Information-Adaptive Image Encoding and Restoration
NAG1-1847

Summary of Research
submitted to
NASA Langley Research Center
Hampton, Virginia 23681

Principal Investigator
Stephen K. Park
Professor and Chair

Co-Principal Investigator
Zia-ur Rahman
Research Assistant Professor

Department of Computer Science
College of William & Mary
PO Box 8795
Williamsburg, VA 23187-8795
ATTN: Grants & Research Administration

October 19, 1998

Abstract: The multiscale retinex with color restoration (MSRCR) has shown itself to be a very versatile *automatic* image enhancement algorithm that simultaneously provides dynamic range compression, color constancy, and color rendition. A number of algorithms exist that provide one or more of these features, but not all. In this paper we compare the performance of the MSRCR with techniques that are widely used for image enhancement. Specifically, we compare the MSRCR with color adjustment methods such as gamma correction and gain/offset application, histogram modification techniques such as histogram equalization and manual histogram adjustment, and other more powerful techniques such as homomorphic filtering and 'burning and dodging'. The comparison is carried out by testing the suite of image enhancement methods on a set of diverse images. We find that though some of these techniques work well for some of these images, only the MSRCR performs universally well on the test set.

3. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Retinex Image Processing: Improved Fidelity To Direct Visual Observation," IS&T Fourth Color Imaging conference: Color Science, Systems, and Applications, Scottsdale, AZ, (November 1996).

   Abstract: Recorded color images differ from direct human viewing by the lack of dynamic range compression and color constancy. Research is summarized which develops the center/surround retinex concept originated by Edwin Land through a single-scale design to a multi-scale design with color restoration (MSRCR). The MSRCR synthesizes dynamic range compression, color constancy, and color rendition, and, thereby, approaches fidelity to direct observation.

4. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multiscale Retinex for Color Image Enhancement," in Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland, (September 1996).

   Abstract: The retinex is a human perception-based image processing algorithm which provides color constancy and dynamic range compression. We have previously reported on a single-scale retinex (SSR) and shown that it can either achieve color/lightness rendition or dynamic range compression, but not both simultaneously. We now present a multi-scale retinex (MSR) which overcomes this limitation for most scenes. Both color rendition and dynamic range compression are successfully accomplished except for some "pathological" scenes that have very strong spectral characteristics in a single band.

5. F. O. Huck, C. L. Fales, and Z. Rahman, "On the Information-Theoretic Assessment of Visual Communication," in Proceedings of the IEEE International Conference on Image Processing, (September 1996).

   Abstract: This paper deals with the extensions of information theory to the assessment of visual communication from scene to observer. The mathematical development rigorously unites the electro-optical design of image gathering and display devices with the digital processing algorithms for image coding and restoration. Results show:

This is the final report for NASA grant NAG-1-1847 and covers the period from 01 July, 1996 to 30 September, 1997. Enclosed are copies of the conference papers and journal articles for which the research was at least partially supported by the grant.

The research during this period was in two primary, related areas. The first was the evaluation of integrated information adaptive image compression. And, the second was the development of a new technique for image enhancement. The NASA Langley Research Center and Science and Technology Corporation (former employers of Z. Rahman) jointly applied for patent on the latter technique in May 1996. The patent was subsequently approved in April 1998, and is awaiting publication.

During the cited time period we continued work on those papers that had been started before the commencement of the grant. Some of these were already in the accepted-for-publication stage but required revisions. Other papers were written and presented during the grant period. A list of the publications is presented below along with the abstracts. A copy of each of the puplications is enclosed except for the book *Visual Communication: An Information Theory Approach* and the two papers that appeared in the October 1996 issue of the *Philosophical Transactions of the Royal Society*. These two papers were awarded the H. J. E. Reid Award for the Outstanding Paper in June 1998.

1 Publications

Publications are listed chronologically in each category.

1.1 Conferences

1. Z. Rahman, F. O. Huck, and C. L. Fales, "Informationally Optimized Image-gathering and Restoration," presented at the IS&T's 50th Annual Conference, Cambridge, MA, (May 1997).

   Abstract: The goal of image gathering and restoration often is to produce the best possible picture in terms of fidelity, sharpness and clarity. However, this goal cannot be attained, at it has been pursued in the past, by treating image gathering and restoration as independent tasks. Instead, in a clean departure from the mores of traditional image processing, we present an approach that rigorously uses modern communication theory to optimally combine the electro-optical design of the image gathering device with the digital processing algorithm for image restoration. Extensive simulations have shown that there exists a strong correlation between the information rate that is produced by the image gathering device and the image quality with which an image can be restored.

2. Z. Rahman, G. A. Woodell, and D. J. Jobson, "A Comparison of the Multiscale Retinex With Other Image Enhancement Techniques," Proceedings of the IS&T's 50th Annual Conference, Cambridge, MA, (May 1997).

- End-to-end system analysis closely correlates with measurable and perceptual performance characteristics, such as data rate and image quality, respectively.
- The goal of producing the best possible image at the lowest possible image data rate can be realized only if (a) the electro-optical design of the image-gathering device is optimized for the maximum-realizable information rate and (b) the image-restoration algorithm properly accounts for the perturbations in the visual communication channel.

6. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multiscale Retinex for Dynamic Range Compression and Color Rendition," *Applications of Digital Image Processing XIX*, Andrew G. Tescher, Ed., Proc. SPIE 2847, (August 1996).

   Abstract: The human vision system performs the tasks of dynamic range compression and color constancy almost effortlessly. The same tasks pose a very challenging problem for imaging systems whose dynamic range is restricted by either the dynamic response of film, in case of analog cameras, or by the analog-to-digital converters, in the case of digital cameras. The images thus formed are unable to encompass the wide dynamic range present in most natural scenes (often $> 500 : 1$). Whereas the human visual system is quite tolerant to spectral changes in lighting conditions, these strongly affect both the film response for analog cameras and the filter responses for digital cameras, leading to incorrect color formulation in the acquired image. Our multiscale retinex, based in part on Edwin Land's work on color constancy, provides a fast, simple, and automatic technique for simultaneous dynamic range compression and accurate color rendition. The retinex algorithm is non-linear, and global—output at a point is also a function of its surround—in extent. A comparison with conventional dynamic range compression techniques such as the application of point non-linearities, e.g. $\log(x, y)$, and global histogram equalization and/or modification shows that the multiscale retinex simultaneously provides the best dynamic range compression and color rendition. The applications of such an algorithm are many; from medical imaging to remote sensing; and from commercial photography to color transmission.

7. Z. Rahman, "Integrated wavelet compression and restoration," *Wavelet Applications in Signal and Image Processing IV*, Michael A. Unser, Akram Aldroubi, Andrew F. Laine, eds., Proc. SPIE 2825, (August 1996).

   Abstract: The performance of wavelet compression algorithms is generally judged solely as a function of the compression ratio and the visual artifacts which are perceivable in the reconstructed image. The problem then becomes one of obtaining the best compression with fewest visible artifacts—a very subjective measure. Our wavelet compression algorithm uses an information theoretic analysis for the design of the compression maps. We have previously shown that maximizing the information for a given visual communication channel also maximizes the visual quality of the restored image. We utilize this to design quantization maps which maximize information for a given compression ratio. Hence we are able to design quantization maps which maximize the restorability of an image—i.e. the information content, the image quality, and the mean-square difference fidelity—for a given compression ratio.

1.2 Journal Articles

1. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A Multi-Scale Retinex For Bridging the Gap Between Color Images and the Human Observation of Scenes," *IEEE Transactions on Image Processing*, Special Issue on Color Processing, (July 1997).

   Abstract: Direct observation and recorded color images of the same scenes are often strikingly different because human visual perception computes the conscious representation with vivid color and detail in shadows, and with resistance to spectral shifts in the scene illuminant. A computation for color images which approaches fidelity to scene observation ¡em¿must¡/em¿ combine dynamic range compression, color consistency—a computational analog for human vision color constancy—and color and lightness tonal rendition. In this paper, we extend a previously designed single scale center/surround retinex to a multi-scale version that achieves simultaneous dynamic range compression/color consistency/lightness rendition. This extension fails to produce good color rendition for a class of images that contain violations of the gray-world assumption implicit to the theoretical foundation of the retinex. Therefore we define a method of color restoration that corrects for this deficiency at the cost of a modest dilution in color consistency. Extensive testing of the multi-scale retinex with color restoration on several test ¡em¿ scenes¡/em¿ and over a hundred images did not reveal any pathological behavior.

2. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and Performance of a Center/Surround Retinex," *IEEE Transactions on Image Processing*, (March 1997).

   Abstract: The last version of Edwin 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 which violate the gray world assumptions (implicit to the retinex) are investigated to provide insight into cases where the retinex fails to produce a good rendition.

3. F. O. Huck, C. L. Fales, and Z. Rahman, "An Information Theory of Visual Communication," *Philosophical Transactions of the Royal Society A: Physical Sciences and Engineering*, (October 1996).

   Abstract: The fundamental problem of visual communication is that of producing the best possible picture at the lowest data rate. We address this problem by extending information

theory to the assessment of the visual communication channel as a whole, from image gathering to image display. The extension unites the two disciplines, the electro-optical design of image-gathering and display devices and the digital processing for image coding and restoration. The mathematical development leads to several intuitively attractive figures of merit for assessing the visual communication channel as a function of the critical limiting factors that constrain its performance. Multiresolution decomposition is included in the mathematical development to optimally combine the economical encoding of the transmitted signal with image gathering and restoration.

Quantitative and qualitative assessments demonstrate that a visual communication channel ordinarily can be expected to produce the best possible picture at the lowest data rate only if the the image-gathering device produces the maximum-realizable information rate and the image-restoration-algorithm properly accounts for the critical limiting factors that constrain visual communication. These assessments encompass (a) the electro-optical design of the image-gathering device in terms of the trade-off between blurring and aliasing in the presence of photodetector and quantization noises, (b) the compression of data transmission by redundancy reduction, (c) the robustness of the image restoration to uncertainties in the statistical properties of the captured radiance filed, and (d) the enhancement of the particular features or, more generally, of the visual quality of the observed images. The 'best visual quality' in this context normally implies a compromise among maximum-realizable fidelity, sharpness, and clarity which depends on the characteristics of the scene and the purpose of the visual communication (e.g. diagnosis versus entertainment).

4. C. L. Fales, F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, "Image Gathering and Digital Restoration," *Philosophical Transactions of the Royal Society* A*: Physical Sciences and Engineering*, (October 1996).

Abstract: This paper seeks to unite two disciplines: the electro-optical design of the image gathering and display devices and the digital processing for image restoration. So far, these two disciplines have remained independent, following strictly separate traditions However, the best possible performance can be attained only when the digital processing algorithm accounts for the critical limiting factors of image gathering and display and the image-gathering device is designed to enhance the performance of the digital-processing algorithm. The following salient advantages accrue:

(a) Spatial detail as fine as the sampling interval of the image-gathering device ordinarily can be restored sharply and clearly.

(b) Even finer spatial detail than the sampling interval can be restored by combining a multiresponse image-gathering sequence with a restoration filter that properly reassembles the within-passband and aliased signal components.

(c) The visual quality produced by traditional image gathering (e.g. television camera) and reconstruction (e.g. cubic convolution) can be improved with a small-kernel restoration operator without an increase in digital processing.

(d) The enhancement of radiance-field transitions can be improved for dynamic range compression (to suppress shadow obscurations) and for edge detection (for computer vision).

1.3 Book

1. F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, (June 1997).

   From the Publishers' catalog: *Visual Communication: An Information Theory Approach* presents an entirely new look at the assessment and optimization of visual communication channels, such as are employed for telephotography and television. The electro-optical design of image gathering and display devices, and the digital processing for image coding and restoration, have remained independent disciplines which follow distinctly separate traditions; yet the performance of visual communication channels cannot be optimized just by cascading image-gathering devices, image-coding processors, and image-restoration algorithms as the three obligatory, but independent, elements of a modern system. Instead, to produce 'the best possible picture at the lowest data rate', it is necessary to jointly optimize image gathering, coding, and restoration.

   Although the mathematical development in *Visual Communication: An Information Theory Approach* is firmly rooted in familiar concepts of communication theory, it leads to formulations that are significantly different from those that are found in the traditional literature on either rate distortion theory or digital image processing. For example, the Wiener filter, which is perhaps the most common image restoration algorithm in the traditional digital image processing literature, fails to fully account for the constraints of image gathering and display. As demonstrated in the book, digitally restored images improve in sharpness and clarity when these constraints are properly accounted for.

   Visual Communication: An Information Theory Approach is unique in its extension of modern communication theory to the end-to-end assessment of visual communication. from scene to observer. As such, it ties together the traditional textbook literature on electro-optical design and digital image processing. This book serves as an invaluable reference for image processing and electro-optical system design professionals and may be used as a text for advanced courses on the subject.

Multiscale Retinex for Dynamic Range Compression and Color Rendition

Z. Rahman, D. J. Jobson and G. A. Woodell

Applications of Digital Image Processing XIX
Andrew G. Tescher, Ed., Proc.
SPIE 2847
(August 1996)

A Multiscale Retinex For Color Rendition and Dynamic Range Compression

Zia-ur Rahman[†]
College of William & Mary, Williamsburg, VA 23187

Daniel J. Jobson and Glenn A. Woodell
NASA Langley Research Center, Hampton, Virginia 23681

Abstract

The human vision system performs the tasks of dynamic range compression and color constancy almost effortlessly. The same tasks pose a very challenging problem for imaging systems whose dynamic range is restricted by either the dynamic response of film, in case of analog cameras, or by the analog-to-digital converters, in the case of digital cameras. The images thus formed are unable to encompass the wide dynamic range present in most natural scenes (often > 500:1). Whereas the human visual system is quite tolerant to spectral changes in lighting conditions, these strongly affect both the film response for analog cameras and the filter responses for digital cameras, leading to incorrect color formulation in the acquired image. Our multiscale retinex, based in part on Edwin Land's work on color constancy, provides a fast, simple, and automatic technique for simultaneous dynamic range compression and accurate color rendition. The retinex algorithm is non-linear, and global—output at a point is also a function of its surround—in extent. A comparison with conventional dynamic range compression techniques such as the application of point non-linearities, e.g. $\log(x, y)$, and global histogram equalization and/or modification shows that the multiscale retinex simultaneously provides the best dynamic range compression and color rendition. The applications of such an algorithm are many; from medical imaging to remote sensing; and from commercial photography to color transmission.

1. Introduction

Human perception excels at constructing a visual representation with vivid color and detail across wide ranging photometric levels caused by lighting variations. In addition human vision computes color so as to be relatively independent of spectral variations in illumination.[1] The images obtained with film and electronic cameras suffer, by comparison, from a loss in clarity of detail and color as light levels drop within shadows, or as distance from a lighting source increases. When the dynamic range of a scene exceeds the camera's dynamic range, there can be irrevocable loss of visual information at both extremes of the scene dynamic range. Improved fidelity of color images to human observation should, therefore, combine dynamic range compression, color constancy, and color and lightness rendition. In this paper we present our initial work in developing a technique, the multiscale retinex with color restoration (MSRCR), which achieves all these goals.

The idea of the retinex was conceived by Edwin Land[2, 3, 4] as a model of the lightness and color perception of human vision. Subsequently Hurlbert[5, 6], and Hurlbert and Poggio[7] studied the properties of the center/surround form of the retinex and other lightness theories and found a common mathematical foundation which possesses some excellent properties but cannot actually compute reflectance for arbitrary scenes. Certain scenes violate the "gray-world" assumption which requires that the average reflectances in the surround be equal in the three spectral color bands. For example, scenes that are dominated by one color—"monochromes"—clearly violate this assumption and are forced to gray (equal values in all spectral channels) by the retinex computation. Hurlbert[5] further showed the lightness problem has a solution that has a center/surround spatial form. This suggests the possibility that the spatial opponency of the center/surround is a general solution to estimating relative reflectances for arbitrary lighting conditions. At the same time it is equally clear that human vision does not determine relative reflectance but rather a context dependent relative reflectance since surfaces in shadow do not appear to be the same lightness as the same surface when lit. Moore et al.[8, 9] took up the retinex problem as a natural implementation for analog

[†]Funded by NASA Langley Research Center Contract #NAS1-19603 to Science and Technology Corporation, and by Grant #NAG1-1847 to the College of William & Mary.

VLSI resistive networks and found that color rendition was dependent on scene content—some scenes worked well, others did not. These studies also pointed out the problems that occur with color Mach bands and the graying out of large uniform zones of color.

The MSRCR builds on the single scale retinex[10] (SSR), and the multiscale retinex[11] (MSR). Both the SSR and the MSR provide very good dynamic range compression but suffer from the graying out which occurs in large areas of uniform color. Hence the overall color/lightness rendition can be poor depending upon the scene. The MSRCR alleviates this problem by using a color restoration function which controls the amount of color saturation for the final rendition. This function provides the color restoration that is needed with the dynamic range compression to approximate the performance of human vision with a computation that is quite automatic and reasonably simple. The MSRCR is extremely useful for enhancing 8-bit color images that suffer from lighting deficiencies commonly encountered in architectural interiors and exteriors, landscapes, and non-studio portraiture applications. Potential benefits for remote sensing applications are improved visibility of color and detail in shadows and low reflectance zones and the diminution of sun angle/atmospheric signal variations that can lead to more resilient and accurate multispectral classification.

2. Multiscale Center/Surround Retinex

The SSR[10, 12, 13] is given by

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

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

$$\mathcal{F}(x,y) = Ke^{-(x^2+y^2)/c^2},$$

where c is the Gaussian surround space constant, or the scale, and K is selected such that

$$\iint \mathcal{F}(x,y)\,dx\,dy = 1.$$

The MSR output is simply the weighted sum of the outputs of several SSRs with different scales. Mathematically,

$$R_{M_i}(x,y) = \sum_{n=1}^{N} w_n R_{n_i}(x,y), \tag{2}$$

where N is the number of scales, $R_{n_i}(x,y)$ is the ith component of the nth scale, $R_{M_i}(x,y)$ is the ith color component of the MSR output, and w_n is the weight associated with the nth scale. The number of scales is application dependent. However, after experimenting with one small scale and one large scale, the need for a third intermediate scale was immediately apparent in order to produce a graceful rendition without visible "halo" artifacts near strong edges. Experimentation shows that assigning equal weights to the scales is adequate for most applications, although a particular scale could be weighted more heavily if a particular feature needs to be enhanced. For instance, weighting the smallest scale heavily can be used to achieve the strongest dynamic range compression but leads to ungraceful edge artifacts and some graying of uniform color zones in the rendition.

To test whether the dynamic range compression of the MSR approaches that of human vision we use test SCENES not just test images, to facilitate the comparison between the processed image and direct observation. An example (Fig. 1) illustrates the complementary strengths and weaknesses of each scale taken separately and the strength of the multiscale synthesis. This image is representative of a number of test scenes (Fig. 2) where for conciseness we show only the multiscale result. The comparison of the unprocessed images to the perception of the scene produces some striking and unexpected results. Compared to recorded images, the color and detail are far more vivid for

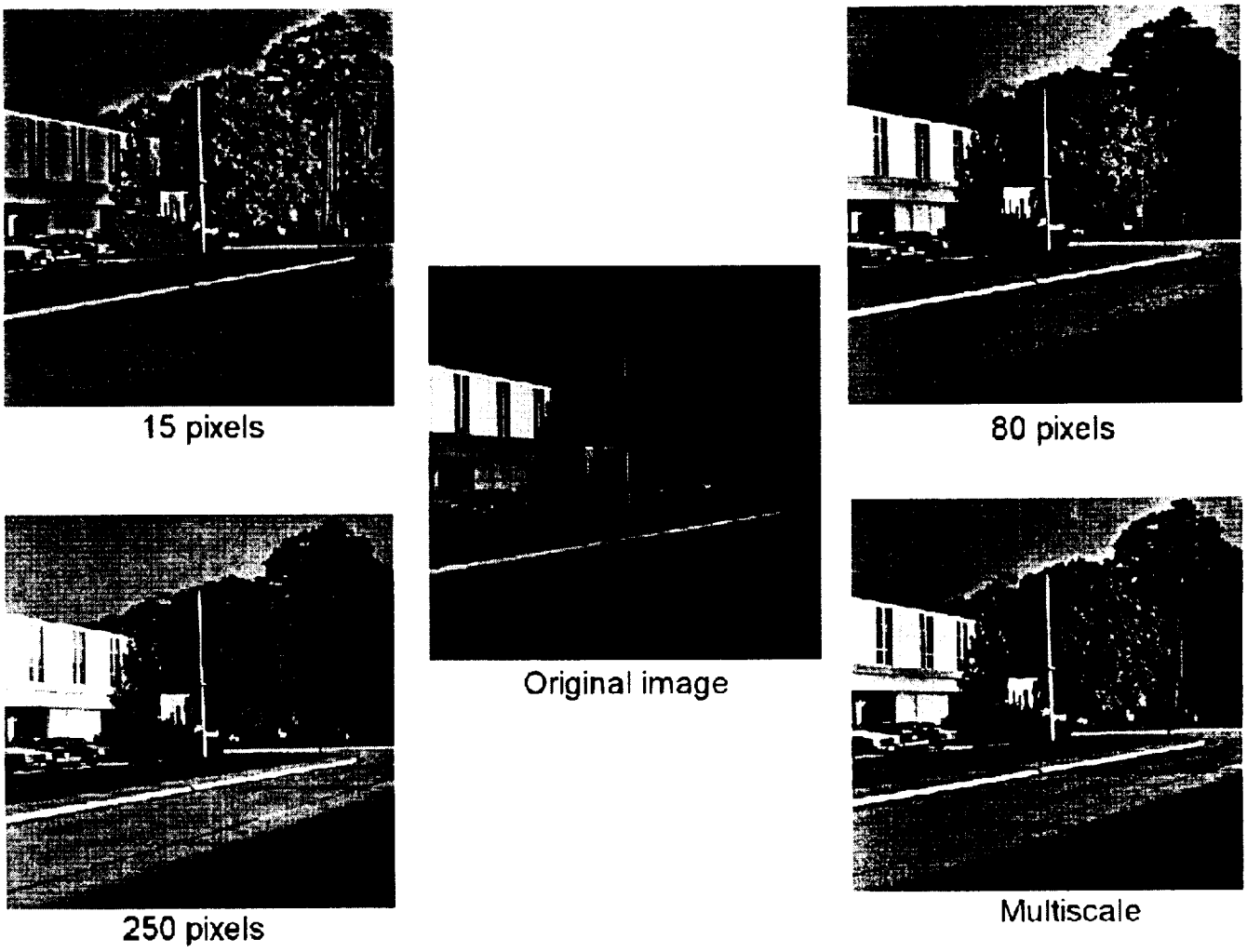


Figure 1: The components of the multiscale retinex which show their complementary information content. The smallest scale is strong on detail and dynamic range compression and weak on tonal and color rendition. The reverse is true for the largest spatial scale. The multiscale retinex combines the strengths of each scale and mitigates the weaknesses of each.

direct observation not only in shadowed regions, but also in the bright zones of the scene. This suggests that human vision is perhaps doing more than just strong dynamic range compression and that enhancements beyond the MSR may be needed to capture the realism of direct viewing.

A sample of image data for surfaces in both sun and shadow indicates a dynamic range compression of 2 : 1 for the MSR compared to the 3 :1 to 5 : 1 measured in our perceptual tests. For the SSR this value is 1.5 : 1 or less. These levels of dynamic range compression are for outdoor scenes which have shadows of large spatial extent. The much higher values of compression that occur for the human visual perception of mixed indoor/outdoor scenes are compared to retinex performance in Fig. 2 (right). The foreground orange book on the grayscale is compressed by approximately 5 : 1 for the MSR while compression for the SSR is only approximately 3 : 1 both relative to the bright building facade in the background. For this case, the compression of human vision is difficult to estimate since both the color and texture of the two surfaces are quite different. Our impression is that the MSR is approaching human vision's performance but not quite reaching it.

The MSR performs well in terms of dynamic range compression but its performance on the pathological classes of

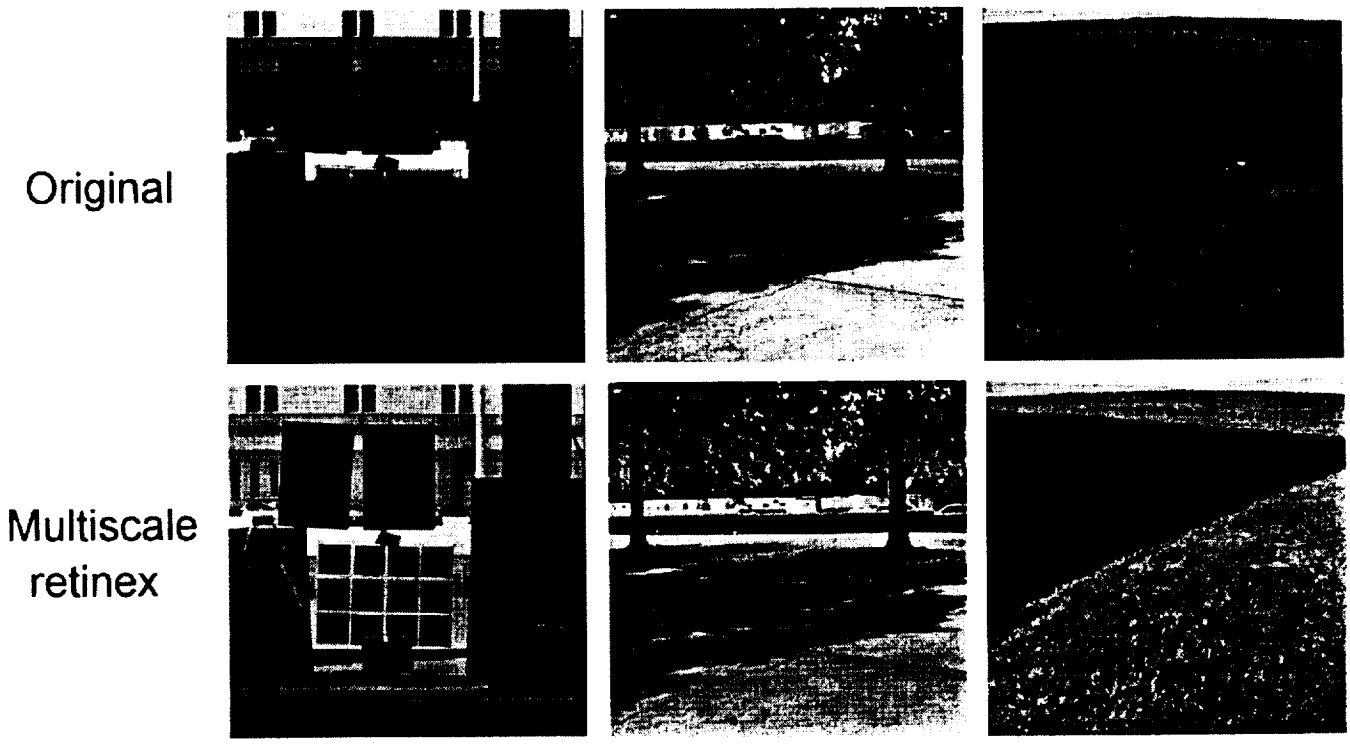


Figure 2: Examples of test SCENES processed with the multiscale retinex prior to color restoration. While color rendition of the left image is good, the other two are "grayed" to some extent. Dynamic range compression and tonal rendition are good for all and compare well with scene observation.

images examined in previous SSR research[10] (Fig. 3 middle row) still needs to be considered. These images represent a variety of regional and global gray-world violations and we can not expect the MSR to handle them effectively. We provide these results as a baseline for comparison with the color restoration which is developed next. All possess notable, and often serious, defects in color rendition. Since we want the MSR to be automatic, and the pathological images cannot be determined a priori, we developed an additional color computation which is universally applied to all post-retinex images to produce a general purpose computation.

3. A Color Restoration Method for the Multi-scale Retinex

The general effect of retinex processing on images with regional or global gray-world violations is a "graying out" of the image either in specific regions or globally. This desaturation of color can, in some cases, be severe (Fig. 3 middle) Therefore we can consider the desired color computation as a color restoration, which should produce good color rendition for images with any degree of graying. More rarely, the gray-world violations can simply produce an unexpected color distortion (Fig.3 top-left). Again we seek a simple computation which also handles these cases. In addition we would like for the correction to preserve a reasonable degree of color constancy since that is one of the basic motivations for the retinex. Color constancy is known to be imperfect in human visual perception, so some level of illuminant color dependency is acceptable provided it is much lower than the physical spectrophotometric variations. Ultimately this is a matter of image quality and color dependency is tolerable to the extent that the visual defect is not visually too strong.

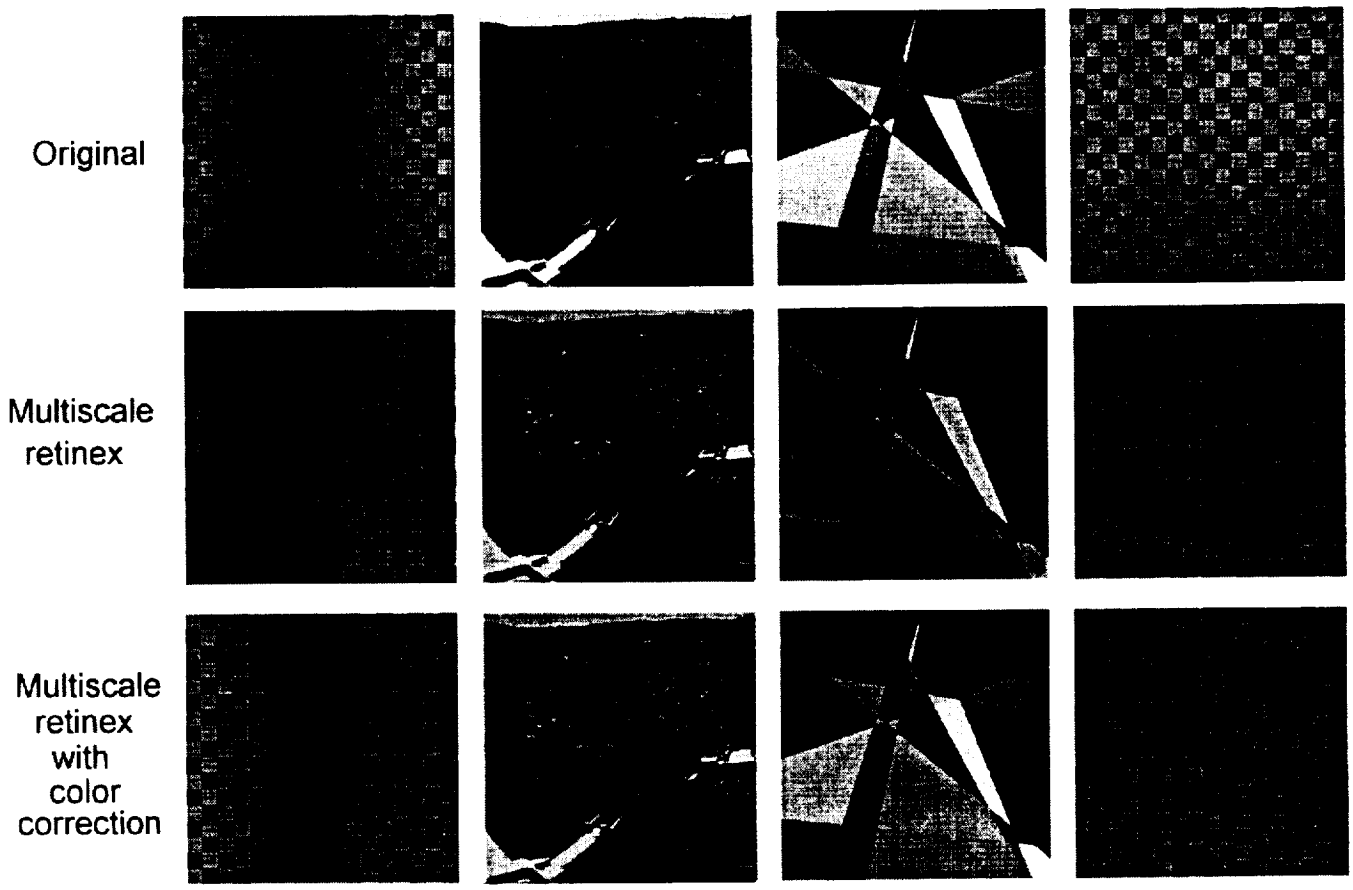


Figure 3: Pathological "gray-world" violations are not handled well by the multiscale retinex alone(middle row) but are treated successfully when color restoration is added (lower row).

