1967). *The Reproduction of Color.* London, John Wiley and Sons.

[Hurlbert, 1986] Hurlbert, A. (1986). Formal connections between lightness algorithms. *J. Opt. Soc. Am.*, A3:1684.

[Hurlbert and Poggio, 1987] Hurlbert, A. and Poggio, T. (1987). Learning a color algorithm from examples. Technical report, MIT Artificial Intelligence Laboratory, Cambridge, MA. AI Memo No. 909.

[Hurvich, 1981] Hurvich, L.M. (1981). *Color Vision.* Sindauer Associates, Sunderland, MA.

[Hutchinson et al., 1988] Hutchinson, J., Koch, C., Luo, J., and Mead, C. (1988). Computing motion using analog and binary resistive networks. *IEEE Computer*, 21:52–63.

[Ingle, 1985] Ingle, D. (1985). The goldfish as a retinex animal. *Science*, 227:651.

[Jacobsen and Gilchrist, 1988] Jacobsen, A. and Gilchrist, A. (1988). The ratio principle holds over a million-to-one range of illumination. *Percep. Psychophys.*, 43:415-424.

[Jameson and Hurvich, 1961] Jameson, D. and Hurvich, L.M. (1961). Complexities of perceived brightness. *Science*, 133:174-179.

[Jameson and Hurvich, 1989] Jameson, D. and Hurvich, L.M. (1989). Essay concerning color constancy. *Ann. Rev. Psychol.*, 40:1-22.

[Jones, 1944] Jones, L. A. (1944). Psychophysics and photography. *J. Opt. Soc. Am.*, 34:66-88.

[Jones and Condit, 1948] Jones, L. A. and Condit, H. R. (1948). The brightness scale of exterior scenes and the computation of the correct photographic exposure. *J. Opt. Soc. Am.*, 31:651-678.

[Judd et al., 1964] Judd, D.B., MacAdam, D.L., and Wysecki, G. (1964). Spectral distribution of typical daylight as a function of color temperature. *J. Opt. Soc. Am.*, 54:1031.

[Kallmann, 1940] Kallmann, H.E. (1940). The gradation of television pictures. *Proc. IRE*, April, 1940, 170-174.

[Kayama et al., 1979] Kayama, Y., Riso, R.R., Bartlett, J.R., and Doty, R.W. (1979). Luxotonic responses of units in macaque striate cortex. *J. Neurophysiol.*, 42:1495-1517.

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

[Land, 1959] Land, E.H. (1959). Color vision and the natural image I. *Proc. Natl. Acad. Sci. USA*, 45:115.

[Land, 1983] Land, E.H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proc. Natl. Acad. Sci. USA*, 80:5163.

[Land, 1986] Land, E.H. (1986). An alternative technique for the computation of the designator in the retinex theory of color vision. *Proc. Natl. Acad. Sci. USA*, 83:3078.

[Land and Daw, 1962] Land, E.H. and Daw, N.W. (1962). Colors seen in a flash of light. *Proc. Natl. Acad. Sci. USA*, 48:1000-1008.

[Land and McCann, 1971] Land, E.H. and McCann, J.J. (1971). Lightness and retinex theory. *J. Opt. Soc. Am.*, 61:1.

[Lappalainen and Hameenaho, 1976] Lappalainen, P. and Hameenaho, V. (1976). On the use of a television system in image reproduction: Part 1 - the image processing system. *IEEE Trans. Broadcasting*, BC-22:109-112.

[Leistner, 1960] Leistner, K. (1960). A sensitometric method for measuring the masking effect of dodging printers. *Phot. Sci. Eng.*, 4:44-47.

[Lennie and D'Zmura, 1988] Lennie, P. and D'Zmura, M. (1988). Mechanisms of color vision. *CRC Crit. Rev. Neurobiol.*, 3:333.

[Lisk, 1970] Lisk, K.G. (1970). A survey of home color television receivers: correlated color temperature, brightness, ambient light. *J. SMPTE*, 79:26-27.

[Livingstone and Hubel, 1984] Livingstone, M.S. and Hubel, D.H. (1984). Anatomy and physiology of a color system in primate primary visual cortex. *J. Neurosci.*, 4:309-364.

[Lowry and Jarvis, 1956] Lowry, E.M. and Jarvis, J.G. (1956). The luminance of subjective black. *J. SMPTE*, 63:411-414.

[Luo et al., 1989] Luo, J., Koch, C., and Mead, C. (1989). An experimental subthreshold, analog CMOS two-dimensional surface interpolation circuit. Neural Information Processing Systems Conference, Denver.

