Properties and Performance of a Center/Surround Retinex

Daniel J. Jobson, Zia-ur Rahman, *Member, IEEE*, and Glenn A. Woodell

***Abstract*—The last version of Land's retinex model for human vision's lightness and color constancy has been implemented and tested in image processing experiments. Previous research has established the mathematical foundations of Land's retinex but has not subjected his lightness theory to extensive image processing experiments. We have sought to define a practical implementation of the retinex without particular concern for its validity as a model for human lightness and color perception. Here we describe the trade-off between rendition and dynamic range compression that is governed by the surround space constant. Further, unlike previous results, we find that the placement of the logarithmic function is important and produces best results when placed after the surround formation. Also unlike previous results, we find best rendition for a "canonical" gain/offset applied after the retinex operation. Various functional forms for the retinex surround are evaluated, and a Gaussian form found to perform better than the inverse square suggested by Land. Images that violate the gray world assumptions (implicit to this retinex) are investigated to provide insight into cases where this retinex fails to produce a good rendition.**

I. Introduction

OF THE MANY visual tasks accomplished so gracefully by human vision, one of the most fundamental and approachable for machine vision applications is lightness and color constancy. While a completely satisfactory definition is lacking, lightness and color constancy refer to the resilience of perceived color and lightness to spatial and spectral illumination variations. Various theories for this have been proposed and have a common mathematical foundation [1]. The last version of Land's retinex [2] has captured our attention because of the ease of implementation and manipulation of key variables, and because it does not have "unnatural" requirements for scene calibration. Likewise, the simplicity of the computation was appealing and initial experiments produced compelling results. This version of the retinex has been the subject of previous digital simulations that were limited because of lengthy computer time involved and was implemented in analog very large-scale integrated circuits (VLSI) to achieve real-time computation [3], [4]. Evidence that this retinex version is an optimal solution to the lightness problem has come from experiments posing Land's Mondrian target, randomly arranged two-dimensional (2-D) gray patches,

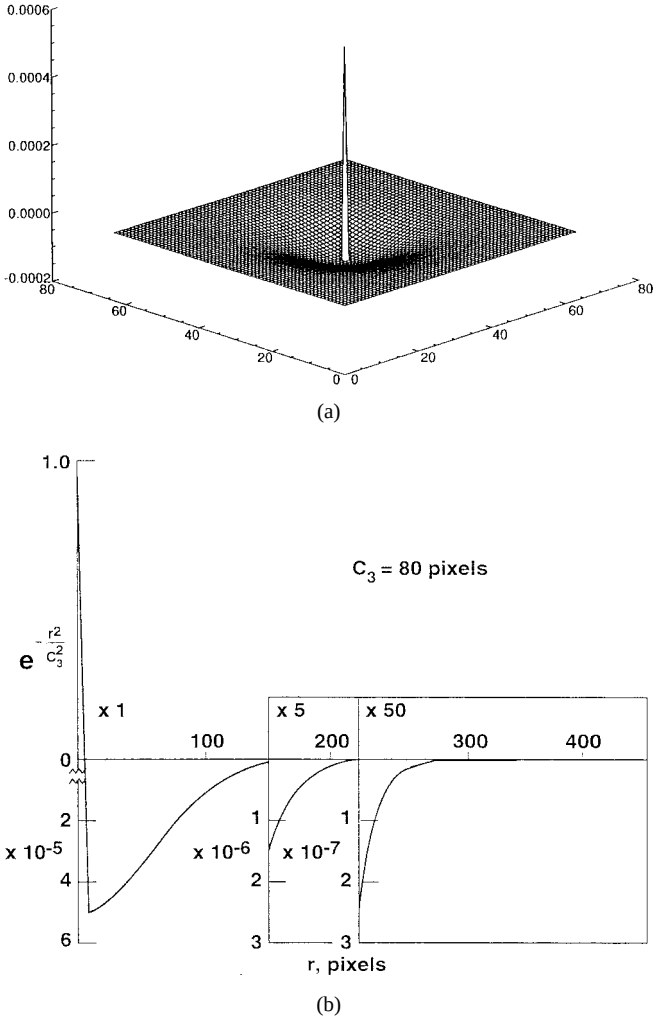


Fig. 1. Spatial form of the center/surround retinex operator. (a) 3-D representation (distorted to visualize surround). (b) Cross-section to illustrate wide weak surround.

as a problem in linear optimization and a learning problem for back propagated artificial neural networks [5], [6].

The utility of a lightness–color constancy algorithm for machine vision is the simultaneous accomplishment of:

1) dynamic range compression;
2) color independence from the spectral distribution of the scene illuminant;
3) color and lightness rendition.

Land's center/surround retinex demonstrably achieves the first two, although Land emphasized primarily the color constancy properties. Well-known difficulties arise, though, for

Manuscript received June 26, 1995; revised May 24, 1996. The work of Z. Rahman was supported by NASA Langley Research Center under Contract NAS1-19 603. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Moncef Gabbouj.

D. J. Jobson and G. A. Woodell are with NASA Langley Research Center, Hampton, VA 23681-0001 USA (e-mail: d.j. jobson@larc.nasa.gov).

Z. Rahman is with Science and Technology Corporation, Hampton, VA 23666 USA.

Publisher Item Identifier S 1057-7149(97)00428-4.

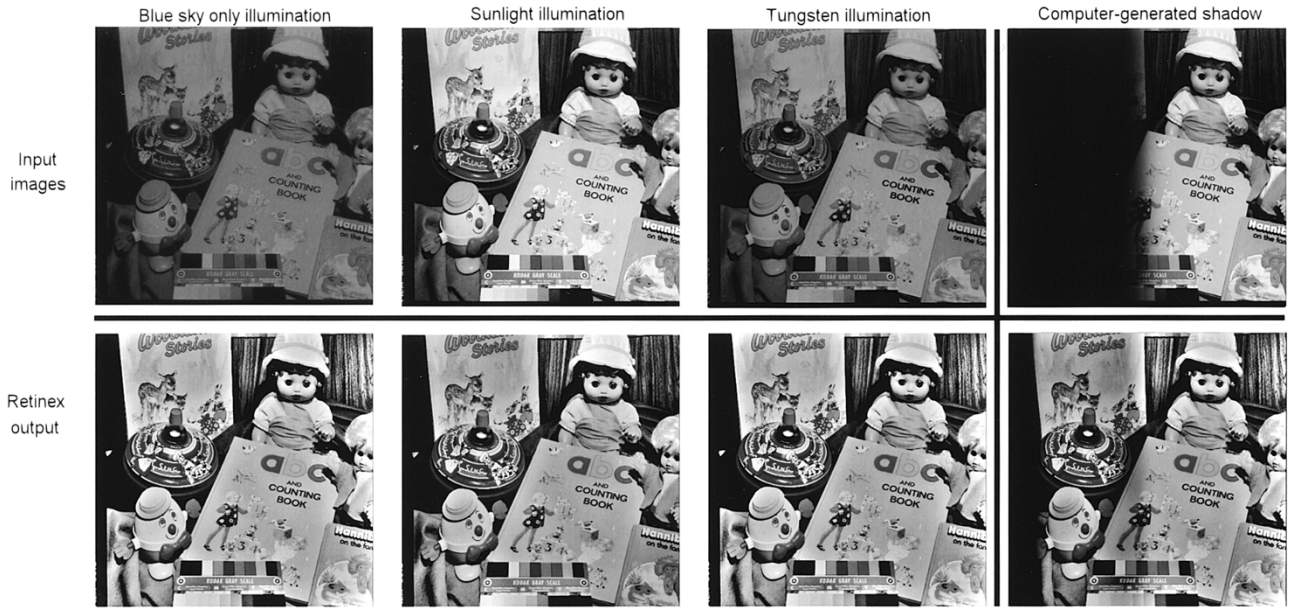


Fig. 2. Demonstration of retinex color constancy and dynamic range compression (prior to optimizing rendition) for a Gaussian surround with small space constant (15 pixels).

color and lightness rendition [1], [3], [6]. These consist of i) lightness and color "halo" artifacts that are especially prominent where large uniform regions abut to form a high contrast edge with "graying" in the large uniform zones in an image, and ii) global violations of the gray world assumption (e.g., an all-red scene) which result in a global "graying out" of the image. Clearly, the retinex (perhaps like human vision) functions best for highly diverse scenes and poorest for impoverished scenes. This is analogous to systems of simultaneous equations where a unique solution exists if and only if there are enough independent equations.