Starting with the foundations of colorimetry[14], the color space is transformed using

$$I_i'(x,y) = \frac{I_i(x,y)}{\sum_{j=1}^{N} I_j(x,y)}, i = 1, ..., N. \tag{3}$$

The color restoration function C(x,y) is then simply

$$C(x,y) \equiv C_i(x,y) = f\left[I_i'(x,y)\right],$$

where $f[]$ represents linearly or non-linearly normalized color space, and controls the saturation of the final rendition. The MSRCR is then given by

$$R_i(x,y) = C(x,y) \sum_{i=1}^{N} W_i \left(\log[I_i(x,y)] - \log[I_i(x,y) * F[x,y]\right). \tag{4}$$

This form provides dynamic range compression, color and lightness constancy, and very good color rendition.

4. Selected Results for Diverse Test Cases

The test images presented here begin with some test scenes. We feel it is fundamental to refer the processed images back to the direct observation of scenes. This is necessary to establish that how well the computation represents

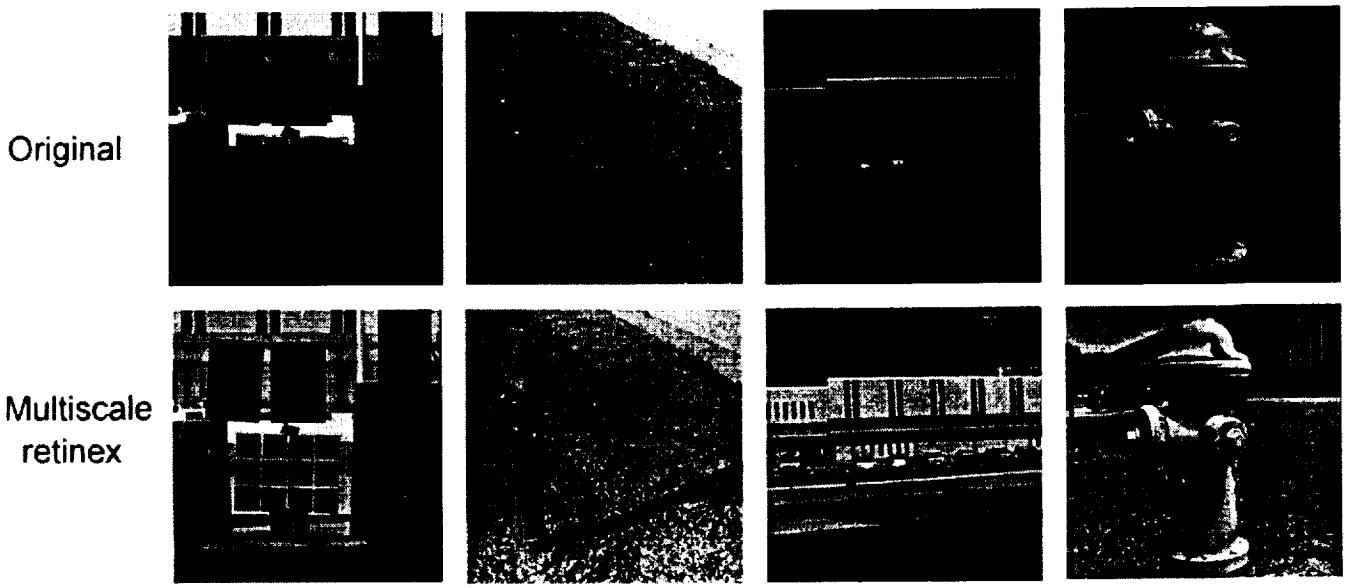


Figure 4: Test SCENES illustrating dynamic range compression, color and tonal rendition, and automatic exposure correction. All processed images compare favorably with direct scene observation with the possible exception of leftmost image which is even lighter and clearer for observation. This scene has the widest dynamic range and suggests that even stronger dynamic range compression may be needed for this case.

a result that is "what you would have seen if you had been there". Clearly we cannot duplicate human vision's peripheral vision which spans almost 180°, but within the narrower angle of most image frames we would like to demonstrate that the computation achieves the clarity of color and detail in shadows, reasonable color constancy and lightness and color rendition that is present in direct observation of scenes. While we cannot yet test performance for scenes that go beyond 8-bit dynamic ranges, these results support the utility of the processing scheme for the enhancement of conventional 8-bit color images. The test scenes are given first (Figs. 4, 5) so that we can describe the degree to which the computation approaches human visual performance. All the test scene images after retinex processing are quite "true to life" compared with direct observation. We did not carefully match camera spatial resolution to observation so some difference in perceived detail is expected and observed. However overall color, lightness, and detail rendering for the multiscale retinex is a good approximation to human visual perception.

5. Discussion

The question which now arises is: What advantages does the MSRCR possess over traditional image enhancement techniques such as histogram equalization, non-linear transforms (gamma correction), and gain/offset manipulation? Again the answer is based on experimental observation, rather than on theory. Each of the traditional techniques is well suited for a certain class of images, where the overall contrast is poor. They almost invariably fail where the image simultaneously contains very bright and very dark areas. They also fail to preserve the color when applied to images where the need for enhancement is not readily observable. The MSRCR successfully overcomes both these weaknesses of the traditional techniques. Figure 6 shows a comparison of the MSRCR with the traditional techniques for two natural scenes. The first contains a typical outdoor scene which has a sharp shadow across the frame. And the second is a *good* image which does not obviously need image enhancement. In both cases, the output of the MSRCR is either better than the original or as good. The same cannot be said of the traditional techniques.

The MSRCR can be applied ex post facto on 8-bit color images to provide image enhancement. The only problem arises when these images have been compressed using lossy methods. Not only does the MSRCR improve

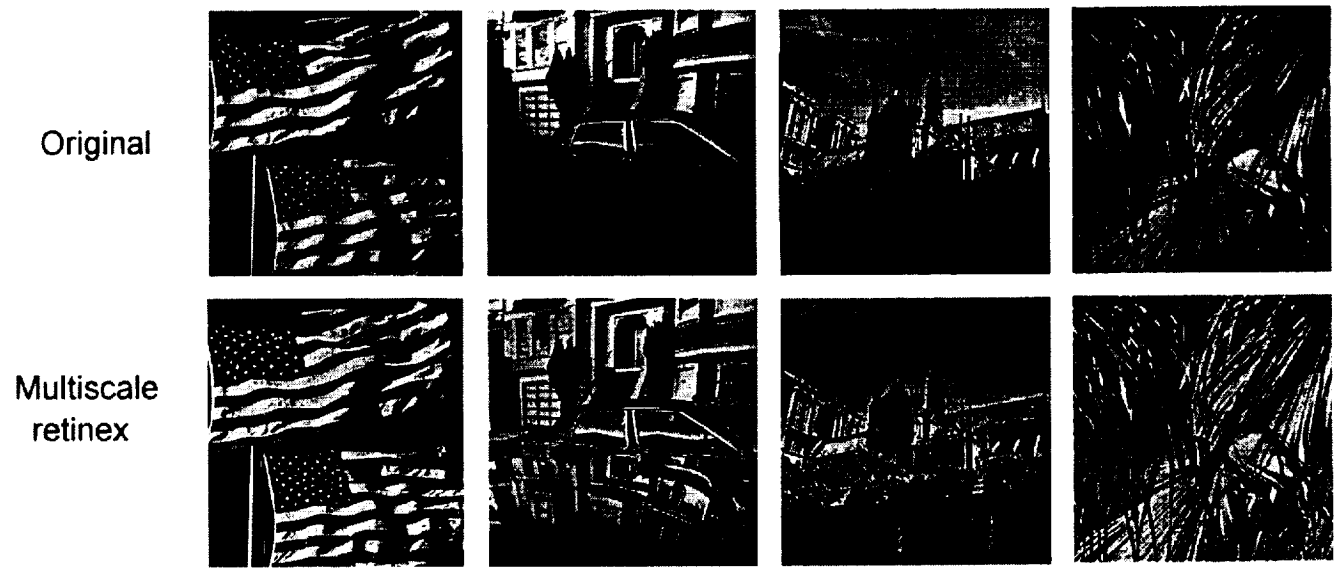


Figure 5: Photographic examples further illustrating graceful dynamic range compression together with tonal and color rendition. The rightmost image shows the processing scheme handling saturated colors quite well and not distorting an image that is quite good in its original form.

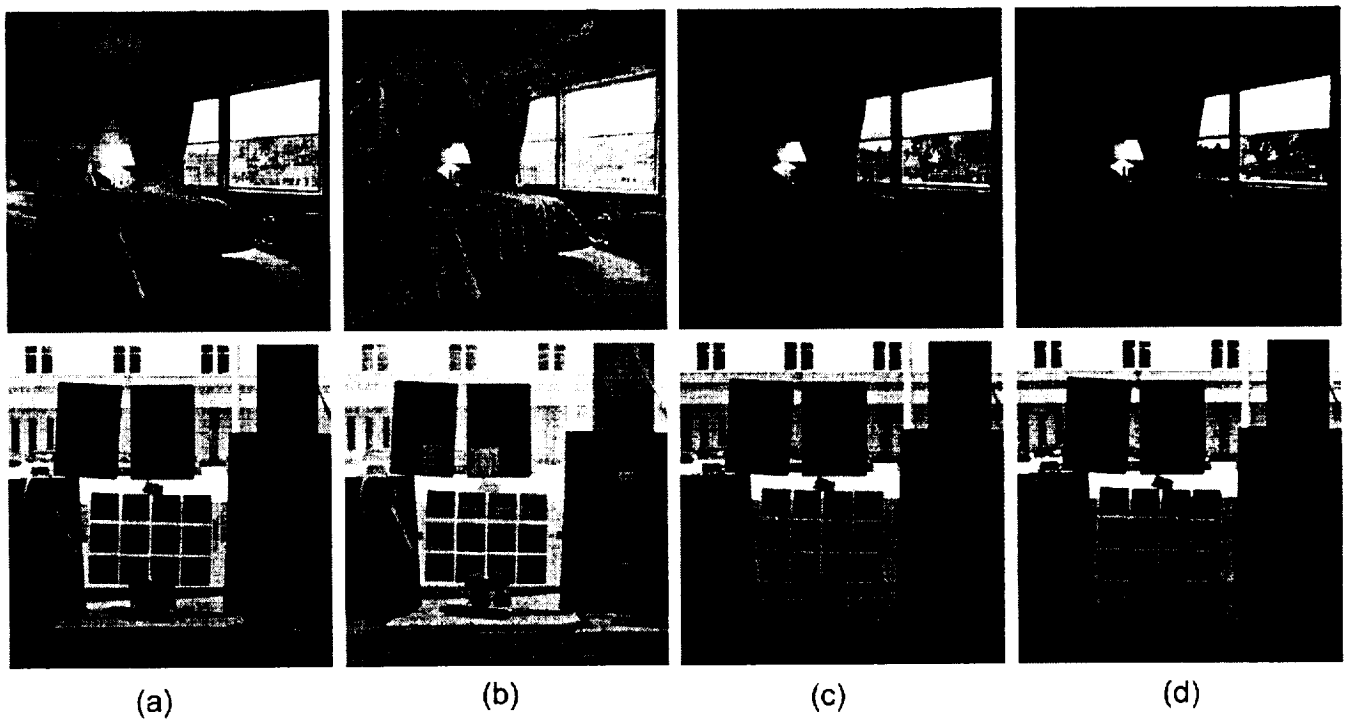


Figure 6: A comparison of image enhancement techniques: (a) MSRCR with 3 scales (b) Histogram Equalization, (c) Gamma correction, and (d) Gain/offset manipulation.

the dynamic range and color, it also enhances the compression artifacts which had been imperceptible before the application. Hence, the retinex is best applied prior to lossy image coding. One obvious advantage that the MSRCR provides for image compression is its ability to compress wide dynamic ranges to 8-bit, or less, per band color output, while preserving, and even enhancing, the details in the scene. The overall effect then is a significant reduction in the number of bits (especially in cases where the original color resolution is higher than 8-bit/band), required to transmit the original without a substantial loss in spatial resolution or contrast quality.

We have encountered many digital images in our testing which are either under- or overexposed. Apparently even with modern photographic auto-exposure controls, exposure errors can and do occur. An additional benefit of the MSRCR is it apparent capability for exposure correction. This is especially beneficial if it is performed before the image is recorded either on film or on disk.

6. Conclusions

The SSR provides a good mechanism for enhancing certain aspects of images and providing dynamic range compression. However, it is limited in its use because it can either provide good tonal rendition or dynamic range compression. The MSR comprised of three scales—small, intermediate, and large—overcomes this limitation and was found to synthesize dynamic range compression, color constancy, and tonal rendition and produce results which compare favorably with human visual perception except for scenes which contain violations of the gray-world assumption. Even when the gray-world violations were not dramatic, some desaturation of color was found to occur. The MSRCR adds a color restoration scheme which produced good color rendition even for severe gray-world violations, but at the expense of a slight sacrifice in color constancy. While there is no firm theoretical or mathematical basis for proving the generality of the MSRCR, we have tested it successfully on numerous diverse scenes and images, including some known to contain severe gray-world violations.

7. Note to readers

Color version of the figures which appear in this paper is avaliable upon request. Please send e-mail to zrahman@cs.wm.edu or us-mail to Zia-ur Rahman, Department of Computer Science, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187-8795.

REFERENCES

1. T. Cornsweet, *Visual Perception.* Orlando, FL: Academic Press, 1970.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.*, vol. 80, pp. 5163–5169, 1983.
3. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3078–3080, 1986.
4. E. Land, "Recent advances in retinex theory," *Vision Research*, vol. 26, no. 1, pp. 7–21, 1986.
5. A. C. Hurlbert, *The Computation of Color.* PhD thesis, Massachusetts Institute of Technology, September 1989.
6. A. C. Hurlbert, "Formal connections between lightness algorithms," *Journal of the Optical Society of America A*, vol. 3, pp. 1684–1693, 1986.
7. A. C. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science*, vol. 239, pp. 482–485, 1988.
8. A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," *IEEE Transactions on Neural Networks*, vol. 2, pp. 237–247, March 1991.

9. 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.), pp. 370–376, San Mateo, CA: Morgan Kaufmann, 1991.

10. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, 1996. Accepted for publication.

11. Z. Rahman, D. Jobson, and G. A. Woodell, "Multiscale retinex for color image enhancement," in *Proceedings of the IEEE International Conference on Image Processing*, IEEE, 1996.

12. Z. Rahman, "Properties of a center/surround Retinex Part One: Signal processing design," *NASA Contractor Report #198194*, 1995.

13. D. J. Jobson and G. A. Woodell, "Properties of a center/surround Retinex Part Two: Surround design," *NASA Technical Memorandum #110188*, 1995.

14. W. D. Wright, *The Measurement of Colour*. London: Hilger and Watts, second ed., 1958.

Integrated Wavelet Compression and Restoration

Z. Rahman

Wavelet Applications in Signal and Image Processing IV
Michael A. Unser, Akram Aldroubi, Andrew F. Laine, eds., Proc
SPIE 2825
(August 1996)

Integrated Wavelet Compression and Restoration

Zia-ur Rahman[†]
College of William & Mary, Williamsburg, VA 23187

Abstract

The performance of wavelet compression algorithms is generally judged solely as a function of the compression ratio and the visual artifacts which are perceivable in the reconstructed image. The problem then becomes one of obtaining the best compression with fewest visible artifacts—a very subjective measure. Our wavelet compression algorithm uses an information theoretic analysis for the design of the compression maps. We have previously shown that maximizing the information for a given visual communication channel also maximizes the visual quality of the restored image. We utilize this to design quantization maps which maximize information for a given compression ratio. Hence we are able to design quantization maps which maximize the restorability of an image—i.e. the information content, the image quality, and the mean-square difference fidelity—for a given compression ratio.

KEY WORDS: Image restoration, image compression, image quality

1. Introduction

Image compression algorithms are generally evaluated in terms of the amount of data compression, a measurable quantity, and the visual quality for this data rate, a subjective quantity. Neglected in the design of compression algorithms and the evaluation process are the effects on the data rate and the quality of the image due to the image acquisition and display systems. Some effort has been made to relate the effects of display into evaluation of image quality[1] but none in incorporating the characteristics of the image gathering system into the design of the compression algorithm. We present a new approach to designing and evaluating image compression algorithms in terms of the information transmitted by the visual communication channel. This approach incorporates the effects of the image gathering device characterized by the signal to noise ratio and the spatial frequency response (SFR) of the combined optics and photodetector array, the quantization due to the analog-to-digital (A/D) converter, and the errors due to the compression process into the design of the quantization maps. These maps are related to visual quality by the amount of information they allow through. Previously,[2-6] we have combined Shannon's information theory[7] with Wiener's restoration filter[8] and with the critical limiting factors that affect a visual communication channel to provide rigorous quantitative metrics for characterizing its design and evaluating its performance in terms of restorability. We now integrate lossy data compression into this framework to optimize data rates[9] and evaluate its effects on image resolution, and hence, quality.

Figure 1 shows the visual communication channel. At the head of the communication channel is an image-gathering device which consists of a lens, a photodetector array, and an analog-to-digital (A/D) converter. This device converts the radiance field incident on it into a quantized, digital signal which is then transmitted. At the tail of the channel is a receiver which resolves the signal and provides the information to an image display device (e.g. video monitor, or a printer) which, in turn, represents this information in a form suitable for interpretation by an observer. Between the output of the A/D converter and the receiver, any number of digital image processing algorithms can be applied for image enhancement and efficient data representation and transmission. Traditional analysis of image compression and restoration algorithm is generally restricted to this stage only, neglecting the analog-to-digital conversion at image acquisition and digital-to-analog conversion at image display. This leads to an inherently incomplete model which results in restorations and compression rates which could have been better if the conversions at acquisition and display had been properly taken into account.

The imaging process injects errors in the original information that is incident on the image gathering device. The combined SFR of the lens and the photosensor array blurs the radiance field; the photosensor array and the

[†]This research was funded by NASA Langley Research Contract #NAS1-19603 to Science and Technology Corporation and by NASA Langley Research Grant #NAG1-1847 to the College of William & Mary.

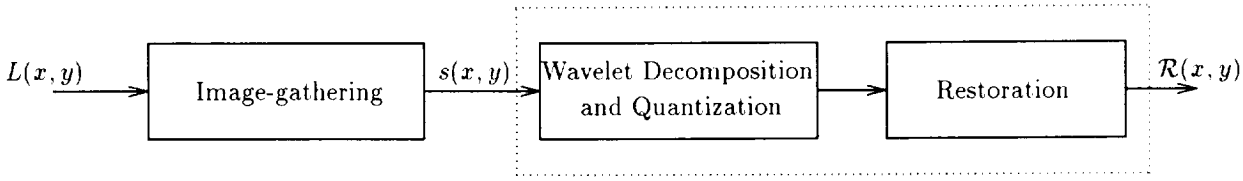


Figure 1: The visual communication channel

A/D converter introduce electronic and quantization noise respectively; and the SFR of the display spot of the image-display device blurs the resolved signal. These errors result in a reduction in the amount of information that is received by the observer when compared with the information which was present at the beginning of the imaging process. Since for a lot of applications (e.g. remote sensing) the *only* information the observer has is this received information, it is advantageous to minimize the channel effects so as to maximize information. A rigorous mathematical analysis provides the framework to evaluate the performance of the communication channel and can thus be used to informationally optimize its design in terms of both resolution (restoration) and compression. We feel that this is *essential* in designing a channel which provides the most information, and hence the highest resolution, for the least data.

2. Mathematics of the visual communication channel

The visual communication channel is divided into five stages: image-gathering, decomposition, quantization, synthesis, and restoration. Though we present the mathematics of each stage individually, each stage builds on the preceding stages and the restoration filters depend upon the end-to-end process.

2.1. Image gathering

The image gathering device converts the incident continuous radiance field $L(x,y)$ into the discrete signal $s(x,y)$ (Figure 2). The combined SFR of the device optics and photosensor array aperture, $\tau_d(x,y)$, blurs the input $L(x,y)$, which is sampled by the rectangular unit sampling lattice, $\underline{\underline{|||}}(x,y)$, and corrupted by the additive noise due to the analog-to-digital (A/D) conversion, $N_{a/d}(x,y)$, and the electronics, $N_e(x,y)$, producing

$$s(x,y) = [K L(x,y) * \tau_d(x,y)]\,\underline{\underline{|||}} + N_e(x,y) + N_{a/d}(x,y), \tag{1a}$$

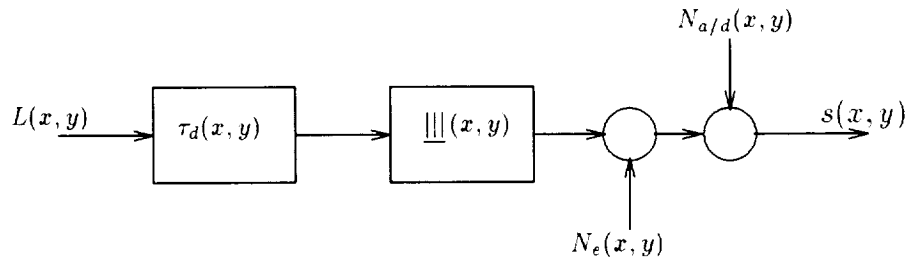


Figure 2: Image gathering: The radiance field $L(x,y)$ is converted to signal $s(x,y)$.

where K is the steady-state linear radiance-to-signal conversion gain. The unit sampling lattice is given by $\underline{\underline{|||}} = \sum_{m,n=-\infty}^{\infty} \delta(x-m, y-n)$. Rewriting in Fourier domain,

$$\tilde{s}(v,\omega) = \left[K\hat{L}(v,\omega)\tilde{\tau}_d(v,\omega)\right] * \hat{\underline{\underline{|||}}} + \tilde{N}_e(v,\omega) + \tilde{N}_{a/d}(v,\omega), \tag{1b}$$

where the notation $\hat{p}(v,\omega)$ refers to the *continuous* Fourier transform of a function $p(x,y)$ and $\tilde{q}(v,\omega)$ refers to the *discrete* Fourier transform of a function $q(x,y)$. The Fourier transform of the sampling lattice $\underline{\underline{|||}}$ is given by $\hat{\underline{\underline{|||}}} = \sum_{m,n=-\infty}^{\infty} \delta(v-m, \omega-n)$, with the associated sampling passband $\hat{B} = \{v,\omega : |v|, |w| \leq 0.5\}$. The probability density function of the noise $N_{a/d}(x,y)$ can be written as

$$p[N_{a/d}(x,y)] = \frac{\kappa}{s_{p_{max}}(x,y) - s_{p_{min}}(x,y)} = \frac{\kappa}{2\sigma_s}, \tag{2}$$

for $s_{p_{max}} = k\sigma_s$ and $s_{p_{min}} = -k\sigma_s$ which specify the range of the signal; κ is the number of quantization levels of the A/D converter; and $\sigma_s^2 = \iint_{\hat{B}} \tilde{\Phi}_s(v,\omega)\, dv d\omega$. The power spectral density (PSD) of the signal $\tilde{\Phi}_s(v,\omega)$ prior to quantization is

$$\begin{aligned} \tilde{\Phi}_s(v,\omega) &= E\left[\tilde{s}(v,\omega)\tilde{s}^*(v,\omega)\right] \\ &= \left[K^2\hat{\Phi}_L(v,\omega)|\tilde{\tau}_d(v,\omega)|^2\right] * \hat{\underline{\underline{|||}}} + \tilde{\Phi}_{N_e}(v,\omega), \end{aligned} \tag{3}$$

where $E[\,]$ is the expectation operator, and * indicates complex conjugation. Assuming that the error within each quantization interval is uncorrelated with errors within other intervals,

$$\tilde{\Phi}_{N_{a/d}}(v,\omega) = \sigma_{N_{a/d}}^2 = \frac{1}{3}\left(\frac{k\sigma_s}{\kappa}\right)^2. \tag{4}$$

2.2. Decomposition

Figure 3 shows signal decomposition, and synthesis and restoration. The signal $s(x,y)$ is decomposed into N parts, at L levels using the discrete wavelet transform.

$$s_{0,1}(x,y) \equiv s(x,y) \tag{5a}$$

$$s_{l,\beta}(x,y) = \left[s_{l-1,1}(x,y) * \tau_{A_{l,\beta}}(x,y)\right] \underline{\underline{|||}}_w, \quad \beta = 1,...,N, l = 1,...,L, \tag{5b}$$

where $s_{l-1,1}(x,y)$ is the "low frequency" band from the previous level, $\tau_{A_{l,\beta}}(x,y)$ are the wavelet analysis filters at the current level and $\underline{\underline{|||}}_w = XY\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\delta(x-mX, y-nY)$ is the downsampler by $X = Y = \sqrt{N}$. In the frequency domain this can be seen as dividing the passband into N segments each occupying $1/N$th the original bandwidth.

$$\tilde{s}_{0,1}(v,\omega) \equiv \tilde{s}(v,\omega) \tag{6a}$$

$$\tilde{s}_{l,\beta}(v,\omega) = \left[\tilde{s}_{l-1,1}(v,\omega)\tilde{\tau}_{A_{l,\beta}}(v,\omega)\right] * \hat{\underline{\underline{|||}}}_w, \quad \beta = 1,...,N, l = 1,...,L, \tag{6b}$$

where the $\tilde{\tau}_{A_{l,\beta}}(v,\omega)$ are, generally, orthogonal, and $\hat{\underline{\underline{|||}}}_w = \sum_{m=0}^{X-1}\sum_{n=0}^{Y-1}(v-\frac{m}{X}, \omega-\frac{n}{Y})$. The signals $\tilde{s}_{l,\beta}$ occupy different frequency bands. Each signal can possess quite different characteristics and hence be amenable to different methods and rates of quantization. This provides a versatile method for efficient signal representation.

Equations 5 show the relationship between the signal $s(x,y)$ and the decomposed signals $s_{l,\beta}(x,y)$ in terms of the wavelet analysis filters and the downsampler. More explicitly, using Equations 1,

$$s_{1,\beta}(x,y) = \left[\left\{\left[KL(x,y) * \tau_d(x,y)\right]\underline{\underline{|||}} + N_e(x,y) + N_{a/d}(x,y)\right\} * \tau_{A_{1,\beta}}(x,y)\right]\underline{\underline{|||}}_w, \tag{7a}$$

$$\tilde{s}_{1,\beta}(v,\omega) = \left[K\hat{L}(v,\omega)\tilde{\tau}_d(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega)\right] * \hat{\underline{\underline{|||}}}' + \left[\tilde{N}_e(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega) + \tilde{N}_{a/d}(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega)\right] * \hat{\underline{\underline{|||}}}_w, \tag{7b}$$

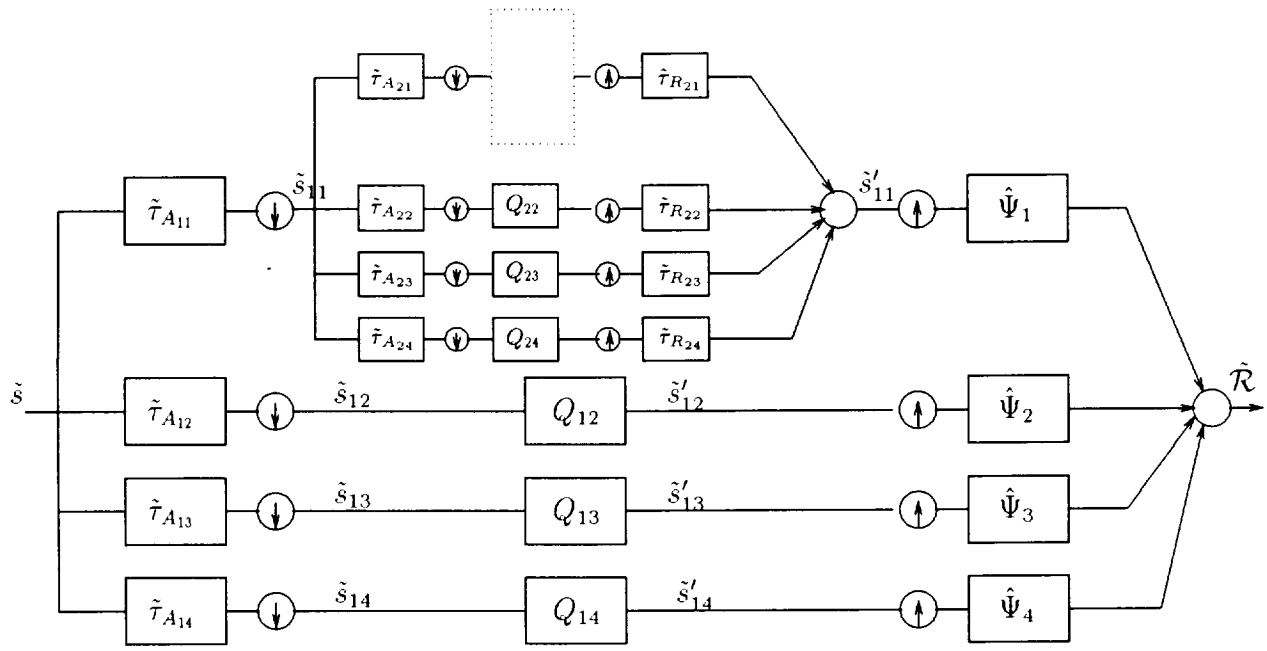


Figure 3: Wavelet decomposition, quantization, and restoration

---

where $\underline{\underline{|||}}' = \sum_{m,n} \delta\left(v - \frac{m}{X},\ \omega - \frac{n}{Y}\right)$, is the combination of the uniform rectangular sampling lattice of the image-gathering device and the downsampler. It has sampling intervals of (X, Y) and associated passband $B_l(v,\omega) = \{v,\omega: |v|, |\omega| \leq 2^{-(l+1)}\}$.