[Mach, 1897] Mach, E. (1897). *Contributions to the analysis of the sensations.* C.M. Williams, translator, Open Court, La Salle, IL.

[Maguire and Bazier, 1975] Maguire, W.M. and Baizer, J.S. (1975). Luminance coding of briefly presented stimuli in area 17 of the Rhesus monkey. *J. Neurophysiol.*, 47:128-137.

[Maloff, 1939] Maloff, I.G. (1939). Gamma and range in television. *RCA Review*, 3:409-417.

[Maloney, 1984] Maloney, L.T. (1984). *Computational approaches to color constancy.* Ph.D. Thesis, Stanford University.

[Maloney, 1986] Maloney, L.T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. Am.*, A3:1673.

[Maloney and Wandell, 1986] Maloney, L.T., and Wandell, B.A. (1986). Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am.*, A3:29.

[Marimont, 1962] Marimont, R.B. (1962). Model for visual response to contrast. *J. Opt. Soc. Am.*, 52:800-806.

[Marimont, 1963] Marimont, R.B. (1963). Linearity and the Mach phenomenon. *J. Opt. Soc. Am.*, 53:400-401.

[Marr, 1974] Marr, D. (1974). The computation of lightness by the primate retina. *Vision Res.*, 14:1377.

[Marr, 1982] Marr, D. (1982). *Vision.* W.H. Freeman, New York.

[Marrocco, 1975] Marrocco, R.T. (1975). Possible neural basis of brightness magnitude estimation. *Brain Res.*, 86:128-133.

[McCann, 1987] McCann, J.J. (1987). Local/global mechanisms for color constancy. *Die Farbe*, 34:275-283.

[McCann, 1989] McCann, J.J. (1989). The role of simple nonlinear operations in modeling human lightness and color sensations. *Proc. SPIE vol. 1077, Human Vision, Visual Processing, and Digital Display*, 1077:355-363.

[McCann, McKee, and Taylor, 1976] McCann, J.J., McKee, S., and Taylor, T. (1976). Quantitative studies in retinex theory. *Vision Res.*, 16:445-458.

[McMann and Goldberg, 1968] McMann, R.H., Jr., and Goldberg, A.A. (1968). Improved signal processing techniques for color television broadcasting. *J. SMPTE*, 77:221-228.

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

[Mees, 1960] Mees, C.E.K. (1960). *The Theory of the Photographic Process, Second Edition.* Macmillan, New York.

[Moon and Cettei, 1938] Moon, P., and Cettei, M.S. (1938). On the reflectance factor of clothing. *J. Opt. Soc. Am.*, 28:277-279.

[Moon and Spencer, 1954] Moon, P. and Spencer, D. E. (1954). The photometric range of outdoor scenes. *J. SMPTE*, 63:237-239.

[Moore and Goodman, 1990] Moore, A. and Goodman, R. (1990). Image smoothing at video rates with analog VLSI. *Proc. IEEE SMC*, Nov.

[Moore et al., 1991] Moore, A., Allman, J., and Goodman, R.M. (1991). A real-time neural system for color constancy. *IEEE Trans. Neural Networks*, 2:237-247.

[Nelson, 1966] Nelson, C.N. (1966). The theory of tone reproduction. In Mees, C.E.K. and James, T.H., editors, *The Theory of the Photographic Process, Third Edition*, MacMillan, New York.

[Nelson, 1977] Nelson, C.N. (1977). Tone and color reproduction. I. Tone reproduction. In James, T.H., editor, *The Theory of the Photographic Process, Fourth Edition*, MacMillan, New York.

[Neuhauser, 1957] Neuhauser, R.C. (1957). Black level - the lost ingredient in television picture fidelity. *J. SMPTE*, 66:597-601.

[Nikoh et al., 1991] Nikoh, H., Kobayashi, M., Kuwajima, T., Ohsawa, K., Kitano, Y., Shimizu, T., and Kanno, K. (1991). New digital signal processor for video camera system. *NEC Res. Dev.*, 32:214-219.

[Normann and Pearlman, 1979] Normann, R.A. and Perlman, I. (1979). The effects of background illumination on the photoreceptors of red and green cones. *J. Physiol.*, 296:491-507.

[Oppenheim, Schafer, and Stockham, 1968] Oppenheim, A.V., Shafer, R.W., and Stockham, T.G., Jr. (1968). Non-linear filtering of multiplied and convolved signals. *Proc. IEEE*, 56:1264.

[Percy and Veal, 1954] Percy, F.T. and Veal, T.G. (1954). Subject lighting contrast for color photographic films in color television. *J. SMPTE*, 63:90-94.

[Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Patt. Anal. Mach. Intell.*, 6:629-638.