The general form of the center/surround retinex (Fig. 1) is similar to the difference-of-Gaussian (DOG) function widely used in natural vision science to model both the receptive fields of individual neurons and perceptual processes. The only extensions required are i) to greatly enlarge and weaken the surround Gaussian (as determined by its space and amplitude constants), and ii) to include a logarithmic function to make subtractive inhibition into a shunting inhibition (i.e., arithmetic division). We have chosen a Gaussian surround form whereas Land opted for a $1/r^2$ function [2] and Moore *et al*. [3] used a different exponential form. These will be compared in Section II. Mathematically, this takes the form

$$R_i(x,y) = \log\ I_i(x,y) - \log\left[F(x,y) * I_i(x,y)\right] \qquad (1)$$

where $I_i(x,y)$ is the image distribution in the ith color spectral band, "$*$" denotes the convolution operation, $F(x,y)$ is the surround function, and $R_i(x,y)$ is the associated retinex output.

This operation is performed on each spectral band to produce Land's triplet values specifying color and lightness. It is readily apparent that color constancy (i.e., independence from single source illuminant spectral distribution) is reasonably complete since

$$I_i(x,y) = S_i(x,y) r_i(x,y) \qquad (2)$$

where $S_i(x,y)$ is the spatial distribution of the source illumination and $r_i(x,y)$, the distribution of scene reflectances (integrated over the spectral band response), so that

$$R_i(x,y) = \log \frac{S_i(x,y) r_i(x,y)}{\overline{S}_i(x,y) \overline{r}_i(x,y)} \qquad (3)$$

where the bars denote the spatially weighted average value. As long as $S_i(x,y) \approx \overline{S}_i(x,y)$, then

$$R_i(x,y) \approx \log \frac{r_i(x,y)}{\overline{r}_i(x,y)}. \qquad (4)$$

The approximate relation is an equality for many cases and, for those cases where it is not strictly true, the reflectance ratio should dominate illumination variations.

Color constancy is demonstrated (Fig. 2) for the extreme cases of blue skylight illumination, direct sunlight only, and tungsten illumination. Actual daylight illumination should fall arbitrarily somewhere between the first two cases. Film and electronic cameras without computational intervention or film selection would produce the top row of images. Dynamic range compression is also readily demonstrated (Fig. 2, right) with computer simulation. Here the original image data is multiplied by a hyperbolic tangent "shadow." Again, cameras without computation produce the upper result (or with a change of f/stop or exposure would bring out the shadowed detail but at the expense of saturating the nonshadowed image zones). Strikingly, color balance is retained across the wide dynamic range encompassed and the highly nonlinear operation of the retinex.

These two examples do, however, point to the difficulty of realizing satisfactory color rendition in contrast to the ease of achieving color constancy and dynamic range compression. Taken together, this discussion indicates the exciting possibilities that motivated us to engage in more extensive investigation.

(a) (b)

(c) (d)

Fig. 3. Examples of serious photographic defects due to spectral and/or spatial illumination variations. (a) "Green" kitchen due to fluorescent illumination. (b) Sodium vapor illumination. (c) Tungsten indoors/daylight outdoors. (d) Obscured foreground.

The need for dynamic range compression and color constancy, especially if both are accomplished simultaneously by a simple real-time algorithm, is well known to photographers. Discrepancies between the photographer's perception through the viewfinder and the captured film image can be quite bizarre (Fig. 3), and require constant vigilance to avoid impossible lighting situations and to carefully select the appropriate film and processing for the illuminant's spectral distribution. The fundamental limit [3] is recognized to be the film or cathode ray tube's (CRT's) narrow dynamic range and static spectral response. Print/display dynamic range constraints of $50:1$ are, however, compatible with the magnitude of scene reflectance variations. Except for extreme cases (snow or lampblack) reflectance variations are only $20:1$ [7] and often much less. Thus, even the extremes of reflectance of $\approx 50:1$ are easily spanned by print/display media. Clearly illumination variations are the culprit which human visual perception has overcome by eye–brain computation. Electronic still cameras have an intrinsically high dynamic range ($> 2000:1$) [8] set by the detector array electronics, and an even higher dynamic range within the detector array proper, since the limiting factor is usually the preamplifier noise added in transferring image signals off-chip or digitization noise added subsequently. Therefore, at least for electronic still cameras,

Fig. 4. Demonstration of improved rendition obtained applying the log response *after* surround formation ($c_3 = 80$ pixels).

we can conclude that sufficient dynamic range is available to retain the full variations of both illumination and reflectance in arbitrary scenes. So it is certainly reasonable to consider either analog [3] implementations of compression/constancy or digital implementation if the initial A/D conversion is done at 10–14 bits (b), rather than the usual 8 b.

Recent advances in high-speed computing led us to reconsider both extensive digital simulations of the retinex and real-time digital implementations for practical use in future electronic camera systems. The hours of computer time previously reported [3] are now reduced to minutes and real-time implementations using specialized digital hardware such as digital signal processing (DSP) chips seem reasonable. In other words, the full image dynamic range is available from current electronic cameras, real-time computation is realizable, and the ultimate bottleneck is only at the first print/display. Obviously, there are image coding aspects to both dynamic range compression and color constancy. We will touch upon these briefly but concentrate primarily on the design of the algorithm to produce combined dynamic range compression/color constancy/color–lightness rendition.

We have seen that the center/surround retinex is both color constant and capable of a high degree of dynamic range compression. It remains, then, to specify an implementation that produces satisfactory rendition and examine alternatives to determine if other design options are equally good or better. Because the retinex exchanges illumination variations for scene reflectance context dependency [9], scene content becomes a major issue especially when it deviates from regionally gray average values—the "gray world" assumption [1]. Therefore, testing with diverse scenes, including random ones, is important to pinpoint possible limits to the generality of this retinex.

Initial image processing simulations revealed the following unresolved implementation issues:

1) the placement of the log function;
2) the functional form of the surround;
3) the space constant for the surround;
4) the treatment of the retinex triplets prior to display.

These will now be explored more comprehensively. The results of testing the optimized algorithm on diverse scenes will then be presented with special emphasis on "gray-world" violations. Finally, the relationship of the algorithm to neurophysiology will be examined briefly.