2.3. Quantization

The wavelet coefficients are quantized in the spatial-frequency domain. Each coefficient, $\tilde{s}_{l,\beta}(v,\omega)$ contains a measurable amount of information $\mathcal{H}_{l,\beta}(v,\omega)$ (§3). McCormick[10] et al., and Huck[4, 11] have shown that maximizing information leads to maximizing the image quality. Hence the coefficients are quantized under the constraint of maximizing total information for the number of bits used to represent the signal. The quantization transforms the wavelet coefficient $\tilde{s}_{l,\beta}(v,\omega)$ into a representation $Q[\tilde{s}_{l,\beta}(v,\omega)]$. The inverse process transforms the $Q[\tilde{s}_{l,\beta}(v,\omega)]$ back to the coefficients plus, perhaps, a noise term.

$$\tilde{s}'_{l,\beta}(v,\omega) = \tilde{s}_{l,\beta}(v,\omega) + \tilde{N}_{Q_{l,\beta}}(v,\omega), \quad \beta = 2, .., N \tag{8a}$$

$$\tilde{s}'_{l,1}(v,\omega) = \sum_{\beta=1}^{N} \tilde{s}'_{l+1,\beta}(v,\omega)\tilde{\tau}_{R_{l+1,\beta}}(v,\omega) \tag{8b}$$

where $\tilde{s}'_{l,\beta}(v,\omega)$ are the dequantized coefficients, and $\tilde{N}_{Q_{l,\beta}}(v,\omega)$ represents the "noise" due to the quantization/dequantization process. The initial conditions are given by

$$\begin{aligned} \tilde{s}'_{1,\beta}(v,\omega) &= \tilde{s}(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega) \\ &= \left[K\hat{L}(v,\omega)\tau_d(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega)\right] * \underline{\underline{|||}}' + \left[\tilde{N}_e(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega)\right] * \underline{\underline{|||}}_w \\ &\quad + \left[\tilde{N}_{a/d}(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega)\right] * \underline{\underline{|||}}_w + \tilde{N}_{Q_{1,\beta}}(v,\omega). \end{aligned} \tag{8c}$$

Assuming uniform quantization for each wavelet coefficient between the range $\pm k\sigma_{l,\beta}(v,\omega)$ we can write the PSD of the coefficient quantization noise as

$$\tilde{\Phi}_{N_{Q_{l,\beta}}}(v,\omega) = \sigma^2_{N_{Q_{l,\beta}}}(v,\omega) = \frac{1}{3}\left(\frac{k\sigma_{l,\beta}(v,\omega)}{\kappa_{l,\beta}(v,\omega)}\right)^2, \tag{9}$$

where $\sigma_{l,\beta}(v,\omega)$ are the ensemble standard deviations for the coefficient at frequency (v,ω) in band β and level l, and $\kappa_{\beta,l}(v,\omega)$ is the number of quantization levels for that frequency location.

2.4. Synthesis

The synthesis filters reconstruct the decomposed signals. When the coefficients are not quantized, this can be done perfectly.[12, 13, 14] With lossy quantization, however, the synthesis filters are generated to minimize the minimum mean square error (MSRE) $\epsilon^2(v,\omega)$ between the decomposed and reconstructed signal. The MSRE is given by

$$\epsilon^2(v,\omega) = E\left[|\tilde{s}_{l,1}(v,\omega) - \tilde{s}'_{l,1}(v,\omega)|^2\right],$$

$\tilde{s}_{l,1}(v,\omega)$ is the input and $\tilde{s}'_{l,1}(v,\omega)$ is the output to level $l+1$. Using Equations 6 and 8,

$$\tilde{\tau}_{R_{l+1,\beta}}(v,\omega) = \frac{\tilde{\Phi}_{S_{l,1}}(v,\omega)\tilde{\tau}^*_{A_{l+1,\beta}}(v,\omega)}{\left[\tilde{\Phi}_{S_{l,1}}(v,\omega)|\tilde{\tau}_{A_{l+1,\beta}}(v,\omega)|^2\right] * \underline{\underline{|||}}_w + \tilde{\Phi}_{N_{Q_{l+1,\beta}}}(v,\omega)}, \tag{10}$$

where $\tilde{\Phi}_{S_{l,1}}(v,\omega)$ is the PSD of $\tilde{s}_{l,1}(v,\omega)$, and $\tilde{\Phi}_{N_{Q_{l+1,\beta}}}(v,\omega)$ is the PSD of the quantization noise in band $l+1$. The $\tilde{\Phi}_{S_{l,\beta}}(v,\omega)$ can be defined by a set of iterative equations

$$\tilde{\Phi}_{S_{0,1}}(v,\omega) = \left[K^2\hat{\Phi}_L(v,\omega)|\tilde{\tau}_d(v,\omega)|^2\right] * \underline{\underline{|||}} + \tilde{\Phi}_{N_{a/d}}(v,\omega) + \tilde{\Phi}_{N_e}(v,\omega) \tag{11a}$$

$$\tilde{\Phi}_{S_{l,\beta}}(v,\omega) = \tilde{\Phi}_{S_{l-1,1}}(v,\omega)|\tilde{\tau}_{A_{l,\beta}}(v,\omega)|^2 * \underline{\underline{|||}}_w \tag{11b}$$

$$\approx \left[\tilde{\Phi}_{S_{0,1}}(v,\omega)|\tilde{\tau}_{A_{l,\beta}}(v,\omega)|^2 \prod_{k=1}^{l-1}|\tilde{\tau}_{A_{k,1}}(v,\omega)|^2\right] * \underline{\underline{|||}}_l, \tag{11c}$$

where $\underline{\underline{|||}}_l = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\delta(v - m/X^l, w - n/Y^l)$.

2.5. Restoration

The Wiener-matrix restoration filters $\hat{\Psi}_\beta(v,\omega)$[6, 15] synthesize the N level-1 outputs $\tilde{s}'_{1,\beta}(v,\omega)$ into a single *continuous* image $\mathcal{R}$ given by

$$\mathcal{R}(v,\omega) = \sum_{\beta=1}^{N} K^{-1}\hat{\Psi}_\beta(v,\omega)\,\tilde{s}'_{1,\beta}(v,\omega) \tag{12}$$

where

$$\hat{\Psi}_\beta(v,\omega) = \hat{\Phi}_L(v,\omega)\tilde{\tau}^*_d(v,\omega)\sum_{\alpha=1}^{N}\tilde{\tau}^*_{A_{1,\alpha}}(v,\omega)\left[\hat{T}^{-1}(v,\omega)\right]_{\alpha\beta}, \tag{13}$$

$$\left[\hat{T}(v,\omega)\right]_{\beta\alpha} = \left[\hat{\Phi}_L(v,\omega)|\tilde{\tau}_d(v,\omega)|^2\tilde{\tau}_{A_{1,\beta}}(v,\omega)\tilde{\tau}^*_{A_{1,\alpha}}(v,\omega)\right] * \underline{\underline{|||}}' + \left[\tilde{\Phi}_{N_e}(v,\omega)\tilde{\tau}_{A_{1,\beta}}(v,\omega)\tilde{\tau}^*_{A_{1,\alpha}}(v,\omega)\right] * \underline{\underline{|||}}_w \tag{14}$$
$$+ \left[\tilde{\Phi}_{N_{a/d}}\tilde{\tau}_{A_{1,\beta}}(v,\omega)\tilde{\tau}^*_{A_{1,\alpha}}(v,\omega) * \underline{\underline{|||}}_w\right] + \tilde{\Phi}_{N_{Q_{1,\beta}}}(v,\omega)\,\delta(\beta,\alpha),$$

$\hat{\Phi}_L(v,\omega), \tilde{\Phi}_{N_e}(v,\omega)$, $\tilde{\Phi}_{N_{a/d}}(v,\omega)$ and $\tilde{\Phi}_{N_{Q_{1,\beta}}}(v,\omega)$ represent the PSD of the radiance field, the photodetector noise, the A/D quantization noise, and the level-1 quantization noise respectively. When $\tilde{\tau}_{A_{l,\beta}}(v,\omega)\tilde{\tau}^*_{A_{l,\alpha}}(v,\omega) = |\tilde{\tau}_{A_{l,\beta}}(v,\omega)|^2\delta(\beta,\alpha)$, the Wiener matrix filter reduces to

$$\Psi_\beta(v,\omega) = \frac{\hat{\Phi}_L(v,\omega)\tilde{\tau}^*_d(v,\omega)\tilde{\tau}^*_{A_{1,\beta}(v,\omega)}}{\tilde{\Phi}_s(v,\omega)|\tilde{\tau}_{A_{1,\beta}}(v,\omega)|^2 * \underline{\underline{|||}}_w + \tilde{\Phi}_{N_{Q_{1,\beta}}}(v,\omega)}. \tag{15}$$

The formulation of the Wiener-matrix filter takes into account the degradations due to not only the blurring and noise[16, 17] but also due to insufficient sampling, the analysis filter response, and quantization. It also suppresses the blurring and raster effects in image display by interpolating the image-gathering lattice on an at least 4 times finer image-display lattice.

3. Information

The information rate, $\mathcal{H}$, is used to evaluate the performance of the visual communication channel. The only information the observer has about the incident radiance field is that contained in the restored image. The degradations due to aliasing and various noise sources appear as artifacts. The information at each level in each decomposed band $\mathcal{H}_{l,\beta}$ is

$$\mathcal{H}_{l,\beta} = \iint_{\tilde{B}_l} \mathcal{H}_{l,\beta}(v,\omega)\, dv d\omega = \frac{1}{2} \iint_{\tilde{B}_l} \log_2 \left[1 + \frac{\tilde{\Phi}_{S_{l-1,1}}(v,\omega)|\tilde{\tau}_{A_{l,\beta}}(v,\omega)|^2}{\tilde{\Phi}_{S_{l-1,\beta}}(v,\omega)|\tilde{\tau}_{A_{l,\beta}}(v,\omega)|^2 * \underline{|||}_a + \tilde{\Phi}_{Q_{N_{l,\beta}}}(v,\omega)} \right] dv d\omega. \qquad (16)$$

where $\underline{|||}_a = \sum_{m=0}^{X-1} \sum_{n=0}^{Y-1} \delta(v - \frac{m}{X}, w - \frac{n}{Y}), m = n \neq 0$, are the sidebands of the down sampler, and the PSDs $\tilde{\Phi}_{S_{l,1}}(v,\omega)$ are given by Equation 11. The total information for the visual communication channel is

$$\mathcal{H} = \sum_{l=1}^{L-1} \sum_{\beta=2}^{N} \mathcal{H}_{l,\beta} + \sum_{\beta=1}^{N} \mathcal{H}_{L,\beta}.$$

4. Quantization Algorithm

Based on Equation 16(a) which provides the amount of information each wavelet coefficient contributes to the total information, a quantization scheme can be devised which either maximizes the total information for a given number of bits, or minimizes the total number of bits given an acceptable level of information loss. The scheme is as follows:

1. Quantize all the coefficients at the maximum rate.
2. For each coefficient, determine the loss in total information when the quantization rate for the coefficient is reduced by 1 bit.
3. For the coefficient which least affects the total information, reduce the quantization rate by 1 bit.
4. Iterate this process until either the total number of bits is exhausted, or the acceptable information loss has been achieved.

Because the power spectral density used in developing these quantization measure is not the actual PSD of the signal, but instead a statistical quantity, the quantization tables thus obtained can be used for a wide range of input scenes, producing excellent results for input radiance fields which closely match the assumed radiance files PSD and good results for others.

5. Simulation Assumptions

Since it is virtually impossible to successfully estimate all the parameters of an image-gathering device from the received signal, we use simulated imagery to test our algorithm. This allows to closely control the system characteristics which affect image quality and compression and observe their effects in a controlled environment.

We use targets made of randomly generated polygons whose mean spatial detail—the average distance between edges—μ, is Poisson distributed and intensity levels are Gaussian distributed with standard deviation σ_L. The targets with mean spatial details $\mu = 1$ and 3 are shown in Figure 4. The associated radiance field $L(x,y)$ is stationary and Gaussian. We assume that both the electronic noise $N_e(x,y)$, and the A/D quantization noise $N_{a/d}(x,y)$, are

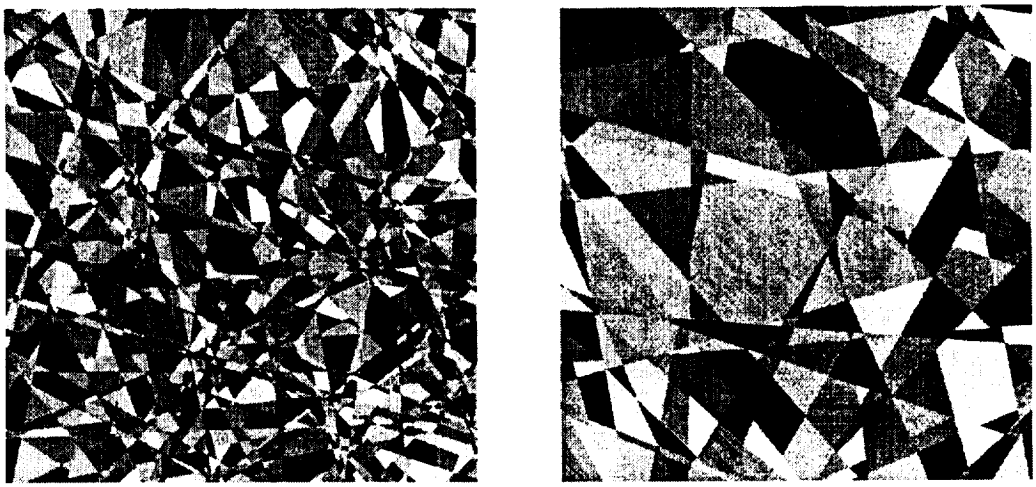


Figure 4: Random polygon targets: (a) $\mu = 1$ (b) $\mu = 3$

Design	ρ_c	$K\sigma_L/\sigma_{N_e}$	$\mathcal{H}_{max}$
1	0.3	256	4.4
2	0.4	64	3.3
3	0.5	16	2.1

Table 1: Informationally optimized visual communication channel designs.

uncorrelated with the radiance field $L(x, y)$. In addition, we model the SFR of the image gathering device with a Gaussian, $\hat{\tau}_d(v, \omega) = \exp\left[-(\frac{\rho}{\rho_c})^2\right]$, where $\rho = v^2 + \omega^2$, and the electro-optical design index ρ_c, which controls the width of the response, and hence the tradeoff between aliasing and blurring, is the point where $\hat{\tau}_d(v, \omega) = 1/e \approx 0.37$. When ρ_c is large there is more aliasing and when it is low, there is more blurring. The restoration filters Ψ_β are generated at at least 4 times finer density than the sampling lattice to suppress the raster effects of the display device[10].

6. Results

In order to design optimally efficient visual communication channels, i.e. channels which transmit the most information for the least data, it is first necessary to look at optimal visual communication channels, i.e channels which transmit the most information given a certain image-gathering device.

Visual communication channels, rather simplistically, are generally characterized only by the signal-to-noise ratio (SNR),* and the SFR of the combined electro-optics. Huck et al.[3, 4, 11] have studied the problem of designing visual communication channels for maximum information throughput in detail. Table 1 specifies three visual communication channel designs, in terms of the SNR and the electro-optical design index ρ_c, that have been informationally optimized. The amount of information each of these designs transmits is also given. We will introduce lossy wavelet compression as an additional constraint on these communication channels and design quantization maps which maximize the information throughput for a given bit rate, or minimize the bit rate for a give acceptable amount of information loss.

The overall effect of lossy quantization on the transmitted signal is the introduction of an additional noise source $\widetilde{N}_{Q_{1,\beta}}(v, \omega)$ (Equations 13-16) which affects the visual quality of the reconstructed image. If the quantization maps are carefully designed, these effects can be minimized. Figure 5 shows quantization maps developed using the

*By this we mean the SNR of the electronic device used for image gathering. We will not consider the effects of channel noise on the integrity of transmission.

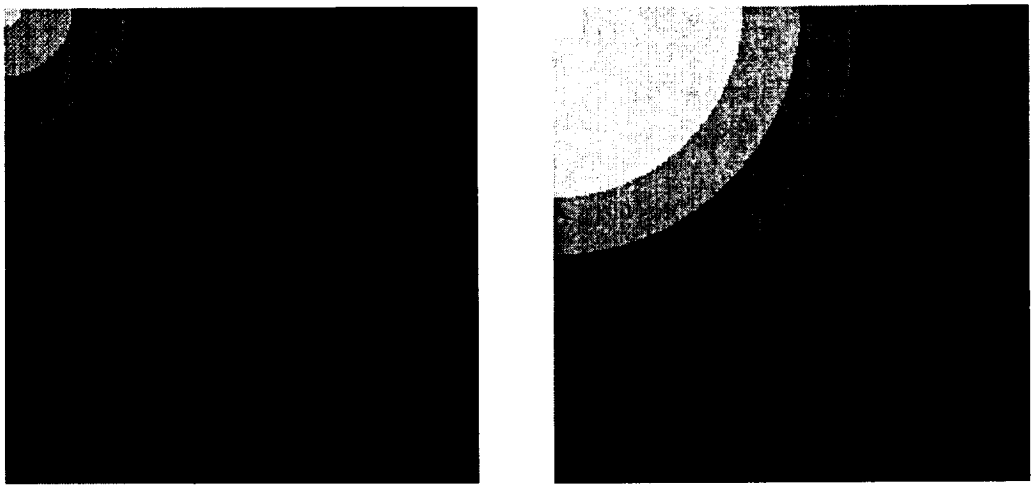


Figure 5: Quantization maps designed with a 5% acceptable loss in information for (a) Design 3 (b) Design 1 (Table 1).

algorithm given in Section 4 for Designs 1 and 3 specified in Table 1. The total acceptable loss in information for these maps is no more than 5%. The priority with which bits are encoded in each band is clearly evident in the quantization maps. Most of the coefficients in the lowest frequency bands contain high information so they are retained in preference to the coefficients in the higher frequency bands which contain less information. Because the bit allocation algorithm (§3) reduces the number of bits for the coefficient based upon the amount of information $\mathcal{H}(v,\omega)$ it contains, the initial reduction in the number of bits does not have a significant impact upon the total information. However, as more and more bits are discarded, the information content of the affected coefficients is higher, and the rate of information reduction increases. It is also interesting to look at the reduction in total information as a function of the reduction in the number of bits in each band. The information content of coefficients in band 4 is very low, hence the number of bits can be reduced by about 15% before any impact is felt on the total information; coefficients in bands 2 and 3 have higher information content and have more of an impact on the total information; and the coefficients in band 1 have the highest information content as is evident by the sharp reduction in $\mathcal{H}$ as the total number of bits for band 1 decreases.

A second point of interest is the the effect on $\mathcal{H}$ as a function of the channel characteristics. The total information for all designs is a monotonically increasing function of the total number of bits being used to represent the signal. The rate of increase, however, is smaller at lower values of the SNR. Thus, for a given number of bits, the reduction in information per bit is generally highest for Design 1, and the lowest for Design 3 which suggests that greater data compression can be achieved for channels that have lower SNRs since the quantization effects will not overwhelm what is already a noisy signal.

Figure 6(b) shows the decomposed images of the targets shown in Figure 4, and Figure 6(c) shows the restored images. The results are shown for an acceptable loss level of 5%. The restorations show some obvious artifacts, more so in the $\mu = 3$ image than in the $\mu = 1$ image. This is because the aliasing artifacts and colored noise are, to a certain extent, masked by the detail in the scene. Conversely, one sees loss of detail in the $mu = 1$ scene due to the loss of high-frequency information in the quantization process. These results point to the necessity of developing better filter banks for the restoration of images.

7. Conclusions

We have presented an integrated treatment of designing quantization maps for a given visual communication channel based upon the metric of maximizing information. Since the design of a compression algorithm should

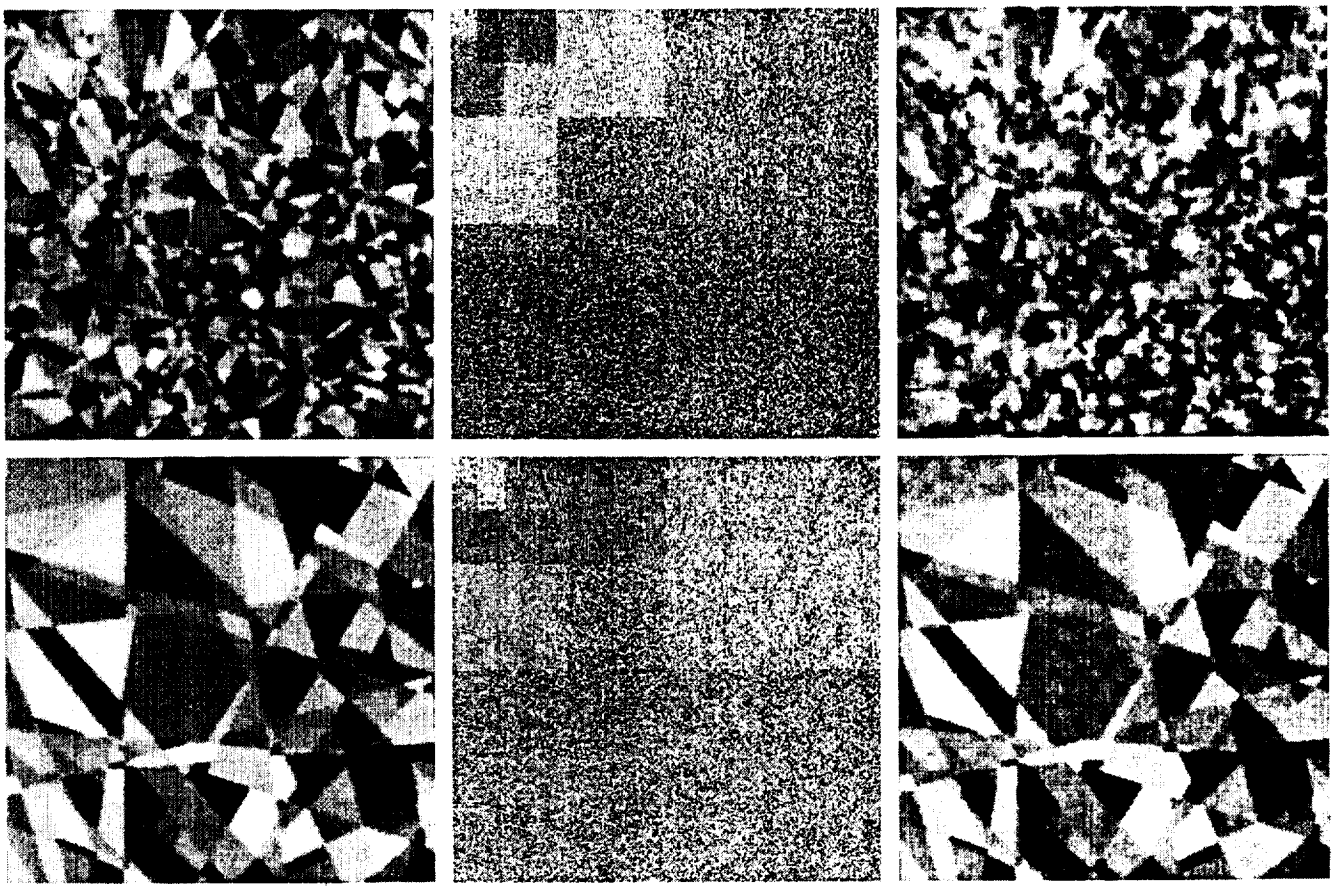


Figure 6: (a) The blurred and noisy signals (output of the image gathering stage). (b) the decomposed signals (output of the wavelet decomposition stage) at $L = 3$, and (c) The expected restored images. Results are shown for restoration done from level-1 components. The original scenes are shown in Figure 4.

maximize visual quality in addition to the compression ratio, it is imperative that an end-to-end system analysis be used in both the design of the compression algorithms and in the design of the evaluation process. Though some work has been done on incorporating the effects of image display on the perceived image quality[1], our development is unique in the sense that it incorporates the effects of the image gathering process into the design of the compression algorithm.

Though the algorithm outlined here is simple conceptually, it is computationally intensive and for that reason the results presented here are for the simplest case where the analysis filters are orthogonal. This reduces the number of computations significantly but at the same time provides good insight into the results which can be expected from this approach. But this also does not fully achieve the image quality that is expected of this algorithm. Current research is looking at improving the analysis/synthesis filters, as well as improving the robustness of the restoration filters. This will lead to both better image quality in terms of human perception, but also restorability in terms of the amount of detail resolved in the displayed images, and improved compression ratios.

REFERENCES

1. A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Perceptually lossless wavelet compression," in *Proceedings of the Combined Industry, Space and Earth Science Data Compression Workshop* (A. B. Kiely and R. L. Renner, eds.), Proc. SPIE 1705, 1996.

2. C. L. Fales and F. O. Huck, "An information theory of image gathering," *Information Sciences*, vol. 57–58, pp. 245–285, 1991.

3. F. O. Huck, C. L. Fales, R. Alter-Gartenberg, and Z. Rahman, "Visual communication: Information and data transmission," *Journal of Visual Communication and Image Representation*, vol. 5, no. 3, 1994.

4. F. O. Huck, C. L. Fales, R. Alter-Gartenberg, and Z. Rahman, "On the assessment of visual communication," in *Handbook of Statistics* (N. Bose and C. R. Rao, eds.), vol. 10, Elsevier Science Publishers, 1993.

5. F. O. Huck, C. L. Fales, R. Alter-Gartenberg, Z. Rahman, and S. E. Reichenbach, "Visual communication: Information and fidelity," *Journal of Visual Communication and Image Representation*, vol. 4, no. 2, 1993.

6. C. L. Fales, F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, "Multiresolution image gathering and restoration," *Journal of Visual Communication and Image Representation*, vol. 3, no. 4, pp. 356–363, 1992.

7. C. Shannon and W. Weaver, *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press, 1964. Originally published in the *Bell System Technical Journal*, 27:379–423 and 28:623–656, 1948.

8. N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* Cambridge, MA: MIT Press, 1949.

9. Z. Rahman, "Information theoretic assessment of visual communication with wavelet coding," in *Visual Information Processing IV* (F. O. Huck and R. D. Juday, eds.), Proc. SPIE, 1995.

10. J. A. McCormick, R. Alter-Gartenberg, and F. O. Huck, "Image gathering and restoration: Information and visual quality," *Journal of the Optical Society of America A*, vol. 6, no. 7, pp. 987–1005, 1989.

11. F. O. Huck, C. L. Fales, J. A. McCormick, and S. K. Park, "Image-gathering system design for information and fidelity," *Journal of the Optical Society of America A*, vol. 5, no. 3, pp. 285–299, 1988.

12. P. P. Vaidyanathan, "Quadrature mirror filter banks, m-band extensions and perfect reconstruction techniques," *IEEE Signal Processing Magazine*, no. 7, 1987.

13. J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 5, pp. 1278–1288, 1986.

14. A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition.* Orlando, FL: Academic Press, 1992.

15. R. Alter-Gartenberg and Z. Rahman, "Multiresolution imaging: Information efficiency and visual quality," in *Visual Information Processing* (F. O. Huck and R. D. Juday, eds.), Proc. SPIE 1705, 1992.

16. M. Ozkan, T. Erdem, I. Sezan, and M. Telkap, "Efficient multiframe Wiener restoration for blurred and noisy image sequences," *IEEE Transactions on Image Processing*, vol. 1, pp. 453–476, 1992.

17. R. C. Gonzalez and P. Wintz, *Digital Image Processing.* Reading, MA: Addison-Wesley, second ed., 1987.

On the Information-Theoretic Assessment of Visual Communication

F. O. Huck, C. L. Fales and Z. Rahman

Proceedings of the IEEE International Conference
on Image Processing
(September 1996)

ON THE INFORMATION-THEORETIC ASSESSMENT OF VISUAL COMMUNICATION

Friedrich O. Huck and Carl L. Fales

NASA Langley Research Center
Hampton, Virginia 23681

Zia-ur Rahman

College of William & Mary
Williamsburg, Virginia 23666

ABSTRACT

This paper deals with the extension of information theory to the assessment of visual communication from scene to observer. The mathematical development rigorously unites the electro-optical design of image gathering and display devices with the digital processing algorithms for image coding and restoration. Results show that:

- End-to-end system analysis closely correlates with measurable and perceptual performance characteristics, such as data rate and image quality, respectively.
- The goal of producing the best possible image at the lowest data rate can be realized only if (a) the electro-optical design of the image-gathering device is optimized for the maximum-realizable information rate and (b) the image-restoration algorithm properly accounts for the perturbations in the visual communication channel.

1. INTRODUCTION

Modern visual communication channels increasingly combine image gathering and display with digital image coding and restoration (Fig. 1). So far, however, the image-gathering devices are still designed to produce the best possible images when reconstructed without the aid of the digital processing, and the image coding and restoration algorithms are still developed and evaluated without fully accounting for the critical constraints of image gathering and display.

The aim of this paper is to summarize some elements of a study [1,2] that rigorously unites the electro-optical design of image gathering and display devices with the digital processing algorithms for image coding and restoration. The study is based on the two classical works that are the foundation of modern communication theory. In one work Shannon [3] introduces the concept of the rate of transmission of information in a noisy channel, and in the other Wiener [4] introduces the concept of the minimum mean-square error restoration of signals corrupted by noise.

Although our mathematical development is firmly rooted in these familiar concepts, it leads to formulations that are significantly different from those that are found in the traditional literature on either rate distortion theory or digital image processing. These differences arise mainly because of two critical factors that this literature has not addressed so far, namely:

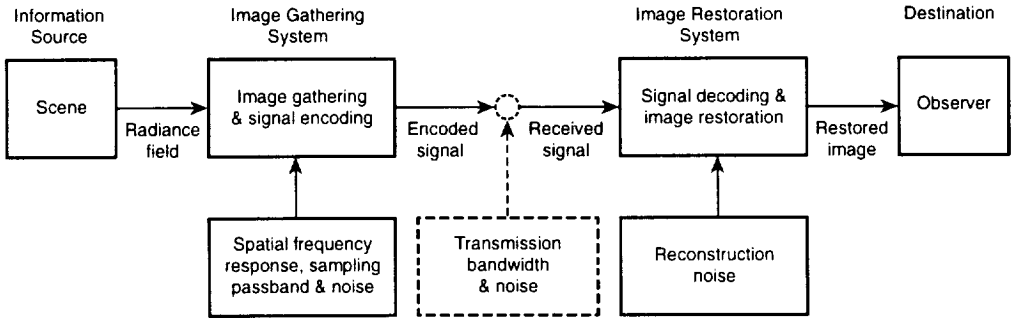


Figure 1. Model of visual communication channel together with the critical limiting factors that constrain its performance.

1. The limitations inherent in the realizability of the spatial frequency response (SFR) of optical apertures and the sampling passband of photodetection mechanisms impose a trade-off between blurring and aliasing on the design of the image-gathering device (Fig. 2). This precludes the treatment of visual communication as a bandwidth-limited process, but, instead, it requires the inclusion of the effects of insufficient sampling.

2. The image-gathering process bars the encoder from unperturbed access to the scene (i.e., the original source). This precludes the application of information theory directly to the scene for the analysis of data compression and rate distortion. Instead, this theory now must account for the perturbations that the image-gathering process causes, namely, the photodetector and quantization noises as well as the blurring and aliasing.