[Ratliff, 1965] Ratliff, F. (1965). *Mach Bands: Quantitative Studies on Neural Networks in the Retina.* Holden-Day, San Francisco.

[Ratliff, 1984] Ratliff, F. (1984). Why Mach bands are not seen at the edges of a step. *Vision Res.*, 24:163-165.

[Ross et al., 1989] Ross, J.M., Morrone, M.C., and Burr, D.C. (1989). The conditions under which Mach bands are visible. *Vision Res.*, 29:699-715.

[Sakmann and Creutzfeldt, 1969] Sakmann, B. and Creutzfeldt, O.D. (1969). Scotopic and mesopic light adaptation in the cat's retina. *Pflugers Arch.*, 313:168-185.

[Schein and Desimone, 1990] Schein, S.J. and Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *J. Neurosci.*, 10:3369.

[Schnapf et al., 1990] Schnapf, J.L., Nunn, B.J., Meister, M., and Baylor, D.A. (1990). Visual transduction in cones of the monkey *Macaca Fascicularis. J. Physiol.*, 427:681-713.

[Schreiber, 1978] Schreiber, W.F. (1978). Image processing for quality improvement. *Proc. IEEE*, 66:1640-1651.

[Seige and Ress, 1987] Seige, P. and Ress, G. (1987). Application of Texas Instruments TC-104 linear charge-coupled-device arrays in spaceborne camera systems. *Opt. Eng.*, 26:1029-1034.

[Shapely and Enroth-Cugell, 1984] Shapely, R.M and Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3:263-346.

[Stevens and Stevens, 1963] Stevens, J.C. and Stevens, S.S. (1963). Brightness function: effects of adaptation. *J. Opt. Soc. Am.*, 53:375-385.

[Stimson and Fee, 1953] Stimson, A., and Fee, E. (1953). Color and reflectance of human flesh. *J. SMPTE*, 60:553-558.

[Stockham, 1982] Stockham, T.G., Jr. (1982). Image processing in the context of a visual model. *Proc. IEEE*, 60:828-842.

[Tanner and Mead, 1987] Tanner, J. and Mead, C. (1987). An integrated optical motion sensor. In Kung, S-Y., Owen, R. and Nash, J., editors, *VLSI Signal Processing II.* IEEE Press, New York.

[Thouless, 1931a] Thouless, R.H. (1931). Phenomenal regression to the real object I. *Brit. J. Psychol.*, 21:339-359.

[Thouless, 1931b] Thouless, R.H. (1931). Phenomenal regression to the 'real' object II. *Brit. J. Psychol.*, 22:1-29.

[Tolhurst, 1972] Tolhurst, D.J. (1972). On the possible existence of edge detector neurones in the human visual system. *Vision Res.*, 12:797-804.

[Ts'o and Gilbert, 1988] Ts'o, D.Y. and Gilbert, C.D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *J. Neurosci.*, 8:1712.

[Virsu and Lee, 1983] Virsu, V. and Lee, B.B. (1983). Light adaptation in the cells of macaque lateral geniculate nucleus and its relation to human light adaptation. *J. Neurophysiol.*, 50:864-878.

[Wallace and Lockhead, 1987] Wallace, W.T., and Lockhead, G.R. (1987). Brightness of luminance changes with gradual changes. *Vision Res.*, 9:1589-1601.

[Walraven et al., 1990] Walraven, J., Enroth-Cugell, C., Hood, D.C., MacLeod, D. I. A., and Schanpf, J.L. (1990). The control of visual sensitivity: receptoral and postreceptoral processes. In Spillman, L. and Werner, J.S., editors, *Visual Perception: The Neurophysiological Foundations*, pages 53-101, Academic Press, New York.

[Wild et al., 1985] Wild, H.M., Butler, S.R., Carden, D., and Kulikowski, J.J. (1985). Primate cortical area V4 important for colour constancy but not wavelength discrimination. *Nature (London)*, 313:133.

[Worthey, 1985] Worthey, J.A. (1985). Limitations of color constancy. *J. Opt. Soc. Am.*, A2:1014.

[Yule, 1967] Yule, J.A.C. (1967). *Principles of Color Reproduction.* John Wiley and Sons, New York.

[Zeki, 1980] Zeki, S.M. (1980). The representation of colours in the cerebral cortex. *Nature (London)*, 284:412.

[Zeki, 1983a] Zeki, S.M. (1983). Colour coding in the cerebral cortex: the reaction of cells in the monkey visual cortex to wavelengths and colours. *Neuroscience*, 9:9.

[Zeki, 1983b] Zeki, S.M. (1983). Colour coding in the cerebral cortex: the responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience*, 9:39.