II. Issues

A. Placement of Log Function

Previous research [3], [6] has largely concluded that the logarithm can be taken before or after the formation of the surround. Processing schemes [3], [6], [10] adhering closely to natural vision science, i.e., an approximate log photoreceptor response, favor placing log response at the photodetection stage prior to any surround formation. Our preliminary testing of this produced rather disappointing results and prompted us to reopen this seemingly decided issue. Initial testing of the postsurround log produced encouraging results with much less emphatic artifacts. Mathematically, we have that

$$R_1 = \log I(x,y) - \log[I(x,y) * F(x,y)] \tag{5}$$

and

$$R_2 = \log I(x,y) - \{[\log I(x,y)] * F(x,y)\} \tag{6}$$

are not equivalent. The discrete convolution $[\log I(x,y) * F(x,y)]$ is, in fact, equivalent to a weighted *product* of $I(x,y)$, whereas the second term in (5) is a weighted *sum*. This is closely related to the difference between the arithmetic mean and the geometric mean except that $F(x,y)$ is selected so that

$$\iint F(x,y)\,dxdy = 1 \tag{7}$$

which does not produce exactly the nth root of n numbers as the geometric mean would. Since the entire purpose of

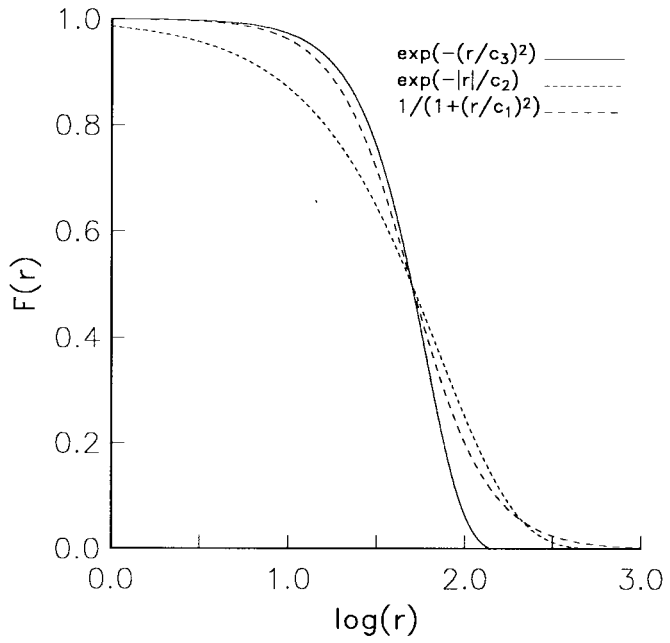


Fig. 5. Comparison of three surround functions—inverse square, exponential, and Gaussian, normalized to equal full-width half-max (FWHM) response. The log(r) scale is necessary for comparison purposes but does diminish the differences between the functions. A linear r scale (if it were graphically feasible) would show very dramatic differences. The space constants are $c_1 = 50$ pixels, $c_2 = 72$ pixels, and $c_3 = 60$ pixels.

the log operation is to produce a point by point ratio to a large regional mean value, (5) seems the desired form and our image processing experiments bear out this preference. A typical example is shown in Fig. 4. While the halo artifact for (6) can be diminished by manipulation of the gain and offset, this results in a significant desaturation of color. In other examples, more severe color distortions occur, which likewise cannot be removed by manipulation of the gain/offset. In addition, a shadow simulation indicates much less dynamic range compression for (6). Therefore, we have selected the (5) form for our testing and optimization. This form is also that given in Land's original presentation [2], though he is quoted as feeling the two forms were equally useful in practice [6].

B. The Surround Function

Land proposed an inverse square spatial surround

$$F(x',y') = 1/r^2 \tag{8}$$

where

$$r = \sqrt{x'^2 + y'^2}$$

which can be modified to be dependent on a space constant as

$$F'(x',y') = \frac{1}{1 + (r^2/c_1^2)}. \tag{9}$$

Moore *et al.* [3] examined an exponential "absolute value"

$$F(x',y') = e^{-|r|/c_2} \tag{10}$$

because it is an approximation to the spatial response of analog VLSI resistive networks, and Hurlbert [6] investigated the Gaussian:

$$F(x',y') = e^{-r^2/c_3^2} \tag{11}$$

because of its widespread use in natural and machine vision modeling. A cross section of these 2-D functions (Fig. 5) shows that for any particular choice of space constant, the inverse square rolls off very rapidly, but ultimately retains a higher response to quite distant image pixels than the exponential and Gaussian forms. At distant values, the exponential ultimately exceeds the Gaussian response, so that in general the inverse square is consistently more "global," the exponential is less so, and the Gaussian is more distinctively "regional."

In initial tests, no space constant for the inverse square surround could be found that achieved reasonable dynamic range compression, i.e., adequate enhancement of shadowed detail. The best performance is shown in Fig. 6. In contrast, both the exponential and Gaussian forms produced good dynamic range compression over a range of space constants. Because the Gaussian offered the most experimental flexibility (good performance over wider range of space constants), it was selected for this implementation. It is likely that the exponential is equally useful and this is clearly of importance for analog VLSI resistive network hardware implementations of retinex computations.

C. Surround Space Constant

While Land proposed the center/surround retinex with a 2–4 pixel diameter for the center (perhaps in keeping with the widely known coarser spatial resolution of purely chromatic vision), a center of only 1 pixel is clearly demanded for general-purpose image processing. Only after segmentation into lightness and chromatic images can the purely chromatic images be made coarser. In contrast, the surround space constant cannot be so clearly defined. Land proposed an inverse square surround with a full width-half maximum (FWHM) of 40° of visual angle. This corresponds to FWHM of about 270 visual pixels (assuming a visual pixel is $\approx 0.015°$). We examined the performance of the Gaussian surround over a wide range of space constants. Since previous research [6] found variations in the space constant with the spatial variation in shadow profiles, a particular concern is the question of an optimum space constant that gives good performance for diverse scenes and lighting conditions.

The image sequence (Fig. 7) established a trade-off that has not been previously studied. In varying the space constant from small to large values, dynamic range compression is sacrificed for improved rendition. The middle of this range ($50 \leq c_3 \leq 100$ pixels) represents a reasonable compromise, where shadows are fairly compensated and rendition achieves acceptable levels of image quality. This is qualitatively compatible with human visual perception in that the treatment of shadows is influenced by their spatial extent. Larger shadows tend to be more compensated (less dark) while smaller shadows appear less compensated (blacker and with less visible internal detail).

While we are not concerned with defining a form of the retinex that accurately models human vision, we must ultimately compare performance to that of human perception in order to meet basic image quality requirements. Our intent, then, is to find a form of the retinex that is functionally equivalent to human visual perception. Since the performance

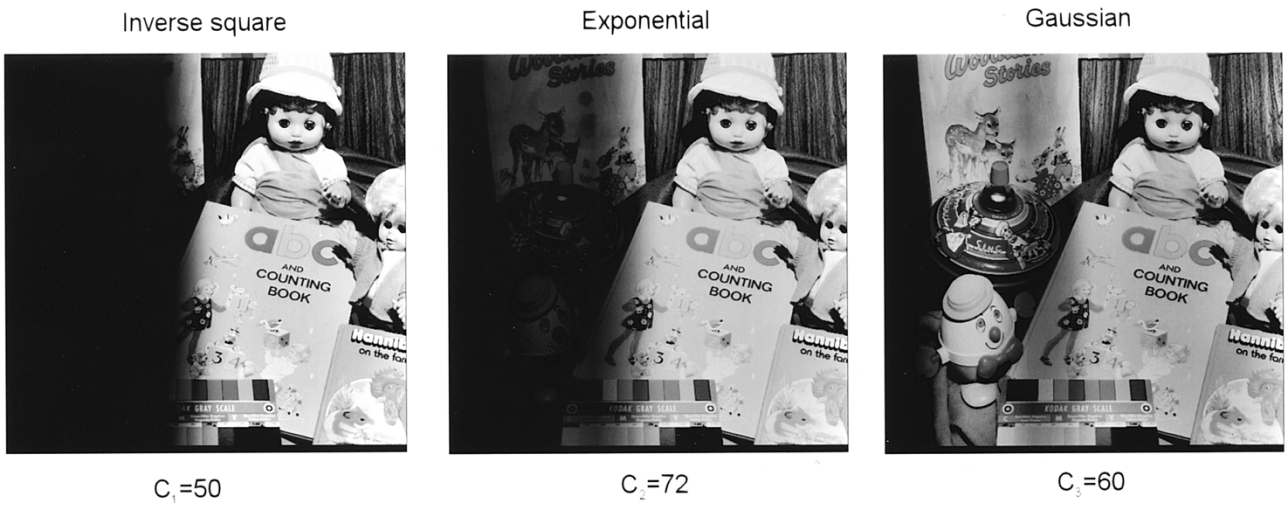


Fig. 6. Comparison of visual performance of three surround functions arranged from left to right in order of increasing dynamic range compression.