2. IMAGE GATHERING AND RESTORATION

The image-gathering process transforms the continuous radiance field $L(x,y)$ that is either reflected or emitted by the scene into the digital signal $s(x,y;\kappa)$, and the image-restoration process transforms this signal into the observed image $R(x,y;\kappa)$. In the Fourier domain, the image-gathering process is defined as

$$\tilde{s}(\upsilon,\omega;\kappa) = [K\hat{L}(\upsilon,\omega)\hat{\tau}(\upsilon,\omega)] * \underline{\underline{|||}} + \tilde{n}_p(\upsilon,\omega) + \tilde{n}_q(\upsilon,\omega;\kappa), \quad (1)$$

where $\hat{L}(\upsilon,\omega)$ is the continuous radiance-field transform, $\hat{\tau}(\upsilon,\omega)$ is the SFR of the image-gathering device, $\tilde{n}_p(\upsilon,\omega)$ and $\tilde{n}_q(\upsilon,\omega;\kappa)$ are the discrete photodetector and quantization noise transforms, and (υ,ω) are the spatial frequencies with units of cycles per sample. The tilde "~" is used instead of the caret "^" whenever the Fourier transformation is discrete and, therefore, the transformed function is periodic in the spatial frequency domain. The function $\underline{\underline{|||}}$ is the Fourier transform of the sampling lattice, as given by

$$\underline{\underline{|||}} = \sum_{m,n} \delta(\upsilon - m, \omega - n)$$

$$= \delta(\upsilon,\omega) + \underline{\underline{|||}}_s(\upsilon,\omega),$$

where $\delta(\upsilon,\omega)$ is the Dirac delta function and $\underline{\underline{|||}}_s$ accounts for the sampling sidebands. The associated sampling passband

$$\hat{B} = \left[(\upsilon,\omega); |\upsilon| < \frac{1}{2}, |\omega| < \frac{1}{2}\right]$$

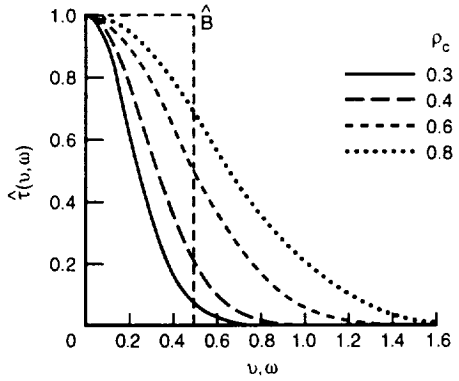


Figure 2. SFRs $\hat{\tau}(\upsilon,\omega)$ of the image-gathering device relative to the sampling passband $\hat{B}$ for unit sampling intervals.

has unit area, i.e., $|\hat{B}| = 1$. The analog-to-digital transformation is done for κ levels with η-bit quantization, where $\eta = \log \kappa$ and log denotes logarithm to base 2. The corresponding image-restoration process is defined as

$$\hat{R}(\upsilon,\omega;\kappa) = K^{-1}[s(x,y;\kappa)\hat{\Psi}(\upsilon,\omega;\kappa) + \hat{N}_r(\upsilon,\omega)], \quad (2)$$

where $\hat{\Psi}(\upsilon,\omega;\kappa)$ is a linear filter that records the digitally processed signal on an interpolation lattice that is sufficiently fine to suppress the blurring and raster effects of the image-display process and $\hat{N}_r(\upsilon,\omega)$ is the transform of the reconstruction noise (e.g., film granularity).

To assess visual communication in terms of information theory, the image-gathering process is constrained to be linear and isoplanatic (spatially invariant), and the radiance-field and noise amplitudes are constrained to be Gaussian, wide-sense stationary, and statistically independent. In addition, we characterize: (a) the radiance field $L(x,y)$ by the power spectral density (PSD)

$$\Phi_L(\upsilon,\omega) = \frac{1}{|A|}\overline{|L(\upsilon,\omega)|^2}$$

of an isoplanatism patch of the scene with area $|A|$, (b) the discrete signal $s(x,y)$ prior to quantization by the PSD

$$\tilde{\Phi}_s(\upsilon,\omega) = \left[K^2\hat{\Phi}_L(\upsilon,\omega)|\hat{\tau}(\upsilon,\omega)|^2\right] * \underline{\underline{|||}} + \tilde{\Phi}_p(\upsilon,\omega), \quad (3)$$

(c) the photodetector noise $n_p(x,y)$ by the PSD $\tilde{\Phi}_p(\upsilon,\omega)$, and (d) the quantization noise $n_q(x,y;\kappa)$ by the PSD

$$\tilde{\Phi}_p(\upsilon,\omega;\kappa) = \left(\frac{\sigma_s}{\kappa}\right)^2. \quad (4)$$

where

$$\sigma_s^2 = \iint_B \tilde{\Phi}_s(v,\omega)\, dv\, d\omega.$$

3. FIGURES OF MERIT

By accounting for the critical constraints of image gathering, we can quantitatively assess visual communication in terms of the following figures of merit:

1. The rate of transmission of information, or information rate, $\mathcal{H}$ that the image-gathering system produces for the radiance field that resides within its field of view, as given by

$$\mathcal{H} = \frac{1}{2} \iint_B \log\left[1 + \frac{\tilde{\Phi}_i(v,\omega)|\tilde{r}(v,\omega)|^2}{\tilde{\Phi}_n(v,\omega;\kappa)}\right] dv\, d\omega, \quad (5)$$

where

$$\tilde{\Phi}_n(v,\omega;\kappa) = \tilde{\Phi}_i(v,\omega)|\tilde{r}(v,\omega)|^2 * \underline{|||}_s + K^{-2}[\tilde{\Phi}_p(v,\omega) + \tilde{\Phi}_q(v,\omega;\kappa)].$$

2. The theoretical minimum data rate $\mathcal{E}$ which is associated with the information rate $\mathcal{H}$, as given by

$$\mathcal{E} = \frac{1}{2} \iint_B \log\left[1 + \frac{\tilde{\Phi}_s(v,\omega)}{\tilde{\Phi}_q(v,\omega;\kappa)}\right] dv\, d\omega. \quad (6)$$

This expression for $\mathcal{E}$ represents the entropy of completely decorrelated data.

3. The maximum-realizable fidelity $\mathcal{F}$ of the digital image that can be restored from the received information, unconstrained by the image-display medium, as given by

$$\mathcal{F} = \sigma_i^{-2} \iint_{-\infty}^{\infty} \tilde{\Phi}_i(v,\omega)\left[1 - 2^{-\tilde{\mathcal{H}}(v,\omega)}\right] dv\, d\omega, \quad (7)$$

where $\tilde{\mathcal{H}}(v,\omega)$ is the spectral distribution of the information rate $\mathcal{H}$ given by the integrand of Eq. (5).

Reference 2 fills in the many details. It also formulates the information rate $\mathcal{H}_o$ and the maximum-realizable fidelity $\mathcal{F}_o$ of the observed image that the image-restoration system produces from the received information on an image-display medium (e.g., film). In addition, it accounts for multiresolution decomposition (wavelet transform) to optimally integrate the economical encoding of the transmitted signal with image gathering and restoration. Finally, Ref. 5 applies the information-theoretic assessment to the electro-optical design of the image-gathering device. It accounts for (a) the f-number, diffraction, and transmittance shading of the objective lens, (b) the sensitivity, aperture shape, and sampling geometry of the photodetection mechanism, and (c) the dynamic-range compression with lateral inhibition in the focal plane.

Figure 3 characterizes the information rate $\mathcal{H}$ as a function of the electro-optical design of the image-gathering device, as specified by the optical-design index ρ_c and the rms signal-to-noise ratio (SNR). The curves show that the preferred SFR is a function of the SNR. This result is intuitively appealing for image restoration. In one extreme, when the SNR is low, one would prefer to avoid substantial blurring because the noise constrains the enhancement of fine spatial detail. In the other extreme, when the SNR is high, one would prefer to avoid substantial aliasing because then the noise no longer contrains this enhancement.

Figure 4 presents an information-entropy $\mathcal{H}(\mathcal{E})$ plot that characterizes the information rate $\mathcal{H}$ versus the associated theoretical minimum data rate $\mathcal{E}$ for η-bit quantization and three informationally optimized designs of the image-gathering device. This plot serves as a useful alternative to the familiar rate-distortion function, which is based on the premise that the encoder has unperturbed access to the original source and, therefore, directly controls the trade-off between distortion and data rate. The curves show that the electro-optical design that increases $\mathcal{H}$ also decreases the associated $\mathcal{E}$ and, thereby, substantially improves the information efficiency $\mathcal{H}/\mathcal{E}$ of the data transmission.

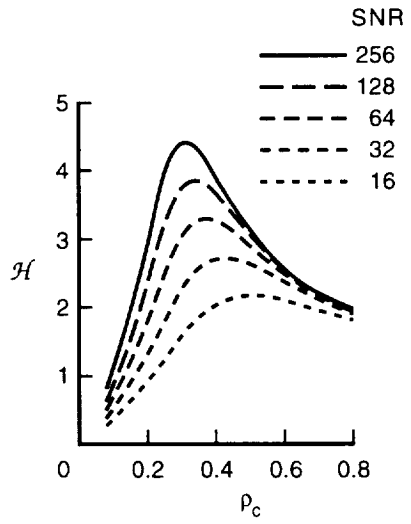


Figure 3. Information rate $\mathcal{H}$ versus optical-design index ρ_c for several SNRs.

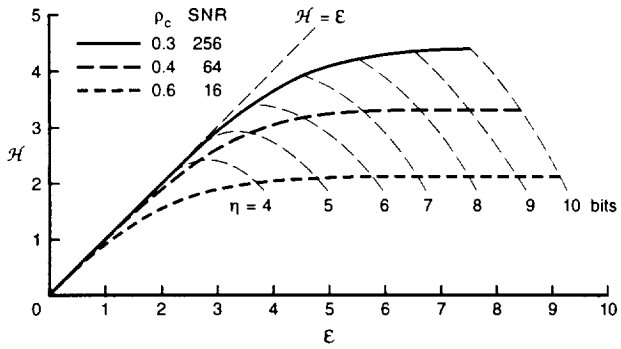


Figure 4. The information-entropy $\mathcal{H}(\mathcal{E})$ plot that characterizes the information rate $\mathcal{H}$ versus the associated theoretical minimum data rate $\mathcal{E}$ for η-bit quantization. The three curves represent informationally optimized designs specified by the optical-design index ρ_c and SNR.

Figure 5 presents images that illustrate the transition from traditional telephotography and television in which images are reproduced without digital processing to modern visual communication systems in which images are reproduced with digital restoration.

4. CONCLUSION

The image-gathering device that is designed to produce the maximum-realizable information rate ordinarily maximizes (a) the *efficiency* of the information transmission (i.e., the ratio of the information rate $\mathcal{H}$ to the theoretical minimum data rate $\mathcal{E}$), (b) the *quality* of the image restoration (i.e., the restorability of images for fidelity, resolution, sharpness, and clarity), and (c) the *robustness* of the image restoration (i.e., the tolerance of the restoration to errors in estimates of the radiance-field statistics). This critical dependence of the efficiency, quality, and robustness of visual communication on the design of the image-gathering device is largely independent of the statistical properties of natural scenes.

REFERENCES

1. C. L. Fales, F. O. Huck, R. Alter-Gartenberg and Z. Rahman, "Image gathering and digital restoration," *Philos. Trans. Roy. Soc., London*, in press.
2. F. O. Huck, C. L. Fales and Z. Rahman, "An information theory of visual communication," *Philos. Trans. Roy. Soc. London*, in press.
3. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (U. Illinois Press, Urbana, 1964).
4. N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Wiley, New York, 1949).
5. F. O. Huck, C. L. Fales, D. J. Jobson and Z. Rahman, "Electro-optical design for efficient visual communication," *Optical Engineering* **34**, 795–813 (1995).

(a) Traditional image gathering and reconstruction

(b) Traditional image gathering with restoration

(c) Informationally optimized image gathering with restoration

Figure 5. Reconstructions and Wiener restorations for two designs of the image-gathering device, the traditional design ($\rho_c = 0.8$, SNR = 16) and an informationally optimized design ($\rho_c = 0.4$, SNR = 64).

Multiscale Retinex for Color Image Enhancement

Z. Rahman, D. J. Jobson and G. A. Woodell

Proceedings of the
IEEE International Conference on Image Processing
Switzerland
(September 1996)

MULTI-SCALE RETINEX FOR COLOR IMAGE ENHANCEMENT

Zia-ur Rahman, Member IEEE

Science and Technology Corporation
101 Research Drive
Hampton, VA 23666

Daniel J. Jobson and Glenn A. Woodell

NASA Langley Research Center
MS 473, 8 N Dryden Street
Hampton, Virginia 23681

ABSTRACT

The retinex is a human perception-based image processing algorithm which provides color constancy and dynamic range compression. We have previously reported on a single-scale retinex (SSR) and shown that it can either achieve color/lightness rendition or dynamic range compression, but not both simultaneously. We now present a multi-scale retinex (MSR) which overcomes this limitation for most scenes. Both color rendition and dynamic range compression are successfully accomplished except for some "pathological" scenes that have very strong spectral characteristics in a single band.

1. INTRODUCTION

A common problem with color imagery—digital or analog—is that of successful capture of the dynamic range and colors seen through the viewfinder onto the acquired image. More often than not, this image is a poor rendition of the actual observed scene. In 1986, Edwin Land presented the last version of his retinex[1] as a model for human color constancy. Hurlbert[2, 3] showed that there is no mathematical solution to the problem of removing lighting variations. Moore[4, 5] implemented a version of the retinex in analog VLSI for real-time dynamic range compression but encountered scene context dependent limitations and hence failed to achieve a generalized implementation. More recently we, inspired by the work of Land, Hurlbert, and Moore decided to delve into this commonly occurring, but surprisingly intractable, problem. Our initial research resulted in the single-scale retinex (SSR) that we have described in detail previously[6, 7, 8]. The SSR shows exceptional promise for dynamic range compression but does not provide good tonal rendition. In fact, a distinct trade-off controlled by the scale of the surround function exists between dynamic range compression and tonal rendition, and one can be improved only at the cost of reducing the other.

This paper describes our initial research in alleviating some of these trade-offs by using a multi-scale retinex (MSR), i.e. a retinex which combines several SSR outputs to produce a single output image which has both good dynamic range compression and color constancy, and good tonal rendition. The tonal rendition, though, is still scene dependent to a certain extent. We will briefly describe the MSR in Section 2. In section 3 we will provide some of the results of applying the MSR to color images and compare our results with other techniques for image enhancement. Finally, in Section 4 we will discuss the future direction for this research.

2. THE MULTI-SCALE RETINEX

The MSR can be compactly written as

$$F_i(x,y) = \sum_{n=1}^{N} W_n \cdot \{\log[S_i(x,y)] - \log[S_i(x,y) * M_n(x,y)]\} \quad (1)$$

where the subscripts $i \in R, G, B$ represent the three color bands, N is the number of scales being used, and W_n are the weighting factors for the scales. The $M_n(x,y)$ are the surround functions given by

$$M_n(x,y) = K_n \exp[-(x^2+y^2)/\sigma_n^2],$$

where the σ_n are the standard deviations of the Gaussian distribution that determine the scale. The magnitude of the scale determines the type of information that the retinex provides: smaller scales providing more dynamic range compression, and larger scales providing more color constancy. The K_n are selected so that $\iint F(x,y)\,dx\,dy = 1$. Each of the expressions within the summation in Eq. 1 represents an SSR.

The SSR has been previously defined[6] to have the following characteristics and properties:

This work was performed under a NASA Langley Research Center Contract #NAS1-19603

1. The functional form of the surround is a Gaussian.

2. The placement of the log function is AFTER surround formation.

3. The post-retinex signal processing is a "canonical" gain-offset rather than an automatic gain-offset.

4. There is a trade-off between dynamic range compression and tonal rendition which is governed by the Gaussian surround space constant. A space constant of 80 pixels was a reasonable compromise between dynamic range compression and rendition.

5. A single scale seemed incapable of simultaneously providing sufficient dynamic range compression and tonal rendition.

6. Violations of the gray-world assumption led to retinexed images which were either "grayed-out" locally or globally or, more rarely, suffered from color distortion.

The MSR combines the dynamic range compression of the small scale retinex with the tonal rendition of the large scale retinex to produce an output which encompasses both.

As stated above, the MSR still suffers from graying-out of uniform zones much as the SSR did. The advantage that the MSR has over the SSR is in the combination of scales which provide both dynamic range compression and tonal rendition at the same time. The overall result of the application of the MSR is still more saturated than human observation, giving the final image a "washed-out" appearance, but it preserves most of the detail in the scene. This "graying" of areas of constant intensity occurs because the retinex processing enhances each color band as a function of its surround. The smaller values in the weaker channels get "pushed" up strongly, making them approximately equal in magnitude to the dominant channel, leading to a graying out of the overall region. Moore[4] encountered this problem in his implementation of the retinex and attempted to resolve it with using variable gains across the color channels. We do not attempt a solution in this paper but provide a detailed solution elsewhere.[9] However, the MSR produces a much better final image in terms of color, and dynamic range than the SSR. Figure 1 shows a comparison of the SSR and the MSR processing. The differences are easier to see in the original color images (see CD-ROM version of paper), but if one looks around the left side of the

Figure 1: (a) Original (b) Single-scale Retinex (c) Multi-scale Retinex

face and in the area just above the right shoulder of the pictured man, one sees details for the MSR which are not evident in the SSR. Also the "haloing" artifacts peculiar to the SSR are eliminated in the MSR.

3. RESULTS

Figure 3 shows a comparison of the MSR with image enhancement methods typically used for dynamic range compression. The scenes are selected to show the effects of MSR processing on "good images" (top row), wide dynamic range compression that is achieved by the MSR (middle row), and color constancy (bottom row). Histogram equalization performs well for the child image, but begins to saturate in both the grass image and the cave image. The logarithmic non-linearity has the poorest performance for all three scenes, though its dynamic range compression capabilities are quite evident in the grass scene. For the MSR processing, the uniform regions in the child scene tend to gray out, but the overall result is still quite good. For the grassy field, the MSR processing compresses the wide dynamic range well and brings out the colors in both the bright and the dark areas very well. For the cave image, the color of the inside rock, and the outside rock formations are both brought out so they agree with actual observation. The CD-ROM version of the proceedings contains the color postscript figures and the comparisons are much easier to make.

The MSR output brings out most of the detail in the black regions but at the cost of enhancing the noise in these regions. This noise is a result of the poor signal-to-noise ratio in these areas. The traditional techniques are also able to enhance the dark regions, but not to the same extent as the MSR. In fact, the MSR achieves a balance between enhancing the darks, yet, at the same time, retaining the colors in the bright regions, as opposed to traditional point non-linearities which tend to enhance the darks at the cost of saturating the brights (Figs. 3(b,c)). Of course, the final rendition in still scene-dependent and can often be grayed-out if the original scene contains large areas of constant intensity (Fig. 3(d)(top row)).

The MSR output is different from existing techniques in that the overall effect of processing is scene dependent but the processing itself is not. In other words, though the overall effect adapts itself to the lighting variations within the scene, the same process, with exactly the same control parameters can be used for any image. This is not true for other adaptive techniques since variations in lighting conditions imply variations in the control parameters.

4. FUTURE RESEARCH

The main direction of further research is to improve the color rendition of the MSR. Though it produces excellent dynamic range compression, the tonal rendition is scene dependent and can be quite poor. Work is already underway on a newer version of the MSR which combines a post-filter with the MSR to produce an MSR which provides very good color rendition with a very slight loss in overall dynamic range compression.

5. REFERENCES

[1] Edwin Land. Recent advances in retinex theory. *Vision Research*, 26(1):7–21, 1986.

[2] Anya C. Hurlbert. Formal connections between lightness algorithms. *Journal of the Optical Society of America A*, 3:1684–1693, 1986.

[3] Anya C. Hurlbert. *The Computation of Color.* PhD thesis, Massachusetts Institute of Technology, September 1989.

[4] Andrew Moore, J. Allman, and R. M. Goodman. A real-time neural system for color constancy. *IEEE Transactions on Neural Networks*, 2(2):237–247, March 1991.

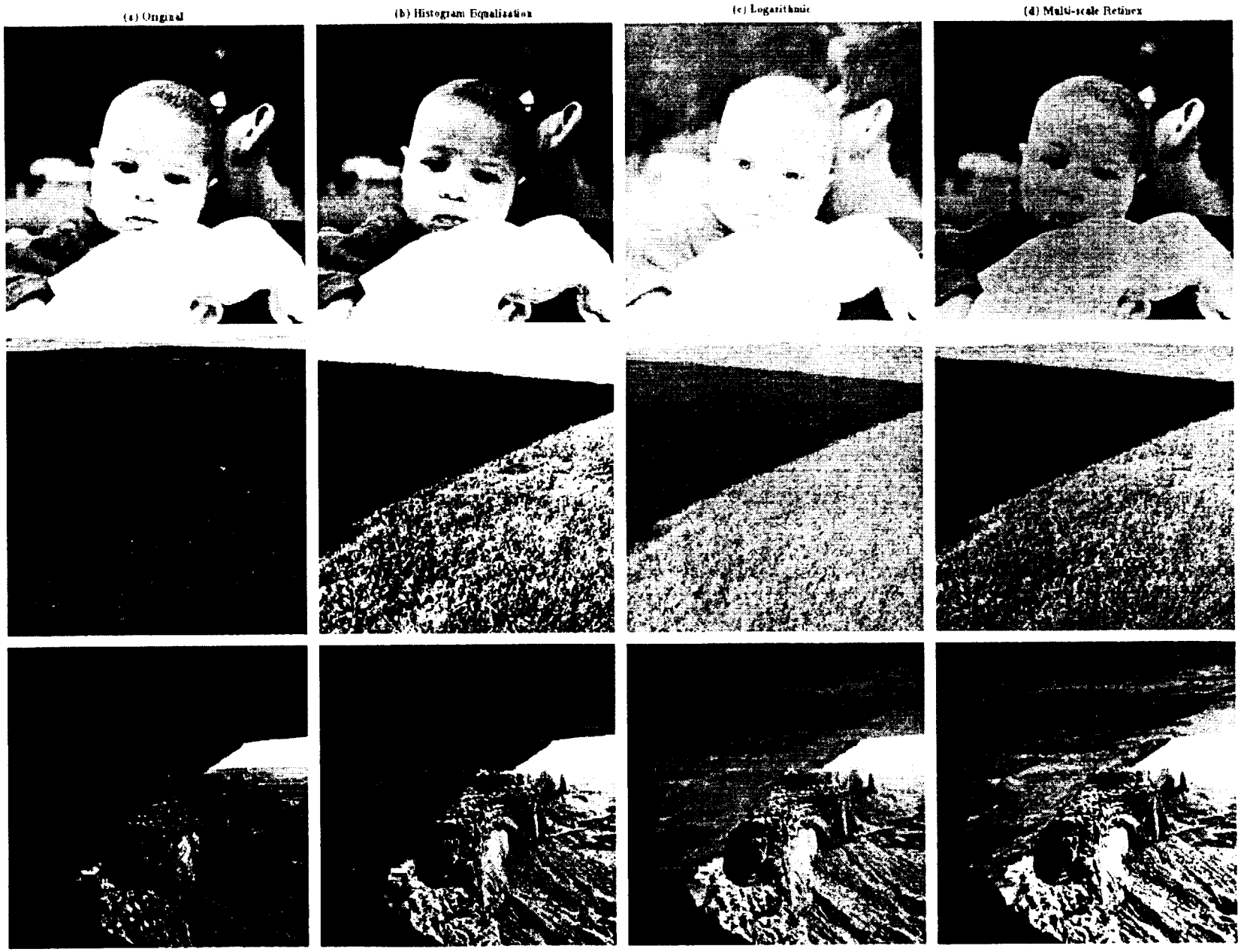
[5] Andrew Moore, G. Fox, J. Allman, and R. M. Goodman. A VLSI neural network for color constancy. In D. S. Touretzky and R. Lippman, editors, *Advances in Neural Information Processing 3*, pages 370–376. Morgan Kaufmann, San Mateo, CA, 1991.

[6] Daniel J. Jobson, Zia-ur Rahman, and Glenn A. Woodell. Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing.* Submitted 1995.

[7] Zia-ur Rahman. Properties of a center/surround Retinex Part One: Signal processing design. *NASA Contractor Report #198194*, 1995.

[8] Daniel J. Jobson and Glenn A. Woodell. Properties of a center/surround Retinex Part Two: Surround design. *NASA Technical Memorandum #110188*, 1995.

[9] Daniel J. Jobson, Zia-ur Rahman, and Glenn A. Woodell. A multi-scale Retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing.* Submitted 1996.


(a) Original
(b) Histogram Equalization
(c) Logarithmic
(d) Multi-scale Retinex

Retinex Image Processing: Improved Fidelity To Direct Visual Observation

D. J. Jobson, Z. Rahman and G. A. Woodell

IS&T Fourth Color Imaging Conference:
Color Science Systems, and Applications
Scottsdale, AZ
(November 1996)

Retinex Image Processing: Improved Fidelity To Direct Visual Observation

Daniel J. Jobson, NASA Langley Research Center, Hampton, VA
Zia-ur Rahman, Science and Technology Corporation, Hampton, VA
Glenn A. Woodell, NASA Langley Research Center, Hampton, VA

Abstract

Recorded color images differ from direct human viewing by the lack of dynamic range compression and color constancy. Research is summarized which develops the center/surround retinex concept originated by Edwin Land through a single-scale design to a multi-scale design with color restoration (MSRCR). The MSRCR synthesizes dynamic range compression, color constancy, and color rendition and, thereby, approaches fidelity to direct observation.

Introduction

A comparison of the recorded color image and the "view through the viewfinder" are strikingly different (Fig. 1) for most everyday scenes due to the presence of shadows. Color and detail in shadows are far more clear in the direct view than in recorded images. We have developed Land's concept of a center/surround retinex[1] to the level of single-scale retinex (SSR) design for which there is a trade-off between dynamic range compression and tonal rendition that is governed by the choice of the surround space constant. Comparison of processed images to direct scene viewing established that no value of an intermediate space constant could simultaneously provide sufficient dynamic range compression and good tonal rendition. The single-scale retinex provided a building block for the construction of a multi-scale retinex which does couple acceptable dynamic range compression with good tonal rendition. Color constancy is excellent for all forms of the retinex but color rendition was elusive as a result of the gray world assumption implicit to the retinex computation. A color restoration was developed and applied after the multiscale retinex in order to overcome this color loss but with a modest dilution in color constancy.

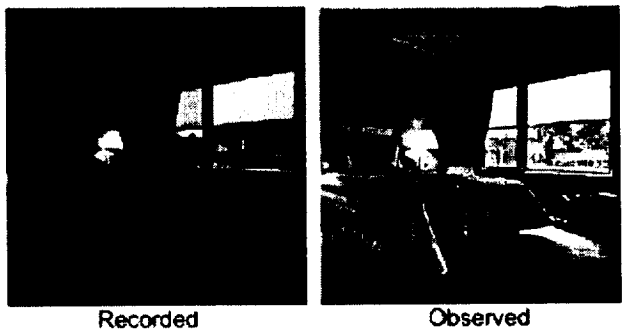


Figure 1. The discrepancy between recorded images and direct observation. Human vision strongly compresses visual information across wide-ranging illumination conditions within a scene.

Methods and Results

Here we briefly highlight results that are described comprehensively elsewhere[2-5]. The design of the single-scale retinex (SSR) consists of: 1) the choice of a surround function, 2) the placement of the log function, and 3) final signal processing prior to display/print. Of the three mathematical functions[1,6,7] previously used for a retinex surround, we found the best visual performance with the Gaussian compared to either the exponential form or the inverse square form originally used by Land. Unlike previous studies, we find that the placement of the log function is quite important in both mathematical and visual terms. We show that its placement after the surround formation is preferable to placement prior to surround formation. Processing after the basic spatial retinex operation was found to be a "canonical" gain-offset applied uniformally to all color bands rather than an auto gain-offset' calculated across the full three band data. These elements lead to a SSR defined as:

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

where $I_i(x,y)$ is the image distribution in the ith color band, * denotes convolution, and F(x,y) is a Gaussian surround function. This is followed by the constant gain-offset applied across all color bands which, thus far, has proven to be universally constant or "canonical" for all images tested. This characteristic provides for general purpose and automatic application of the method and for simple construction of a multi-scale retinex as:

$$R_{m,i}(x,y) = \sum_{k=1}^{n} c_k R_{k,i}(x,y) \quad (2)$$

for the kth surround space constant. The design of the multi-scale retinex was found to require a minimum of three scales for image frame sizes of about 512x512 pixels. A comparison of direct viewing of scenes to scene photometry established that dynamic range compression for human vision is typically 5:1 or so for outdoor scenes with shadows and easily achieves 500:1 for mixed interior/exterior scenes. From this it is evident that everyday scenes often exceed the 255:1 (8-bit) dynamic range of most color imaging systems and that wide dynamic range color imaging, together with the retinex, or other compressive processing, is essential if recorded color images are to approach the quality of observation. The use of test scenes together with a battery of diverse digital images revealed that the violations of the gray world assumption implicit to the retinex were a common occurrence both zonally and globally in images. The degree of impact on color rendition ranges from slight desaturation of color to rather severe graying for the extreme cases of "monochromatic" scenes. Therefore a color restoration that could be universally applied was developed because scene content is not predictable. Thus the MSRCR is given by:

$$R'_{m,i}(x,y) = R_{m,i}(x,y) * I'_i(x,y) \quad (3)$$

where the color restoration, $I'_i(x,y)$, is:

$$I'_i(x,y) = \log \left[C \frac{I_i(x,y)}{\sum_{i=1}^{3} I_i(x,y)} \right] \quad (4)$$

The current form of the MSRCR does compare favorably with direct viewing by synthesizing dynamic range compression and color constancy with color and tonal rendition.

Applications

We isolate two applications of the MSRCR to illustrate a wider range of applications- aerospace image enhancements and digital photoprocessing. The MSRCR can be used to advantage in both space operations and remote sensing (Fig.2). For the former, the often dramatic lighting variations present in space operations can be ameliorated and better visual information achieved. For remote sensing, the MSRCR brings out the visual information present in large shadow zones and large zones of low reflectance, such as water areas. An example of an enhancement for improved documentation of aeronautical research is also shown. The automatic correction of low exposure images is evident and is useful for digital photoprocessing.

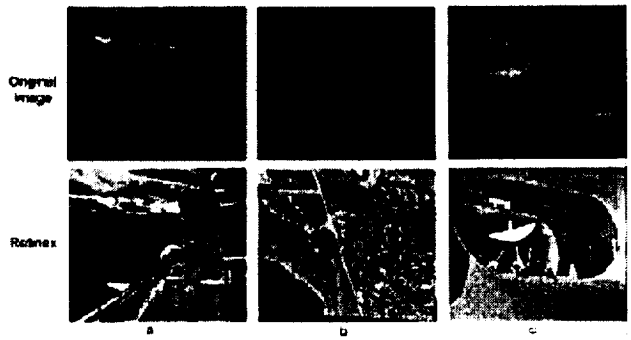


Figure 2. Enhancement of aerospace images using the MSRCR:
a) Shuttle operations, b) remote sensing, and c) aeronautical research documentation.

The MSRCR can be used as a "digital darkroom", allowing burning and dodging of areas that would have been extremely labor intensive if not impossible using traditional darkroom techniques. Although there are software packages that allow the selective lightening and darkening of specific areas of digitized images, in the cases below, it would be impractical because of the degree of detail required in the selection of these areas and the different changes required for each selection.

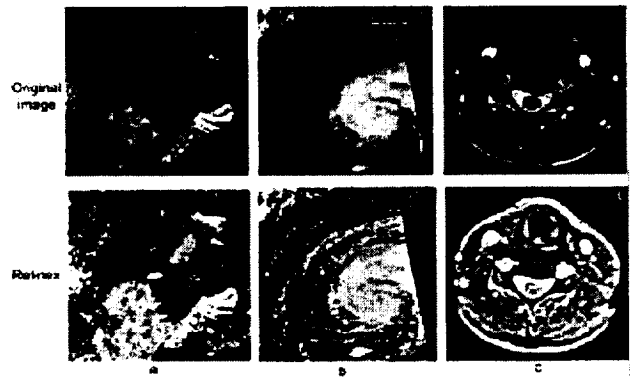


Figure 3. The MSRCR as a digital photo-processing method: graceful automatic "burning and dodging":
a) Underwater image from traditional film camera, b) Image from satellite data, c) digitized image from nuclear magnetic resonance (NMR) film.