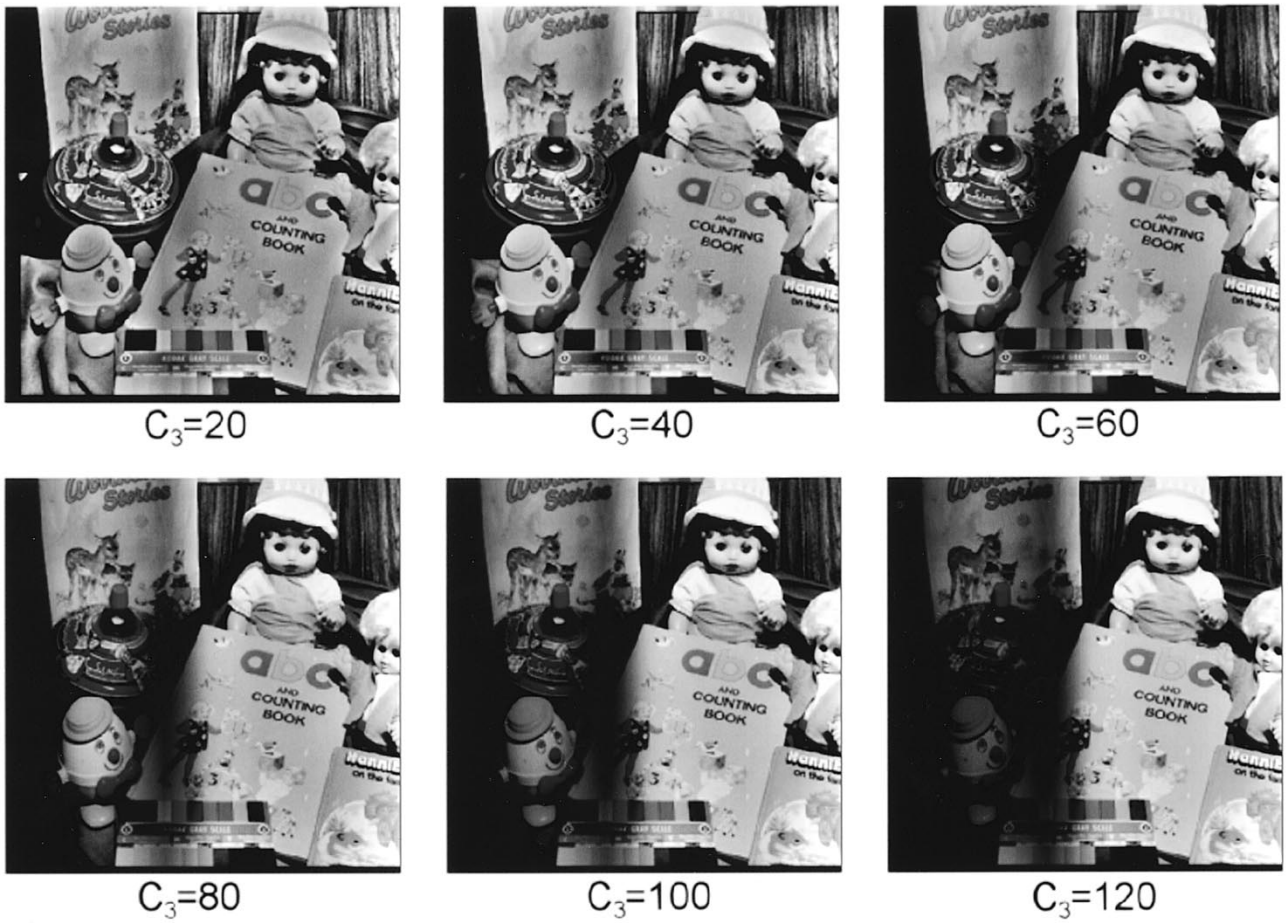


Fig. 7. Trade-off between dynamic range compression and color rendition for the Gaussian surround. Small space constants produce excellent dynamic range compression, while large constants produce the best rendition.

of human vision for complex natural images has not been comprehensively defined, we are left with purely subjective assessments of image quality. Since the retinex is, to some extent, compensating for lighting variations and approximating a "reflectance world," there are two directions available for assessment. First is the psychophysical comparison between the human observation *of the scene* to the processed and displayed image. Second is the quantitative comparison of the processed/displayed image to the measured scene reflectance values. The latter approach is replete with problems since lighting variations are clearly not completely removed by human visual perception. If, however, we pursue additional computation to segment lightness and chromatic images, the chromatic images are likely to be measures of relative spectral reflectance ratios that can be compared with scene reflectances to establish a figure-of-merit. Here, we will rely only on the

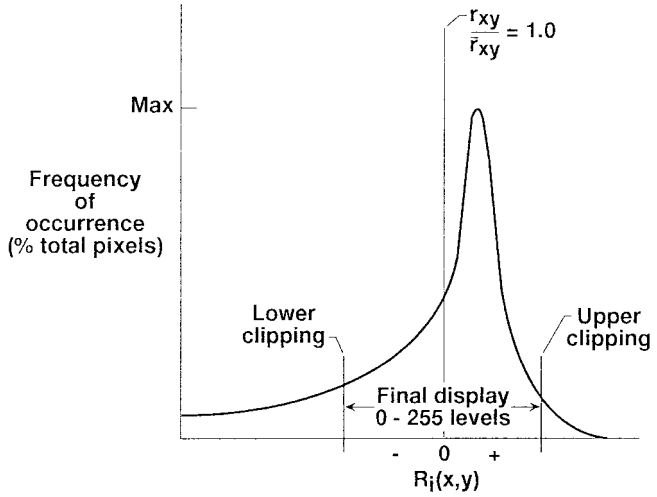


Fig. 8. Schematic of a characteristic retinex histogram illustrating the final gain/offset selection applied uniformly to the three color subimages.

first comparison, since we are examining the overall utility of the computation to enable electronic imagery to be as resilient as the human observer of the same scene and not lose or distort major semantic information that would have been obtained by direct observation. The second approach will, perhaps, become more important in scientific data analysis such as multispectral classification in remote sensing imaging.

D. Treatment of Retinex Output Prior to Display

During initial experiments, we were surprised to find a characteristic form for the histograms of diverse scenes after the retinex operation (Fig. 8). Exceptions were for severe violations of the "gray world" assumption, e.g., an all-red scene. These violations are explored in a subsequent section, so here we will examine a natural image with reasonable scene diversity.