References

1. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," Proc. Nat. Acad. Sci., vol. 83, pp. 3078--3080, 1986.

2. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," To appear in IEEE Transactions on Image Processing.

3. D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes, " IEEE Transactions on Image Processing. Submitted 1996.

4. Z. Rahman, "Properties of a center/surround Retinex Part One: Signal processing design," NASA Contractor Report 198194, 1995.

5. D. J. Jobson and G. A. Woodell, "Properties of a center/surround Retinex Part Two: Surround design," NASA Technical Memorandum 110188, 1995.

6. A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," IEEE Transactions on Neural Networks, vol. 2, pp. 237--247, March 1991.

7. A. C. Hurlbert, The Computation of Color. PhD thesis, Massachusetts Institute of Technology, September 1989.

Correspondence: Dan Jobson, MS/473 NASA Langley Research Center, Hampton, VA 23681-0001
e-mail: d.j.jobson@larc.nasa.gov
internet: http://dragon.larc.nasa.gov/viplab/projects/retinex/retinex.html

Informationally Optimized Image-Gathering and Restoration

Z. Rahman, F. O. Huck and C. L. Fales

Presented at the IS&T's 50th Annual Conference
Cambridge, MA
(May 1997)

Informationally Optimized Image Gathering and Restoration

Zia-ur Rahman[†], *Carl L. Fales and* [‡]*Friedrich O. Huck*[‡]
[†]*College of William & Mary,* [‡]*NASA Langley Research Center*

Abstract

The goal of image gathering and restoration often is to produce the best possible picture in terms of fidelity, sharpness and clarity. However, this goal cannot be attained, at it has been pursued in the past, by treating image gathering and restoration as independent tasks. Instead, in a clean departure from the mores of traditional image processing, we present an approach that rigorously uses modern communication theory to optimally combine the electro-optical design of the image gathering device with the digital processing algorithm for image restoration. Extensive simulations have shown that there exists a strong correlation between the information rate that is produced by the image gathering device and the image quality with which an image can be restored.

Introduction

Modern visual communication channels increasingly combine image gathering and display with digital image coding and restoration (Fig. 1). So far, however, the image-gathering devices are still designed to produce the best possible images when reconstructed without the aid of the digital processing, and the image restoration algorithms are still developed and evaluated without fully accounting for the critical constraints of image gathering and display.

The aim of this paper is to summarize some elements of a study[1, 2, 3] that rigorously unites the electro-optical design of image gathering and display devices with the digital processing algorithms for image coding and restoration. In particular, this paper will present the informationally optimized image-gathering designs that maximize image quality in terms of clarity and sharpness of fine detail. The study is based on the two classical works that are the foundation of modern communication theory. In one work Shannon[4] introduces the concept of the rate of transmission of information in a noisy channel, and in the other Wiener[5] introduces the concept of the minimum mean-square error restoration of signals corrupted by noise.

Although our mathematical development is firmly rooted in these familiar concepts, it leads to formulations that are significantly different from those that are found in the traditional literature on digital image processing. One fundamental difference, which we address in this summary paper, arises primarily because the limitations inherent in the realizability of the spatial frequency response (SFR) of optical apertures and the sampling passband of photodetection mechanisms. The limitations inevitably impose a trade-off between blurring and aliasing on the design of the image-gathering device (Fig. 2). This precludes the treatment of visual communication as strictly a bandwidth-limited process. Instead, it requires the inclusion of the effects of insufficient sampling both in the end-to-end analysis of the visual communication channel, and in the development of the restoration algorithm.

Image gathering and restoration

The image-gathering process transforms the continuous radiance field $L(x, y)$ that is either reflected or emitted by the scene into the digital signal $s(x, y; \kappa)$,

$$s(x, y; \kappa) = [KL(x, y) * \tau(x, y)] \underline{|||} + n_p(x, y) + n_{a/d}(x, y), \tag{1}$$

where $\tau(x, y)$ represents the spatial response of the image-gathering device, $n_p(x, y)$ is the discrete photodetector noise, $n_{a/d}$ is the analog-to-digital (A/D) conversion noise, and κ represents the number of levels used for the A/D conversion. In the Fourier domain, the image-gathering process is defined as

$$\tilde{s}(\upsilon, \omega; \kappa) = [K\hat{L}(\upsilon, \omega)\hat{\tau}(\upsilon, \omega)] * \underline{|\hat{|}|} + \tilde{n}_p(\upsilon, \omega) + \tilde{n}_{a/d}(\upsilon, \omega; \kappa), \tag{2}$$

where $\hat{L}(\upsilon, \omega)$ is the continuous radiance-field transform, $\hat{\tau}(\upsilon, \omega)$ is the SFR of the image-gathering device, $\tilde{n}_p(\upsilon, \omega)$ and $\tilde{n}_{a/d}(\upsilon, \omega; \kappa)$ are the discrete photodetector and analog-to-digital (A/D) conversion noise transforms, and (υ, ω) are the spatial frequencies with units of cycles per sample. The tilde "~" is used instead of the caret "^" whenever the Fourier transformation is discrete and, therefore, the transformed function is periodic in the spatial frequency do-

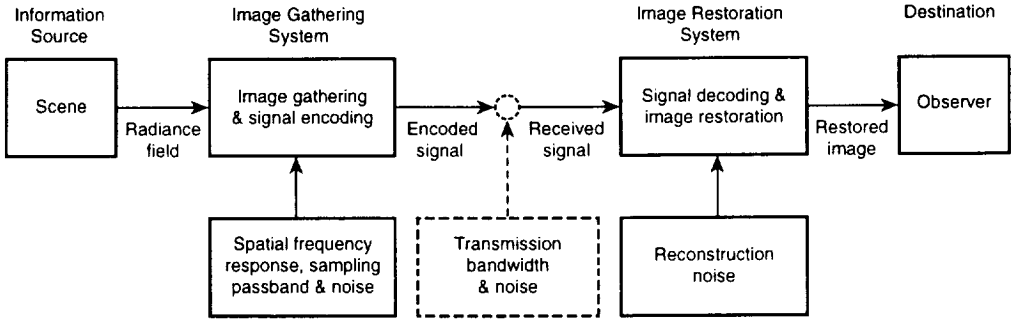


Figure 1. Model of visual communication channel together with the critical limiting factors that constrain its performance.

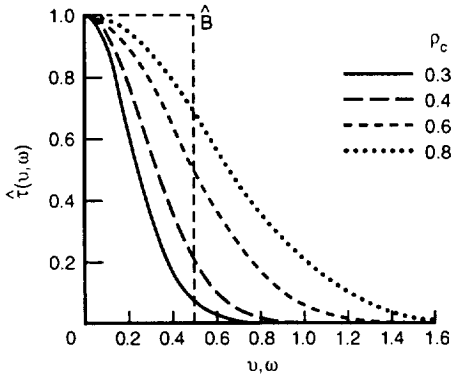


Figure 2. SFRs $\hat{\tau}(\upsilon,\omega)$ of the image-gathering device relative to the sampling passband $\hat{B}$ for unit sampling intervals.

main. The function $\hat{\underline{\underline{|||}}}$ is the Fourier transform of the rectangular sampling lattice with unit intervals, and is given by

$$\hat{\underline{\underline{|||}}} = \sum_{m,n} \delta(\upsilon - m, \omega - n) \quad (3)$$

$$= \delta(\upsilon,\omega) + \hat{\underline{\underline{|||}}}_s(\upsilon,\omega), \quad (4)$$

where $\delta(\upsilon,\omega)$ is the Dirac delta function and $\hat{\underline{\underline{|||}}}_s$ accounts for the sampling sidebands. The associated sampling passband

$$\hat{B} = \left[(\upsilon,\omega); |\upsilon| < \frac{1}{2}, |\omega| < \frac{1}{2}\right]$$

has unit area, i.e., $|\hat{B}| = 1$. The analog-to-digital transformation is done for κ levels with η-bit quantization, where $\eta = \log_2 \kappa$. The image-restoration process transforms this signal into the observed image $R(x,y;\kappa)$. The corresponding image-restoration process in the spatial frequency domain is defined as

$$\hat{R}(\upsilon,\omega;\kappa) = K^{-1} s(x,y;\kappa)\hat{\Psi}(\upsilon,\omega;\kappa) + \hat{N}_r(\upsilon,\omega), \quad (5)$$

where $\hat{\Psi}(\upsilon,\omega;\kappa)$ is a linear filter that records the digitally processed signal on an interpolation lattice that is sufficiently fine to suppress the blurring and raster effects of the image-display process and $\hat{N}_r(\upsilon,\omega)$ is the transform of the reconstruction noise (e.g., film granularity).

To assess visual communication in terms of information theory, the image-gathering process is constrained to be linear and isoplanatic (spatially invariant), and the radiance-field and noise amplitudes are constrained to be Gaussian, wide-sense stationary, and statistically independent. In addition, we characterize: (a) the radiance field $L(x,y)$ by the power spectral density (PSD)

$$\hat{\Phi}_L(\upsilon,\omega) = \frac{1}{|A|}\overline{|\hat{L}(\upsilon,\omega)|^2}$$

of an isoplanatism patch of the scene with area $|A|$, (b) the discrete signal $s(x,y)$ prior to A/D conversion by the PSD

$$\tilde{\Phi}_s(\upsilon,\omega) = \left[K^2\hat{\Phi}_L(\upsilon,\omega)|\hat{\tau}(\upsilon,\omega)|^2\right] * \hat{\underline{\underline{|||}}} + \tilde{\Phi}_p(\upsilon,\omega), \quad (6)$$

(c) the photodetector noise $n_p(x,y)$ by the PSD $\tilde{\Phi}_p(\upsilon,\omega)$, and (d) the A/D conversion noise $n_{a/d}(x,y;\kappa)$ by the PSD

$$\tilde{\Phi}_{a/d}(\upsilon,\omega;\kappa) = \left(\frac{\sigma_s}{\kappa}\right)^2, \quad (7)$$

where

$$\sigma_s^2 = \iint_{\hat{B}} \tilde{\Phi}_s(\upsilon,\omega)\, d\upsilon\, d\omega.$$

Figures of merit

By accounting for the critical constraints of image gathering, we can quantitatively assess visual communication in terms of the following figures of merit:

1. The rate of transmission of information, or information rate, $\mathcal{H}$ that the image-gathering system produces for the radiance field that resides within its

field of view, as given by

$$\mathcal{H} = \frac{1}{2} \iint_{\hat{B}} \log_2 \left[1 + \frac{\hat{\Phi}_L(v,\omega)|\hat{\tau}(v,\omega)|^2}{\hat{\Phi}_n(v,\omega;\kappa)} \right] dv\, d\omega, \tag{8}$$

where

$$\hat{\Phi}_n(v,\omega;\kappa) = \hat{\Phi}_L(v,\omega)|\hat{\tau}(v,\omega)|^2 * \underline{\hat{|||}}_s + K^{-2}[\tilde{\Phi}_p(v,\omega) + \tilde{\Phi}_{a/d}(v,\omega;\kappa)]. \tag{9}$$

2. The maximum-realizable fidelity $\mathcal{F}$ of the digital image that can be restored from the received information, unconstrained by the image-display medium, as given by

$$\mathcal{F} = \sigma_L^{-2} \int_{-\infty}^{\infty} \hat{\Phi}_L(v,\omega) \left[1 - 2^{-\hat{\mathcal{H}}(v,\omega)} \right] dv\, d\omega, \tag{10}$$

where $\hat{\mathcal{H}}(v,\omega)$ is the spectral distribution of the information rate $\mathcal{H}$ given by the integrand of Eq. 8.

Figure 3 characterizes the information rate $\mathcal{H}$ as a function of the electro-optical design of the image-gathering device, as specified by the optical-design index ρ_c and the root-mean-square (rms) signal-to-noise ratio (SNR) for a radiance field with mean spatial detail equal to the sampling interval. The curves show that the SFRs that maximize information $\mathcal{H}$ are a function of the SNR. This result is intuitively appealing for image restoration. In one extreme, when the SNR is low, one would prefer to avoid substantial blurring—SFR extends well beyond the sampling passband, hence, ρ_c is large—because the noise constrains the enhancement of fine spatial detail. In the other extreme, when the SNR is high, one would prefer to avoid substantial aliasing—SFR remains mostly inside the sampling passband, hence, ρ_c is small—because then the noise no longer constrains this enhancement.

Using the curves shown in Fig. 3, visual communication channels can be specified, in terms of their SNR and the SFR of the image gathering device, that maximize the information throughput. Huck, et al[1, 3] show that visual communication channels that are designed to maximize the information throughput also maximize the quality of the restored image in terms of sharpness and clarity of the fine detail. Table 1 lists the electro-optical designs specified by an SNR and an SFR parameterized by the index ρ_c that maximize information throughput. Conventional image gathering typically has an SFR with $\rho_c = 0.80$.

Figure 4 presents images that illustrate the transition from traditional telephotography and television in which images are reproduced without digital processing to informationally optimized visual communication systems in which images are reproduced with digital restoration.

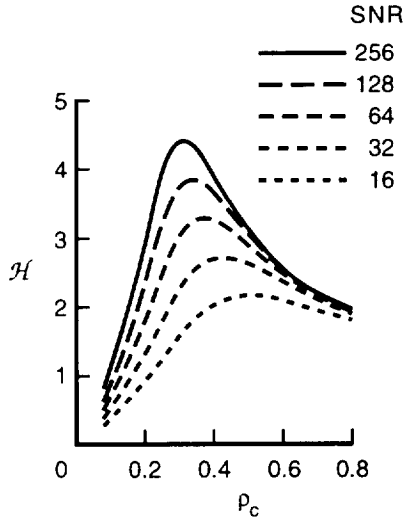


Figure 3. Information rate $\mathcal{H}$ versus optical-design index ρ_c for several SNRs.

Design	SNR	ρ_c
1	256	0.30
2	64	0.40
3	16	0.60

Table 1: Channel designs that maximize information throughput.

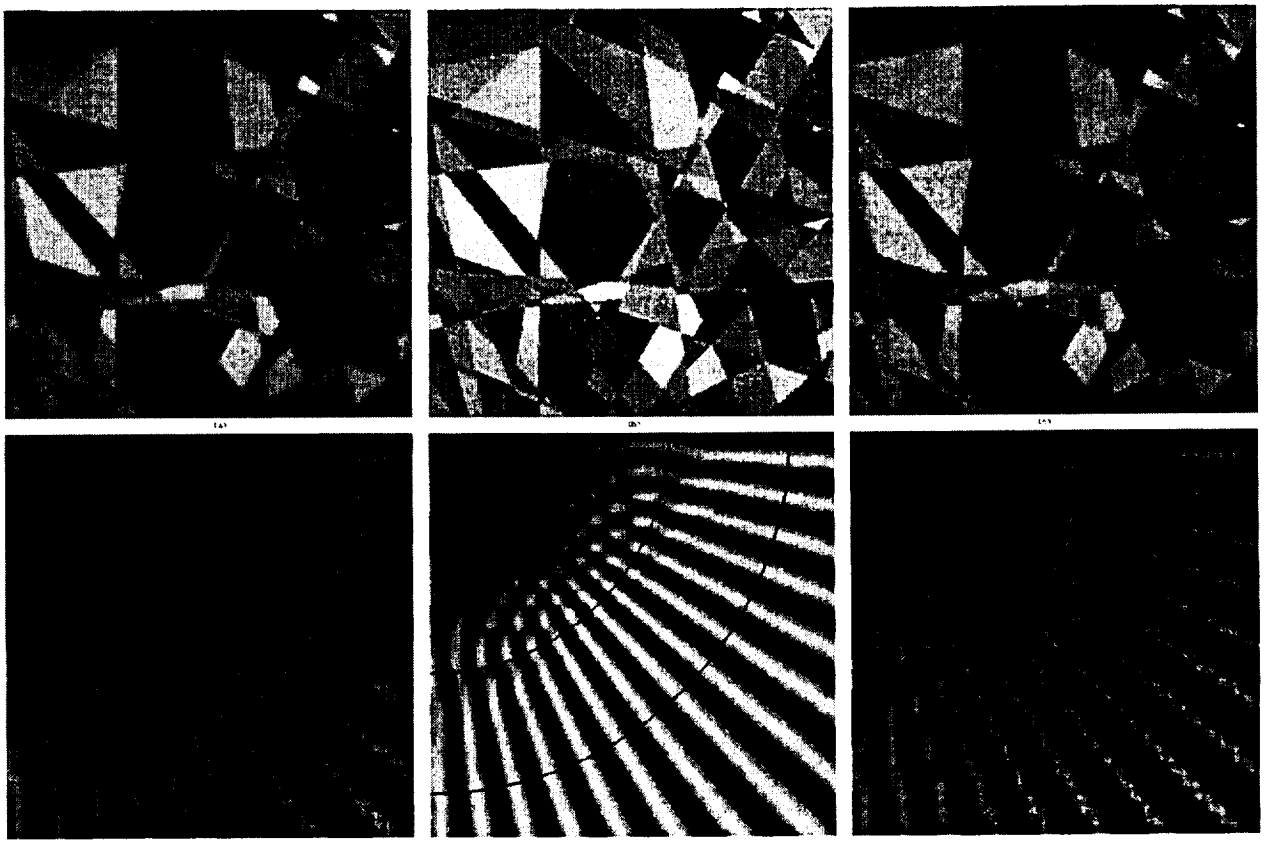
Conclusion

The image-gathering device that is designed to produce the maximum-realizable information rate maximizes both the *quality* of the image restoration (i.e., the restorability of images for fidelity, resolution, sharpness, and clarity), and the *robustness* of the image restoration (i.e., the tolerance of the restoration to errors in estimates of the radiance-field statistics). This critical dependence of the quality and robustness of visual communication on the design of the image-gathering device is largely independent of the statistical properties of natural scenes.

References

[1] F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A*, vol. 354, pp. 2193–2248, Oct. 1996.

[2] C. L. Fales, F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, "Image gathering and digital restoration," *Philosophical Transactions of the Royal Society of London A*, vol. 354, pp. 2249–2287, Oct. 1996.



(a) (b) (c)

Figure 4: Reconstructions and Wiener restorations for two designs of the image gathering device; (a) Traditional design (SNR=16; $\rho_c = 0.80$) (b) Information optimized image gathering and restoration (SNR=64; $\rho_c = 0.40$) (c) Traditional Wiener restoration (SNR=64; $\rho_c = 0.80$).

[3] F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*. Boston, MA: Kluwer Academic Publishers, 1997. In print.

[4] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1964. Originally published in the *Bell System Technical Journal*, 27:379–423 and 28:623–656, 1948.

[5] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. Cambridge, MA: MIT Press, 1949.

A Comparison of the Multiscale Retinex with Other Image Enhancement Techniques

Z. Rahman, G. A. Woodell and D. J. Jobson

Proceedings of the IS&T's 50th Annual Conference
Cambridge, MA
(May 1997)

A Comparison of the Multiscale Retinex With Other Image Enhancement Techniques

Zia-ur Rahman[†], *Glenn A. Woodell*[‡] *and Daniel J. Jobson*[‡]
[†]*College of William & Mary,* [‡]*NASA Langley Research Center*

Abstract

The multiscale retinex with color restoration (MSRCR) has shown itself to be a very versatile *automatic* image enhancement algorithm that simultaneously provides dynamic range compression, color constancy, and color rendition. A number of algorithms exist that provide one or more of these features, but not all. In this paper we compare the performance of the MSRCR with techniques that are widely used for image enhancement. Specifically, we compare the MSRCR with color adjustment methods such as gamma correction and gain/offset application, histogram modification techniques such as histogram equalization and manual histogram adjustment, and other more powerful techniques such as homomorphic filtering and 'burning and dodging'. The comparison is carried out by testing the suite of image enhancement methods on a set of diverse images. We find that though some of these techniques work well for some of these images, only the MSRCR performs universally well on the test set.

Introduction

The Multiscale Retinex[1] (MSR) is a generalization of the single-scale retinex[2-4] (SSR), which, in turn, is based upon the last version of Land's center/surround retinex[5]. The current version of the MSR combines the retinex dynamic range compression and color constancy with a color 'restoration' filter that provides excellent color rendition[6-8]. This version of the MSR is called the Multiscale Retinex with Color Restoration (MSRCR). The MSRCR has been tested with a very large suite of images and has consistently proven to be better than any conventional image enhancement technique. In this paper we present a comparison of the MSRCR with several of the most popular image enhancement methods. These include point transforms such as automatic gain/offset, non-linear gamma correction, non-linear intensity transforms such as the logarithmic transform or the 'square-root' transform; and global transforms such as histogram equalization[9], homomorphic filtering[10], and manual 'burning and dodging.'

State-of-the-art Techniques

In this section we briefly describe the characteristics of some of the state-of-the-art techniques most commonly used for image enhancement.

Gain/offset correction

One of the most common methods of enhancing an image is the application of a gain and an offset to stretch the dynamic range of an image. This is a linear operation and hence has limited success on scenes that encompass a much wider dynamic range than that that can be displayed. In this case, loss of detail occurs due to saturation and clipping as well as due to poor visibility in the darker regions of the image. For a scene with dynamic range between r_{max} and r_{min}, and a display medium with dynamic range d_{max}, this transform can be represented by

$$I_i'(x,y) = \frac{d_{max}}{r_{max} - r_{min}} \cdot (I_i(x,y) - r_{min}), \quad (1)$$

where I_i is the ith input band, and I_i' is the ith output band. This particular transform will transform the scene to completely fill the dynamic range of the display medium. This does not imply, however, that this process will provide a good visual representation of the original scene.

Non-linear Point Transforms

Another well known method used for providing dynamic range compression is the application of non-linear transforms such as the gamma non-linearity, the logarithm function, and the power-law function to the original image. These functions are typically biased toward increasing the 'visibility' in the 'dark' regions by sacrificing the visibility in the 'bright' areas. The output of such filters can be described by

$$I_i'(x,y) = P\left[I_i(x,y)\right], \quad (2)$$

where $P[\,]$ represents the point non-linearity. A typical point non-linearity is illustrated in Fig. 1.

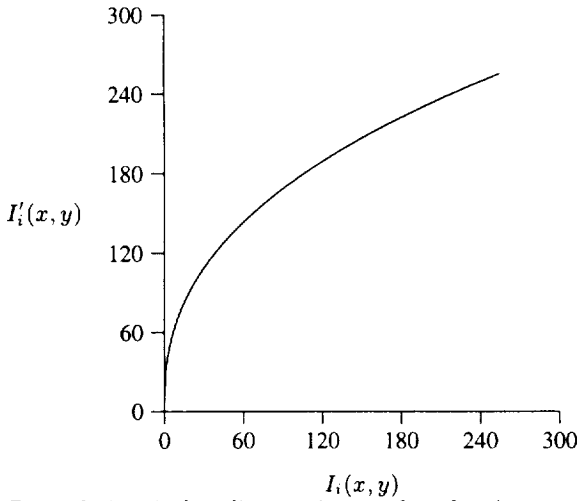


Figure 1: A typical nonlinear point transform function.

Histogram Equalization

A global technique that works well for a wide variety of images is histogram equalization. This technique is based on the idea of remapping the histogram of the scene to a histogram that has a near-uniform probability density function. This results in reassigning dark regions to brighter values and bright regions to darker values. Histogram equalization works well for scenes that have unimodal or weakly bi-modal histograms (i.e. very dark, or very bright), but not so well for those images with strongly bi-modal histograms (i.e. scenes that contain very dark and very bright regions).

Homomorphic Filtering

The technique that most resembles ours conceptually and functionally is homomorphic filtering[10]. The general idea of homomorphic filtering is shown in Fig. 2. The image is first passed through a logarithmic non-linearity that provides dynamic range compression. It is then Fourier transformed, and its representation in the spatial frequency domain is modified by applying a filter that provides contrast enhancement. The modified image is then inverse Fourier transformed and is passes through an exponential non-linearity that 'reverses' the effects of the logarithmic nonlinearity.* Mathematically,

$$s_i(x,y) = \ln[I_i(x,y)] \quad (3)$$

$$s_i'(v,\omega) = \mathcal{F}[s_i(x,y)] \quad (4)$$

$$s_i''(v,\omega) = s'(v,\omega)\mathcal{H}(v,\omega) \quad (5)$$

$$s_i'''(x,y) = \mathcal{F}^{-1}[s_i''(v,\omega)] \quad (6)$$

$$I_i'(x,y) = \exp[s_i'''(x,y)], \quad (7)$$

where $\mathcal{F}[]$, and $\mathcal{F}^{-1}[]$ represent the Fourier and the inverse Fourier transforms respectively, and $\mathcal{H}$ represents the homomorphic filter. It is in its final exponential transform that the homomorphic filter differs the most from the MSRCR. MSRCR does not apply a final inverse transform to go back to the original domain!

Manual Image Enhancement

As both professional and amateur photographers face the limitations of the narrow dynamic range in current printing technology, and the inadequate performance of image enhancement algorithms, more and more attention is being focused on manual enhancement methods. One such technique is 'burning-and-dodging' where different regions of an image are interactively modified by a user†. The burn and dodge tool provides the capability of modifying the color content of a region by using tools of varying sizes and shapes that work as electronic "scrims."

Multiscale Retinex with Color Restoration

The general form of the MSRCR can be summarized by the following equation:

$$\mathcal{R}_{M_i}(x,y) = \sum_{s=1}^{S} w_s(\log[I_i(x,y)] - \log[I_i(x,y) * M_s(x,y)]), \quad i = 1,...,N \quad (8)$$

where $\mathcal{R}_{M_i}$ is the ith band of the MSRCR output, S is the number of scales being used, w_s is the weight of the scale, I_i is the ith band of the input image,, and N is the number of bands in the input image. The surround function M_s is defined by

$$M_s(x,y) = K \exp\left[\sigma_s^2/(x^2+y^2)\right],$$

where σ_s is the standard deviation of the sth surround function, and $\iint K \exp\left[\sigma_s^2/(x^2+y^2)\right]\, dx\, dy = 1$. The number of scales, S, and the widths of the surround functions, σ_s, are image independent‡. In other words, these have been chosen to maximize enhancement for a large§ number of images. Once the constants have been selected, then the process is truly automatic and independent of the variations in scene statistics.

*A modified color version of the homomorphic filter was proposed by Faugeras[11] in 1979. Our implementation simply applies the black and white version of the homomorphic filter to each band of the color image and combines the results to form a color output image.

†Adobe Photoshop 4.0, a commercial photo manipulation software package, provides a burn and dodge tool.

‡Typically for 512 × 512 images. The σ_s may change with the dimension of images.

§We have not yet found an exception after having processed 1000+ images!

Figure 2: Homomorphic filtering[9]

Comparison

We have compared the MSRCR with all of the image enhancement techniques described above. We present the results in Figs. 3, and 4. We present the comparison with manual burning and dodging separately.

Point operations

Figure 3 shows a collage of images that compares the output of the MSRCR with the point transforms. As can be seen, the MSRCR provided the best overall visual quality in each case. The techniques such as histogram equalization perform well for a wide range of scenes, but they also fail for a large set. The MSRCR outperforms the other methods universally.

Homomorphic filtering

Figure 4 shows a comparison of the MSRCR with homomorphic filtering. The homomorphic filter consistently provided excellent dynamic range compression but is lacking in final color rendition. The output of the homomorphic filter in effect appears extremely hazy compared with the output of the MSRCR though the dynamic range compression of the two methods appears to be comparable.

Manual Burning and Dodging

Figure 5 shows a comparison of the MSRCR with the results obtained by using manual burning and dodging. The manually processed image shows an improvement over the original as far as the information and detail in the dark areas is concerned but it lacks the vividness and color saturation that the MSRCR image retains and even enhances. There is obvious streaking from the very local operation of the tool stroke—this could be eliminated but only at the expense of adding considerably to the total processing time. In the high detail areas where there are sharp differences in reflectance, a tool with size approaching that of a single pixel would be required to bring out all the details. Since the time needed for enhancing a region is roughly in inverse proportion to the size of the tool being used for the processing, this suggests that a very large amount of time would be needed to perform such an enhancement. On a scene-by-scene basis, the time and effort required for manual manipulation can be reasonable; but the MSRCR produces images that are equivalent or better in quality at a fraction of the time. Because the visual quality of manual burning and dodging is solely limited by the patience and time commitment of the user, the case shown is perhaps typical of the performance achieved by the persistent non-specialist.

Conclusions

We have provided a brief description of the most commonly used image enhancement techniques and compared their operation with the multiscale retinex with color restoration. We have shown that the MSRCR outperforms these techniques in all cases in terms of dynamic range compression achieved, and the rendition of the final color image. The automatic nature of the process also enables us to use the same set of parameters 'blindly' for each and every image that is encountered. Of course, there are a few images for which the MSRCR has sub-par performance. But these are fairly rare and generally relate to defects in the original image data—such as preferential clipping of a spectral band. We are currently investigating methods to detect such scenes and adaptively adjust the MSRCR to correct for these sub-par performances.

References

[1] Z. Rahman, D. Jobson, and G. A. Woodell, "Multiscale retinex for color image enhancement," in *Proceedings of the IEEE International Conference on Image Processing*, IEEE, 1996.

[2] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing: Special Issue on Color Processing*, vol. 6, pp. 451–462, March 1996.

[3] Z. Rahman, "Properties of a center/surround Retinex Part One: Signal processing design," *NASA Contractor Report #198194*, 1995.

[4] D. J. Jobson and G. A. Woodell, "Properties of a center/surround Retinex Part Two: Surround design," *NASA Technical Memorandum #110188*, 1995.

[5] E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.*, vol. 83, pp. 3078–3080, 1986.

[6] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing: Special Issue on Color Processing*, July 1997.

Figure 3: A comparison of the MSRCR with point operations. Top row: original; second row: histogram equalization; third row: gain/offset; fourth row: gamma non-linearity; bottom row: MSRCR

(a) Original (b) Homomorphic filter (c) MSRCR

Figure 4: A comparison of the MSRCR with images enhanced by homomorphic filtering. The dynamic range compression achieved by the two methods is comparable, but the MSRCR produces images that possess much better contrast and sharper colors.

(a) Original (b) Manual burning and dodging (c) MSRCR

Figure 5: Comparison of the MSRCR with manual 'burning-and-dodging.' The manually enhanced image was produced using the burning and dodging tool provided in Adobe Photoshop 4.0. Circular tools with soft edges were used to modify the color content of different regions. The total time to produce this enhanced image was 20 minutes. The MSRCR image took 45 seconds on a PentiumPro 200MHz machine.

[7] Z. Rahman, D. Jobson, and G. A. Woodell, "Multiscale retinex for color rendition and dynamic range compression," in *Applications of Digital Image Processing XIX* (A. G. Tescher, ed.), Proc. SPIE 2847, 1996.

[8] D. Jobson, Z. Rahman, and G. A. Woodell, "Retinex image processing: Improved fidelity for direct visual observation," in *Proceedings of the IS&T Fourth color Imaging Conference: Color Science, Systems, and Applications*, pp. 124–126, IS&T, 1996.

[9] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, second ed., 1987.

[10] T. G. Stockham, Jr., "Image processing in the context of a visual model," *Proceedings of the IEEE*, vol. 60, no. 7, pp. 828–842, 1972.

[11] O. D. Faugeras, "Digital color image processing within the framwork of a human visual model," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, pp. 380–393, Aug. 1979.

An Information Theory of Visual Communication

F. O. Huck, C. L. Fales and Z. Rahman

Philosophical Transactions of the Royal Society A:
Physical Sciences and Engineering
(October 1996)

ISSN 1364-503X

The Royal Society

Philosophical Transactions:

Mathematical, Physical and Engineering Sciences

Volume **354** Pages 2193–2287 Number **1716** 15 October 1996

Philosophical Transactions of the Royal Society of London SERIES A

THE ROYAL SOCIETY

Philosophical Transactions: Mathematical, Physical and Engineering Sciences

1364-503X(199610)354:1716

Series A Volume 354 Number 1716 15 October 1996

CONTENTS

This issue was produced by using the TEX typesetting system

Published in Great Britain by the Royal Society, 6 Carlton House Terrace, London SW1Y 5AG
Printed in Great Britain for the Royal Society by the University Press, Cambridge

An information theory of visual communication

BY FRIEDRICH O. HUCK[1], CARL L. FALES[1] AND ZIA-UR RAHMAN[2]

[1]*NASA Langley Research Center, Hampton, VA 23681, USA*

[2]*Science and Technology Corporation, Hampton, VA 23666, USA*

Contents

The fundamental problem of visual communication is that of producing the best possible picture at the lowest data rate. We address this problem by extending information theory to the assessment of the visual communication channel as a whole, from image gathering to display. The extension unites two disciplines, the electro-optical design of image gathering and display devices and the digital processing for image coding and restoration. The mathematical development leads to several intuitively attractive figures of merit for assessing the visual communication channel as a function of the critical limiting factors that constrain its performance. Multiresolution decomposition is included in the mathematical development to optimally

Phil. Trans. R. Soc. Lond. A (1996) **354**, 2193–2248
Printed in Great Britain

Image Gathering and Digital Restoration

C. L. Fales, F. O. Huck, R. Alter-Gartenberg and Z. Rahman

Philosophical Transactions of the Royal Society A:
Physical Sciences and Engineering
(October 1996)

Image gathering and digital restoration

By Carl L. Fales[1], Friedrich O. Huck[1], Rachel Alter-Gartenberg[2] and Zia-ur Rahman[3]

[1]*NASA Langley Research Center, Hampton, VA 23681, USA*
[2]*College of William and Mary, Williamsburg, VA 23187, USA*
[3]*Science and Technology Corporation, Hampton, VA 23666, USA*

Contents

This paper seeks to unite two disciplines: the electro-optical design of image gathering and display devices and the digital processing for image restoration. So far, these two disciplines have remained independent, following distinctly separate traditions. However, the best possible performance can be attained only when the digital processing algorithm accounts for the critical limiting factors of image gathering and display and the image-gathering device is designed to enhance the performance of the digital-processing algorithm. The following salient advantages accrue:

1. Spatial detail as fine as the sampling interval of the image-gathering device ordinarily can be restored sharply and clearly.

2. Even finer spatial detail than the sampling interval can be restored by combining a multiresponse image-gathering sequence with a restoration filter that properly reassembles the within-passband and aliased signal components.

3. The visual quality produced by traditional image gathering (e.g. television camera) and reconstruction (e.g. cubic convolution) can be improved with a small-kernel restoration operator without an increase in digital processing.

4. The enhancement of radiance-field transitions can be improved for dynamic-range compression (to suppress shadow obscurations) and for edge detection (for computer vision).

Phil. Trans. R. Soc. Lond. A (1996) **354**, 2249–2287
Printed in Great Britain

Properties and Performance of a Center/Surround Retinex

D. J. Jobson, Z. Rahman and G. A. Woodell

IEEE Transactions on Image Processing
(March 1997)

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

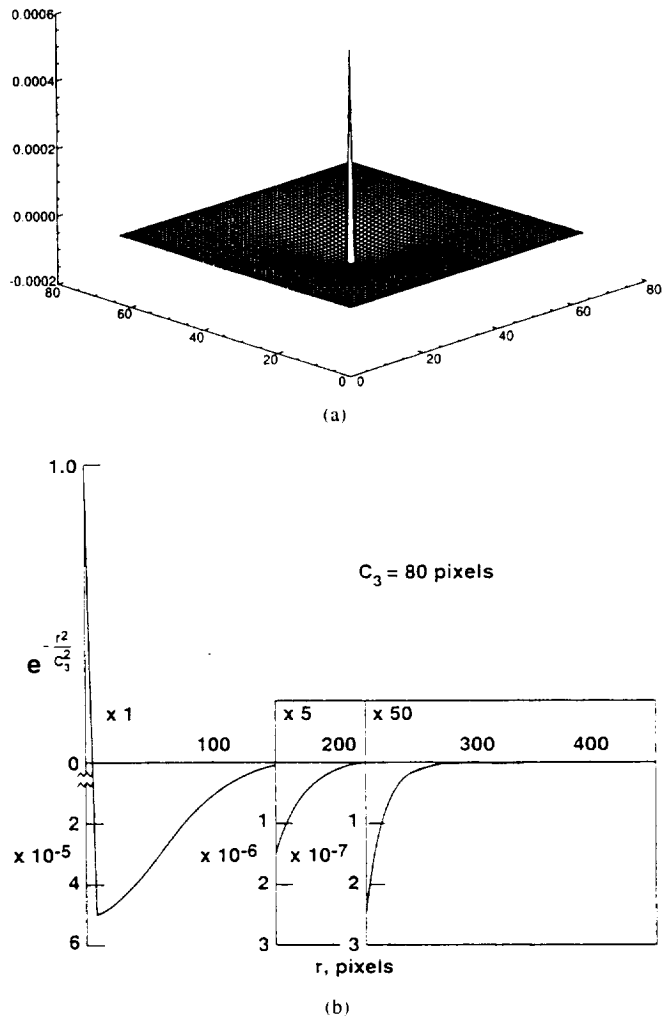


Fig. 1. Spatial form of the center/surround retinex operator. (a) 3-D representation (distorted to visualize surround). (b) Cross-section to illustrate wide weak surround.

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 NAS1-19603. 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.

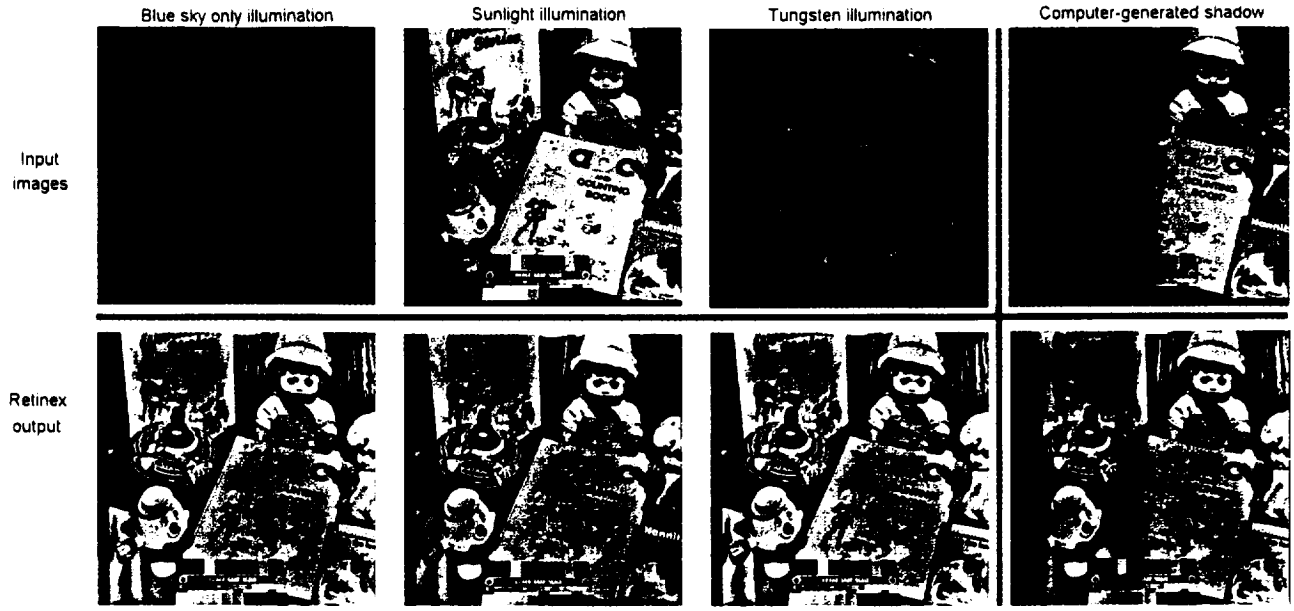


Fig. 2. Demonstration of retinex color constancy and dynamic range compression (prior to optimizing rendition) for a Gaussian surround with small space constant (15 pixels).

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

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

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.

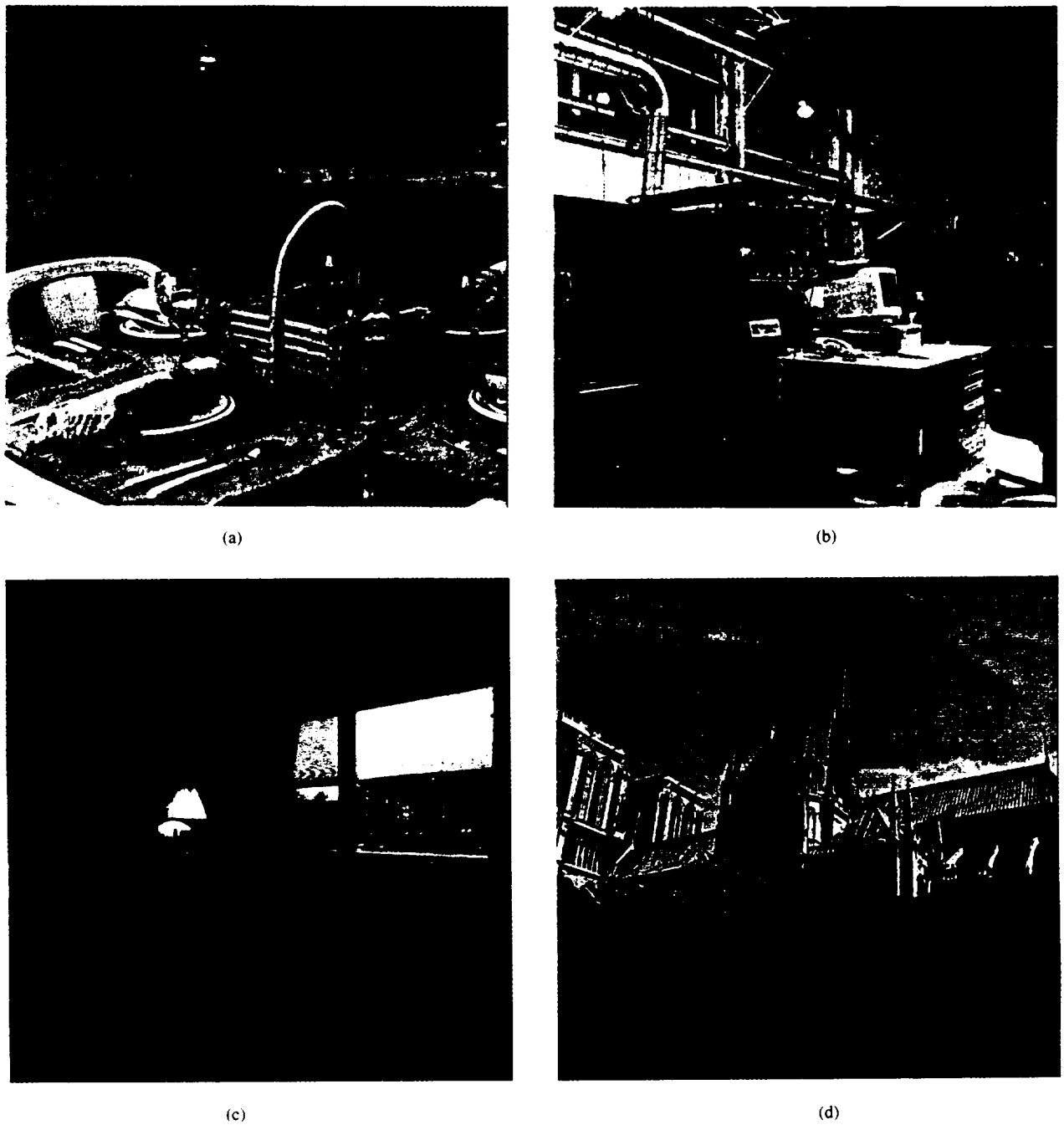


Fig. 3. Examples of serious photographic defects due to spectral and/or spatial illumination variations. (a) "Green" kitchen due to fluorescent illumination. (b) Sodium vapor illumination. (c) Tungsten indoors/daylight outdoors. (d) Obscured foreground.

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,

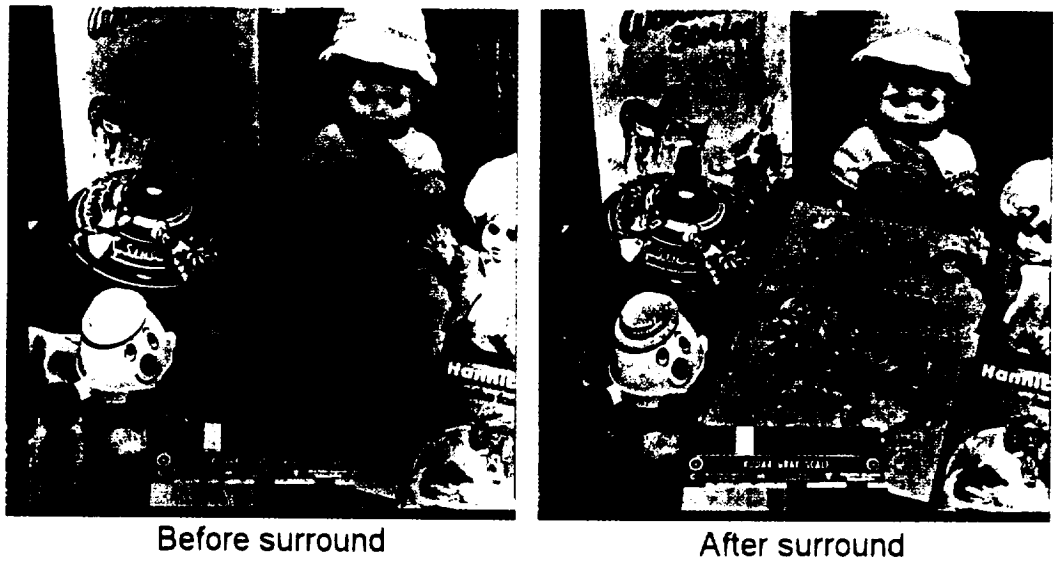


Fig. 4. Demonstration of improved rendition obtained applying the log response *after* surround formation ($\sigma_3 = 80$ pixels).

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

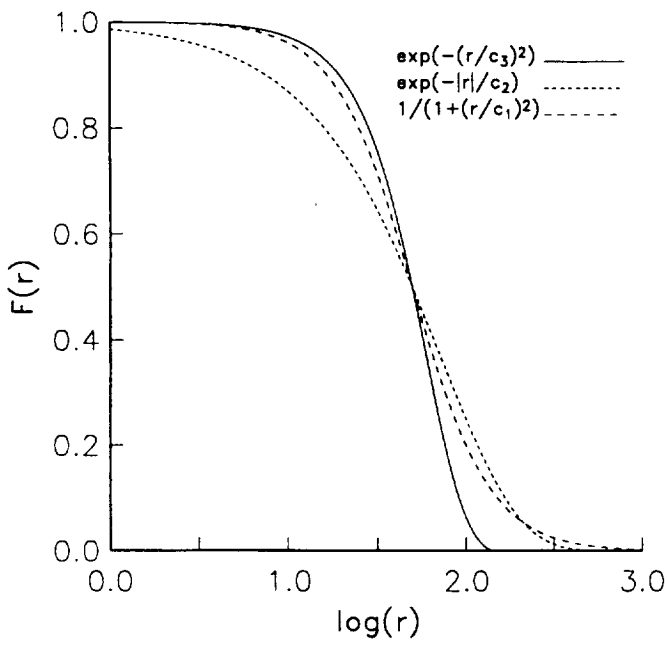


Fig. 5. Comparison of three surround functions—inverse square, exponential, and Gaussian, normalized to equal full-width half-max (FWHM) response. The log(r) scale is necessary for comparison purposes but does diminish the differences between the functions. A linear r scale (if it were graphically feasible) 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

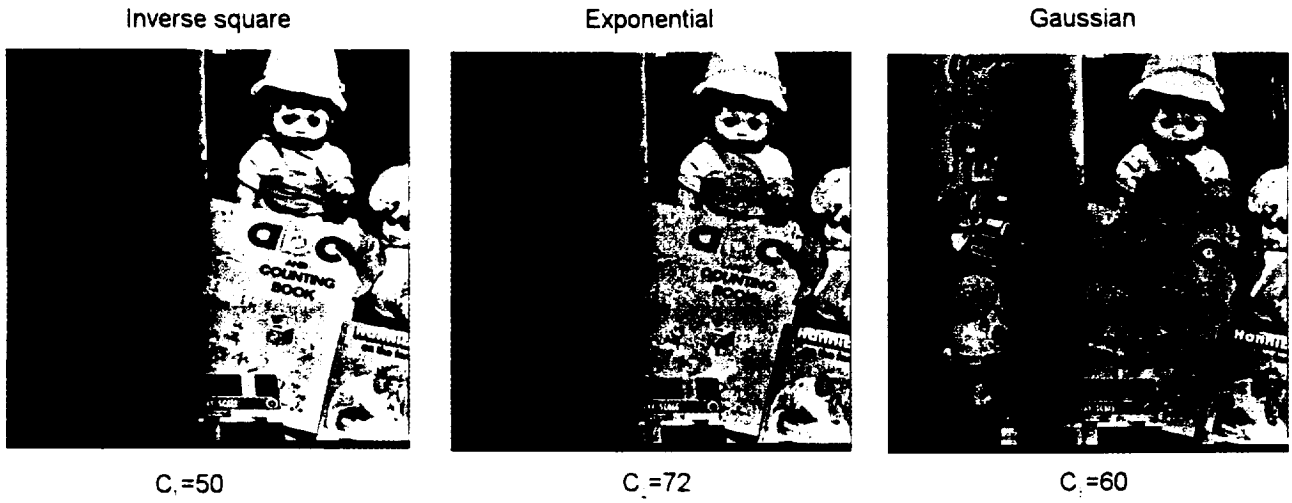


Fig. 6. Comparison of visual performance of three surround functions arranged from left to right in order of increasing dynamic range compression.

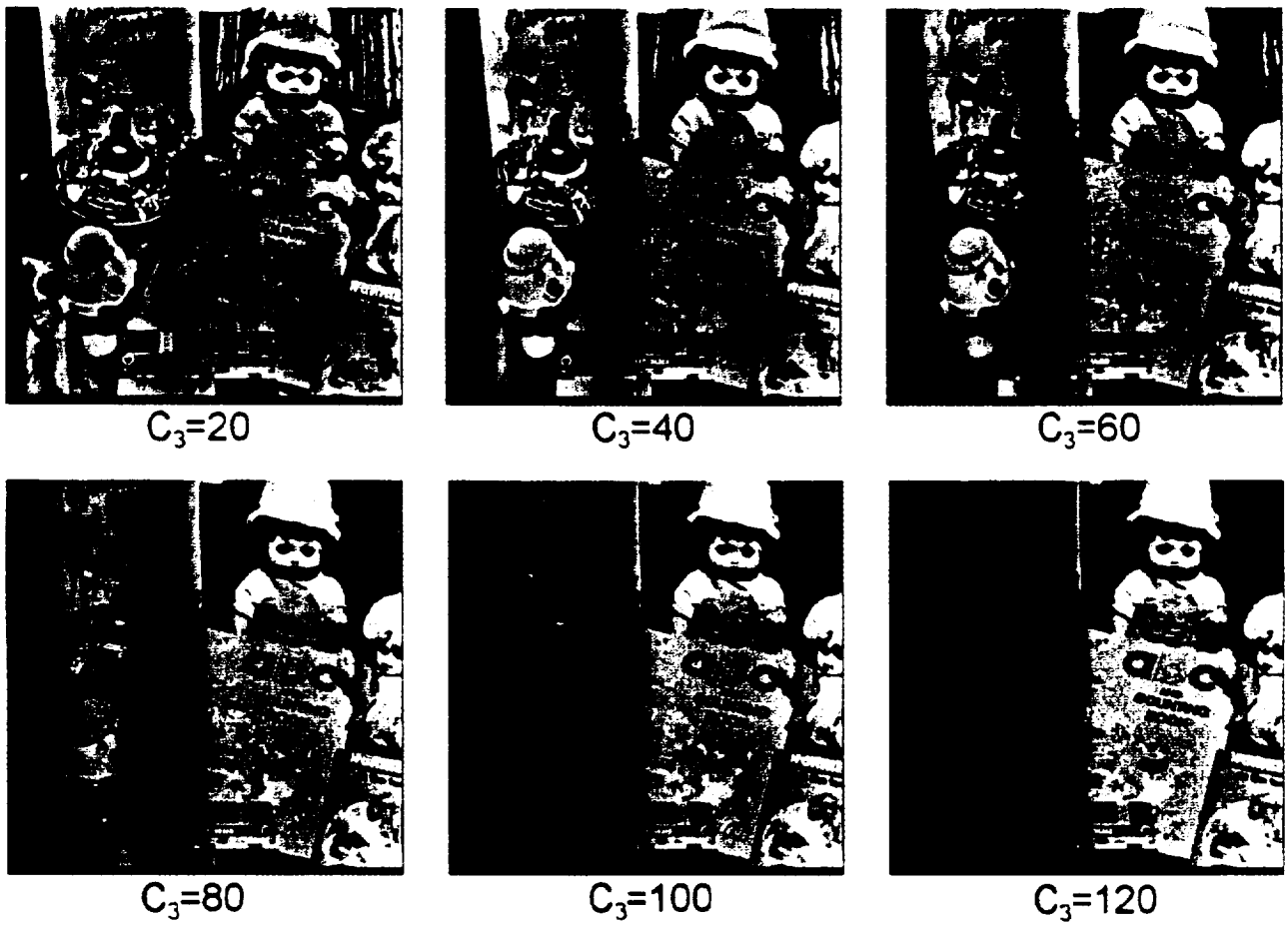


Fig. 7. Trade-off between dynamic range compression and color rendition for the Gaussian surround. Small space constants produce excellent dynamic range compression, while large constants produce the best rendition.

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

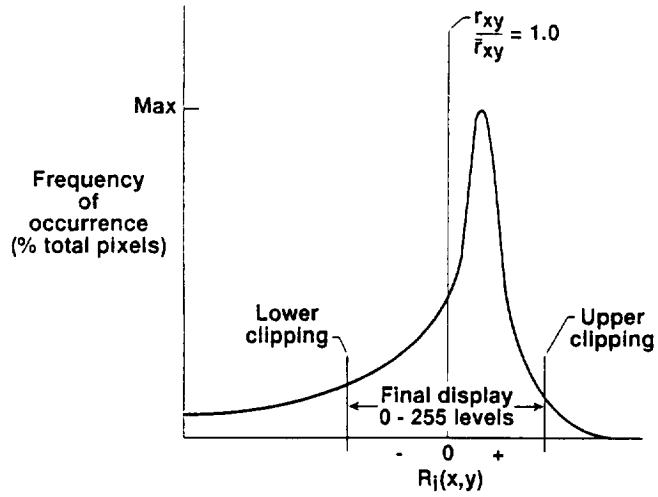


Fig. 8. Schematic of a characteristic retinex histogram illustrating the final gain/offset selection applied uniformly to the three color subimages.

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

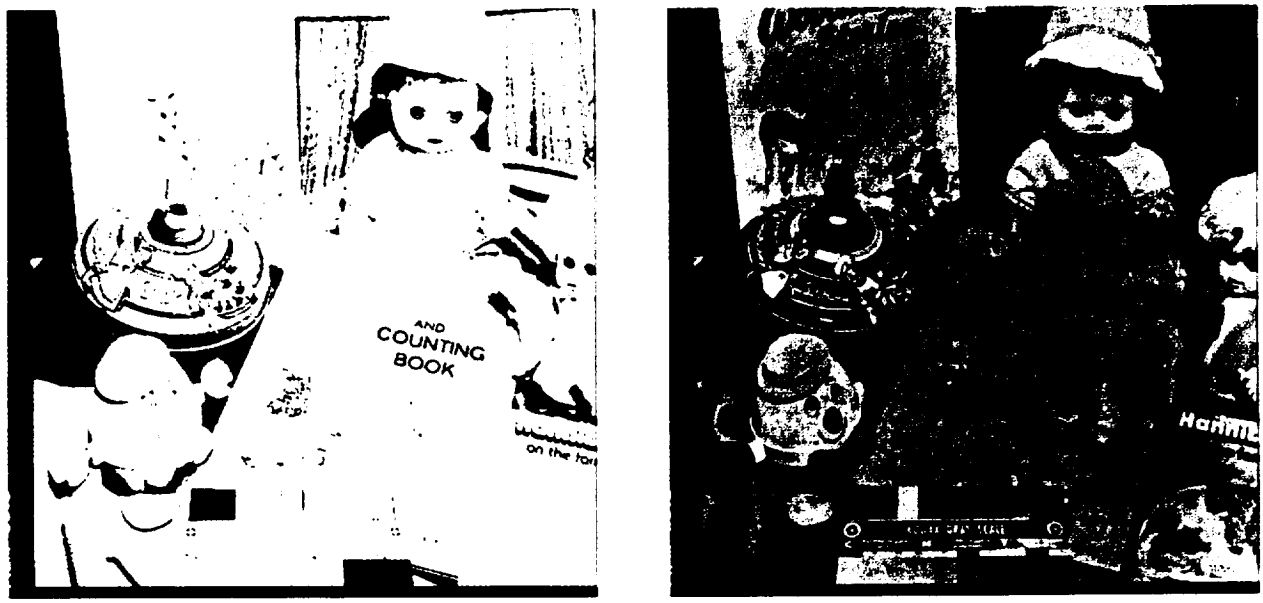


Fig. 9. Comparison of the visual performance of auto gain/offset versus "canonical" gain/offset. The auto gain/offset is selected on the absolute maximum and minimum values in all three color bands and applied uniformly to all three as a global operation. The "canonical" gain/offset accepts some clipping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

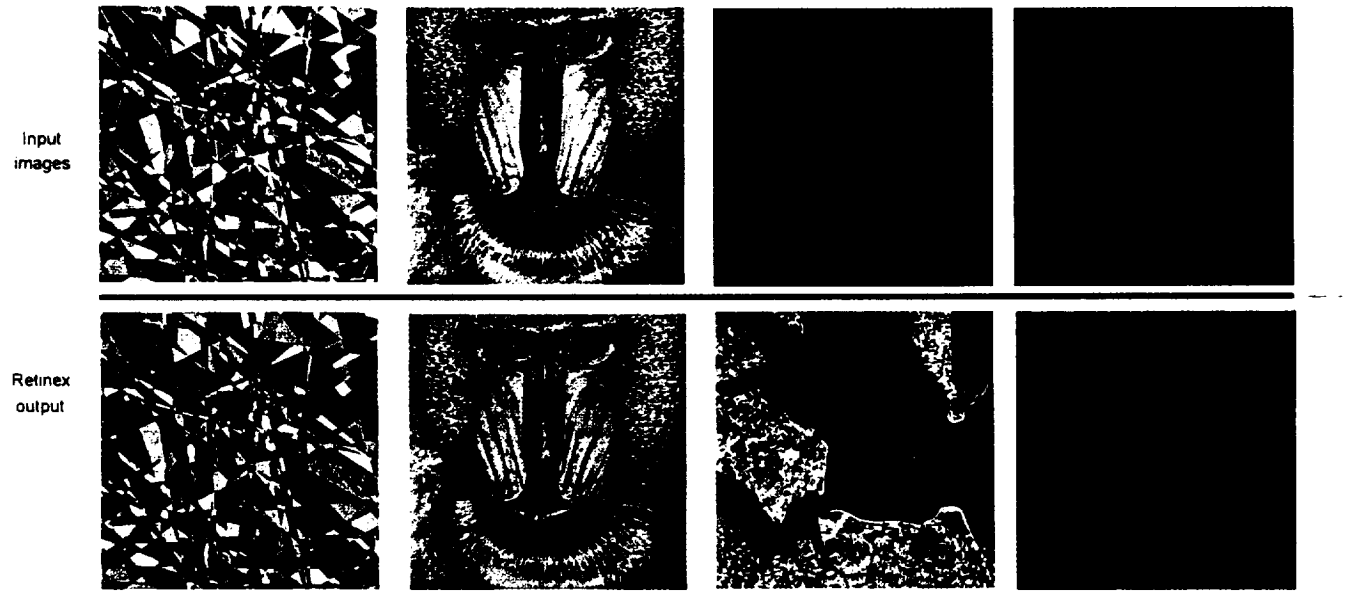


Fig. 10. Results for diverse test images—stochastic and deterministic; computer-generated; natural; and false-color ($c_3 = 80$ pixels).

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.