Land's proposal [2] of the center/surround retinex does not explicitly address the issue of a final treatment with the possible implication that none is necessary. On the other hand, Moore *et al.* [3] advocate the automatic gain/offset approach, whereby the triplet retinex values are adjusted by the absolute maximum and minimum found across all values in all the color bands. Our own empirically derived approach (Fig. 8) differs from either of these in that a constant gain/offset is selected for best color rendition. This results in actually clipping some of both the highest and lowest signal transitions. Little information is lost because the retinex output signals form, to a large degree, a contrast image (being in essence a ratio). This constant gain/offset has thus far proven to be independent of image/scene content. Our approach, otherwise, agrees with Moore *et al.* in that a final gain and offset is uniformly applied to all pixels in all three color bands. A comparison of these two approaches is illustrated (Fig. 9) to underline the considerable visual differences encountered. We speculate that the significant deviations from the characteristic histogram that occur for gross violations of the gray-world assumption could be used to detect errors. The gain/offset appears to be invariant from image to image, so that we have the sense that it is canonical and, therefore, satisfies the original intent of Land to produce a general computation that applies to most images. The term "canonical" refers to the post-retinex gain/offset being general constants that do not vary either from image-to-image or between band-to-band.

E. Summary

The specific implementation we have defined from preliminary testing is a center/surround operation with the following characteristics:

1) the spatial extent of the center is the individual pixel, which can be thought of as a small Gaussian defined by the optical blur function of the imaging optics;
2) the form of the surround is Gaussian;
3) The spatial extent of the surround is that for a Gaussian space constant of about 80 pixels (which corresponds to an FWHM spread of 210 pixels);
4) the logarithm is applied after surround formation by 2-D spatial convolution;
5) a "canonical" gain/offset is applied to the retinex output which, in signal terms, clips some of the highest and lowest signal excursions. The gain and offset are general constants that do not vary either from image to image or between color bands.

Our implementation differs from previous ones in that Land [2] proposed an inverse square surround while Moore *et al.* [3] and Hurlbert [6] concentrated on placement of the log prior to surround formation (or else considered placement as interchangeable). Finally, Moore *et al.* specified an automatic gain/offset process rather than the canonical one used here. All of these differences were shown to result in significant visual effects on processed images.

III. Results

Because the mathematics, though simple, involve a nonlinearity coupled to large-scale spatial interactions, the performance on complex images is not predictable. The only recourse is to apply the method to diverse images in hopes of exposing limitations and distortions. The performance on images not meeting the regional gray-world assumption is examined to attempt to define ways to detect and minimize or correct errors if they occur in some systematic fashion. It should be clear that while the dynamic range compression and color constancy are readily achievable, the goal of rendition poses a great challenge.

Rendition is as difficult to define as it is to achieve. Our working definition is that rendition means producing a resultant displayed image that is convincingly like what a human observer would see when examining the same *scene* as the camera does. Therefore rendition means fidelity both to the scene and to human perception. This is by necessity a qualitative criterion because, while we can quantify the scene, the current state of color psychophysics does not provide the ability to quantify color perception in complex scenes. Our working criteria is to compare the original and processed images visually and, where possible, to compare the

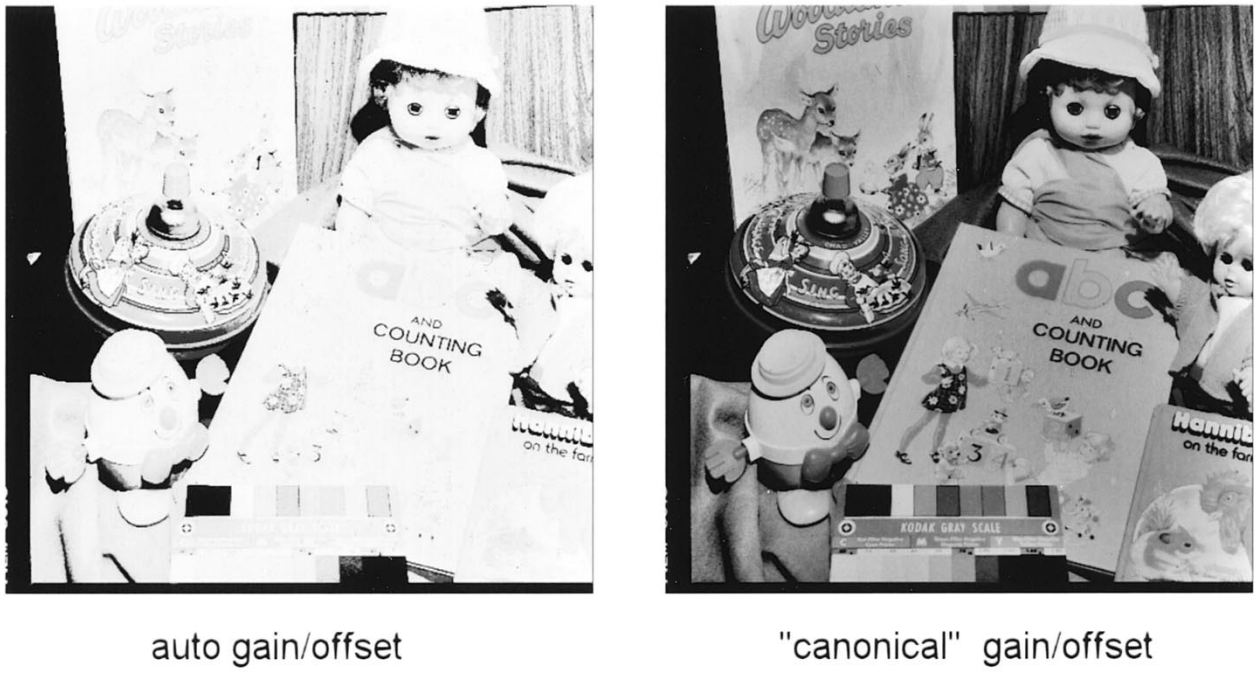


Fig. 9. Comparison of the visual performance of auto gain/offset versus "canonical" gain/offset. The auto gain/offset is selected on the absolute maximum and minimum values in all three color bands and applied uniformly to all three as a global operation. The "canonical" gain/offset accepts some clipping of extreme high and low values but provides superior rendition with minimal loss of visual information ($c_3 = 80$).

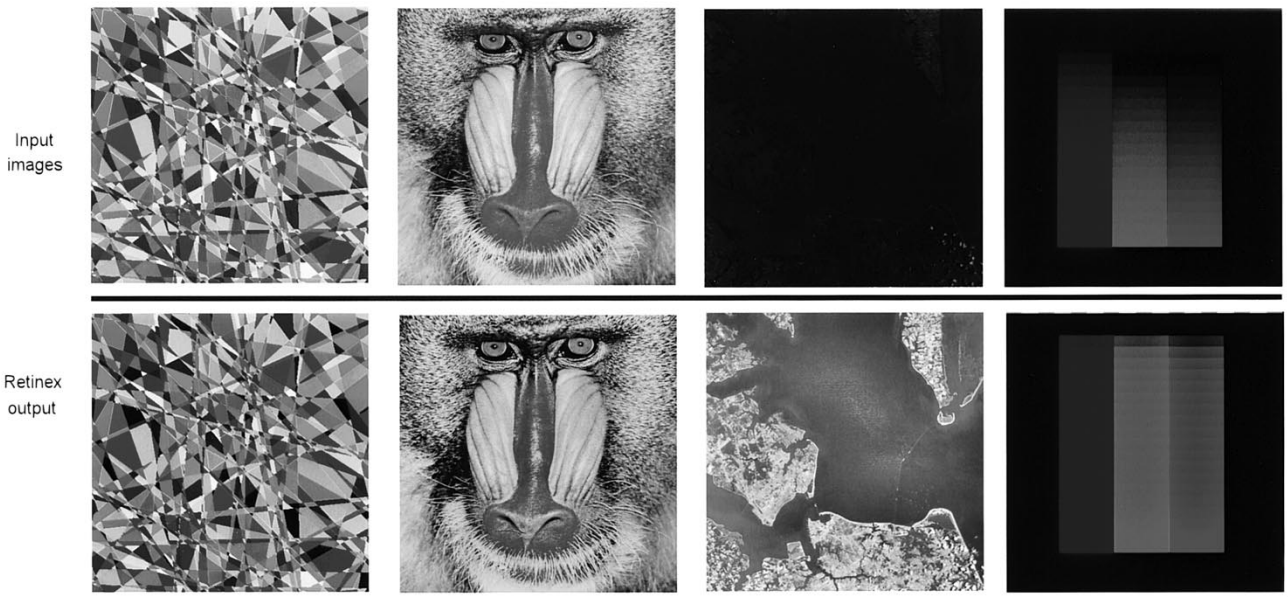


Fig. 10. Results for diverse test images—stochastic and deterministic; computer-generated; natural; and false-color ($c_3 = 80$ pixels).