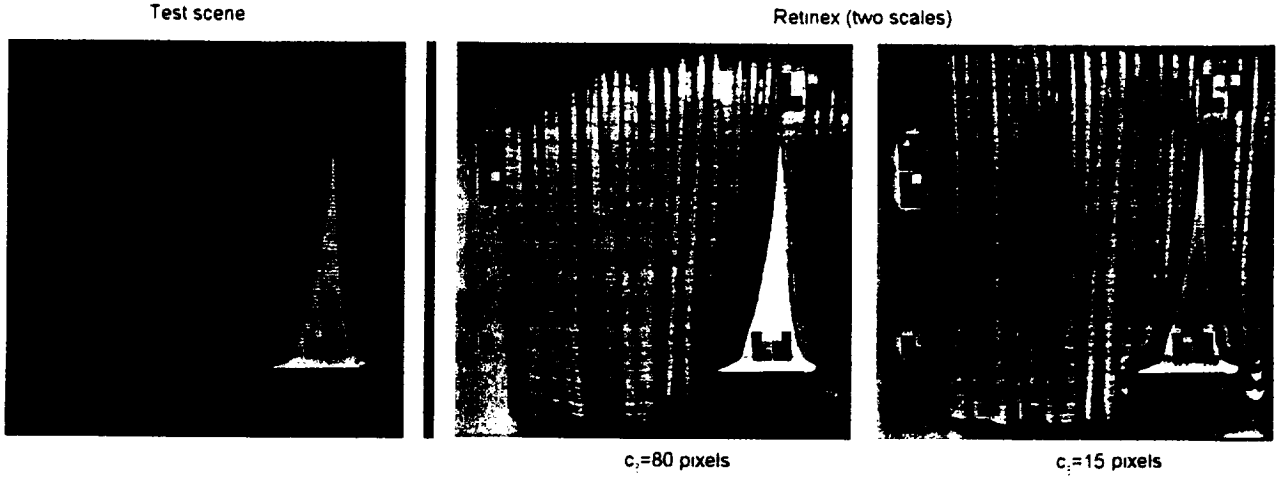


Fig. 11. Results for a test *scene* with multiple color illuminants and familiar test targets that allows comparison of retinex performance to direct human observation of the scene. Test targets on the left are predominantly in fluorescent illumination, while the middle set is in daylight, and the ones 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

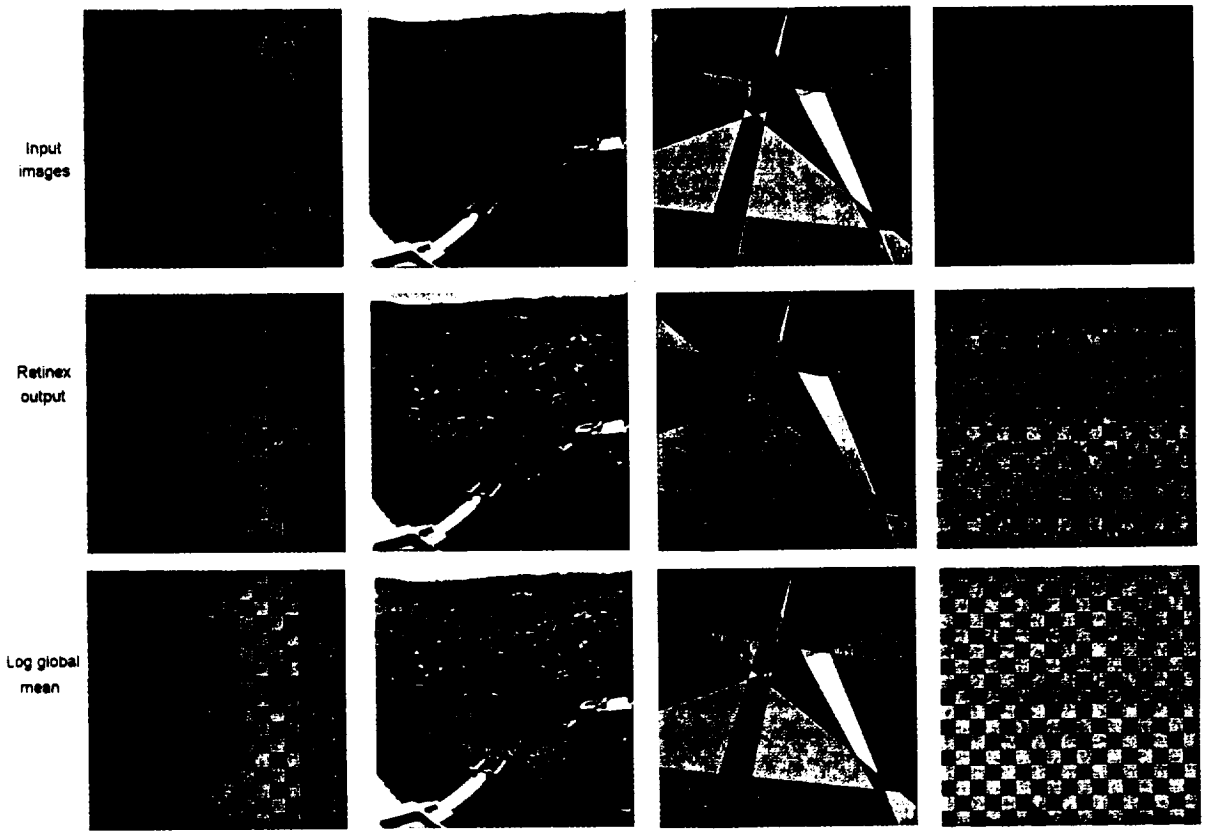


Fig. 12. Results for images with noticeable artifacts due to zonal or global violations of the gray-world assumption and the substitution of log global mean values for log surround to correct regional gray-world violations ($c_3 = 80$ pixels).

[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

$$R_{GR} = \log S_G r_G - \log \overline{S_R r_R} \quad (12)$$

$$R_{RG} = \log S_R r_R - \log \overline{S_G r_G} \quad (13)$$

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

$$R_{GR}^{DO} = R_{GR} + R_{RG} \quad (14)$$

$$= \log S_G r_G + \log S_R r_R - \log \overline{S_R r_R} - \log \overline{S_G r_G} \quad (15)$$

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

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

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

Likewise, for a blue yellow double-opponent form:

$$R_{BY} = \log S_B r_B - \log \overline{S_Y r_Y} \quad (17)$$

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

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

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

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

Analogously, for the "green minus red" case

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

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

$$R_L = \log c_3 S_B r_B + \log c_4 S_G r_G + \log c_5 S_R r_R \quad (21)$$

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

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

Again, log placement is important, since

$$\log \overline{S_B r_B} \neq \overline{\log S_B r_B} \quad (23)$$

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

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

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.

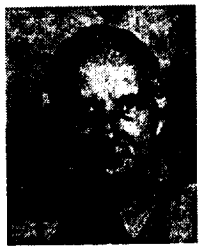
An implementation in analog VLSI or digital VLSI computer chips is an exciting possibility for realizing "smart" cameras of the future.

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.

A Multi-Scale Retinex For Bridging the Gap Between Color Images and the Human Observation of Scenes

D. J. Jobson, Z. Rahman and G. A. Woodell

IEEE Transactions on Image Processing
Special Issue on Color Processing
(July 1997)

A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes

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

***Abstract*— Direct observation and recorded color images of the same scenes are often strikingly different because human visual perception computes the conscious representation with vivid color and detail in shadows, and with resistance to spectral shifts in the scene illuminant. A computation for color images that approaches fidelity to scene observation *must* combine dynamic range compression, color consistency—a computational analog for human vision color constancy—and color and lightness tonal rendition. In this paper, we extend a previously designed single-scale center/surround retinex to a multiscale version that achieves simultaneous dynamic range compression/color consistency/lightness rendition. This extension fails to produce good color rendition for a class of images that contain violations of the gray-world assumption implicit to the theoretical foundation of the retinex. Therefore, we define a method of color restoration that corrects for this deficiency at the cost of a modest dilution in color consistency. Extensive testing of the multiscale retinex with color restoration on several test scenes and over a hundred images did not reveal any pathological behavior.**

I. Introduction

A COMMON (and often serious) discrepancy exists between recorded color images and the direct observation of scenes (see Fig. 1). Human perception excels at constructing a visual representation with vivid color and detail across the wide ranging photometric levels due to lighting variations. In addition, human vision computes color so as to be relatively independent of spectral variations in illumination [1]; i.e., it is color constant. The recorded images of film and electronic cameras suffer, by comparison, from a loss in clarity of detail and color as light levels drop within shadows, or as distance from a lighting source increases. Likewise, the appearance of color in recorded images is strongly influenced by spectral shifts in the scene illuminant. We refer to the computational analog to human vision color constancy as color consistency. When the dynamic range of a scene exceeds the dynamic range of the recording medium, there is an irrevocable loss of visual information at the extremes of the scene dynamic range. Therefore, improved fidelity of color images to human observation demands i) a computation that synthetically combines dynamic range compression, color consistency, and color and lightness rendition, and ii) wide dynamic range color imaging systems. The multiscale retinex (MSR) approaches the first of these goals. The design of the computation is tailored to visual perception by comparing the measured photometry of scenes with the performance of visual perception. This provides a rough quantitative measure of human vision's dynamic range compression—approaching 1000 : 1 for strong illumination variations of bright sun to deep shade.

The idea of the retinex was conceived by Land [2] as a model of the lightness and color perception of human vision. Through the years, Land evolved the concept from a random walk computation [3] to its last form as a center/surround spatially opponent operation [4], which is related to the neurophysiological functions of individual neurons in the primate retina, lateral geniculate nucleus, and cerebral cortex. Subsequently, Hurlbert [5]–[7] studied the properties of this form of retinex and other lightness theories and found that they share a common mathematical foundation but cannot actually compute reflectance for arbitrary scenes. Certain scenes violate the "gray-world" assumption—the requirement that the average reflectances in the surround be equal in the three spectral color bands. For example, scenes that are dominated by one color—"monochromes"—clearly violate this assumption and are forced to be gray by the retinex computation. Hurlbert further studied the lightness problem as a learning problem for artificial neural networks and found that the solution had a center/surround spatial form. This suggests the possibility that the spatial opponency of the center/surround is, in some sense, a general solution to estimating relative reflectances for arbitrary lighting conditions. At the same time, it is equally clear that human vision does not determine relative reflectance, but rather a context-dependent relative reflectance since the same surfaces in shadow and light do not appear to be the same. Moore *et al.* [8], [9] took up the retinex problem as a natural implementation for analog very large scale integration (VLSI) resistive networks and found that color rendition was dependent on scene content—whereas some scenes worked well, others did not. These studies also pointed out the problems that occur due to color Mach bands and the graying-out of large uniform zones of color.

We have previously defined a single-scale retinex [10] (SSR) that can either provide dynamic range compression (small scale), or tonal rendition (large scale), but not both simultaneously. The multiscale retinex with color restoration (MSRCR) combines the dynamic range compression of the small-scale retinex and the tonal rendition of the large scale

Manuscript received March 21, 1996; revised January 28, 1997. The work of Z. Rahman was supported by NASA Contract NAS1-1903 and NASA Grant NAG1-1847. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. H. Joel Trussell.

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

Z. Rahman was with Science and Technology Corporation, Hampton, VA USA. He is now with the Department of Computer Science, College of William and Mary, Williamsburg, VA 23187 USA.

Publisher Item Identifier S 1057-7149(97)04726-X.

Fig. 1. Illustration of the discrepancy between color images and and perception. The right image is a much closer representation of the visual impression of the scene.

retinex with a universally applied color restoration. This color restoration is necessary to overcome the problems that the MSR has in the rendition of scenes that contain gray-world violations. It merges all the necessary ingredients to approximate the performance of human vision with a computation that is quite automatic and reasonably simple. These attributes make the MSRCR attractive for smart camera applications, in particular for wide dynamic range color imaging systems. For more conventional applications, the MSRCR is useful for enhancing 8-b color images that suffer from lighting deficiencies commonly encountered in architectural interiors and exteriors, landscapes, and nonstudio portraiture.

Most of the emphasis in previous studies has been on the color constancy property of the retinex, but its dynamic range compression is visually even more dramatic. Since we want to design the retinex to perform in a *functionally* similar manner to human visual perception, we begin with a comparison of the photometry of scenes to their perception. This defines (at least in some gross sense) the performance goal for the retinex dynamic range compression.

An apparent paradox has been brought to our attention by a colleague as well as a reviewer. This paradox is so fundamental that it requires careful consideration before proceeding. The question, simply stated, is why should recorded images need dynamic range compression, since the compression of visual perception will be performed when the recorded image is observed? First we must state categorically that recorded images with significant shadows and lighting variations *do need compression*. This has been our experience in comparing the perception of recorded images with direct observation for numerous scenes. Therefore, we have to conclude that the dynamic range compression for perception of the recorded images is substantially weaker than for the scene itself. Fig. 1 is a case in point. There is no linear representation of this image, such as the viewing of the image on a gamma-corrected cathode ray tube (CRT) display, which even comes close to the dynamic compression occurring during scene observation. The same is true for all scenes we have studied with major lighting variations. We offer the possible explanation that weak dynamic range compression can result from the major differences in angular extent between scene and image viewing. Image frames are typically about 40° in angular extent for a 50 mm film camera. These same frames are usually viewed with about a 10° display or photographic print. Furthermore, the original 40° frame is taken out of the larger context, which would be present when observing the scene directly. The dynamic range compression of human vision is strongly dependent upon the angular extent of visual phenomena. Specifically, compression is much stronger for large shadow zones than for smaller ones. We feel that this a plausible resolution for this apparent paradox, and are certainly convinced by considerable experience that recorded images do need computational dynamic range compression for scenes that contain significant lighting variations. Likewise, this explanation applies to color consistency.

Since the nonlinear nature of the MSR makes it almost impossible to prove its generality, we provide the results of processing many test images as a measure of confidence in its general utility and efficacy. Results obtained with test scenes—i.e., where direct observation of the subject of the image is possible—are given more weight because the performance of the computation can be compared directly to observation of the scene.

II. The Photometry of Scenes Compared to Perception

We approached learning more about the dynamic range compression in human vision by exploring the perceptual and photometric limits. We did this by selecting and measuring

TABLE I
PHOTOMETRY OF SCENES

	cd/m^2
Visual saturation (white clouds near sun)	49,000
Just below saturation (clouds further from sun)	18,000–37,000
Outdoor building facade—bright sun	7000–13,000
Blue sky—morning	4600
Concrete sidewalk in	
sun	3200
shadow	570
deep shadow	290
Interior conference room-fluorescent lighting	
Floor/walls	36–140
Shadows	4–18
Interior conference room-unlit but by window	
Walls	29
Shadows	6
Inside open closet	1

scenes with increasingly emphatic lighting variations and then examining the point at which dynamic range compression gives way to loss of visual information. In other words, we looked for the dynamic range extremes at which human vision either saturates or clips the signals from very dark zones in a scene. We used a photographic spotmeter for the photometric measurements. In addition, we attempted to calibrate the perceptual lightness difference that occurs when the same surface is viewed in direct sunlight and in shadow. To quantify this difference, we compared the perceived lightness under both conditions to a reference gray-scale in direct sun and asked the question: Which gray scales match the surface in sun and shadow? Whereas the extreme measurements provide information about where dynamic range compression becomes lossy, the sun/shadow/gray-scale matches give some measure of the dynamic range compression taking place within more restricted lighting changes.

The results of the photometric measurements are given in Table I. The conditions shown are representative of the wide dynamic range encountered in many everyday scenes. Scene visibility is good except under the most extreme lighting conditions. On the low end, visibility is quite poor at 1 candles/m^2 (cd/m^2) luminance but improves rapidly as light levels approach 10 cd/m^2. Detail and color are quite easily visible across the range of 10–10 000 cd/m^2, even when all occur together in a scene. We can therefore conclude that dynamic range compression within a scene can approach 1000 : 1, but becomes lossy for wider ranges. For low luminance, color and detail are perceptually hazy with a loss of clarity; and for extremely low levels of luminance (approaching 10 000 : 1 when compared with direct sunlight), all perception of color and detail is lost.

We can also quantitatively estimate from this data the difference between perception and photometry for a very commonly encountered case: objects in sun and shadow. The drop in light level usually associated with a shadow is between 10–20% of the sunlit value, depending on the depth of the shadow. We compared the perceived drop in lightness to a reflectance gray-scale and concluded that the perceptual decrease is only about 50% of the sunlit lightness value. This clearly demonstrates the large discrepancy between recorded images and perception, even for conditions that do not encompass a very wide dynamic range. This data implies that for 10 : 1 changes in lighting, the perception of these changes is about 3–5 : 1 to minimize the impact of lighting on the scene representations formed by consciousness. Hence, as simple and ubiquitous an event as a shadow immediately introduces a major discrepancy between recorded images and visual perception of the same scene. This sets a performance goal derived from human visual perception with which to test the retinex. Clearly, a very strong nonlinearity exists in human vision, although our experiments can not define the exact form of this neural computation.

III. CONSTRUCTION OF A MULTISCALE CENTER/SURROUND RETINEX

The single-scale retinex [10]–[12] is given by

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

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

$$F(x,y) = Ke^{-r^2/c^2}$$

where c is the Gaussian surround space constant, and K is selected such that

$$\iint F(x,y)\,dx\,dy = 1.$$

The MSR output is then simply a weighted sum of the outputs of several different SSR outputs. Mathematically,

$$R_{\mathrm{MSR}_i} = \sum_{n=1}^{N} w_n R_{n_i} \quad (2)$$

where N is the number of scales, R_{n_i} is the ith component of the nth scale, R_{MSR_i} is the ith spectral component of the MSR output, and w_n is the weight associated with the nth scale. The only difference between $R(x,y)$ and $R_n(x,y)$ is that the surround function is now given by

$$F_n(x,y) = Ke^{-r^2/c_n^2}.$$

A new set of design issues emerges for the design of the MSR in addition to those for the SSR [10]. This has primarily to do with the number of scales to be used for a given application, and how these realizations at different scales should be combined. Because experimentation is our only guide in resolving these issues, we conducted a series of tests starting with only two scales and adding further scales as needed. After experimenting with one small scale ($c_n < 20$) and one large scale ($c_n > 200$), the need for a third intermediate scale was immediately apparent in order to produce a graceful rendition without visible "halo" artifacts near strong edges. Experimentation showed that equal weighting of the scales—$w_n = 1/3, n = 1,2,3$—was sufficient for most

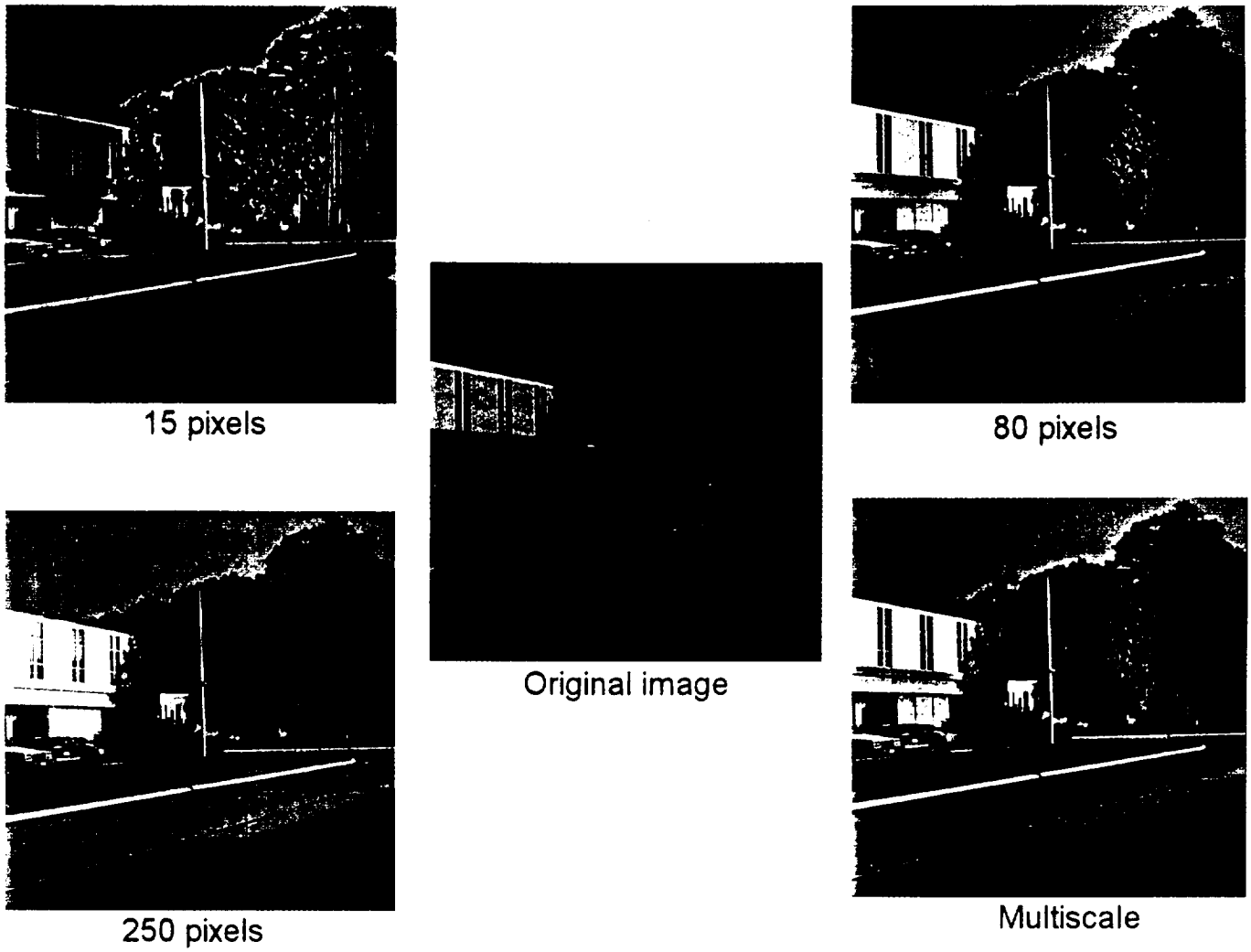


Fig. 2. Components of the multiscale retinex that show their complementary information content. The smallest scale is strong on detail and dynamic range compression and weak on tonal and color rendition. The reverse is true for the largest spatial scale. The multiscale retinex combines the strengths of each scale and mitigates the weaknesses of each.

applications. Weighting the smallest scale heavily to achieve the strongest dynamic range compression in the rendition leads to ungraceful edge artifacts and some graying of uniform color zones.

To test whether the dynamic range compression of the MSR approaches that of human vision, we used test *scenes* that we had observed in addition to test images that we had obtained from other test sources. The former allowed us to readily compare the processed image to the direct observation of the scene. Fig. 2 illustrates the complementary strengths and weaknesses of each scale taken separately and the strength of the multiscale synthesis. This image is representative of a number of test scenes (see Fig. 3) where for conciseness we show only the multiscale result.

The comparison of the unprocessed images to the perception of the scene produced some striking and unexpected results. When direct viewing was compared with the recorded image, the details and color were far more vivid for direct viewing not only in shadowed regions, but also in the bright zones of the scene! This suggests that human vision is doing even more image enhancement than just strong dynamic range compression, and the MSR may ultimately need to be modified to capture the realism of direct viewing. Initially, we tackle the dynamic range compression, color consistency, and tonal/color rendition problems, while keeping in mind that further work may be necessary to achieve full realism.

A sample of image data for surfaces in both sun and shadow indicates a dynamic range compression of 2 : 1 for the MSR compared to the 3–5 : 1 measured in our perceptual tests. For the SSR ($c_1 = 80$) this value is 1.5 : 1 or less. These levels of dynamic range compression are for outdoor scenes where shadows have large spatial extent. Shadows of small spatial extent tend to appear "darker" and are more likely to be clipped in recorded images. Fig. 3 shows a high dynamic range indoor/outdoor scene. The foreground orange book on the gray-scale is compressed by approximately 5 : 1 for the MSR while compression for the SSR is only about 3 : 1, both relative to the bright building facade in the background.

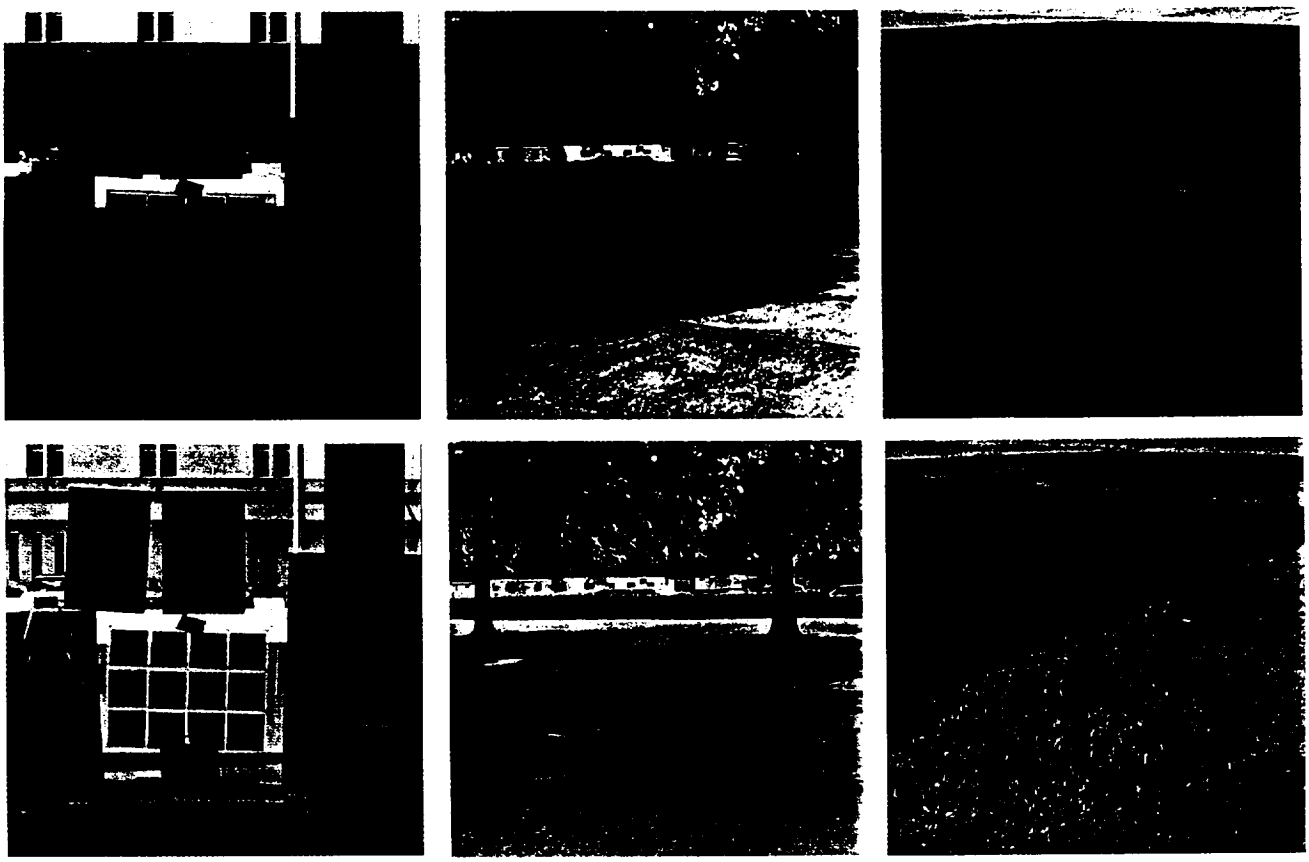


Fig. 3. Examples of test scenes processed with the multiscale retinex prior to color restoration. While color rendition of the left image is good, the other two are "grayed" to some extent. Dynamic range compression and tonal rendition are good for all and compare well with scene observation. Top row: Original. Bottom row: Multiscale retinex.

The compression for human vision is difficult to estimate in this case, since both the color and texture of the two surfaces are quite different. Our impression from this analysis is that the MSR is approaching human vision's performance in dynamic range compression but not quite achieving it. For scenes with even greater lighting dynamics than these, we can anticipate an even higher compression for the MSR to match human vision. However, we are currently unable to test this hypothesis because the conventional 8-b analog-to-digital converters of both our solid-state camera and slide film/optical scanner digitizer restrict the dynamic range with which the image data for such scenes can be acquired. Solid state cameras with 12-b dynamic range and thermoelectrically cooled detector arrays with 14-b dynamic range are, however, commercially available, and can be used for examining the MSR performance on the wider dynamic range natural scenes. Even for the restricted dynamic range shown in Fig. 3 (left), it is obvious that limiting noise has been reached, and that much wider dynamic range image acquisition is essential for realizing a sensor/processing system capable of approximating human color vision.

For the conventional 8-b digital image range, the MSR performs well in terms of dynamic range compression, but its performance on the pathological classes of images examined in previous SSR research [10] must still be examined. Fig. 4 shows a set of images that contain a variety of regional and global gray-world violations. The MSR, as expected, fails to handle them effectively—all images possessing notable, and often serious, defects in color rendition (see Fig. 4, middle row). We only provide these results as a baseline for comparison with the color restoration scheme, presented in the next section, that overcomes these deficiencies of the MSR.

IV. A Color Restoration Method for the Multiscale Retinex

The general effect of retinex processing on images with regional or global gray-world violations is a "graying out" of the image, either globally or in specific regions. This desaturation of color can, in some cases, be severe (see Fig. 4, middle). More rarely, the gray-world violations can simply produce an unexpected color distortion (see Fig. 4, top left). Therefore, we consider a color restoration scheme that provides good color rendition for images that contain gray-world violations. We, of course, require the restoration to preserve a reasonable degree of color consistency, since that is one of the prime objectives of the retinex. Color constancy is known to be imperfect in human visual perception, so some level of illuminant color dependency is acceptable, provided it is much lower than the physical spectrophotometric variations. Ultimately, this is a matter of image quality, and

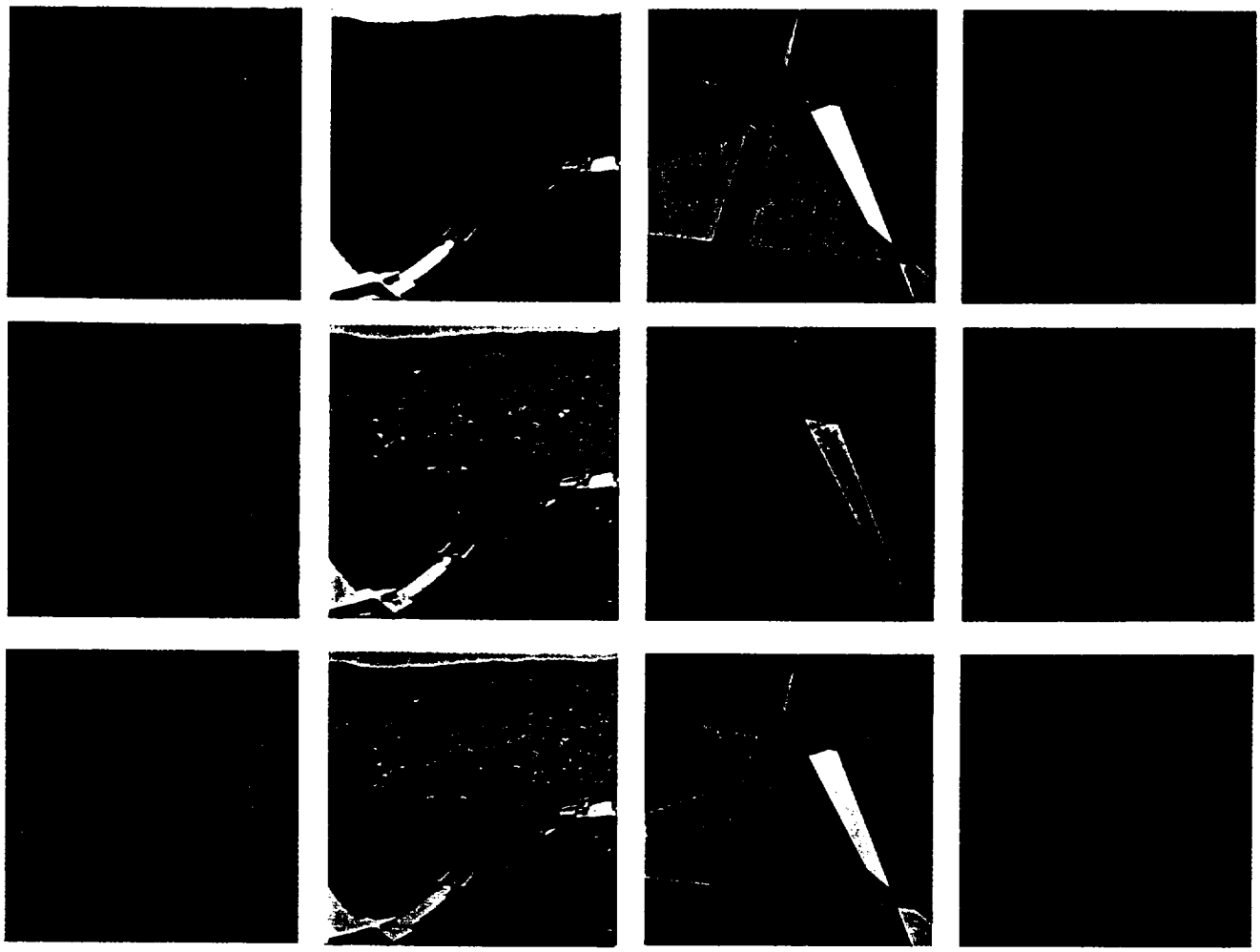


Fig. 4. Pathological "gray-world" violations are not handled well by the multiscale retinex alone (middle row), but are treated successfully when color restoration is added (lower row). Top row: Original.

color dependency is tolerable to the extent that the visual defect is not visually too strong.

We begin by considering a simple colorimetric transform [13], even though it is often considered to be in direct opposition to color constancy models. It is also felt to describe only the so-called "aperture mode" of color perception, i.e., restricted to the perception of color lights rather than color surfaces [14]. The reason for this choice is simply that it is a method for creating a relative color space, and in so doing becomes less dependent than raw spectrophotometry on illuminant spectral distributions. This starting point is analogous to the computation of chromaticity coordinates where

$$I_i'(x,y) = I_i(x,y) / \sum_{i=1}^{S} I_i(x,y) \tag{3}$$

for the ith color band, and S is the number of spectral channels. Generally, $S = 3$, using the red–green–blue (RGB) color space. The modified MSR that results is given by

$$R_{\mathrm{MSRCR}_i}(x,y) = C_i(x,y) R_{\mathrm{MSR}_i}(x,y) \tag{4}$$

where

$$C_i(x,y) = f[I_i'(x,y)]$$

is the ith band of the color restoration function (CRF) in the chromaticity space, and R_{MSRCR_i} is the ith spectral band of the multiscale retinex with color restoration. In a purely empirical manner, we tried several linear and nonlinear color restoration functions on a range of test images. The function that provided the best overall color restoration was

$$\begin{aligned} C_i(x,y) &= \beta \log[\alpha I_i'(x,y)] \\ &= \beta \left\{ \log[\alpha I_i(x,y)] - \log\left[\sum_{i=1}^{S} I_i(x,y)\right] \right\} \end{aligned} \tag{5}$$

where β is a gain constant, and α controls the strength of the nonlinearity. In the spirit of a preserving a canonical computation, we determined that a single set of values for β and α worked for all spectral channels. The final MSRCR output is obtained by using a "canonical" gain/offset to transition between the logarithmic domain and the display domain. Looking at the forms of the CRF of (5) and the SSR of (1), we conjecture that the CRF represents a spectral analog to the spatial retinex. This mathematical and philosophical

TABLE II
LIST OF CONSTANTS USED FOR ONE PARTICULAR IMPLEMENTATION OF THE MSRCR ON A DEC ALPHA 3000, USING THE VMS F77 COMPILER

Constant	N	c_1	c_2	c_3	G	b	α	β	w_n
Value	3	15	80	250	192	-30	125	46	1/3

symmetry is intriguing, since it suggests that there may be a unifying principle at work. Both computations are nonlinear, contextual, and highly relative. We can speculate that the visual representation of wide dynamic range scenes must be a compressed mesh of contextual relationships for lightness and color representation. This sort of information representation would certainly be expected at more abstract levels of visual processing such as form information composed of edges, links, and the like, but is surprising for a representation so closely related to the raw image. Perhaps in some way this front-end computation can serve later stages in a presumed hierarchy of machine vision operations that would ultimately need to be capable of such elusive goals as resilient object recognition.

The bottom row in Fig. 4 shows the results of applying the CRF to the MSR output for pathological images. The MSRCR provides the necessary color restoration, eliminating the color distortions and gray zones evident in the MSR output. The challenge now is to prove the generality of this computation. Since there is not a mathematical way to do this, we have tested the computation on several hundred highly diverse images without discovering exceptions. Unfortunately, space considerations allow us to present only a very small subset of all the images that we have tested.

V. SELECTED RESULTS FOR DIVERSE TEST CASES

Extensive testing indicates that the gain constant α for the CRF and the final gain/offset adjustment required to transition from the logarithmic to the display domain are independent of the spectral channel and the image content. This implies that the method is general or "canonical," and can be applied automatically to most (if not all) images without either interactive adjustments by humans or internal adjustments such as an auto-gain. This final version of the MSRCR can then be written as

$$R_{\mathrm{MSRCR}_i}(x,y) = G[C_i(x,y)\{\log I_i(x,y) - \log[I_i(x,y) * F_n(x,y)]\} + b] \quad (6)$$

where G and b are the final gain and offset values, respectively. The constants G and b intrinsically depend upon the implementation of the algorithm in software. Table II gives a list of the constants used to produce all the outputs in this paper.

We must again emphasize that the choice of the all constants merely represents a particular implementation that works well for a wide variety of images. In no way do we mean to imply that these constants are optimal or "best case" for all possible implementations of this algorithm. The choice of the surround space constants, c_ns, in particular does not seem to be critical. Instead, the choice seems to only need to provide reasonable coverage from local to near global. Likewise, the choice of using three scales was made empirically to provide the minimum number of scales necessary for acceptable performance.

The test images presented here begin with some test scenes since we feel it is fundamental to refer the processed images back to the direct observation of scenes. This is necessary to establish how well the computation represents an observation. Clearly, we cannot duplicate human vision's peripheral vision which spans almost 180°, but within the narrower angle of most image frames, we would like to demonstrate that the computation achieves the clarity of color and detail in shadows, reasonable color constancy and lightness and color rendition that is present in direct observation of scenes. The test scenes (see Fig. 5) compare the degree with which the MSRCR approaches human visual performance. All four of the MSRCR outputs shown in Fig. 5 are quite "true to life" compared to direct observation, except for the leftmost, which seems to require even more compression to duplicate scene perception. This image was scanned from a slide and digitized to 8-b/color. The other three images were taken with a Kodak DCS200C CCD detector array camera. In none of the cases could a gamma correction produce a result consistent with direct observation. Therefore, we conclude that the MSRCR is not correcting simply for a CRT display nonlinearity, and that far stronger compression than gamma correction is necessary to approach fidelity to visual perception of scenes with strong lighting variations. We did not match camera spatial resolution to observation very carefully, so some difference in perceived detail is expected and observed. However, overall color, lightness, and detail rendering for the MSRCR is a good approximation to human visual perception.

The rest of the selected test images (Figs. 6–8) were acquired from a variety of sources (see acknowledgments) and provide as wide a range of visual phenomena as we felt could be presented within the framework of this paper. Little comment is necessary and we will leave the ultimate judgment to the reader. Some images with familiar colors and no strong lighting defects are included to show that the MSRCR does not introduce significant visual distortions into images that are without lighting variations. The white stripes of the American flag in Fig. 6(a) show a shift toward blue-green in the MSRCR output. This is, perhaps, analogous to the simultaneous color contrast phenomena of human perception. Moore *et al.* [8] noted a similar effect in their implementation of a different form of the retinex. The Paul Klee painting in Fig. 7(b) is included as a test of the subtlety of tonal and color rendition. Some of the test images with strong shadows zones where one or two color channels are preferentially clipped do exhibit a color distortion. This is due to the rather limited dynamic range of the "front-end" imaging/digitization, and is not an artifact of the computation. Even for these cases, the MSRCR produces far more visual information and is more "true-to-life" than the unprocessed image. The set of space images are included to show the application of the MSRCR to both space operations imagery and remote sensing applications.

A further test is worthwhile in assessing the impact of the CRF on color consistency. The CRF, as expected, dilutes color consistency, as shown in Fig. 9. However, the residual color dependency is fairly weak and the visual impression of color shift is minimal especially in comparison with the dramatic shifts present in the unprocessed images.

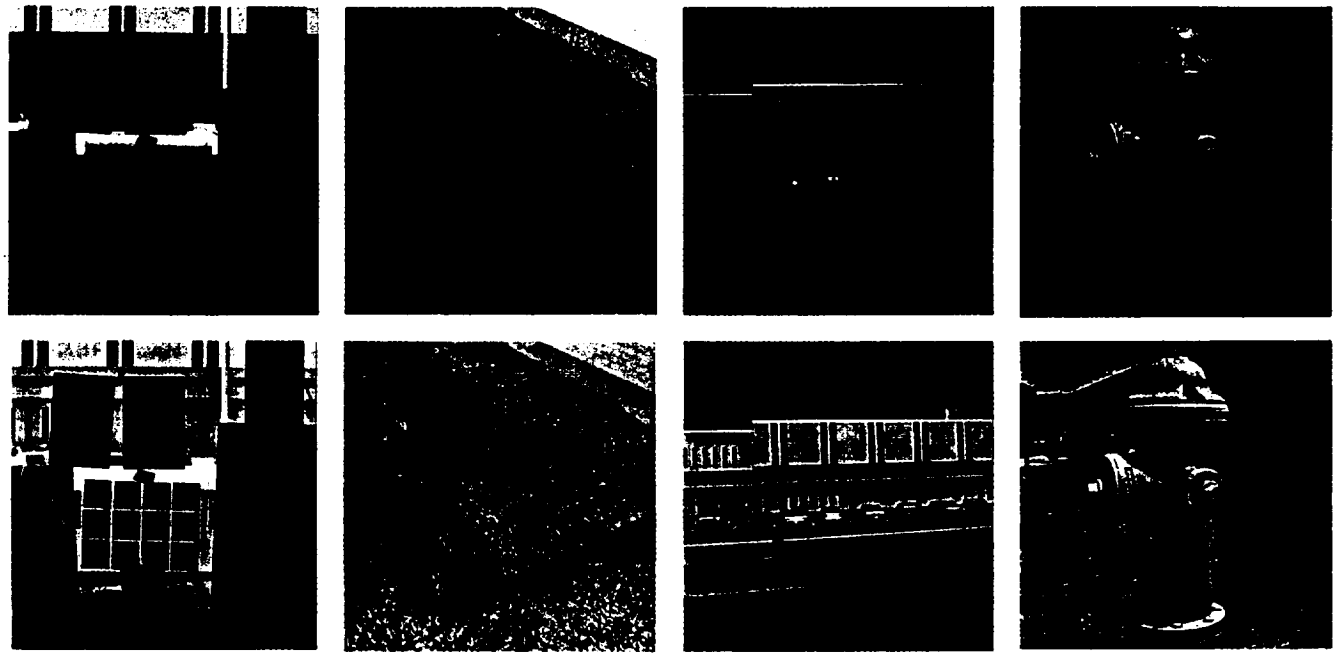


Fig. 5. Test scenes illustrating dynamic range compression, color, and tonal rendition, and automatic exposure correction. All processed images compare favorably with direct scene observation with the possible exception of leftmost image, which is even lighter and clearer for observation. This scene has the widest dynamic range and suggests that even stronger dynamic range compression may be needed for this case. Top row: Original. Bottom row: Multiscale retinex.

Fig. 6. Photographic examples further illustrating graceful dynamic range compression together with tonal and color rendition. The rightmost image shows the processing scheme handling saturated colors quite well and not distorting an image that is quite good in its original form. Top row: Original. Bottom row: Multiscale retinex.

VI. Discussion

While we have not yet conducted an extensive performance comparison of the MSRCR to other image enhancement methods, we have done some preliminary tests of the MSRCR relative to the simpler image enhancement methods—histogram equalization, gamma correction, and gain/offset manipulation [15], and point logarithmic nonlinearity [16]. Overall, the performance of the retinex is consistently good, while performance for the others is quite variable. In particular, the retinex excels when there are major zones of both high and low light levels. The traditional methods that we have compared against are all point operations on the image,

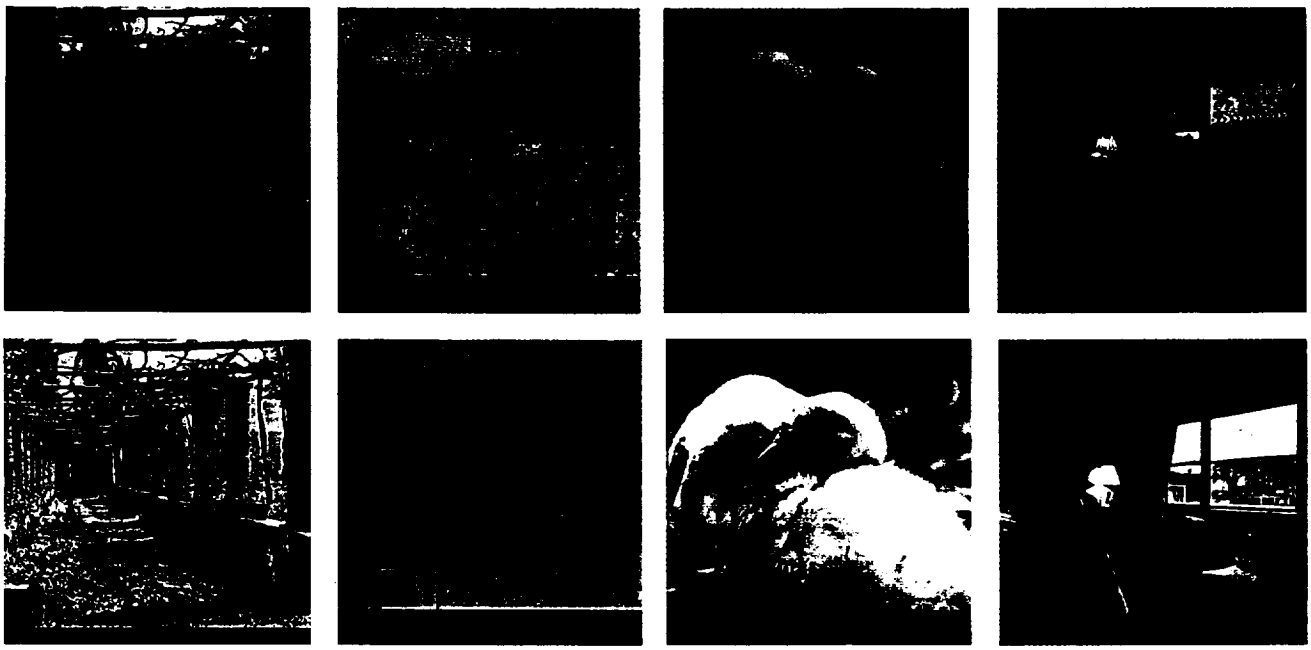


Fig. 7. Miscellaneous examples illustrating fairly dramatic dynamic range compression as well one for subtlety of color rendition (second from leftmost—painting by Paul Klee). Top row: Original. Bottom row: Multiscale retinex.

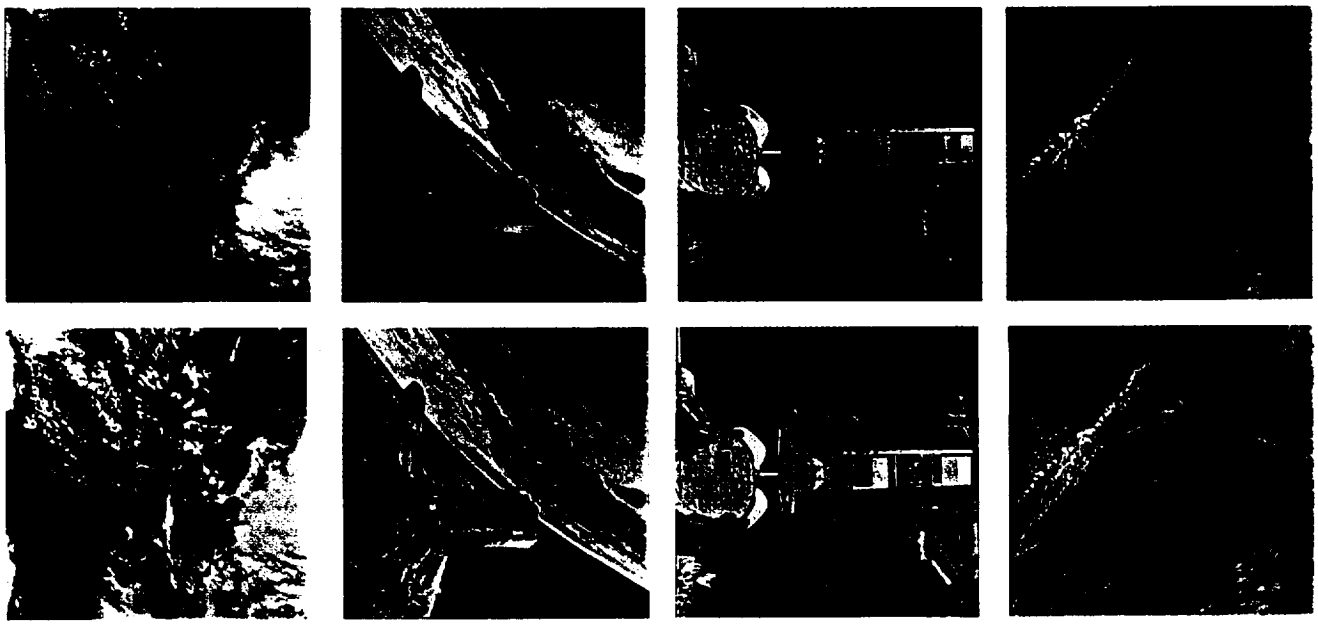


Fig. 8. Selection of space images to show enhancement of space operations imagery and remote sensing data. Top row: Original. Bottom row: Multiscale retinex.

whereas unsharp masking [17] and homomorphic filtering [17], [18] are spatial operations more mathematically akin to center/surround operation of the retinex. Unsharp masking is a linear subtraction of a blurred version of the image from the original and is generally applied using slight amounts of blurring. For a given space constant for the surround, we would expect the retinex to be much more compressive. It is not clear that unsharp masking would have any color constancy property, since the subtraction process in the linear domain is essentially a highpass filtering operation and not a ratio that provides the color constancy of the retinex.

Homomorphic filtering is perhaps the closest computation to the MSRCR and in one derivation [19] has been applied to color vision. Both its original form and the color form rely upon a highpass filtering operation that takes place after the dynamic range of the image is compressed with a point log-

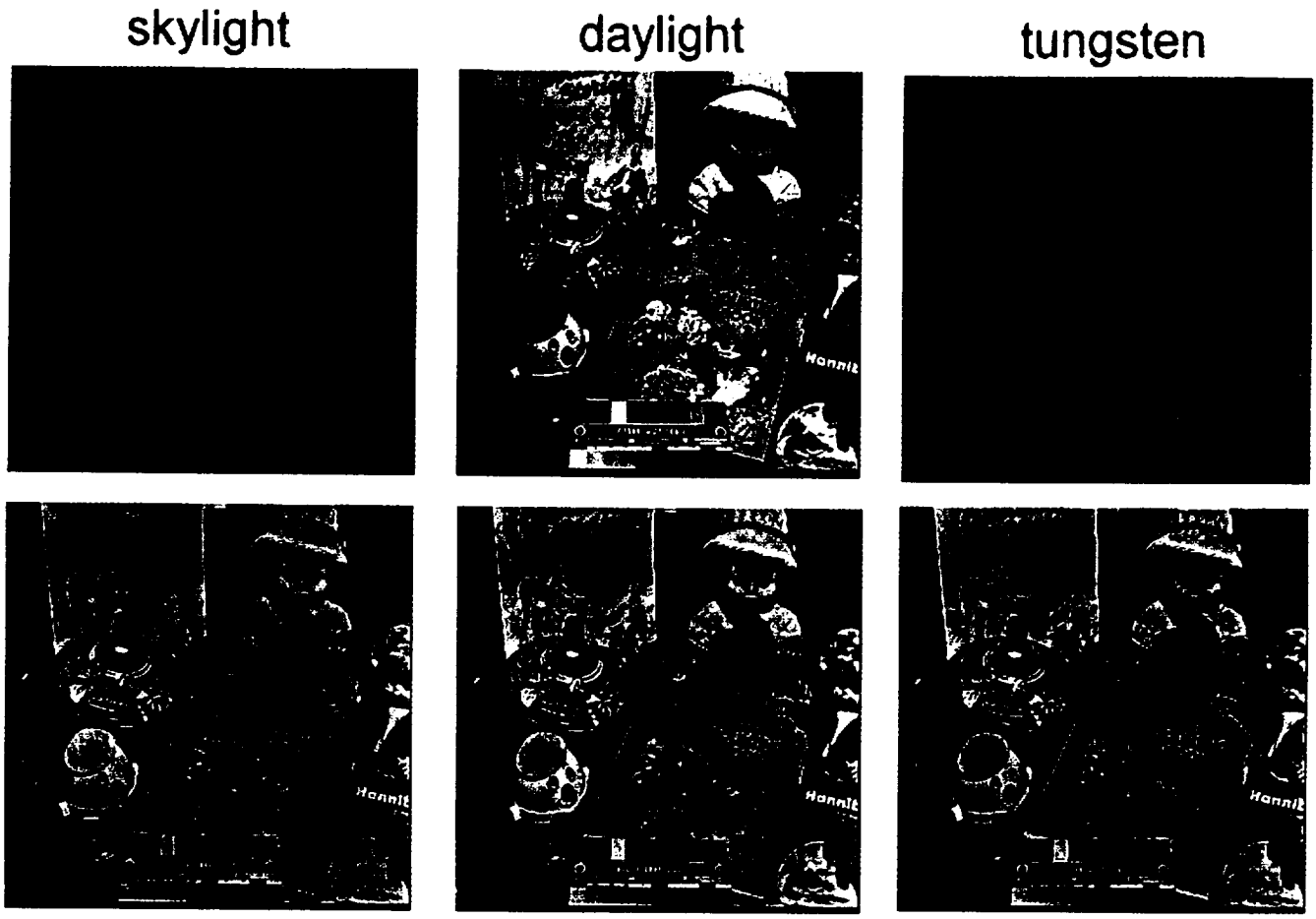


Fig. 9. Toy scene revisited. A test of the dilution of color consistency by the color restoration. While color consistency was shown previously to be near perfect for the SSR and MSR, some sacrifice of this was necessary to achieve color rendition. While slight changes in color can be seen, color consistency is still quite strong relative to the spectrophotometric changes seen in the original images (top row). The blues and yellows are in the color restored multiscale retinex (bottom row) are the most affected by the computer simulated spectral lighting shifts, but the effect is visually weak and most colors are not visibly affected.

arithmic nonlinearity. An inverse exponentiation then restores the dynamic range to the original display space. The color vision version adds an an opponent-color/achromatic transformation after the application of the logarithmic nonlinearity. We have found that the application of the logarithmic nonlinearity before spatial processing gives rise to emphatic "halo" artifacts and have also shown that it is quite different visually and mathematically from the application of the log after the formation of the surround signal [10]. Because of the nonlinearities in both the MSRCR and homomorphic filtering, a straightforward mathematical comparison is not possible. We do, however, anticipate significant performance differences between the two in terms of dynamic range compression, rendition, and, for the color vision case, color consistency. Another major difference between the MSRCR and homomorphic filtering is in the application of the inverse function in homomorphic filtering. The analogous operation in the MSRCR is the application of the final gain/offset. Obviously, the two schemes use quite different techniques in going from the nonlinear logarithmic to the display domain. We conjecture that the application of the inverse log function in the retinex computation would undo some of the compression it achieves.

One of the most basic issues for the use of this retinex is the trade-off between the advantages versus the introduction of context dependency on local color and lightness values. Our experience is that the gains in visual quality, which can be quite substantial, outweigh the relatively small context dependency. The context dependencies are perhaps of most concern in remote sensing applications. The strongest context dependencies occur for the dark regions that are low because of low scene reflectances—for example, large water areas in remote sensing data adjacent to bright land areas. The large zones of water are greatly enhanced and subtle patterns in them emerge. The retinex clearly distorts radiometric fidelity in favor of visual fidelity. The gains in visual information, we hope, have been demonstrated adequately in our results. Even for specific remote sensing experiments where radiometric fidelity is required, the retinex may be a necessary auxiliary tool for the visualization of overall patterns in low signal zones. Visual information in darker zones that may not be detected with linear representations which preserve radiometry will "pop out" with a clarity limited only by the dynamic range of the sensor front-end and any intervening digitization scheme employed prior to the retinex. This may be especially useful

in visualizing patterns in remote sensing images covering land and water. Water has a much lower reflectance than land especially for false-color images including a near-infrared channel. The ability of the MSRCR to visualize features within both land and water zones simultaneously should be useful in coastal zone remote sensing.

The retinex computation can be applied ex post facto on 8-b color images and all of the results presented here represent this application. We have noticed only one problem with this—that the retinex can and will enhance artifacts introduced by lossy coding schemes, most notably lossy JPEG. Hence, the retinex is best applied prior to lossy image coding. One obvious advantage that the MSRCR provides for image compression is its ability to compress wider dynamic ranges to 8-bit or less per band color output, while preserving, and even enhancing, the details in the scene. The overall effect then is a significant reduction in the number of bits (especially in cases where the original color resolution is higher than 8-b/band) required to transmit the original without a substantial loss in spatial resolution or contrast quality.

The greatest power and advantage of the retinex is as a front-end computation, especially if the camera is also capable of wider than 8-b dynamic range. We have seen from scene photometry that 10–12-b dynamic ranges are required to encompass everyday scenes. Obviously, the retinex is most powerful as a front-end computation if it can be implemented within a sensor or between the sensor and coding/archival storage. We have not tested this retinex on wide dynamic range images, since we do not yet have access to an appropriate camera, therefore for wider dynamic range images some modifications in the processing may be anticipated. This may involve adding more scales, especially smaller ones, to provide a greater but still graceful dynamic range compression.

We have encountered many digital images in our testing that are underexposed. Apparently even with modern photographic autoexposure controls, exposure errors can and do occur. An additional benefit of the MSRCR is it capacity for exposure correction. Again, this is especially beneficial if it is performed as a front-end computation.

We do have the sense from our extensive testing thus far that the MSRCR approaches the high degree of dynamic range compression of human vision but may not quite achieve a truly comparable level of compression. Our impressions of the test scene cases is that direct observation is still more vivid in terms of color and detail than the processed images. This could be due to limitations in display/print media, or it could be that the processing scheme should be further designed to produce an even more emphatic enhancement. Further experimentation comparing test scenes to processed images and an accounting for display/print transfer characteristics will be necessary to resolve this remaining question and refine the method if necessary in the direction of greater enhancement of detail and color intensity. The transfer characteristics of print/display media deserve further investigation since most CRT's and print media have pronounced nonlinear properties. Most CRT's have an inverse "gamma" response [17] and the specific printer that we have used (Kodak XLT7720 thermal process) has a nonlinear response. For the printed results shown, we used a modest gamma correction ($\gamma = 1.2$). While this does not represent an accurate inverse that linearizes the printer transfer function, it does capture the the visual information with a reasonable good and consistent representation. Obviously no matter how general purpose the MSRCR is, highest quality results will still need to account for the specifics of print/display media especially since these are so often nonlinear.

VII. Conclusions

The MSR, comprised of three scales (small, intermediate, and large), was found to synthesize dynamic range compression, color consistency, and tonal rendition, and to produce results that compare favorably with human visual perception, except for scenes that contain violations of the gray-world assumption. Even when the gray-world violations were not dramatic, some desaturation of color was found to occur. A color restoration scheme was defined that produced good color rendition even for severe gray-world violations, but at the expense of a slight sacrifice in color consistency. In retrospect, the form of the color restoration is a virtual spectral analog to the spatial processing of the retinex. This may reflect some underlying principle at work in the neural computations of consciousness; perhaps, even that the visual representation of lightness, color, and detail is a highly compressed mesh of contextual relationships, a world of relativity and relatedness that is more often associated with higher levels of visual processing such as form analysis and pattern recognition.

While there is no firm theoretical or mathematical basis for proving the generality of this color restored MSR, we have tested it successfully on numerous diverse scenes and images, including some known to contain severe gray-world violations. No pathologies have yet been observed. Our tests were, however, confined to the conventional 8-b dynamic range images, and we expect that some refinements may be necessary when the wider dynamic range world of 10–12-b images is engaged.

Acknowledgment

The following World Wide Web sites provided the test images used for evaluating the performance of the MSRCR: Kodak Digital Image Offering at www.kodak.com/digitalImaging/samples/imageIntro. shtml; Monash University, Australia, DIVA Library at www.arts.monash.edu.au/visarts/diva/gardens.html; NASA Langley Research Center, LISAR Image Library at lisar.larc.nasa.gov/LISAR/BROWSE/ldef.html; and NASA Lyndon B. Johnson Space, Center Digital Image Collection at images.jsc.nasa.gov/html/shuttle.htm; Webmuseum, Paris at sunsite.unc.edu/louvre. The toy scene image is available from numerous sources.

References

[1] T. Cornsweet, *Visual Perception*. Orlando, FL: Academic, 1970.
[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] ——, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.*, vol. 80, pp. 5163–5169, 1983.
[4] ——, "Recent advances in retinex theory," *Vis. Res.*, vol. 26, pp. 7–21, 1986.
[5] A. C. Hurlbert, "The computation of color," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, Sept. 1989.
[6] ——, "Formal connections between lightness algorithms," *J. Opt. Soc. Amer. A*, vol. 3, pp. 1684–1693, 1986.
[7] A. C. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science*, vol. 239, pp. 482–485, 1988.
[8] 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.
[9] 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.
[10] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Processing*, vol. 6, pp. 451–462, Mar. 1997.
[11] Z. Rahman, "Properties of a center/surround retinex, part 1: Signal processing design," *NASA Contractor Rep. 198194*, 1995.
[12] D. J. Jobson and G. A. Woodell, "Properties of a center/surround retinex, part 2: Surround design," NASA Tech. Memo. 110188, 1995.
[13] P. K. Kaiser and R. M. Boynton, *Human Color Vision*, 2nd ed. Washington, DC: Opt. Soc. Amer., 1996.
[14] P. Lennie and M. D. D'Zmura, "Mechanisms of color vision," *CRC Crit. Rev. Neurobiol.*, vol. 3, pp. 333–400, 1988.
[15] Z. Rahman, D. Jobson, and G. A. Woodell, "Multiscale retinex for color rendition and dynamic range compression," in *Proc. SPIE 2847, Applications of Digital Image Processing XIX*, A. G. Tescher, Ed., 1996.
[16] ——, "Multiscale retinex for color image enhancement," in *Proc. IEEE Int. Conf. Image Processing*, 1996.
[17] J. C. Russ, Ed., *The Image Processing Handbook*. Boca Raton, FL: CRC, 1995.
[18] A. Oppenheim, R. Schafer, and T. Stockham, Jr., "Non-linear filtering of multiplied and convolved signals," *Proc. IEEE*, vol. 56, pp. 1264–1291, Aug. 1968.
[19] O. D. Faugeras, "Digital color image processing within the framework of a human vision model," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 380–393, Aug. 1979.

Daniel J. Jobson (M'97) 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, the design and testing of multispectral sensors, and the 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 compression/lightness constancy of human vision.

Zia-ur Rahman (M'87) received the B.A. degree in physics from Ripon College, Ripon, WI, in 1984 and the M.S. and Ph.D. degree 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 Assistant Professor with the Department of Computer Science, College of William and Mary, Williamsburg, VA. Prior to that, he was a research scientist with the Science and Technology Corporation, and worked 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 an emphasis in 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 aboard the Space Shuttle in 1985 to study the effect of gravity-induced convection. 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.

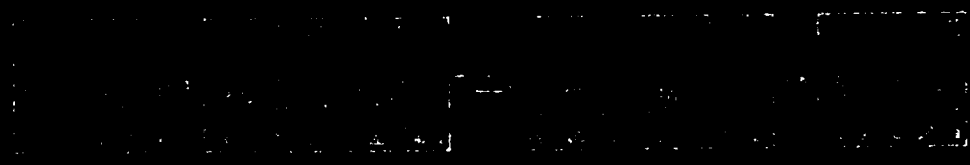
Visual Communication: An Information Theory Approach

F. O. Huck, C. L. Fales and Z. Rahman

Kluwer Academic Publishers
(June 1997)

VISUAL COMMUNICATION

An Information Theory Approach



Friedrich O. Huck
Carl L. Fales
Zia-ur Rahman

KLUWER ACADEMIC PUBLISHERS

The fundamental goal of *Visual Communication* is to produce the best possible picture at the lowest data rate. This goal cannot be reached, as it has been pursued in the past, by treating image gathering, coding and restoration as separate and independent tasks. Instead, in a clear departure from the mores of the traditional image processing literature, this monograph rigorously extends modern communication theory to the integration of the two disciplines that are involved: the electro-optical design of image gathering and display devices and the digital processing for image coding and restoration. Extensive simulations demonstrate that this approach establishes, for the first time, a close correlation between predicted and actual performance.

ISBN 0-7923-9956-0

9 780792 399568