processed image *to the scene*. While quantitative measures do exist for comparing input/output images, these do not capture the essential quality of visual significance. An abundance of psychophysical research underlines the central role of context in visual significance as well as the type of visual phenomena. We would like to admit and accept any distortion that the eye–brain does not find disturbing or perceptible. So we are left at this time with a reliance upon only visual perception in assessing the rendition in these experiments in retinex processing. From our own visual experience, we make the following statements about human visual perception:

1) The dynamic range compression of shadows is related to the visual extent of the shadow. Larger shadows are more compressed than smaller shadows, i.e., the surfaces in larger shadows are lighter than those same surfaces in much smaller shadows.
2) Lightness constancy seems less strong than color constancy. Hue and saturation of colors seem less affected by lighting variations, than absolute gray scale. In complex natural scenes the perception of color within shadows is not affected significantly, but the perception of lightness is.

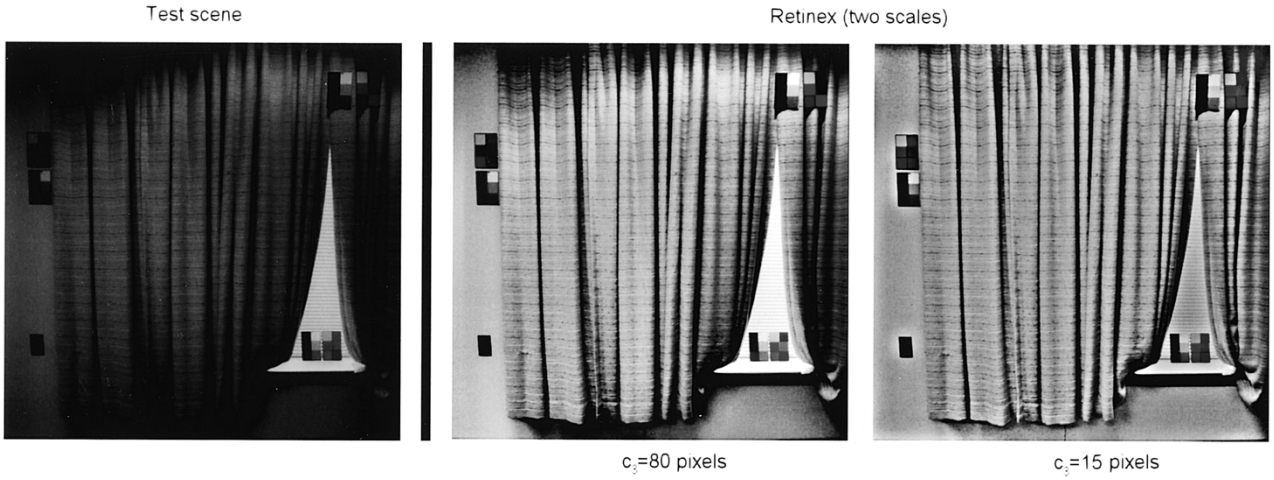


Fig. 11. Results for a test *scene* with multiple color illuminants and familiar test targets that allows comparison of retinex performance to direct human observation of the scene. Test targets on the left are predominantly in fluorescent illumination, while the middle set is in daylight, and the ones on the right are in tungsten light. For direct human observation, there is no shadow perceived at the top, no greenish overall tint observed (due to fluorescent illumination), and excellent color constancy of test targets.

From psychophysical research [11], we add:

3) Lightness constancy is primarily the preservation of *relative* gray-level relationships even though the sensations of absolute lightness slide up and down to some extent with lighting variations.

Therefore, we look for these same effects in the results of retinex processing.

All the images used in subsequent testing are 512×512 pixels, three spectral bands, with 8 b per band. The test *scene* image (Fig. 11) was acquired using Ektachrome slide film and then digitized using a high-resolution slide scanner. All color prints were printed on a Kodak XLT7720 continuous-tone printer with $\gamma = 1.5$ to compensate for the printer's nonlinear transfer function.

We begin by showing a range of diverse images (Fig. 10) for which this retinex produces good results, and include a false color LANDSAT image (i.e., green, red, and infrared are spectrally translated to blue, green, and red). Also included is a computer generated stochastic image. This image is constructed as a Poisson distribution of edges around a selectable mean spatial detail-parameter and intensity levels [12] that are Gaussian distributed in the three color bands. The case shown is for a high degree of spatial detail. The low signal values of the original LANDSAT image are accurately portrayed.

While these results are encouraging and support the hypothesis that this retinex performs well on a wide array of images, we felt it necessary to go further and construct a test scene (Fig. 11) that combines mixed color illuminants, variations, and familiar colors in multiple locations. This test allows us to compare the processed image to the *scene* and to an extent convey this comparison to the reader and for a case with visually severe defects. Our direct observation of this scene does not contain any sense of the shadow at the top of the image, and color constancy of the color charts and gray scales is almost complete. Likewise, direct observation contains no sense of the greenish tint that dominates the raw image and is due to the predominant fluorescent illumination. The optimized single scale retinex result falls short of human observation but succeeds in producing the correct beige scene color and some dynamic range compression of the shadow. The defects in the single scale retinex are the imperfect local color constancy (some of which is due to insufficient dynamic range in the raw image) and insufficient dynamic range compression of the shadow. A much smaller scale retinex ($c_3 = 15$ pixels) produces excellent dynamic range compression and local color constancy (to the limit of the original image). This suggests that the two scales produce complementary visual information and that a multiple-scale retinex should more closely approach the performance of human vision. This experiment dramatically convinced us of the importance of test *scenes* and comparison to direct human observation. Without that, we would have had no way of knowing that the prominent shadow was not really evident to the human observer or that local color constancy of the test targets was so perfect for human perception. The retinex processing seems capable of producing a rendition far closer to our direct observation than the unprocessed image.

We also explored test images (Fig. 12) with zonal and global "gray-world" violations, i.e., spatially averaged relative spectral reflectance values are clearly not equal in the three color spectral bands. Mathematically, it is clear that errors are produced by retinex processing for these cases, but we wished to understand the visual impact of these errors for a variety of cases. The common thread in these retinex images is that "middle gray" is an error and transmits the message—"local equals regional context." An intuitive remedy seems to be to expand "middle gray" regions to larger space constants and, ultimately, to replace the log surrounds with the log of the global means (Fig. 12, bottom). This does correct for zonal gray-world violations but clearly not for global violations (Mars surface and green checkerboard images). The Mars surface image is especially instructive as a near-global gray-world violation. The correct color appears only at chromatic edges but not at lightness edges. This suggests the possible benefit of a chromatic/lightness segmentation and a "filling in" operation

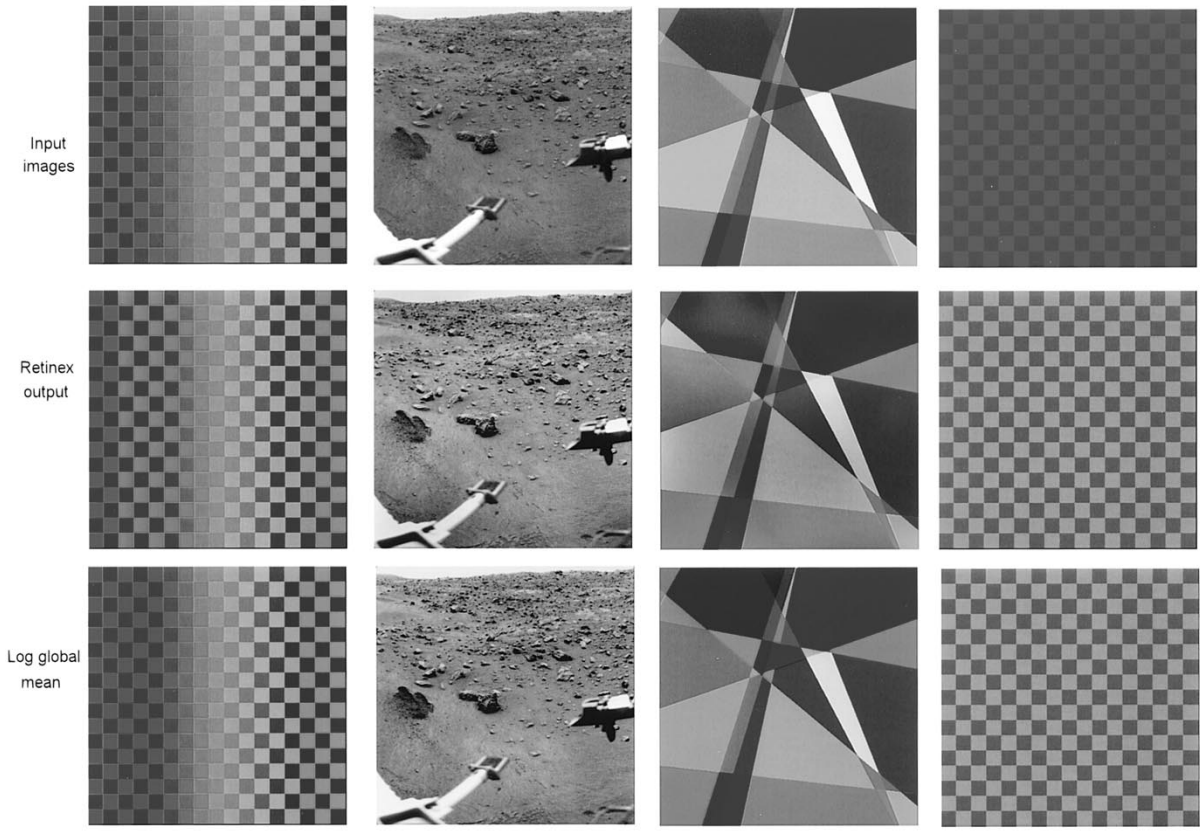


Fig. 12. Results for images with noticeable artifacts due to zonal or global violations of the gray-world assumption and the substitution of log global mean values for log surround to correct regional gray-world violations ($c_3 = 80$ pixels).

[6] from chromatic edges only. The truly global gray-world violation of the green checkerboard image suggests that upon detection of no chromatic edges the processing should retreat to log image (equivalent to human color perception's "aperture mode" [13]). Ultimately, we expect there to be a way to detect and correct these error cases. Since the retinex can most fundamentally be understood as exchanging illumination dependencies for contextual dependency, we anticipate that the solution to this problem lies in the analysis of the large zonal context at the scale of the surround function. This is obviously the central problem that limits the retinex's general application, but our results also indicate that (in the form we tested) the retinex performs well on a rather wide array of both natural and computer-generated images.

We do feel, however, that the final treatment of the retinex triplets prior to display would benefit from some additional "tinkering," such as a fairly restrained nonlinear intensity transformation. The "canonical" gain/offset is surprising in view of the fact that the retinex output is proportional to the log of reflectance ratios.

IV. Discussion

Our findings raise several questions with respect to both natural and machine vision. Perhaps the most interesting is the placement of the log function after surround formation. This is completely contrary to the measured approximate logarithmic response of cone photoreceptors and the design of Mead's silicon retina [10], which was based on those measurements. An examination of recent measurements of primate cones [14] reveals that, while the electrical probe is sampling a single cone, the cone is intact in a small patch of retina. Therefore, we wonder to what degree the measurements may reflect a "network" response rather than just the cone response.

Another possible explanation is that other higher level nonlinear operations might serve to diminish or correct the emphatic halo artifact of the initial log response. The filling-in mechanism [6] is a possible candidate for this, and could also be responsible for correcting errors due to gray-world violations. In any event, it seems reasonable to reconsider a linear photoresponse for machine vision applications especially in view of the wide dynamic range available from current charge-coupled device (CCD) detector arrays. Clearly for dynamic range compression, the log function must be applied prior to any significant bottleneck. For the retina, this bottleneck appears to be the ganglion cells that transmit from the retina to the lateral geniculate nucleus of the brain.

It has not been possible to fully reconcile Land's retinex with the neurophysiology of the primate retina. Receptive fields are invariably found to be spectrally opponent. Mathematically (and regardless of log placement), this is clearly not color constant since the spectral ratios of the illuminant variables do not cancel. Land [15] proposed that linear transformations, much like television's red–green–blue (RGB) to hue–saturation–value (HSV), were a workable resolution which, in combination with his center/surround, does result in

a color constant system as

$$R_{GR} = \log\ S_G r_G - \log\ \overline{S_R r_R} \tag{12}$$

$$R_{RG} = \log\ S_R r_R - \log\ \overline{S_G r_G} \tag{13}$$

where R_{GR} is the "green minus red" spectrally opponent retinex, and R_{RG} is the "red minus green" opponent form. These can be combined to form

$$R_{GR}^{DO} = R_{GR} + R_{RG} \tag{14}$$

$$= \log\ S_G r_G + \log\ S_R r_R - \log\ \overline{S_R r_R} - \log\ \overline{S_G r_G} \tag{15}$$

where R_{GR}^{DO} is the double opponent "green minus red" retinex, which is color constant because

$$R_{GR}^{DO} = \log\ \frac{r_G r_R}{\overline{r}_G \overline{r}_R} \tag{16}$$

when $S_i = \overline{S}_i$ for the ith spectral channel.

Likewise, for a blue yellow double-opponent form:

$$R_{BY} = \log\ S_B r_B - \log\ \overline{S_Y r_Y} \tag{17}$$

where R_{BY} is a "blue minus yellow" spectral opponency retinex, and

$$\overline{S_Y r_Y} = c_1 \overline{S_R r_R} + c_2 \overline{S_G r_G} \tag{18}$$

and c_1, c_2 are weighting constants. Note that the placement of the log is important here, since

$$\log\ \overline{S_Y r_Y} \neq \overline{\log\ c_1 S_R r_R} + \overline{\log\ c_2 S_G r_G}. \tag{19}$$

Analogously, for the "green minus red" case

$$R_{BY}^{DO} = \log\ \frac{r_B r_Y}{\overline{r}_B \overline{r}_Y} \tag{20}$$

where R_{BY}^{DO} is the double opponent "blue minus yellow" retinex, which is likewise color constant. For the color constant lightness channel, the lightness "center," R_L is given by

$$R_L = \log\ c_3 S_B r_B + \log\ c_4 S_G r_G + \log\ c_5 S_R r_R \tag{21}$$

and the lightness surround, $\overline{R}_L$, is given by

$$\overline{R}_L = \log\ c_3 \overline{S_B r_B} + \log\ c_4 \overline{S_G r_G} + \log\ c_5 \overline{S_R r_R}. \tag{22}$$

Again, log placement is important, since

$$\log\ \overline{S_B r_B} \neq \overline{\log\ S_B r_B} \tag{23}$$

which leads to a color constant lightness center/surround retinex as

$$R_L - \overline{R}_L = \log\ \frac{r_B r_G r_R}{\overline{r}_B \overline{r}_G \overline{r}_R}. \tag{24}$$

Previous work [3], [6], [15] indicates that perceptual color constancy is consistent with noncolor constant early vision signals up to and perhaps including the striate cortex with the first clearcut evidence of color constancy in V4 cortex (downstream in the processing pathways from the striate cortex but prior to the full perceptual constructs, which appear to occur in the inferotemporal and parietal-cortices).

On the whole, we are impressed by the performance of this retinex on wide ranging natural and test images even with the shortcomings of the gray-world assumption that show up as a significant perceptual distortion in certain of our test images. We are encouraged that these "error" cases appear to be detectable, and therefore may be minimized or corrected by some simple extension of this retinex. We feel that this extension can be based upon the fundamental mechanism of the retinex, which is to exchange illumination variations for context relationships and is likely to require a multiple scale approach.

While we have not yet explored the relationship of retinex operations to image coding schemes, there is certainly an important connection. To the extent that a retinex operation is a general "front-end" computation for implementation in cameras, the retinex outputs become the inputs for image coding. The dynamic range compression aspect of the retinex does restrict signal variances as well as preserving scene information that would otherwise be lost to saturation or dark clipping. Dynamic range compression has been found to be broadly beneficial [12] for image coding.

V. Conclusions

In the course of defining a specific form for the center/surround retinex, we encountered several fundamental issues that had not been fully resolved by previous investigations. These were:

1) placement of the log function;
2) functional form of the surround;
3) size of the surround space constant;
4) treatment of the retinex outputs prior to final display.

The examination of these issues with experiment image processing led us to define a specific retinex that is different from previous versions. Our version consists of:

1) placement of the log function *after* surround formation;
2) use of the Gaussian form for the surround (although an exponential form is also a good choice);
3) a space constant of about 80 pixels as a reasonable compromise between dynamic range compression and rendition. (Better rendition can be achieved with even larger space constants, but at the expense of detail in shadow zones. This trade-off between compression and rendition is a property of the retinex);
4) a "canonical" gain/offset for the final treatment of retinex output signals.

It remains to generalize the retinex processing to handle gray-world violations and refine the final treatment of the retinex outputs. Even so, we are encouraged by the overall performance of this retinex—that it combines dynamic range compression, color constancy, and lightness/color rendition. The trade-off between dynamic range compression and color rendition, that is governed by the surround space constant, suggests a multiscale approach to generalizing retinex processing.

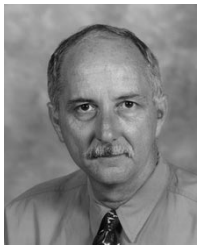
An implementation in analog VLSI or digital VLSI computer chips is an exciting possibility for realizing "smart" cameras of the future.

ACKNOWLEDGMENT

The authors thank F. Huck for his continuing encouragement to explore nonlinear vision processing research and W. Gerdes for his unflagging assistance with the inevitable computer snarls. Both of these co-workers at NASA Langley Research Center have contributed in indirect but highly important ways to this work.

REFERENCES

[1] A. C. Hurlbert, "Formal connections between lightness algorithms," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1684–1693, 1986.
[2] E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," in *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3078–3080, 1986.
[3] A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," *IEEE Trans. Neural Networks*, vol. 2, pp. 237–247, Mar. 1991.
[4] A. Moore, G. Fox, J. Allman, and R. M. Goodman, "A VLSI neural network for color constancy," in *Advances in Neural Information Processing 3* , D. S. Touretzky and R. Lippman, Eds. San Mateo, CA: Morgan Kaufmann, 1991, pp. 370–376.
[5] A. C. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science*, vol. 239, pp. 482–485, 1988.
[6] A. C. Hurlbert, "The computation of color," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, Sept. 1989.
[7] D. E. Bowker, R. E. Davis, D. L. Myrick, K. Stacy, and W. L. Jones, "Spectral reflectances of natural targets for use in remote sensing studies," *NASA Ref. Pub.*, June 1985.
[8] R. H. Dyck, "Design, fabrication, and performance of CCD imagers," in *VLSI Electronics Microstructures Science*, vol. 3, N. G. Einspruch, Ed. New York: Academic, pp. 65–107, 1982.
[9] D. H. Brainard and B. A. Wandell, "An analysis of the retinex theory of color vision," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1651–1661, 1986.
[10] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.
[11] C. J. Bartleson and E. J. Breneman, "Brightness perception in complex fields," *J. Opt. Soc. Amer. A*, vol. 57, pp. 953–957, July 1967.
[12] F. O. Huck, C. L. Fales, R. Alter-Gartenberg, and Z. Rahman, "On the assessment of visual communication," in *Handbook of Statistics*, vol. 10, N. Bose and C. R. Rao, Eds. New York: Elsevier, 1993.
[13] P. Lennie and M. D. D'Zmura, "Mechanisms of color vision," in *CRC Critic. Rev. Neurobiol.*, vol. 3, pp. 333–400, 1988.
[14] D. A. Baylor, B. J. Nunn, and J. L. Schnapf, "Spectral sensitivity of cones of the monkey, *Macaca fascicularis*," *J. Physiol.*, vol. 390, pp. 145–160, 1987.
[15] E. Land, "Recent advances in retinex theory," *Vis. Res.*, vol. 26, no. 1, pp. 7–21, 1986.



Daniel J. Jobson received the B.S. degree in physics from the University of Alabama, Tuscaloosa, in 1969.

He is a senior research scientist at NASA Langley Research Center, Hampton, VA. His research has spanned topics including the design and calibration of the Viking/Mars lander camera, the colorimetric and spectrometric characterization of the two lander sites, design and testing of multispectral sensors, and analysis of coastal and ocean properties from remotely sensed data. For the past several years, his research interest has been in visual information processing with emphasis on machine vision analogs for natural vision, focal-plane processing technology, and nonlinear methods that mimic the dynamic-range–lightness constancy of human vision.

Zia-ur Rahman (M'87) received the B.A. degree in physics from Ripon College, WI, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from the University of Virginia, Charlottesville, in 1986 and 1989, respectively. His graduate research focused on using neural networks and image processing techniques for motion detection and target tracking.

He is a research scientist with the Science and Technology Corporation, and is presently working under contract to NASA Langley Research Center, Hampton, VA, on advanced concepts in information processing for high-resolution imaging and imaging spectrometry. Currently, he is involved in conducting research in multidimensional signal processing, with emphasis on data compression and feature extraction methods. This work supports a NASA project for providing readily accessible, inexpensive remote-sensing data.

Dr. Rahman is a member of SPIE and INNS.

Glenn A. Woodell graduated from the NASA apprentice school in 1987 in materials processing.

He is a research technician at NASA Langley Research Center, Hampton, VA. His work has included semiconductor crystal growth experiments flown in space aboard the Space Shuttle in 1985 to study the effect of gravity-induced convection, and is working on a second experiment to be flown in February 1996. His research has included demarcation, calculation, and visualization of crystal growth rates and real-time gamma ray visualization of the melt-solid interface and the solidification process. He has recently become involved in research on nonlinear image processing methods as analogs of human vision.