Retinex Processing for Automatic Image Enhancement

Zia-ur Rahman†, Daniel J. Jobson‡, Glenn A. Woodell‡

†College of William & Mary, Department of Computer Science, Williamsburg, VA 23187.
‡NASA Langley Research Center, Hampton, Virginia 23681.

ABSTRACT

In the last published concept (1986) for a Retinex computation, Edwin Land introduced a center/surround spatial form, which was inspired by the receptive field structures of neurophysiology. With this as our starting point we have over the years developed this concept into a full scale automatic image enhancement algorithm—the Multi-Scale Retinex with Color Restoration (MSRCR) which combines color constancy with local contrast/lightness enhancement to transform digital images into renditions that approach the realism of direct scene observation. The MSRCR algorithm has proven to be quite general purpose, and very resilient to common forms of image pre-processing such as reasonable ranges of gamma and contrast stretch transformations. More recently we have been exploring the fundamental scientific implications of this form of image processing, namely: (i) the visual inadequacy of the linear representation of digital images, (ii) the existence of a canonical or statistical ideal visual image, and (iii) new measures of visual quality based upon these insights derived from our extensive experience with MSRCR enhanced images. The lattermost serves as the basis for future schemes for automating visual assessment—a primitive first step in bringing visual intelligence to computers.

1. INTRODUCTION

The idea of the retinex was conceived by Edwin Land[1] as a model of the lightness and color perception of human vision. Through the years, Land[2,3] evolved the concept from a random walk computation to its last form as a center/surround spatially opponent operation[3] which is related to the neurophysiological functions of individual neurons in the primate retina, lateral geniculate nucleus, and cerebral cortex. Subsequently Hurlbert[4–6] studied the properties of this form of 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 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 produced was 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.[7,8] took up the retinex problem as a natural implementation for analog VLSI resistive networks and found that color rendition was dependent on scene content. In each study, the consistent theoretical viewpoint was to perform all spatial processing within each spectral band and prohibit any interactions between spatial and spectral processing. This restriction provides very strong global color constancy

In our research[9–16] we do not use the retinex theory as a model for human vision color constancy. Rather, we use it as a platform for synthesizing local contrast improvement, color constancy, and lightness/color rendition as a goal for digital image enhancement. The intent is to transform the visual characteristics of the recorded digital image so that the rendition of the transformed image approaches that of the direct observation of scenes. Special emphasis is placed on increasing the local contrast in dark zones of the recorded image so that it would match our perception of wide dynamic range scenes, e.g., scenes which contain objects that are partly in sunlight

ZR: zrahman@cs.wm.edu; DJJ: d.j.jobson@larc.nasa.gov; GAW: g.a.woodell@larc.nasa.gov

and partly in shadow. Basic study of the properties of the center/surround retinex led us in the direction of using a Gaussian surround used by Hurlbert[4–6] as opposed to the $1/r^2$ surround originally proposed by Land[2, 3] or the exponential surround used by Moore[7, 8] for analog VLSI resistive networks. Since the width of the surround affects the rendition of the processed image, multiple scale surrounds were found to be necessary to provide a visually acceptable balance between dynamic range compression and graceful tonal rendition. This is discussed in more detail in Section 2.

The final visual defect in performance was the color "graying" due to global and regional violations of the gray-world assumption intrinsic to retinex theory. A color restoration was essential for correcting this and took the form of a log spectral operation similar to the log spatial operation of the center/surround. This produces an interaction between spatial and spectral processing and results in a tradeoff between strength of color constancy and color rendition. The color restoration yields a modest relaxation in color constancy perhaps comparable to human color vision's perceptual performance. This is discussed in more detail in Section 3.

In the course of our experiments, we noted that the commonly accepted linear representation of a scene's radiometric characteristics often fails to encompass its full dynamic range, resulting in images that either have saturated bright regions to compensate for the dark regions, or clipped dark regions in order to compensate for the bright regions. A nonlinear representation such as the MSRCR provides th e necessary dynamic range compression needed to encompass the full dynamic range of the scene. Section 4 lays out these ideas in more detail.

2. THE MULTI-SCALE RETINEX

The basic form of the Multi-scale retinex (MSR) is given by

$$R_i(x_1,x_2) = \sum_{k=1}^{K} W_k \left(\log\ I_i(x_1,x_2) - \log\left[F_k(x_1,x_2) * I_i(x_1,x_2)\right]\right) \quad i = 1,\cdots,N \tag{1}$$

where the sub-index i represents the i^{th} spectral band, N is the number of spectral bands—$N = 1$ for grayscale images and $N = 3$ for typical color images. In the latter case, $i \in R, G, B$—I is the input image, R is the output of the MSR process, F_k represents the k^{th} surround function, W_k are the weights associated with F_k, K is the number of surround functions, or scales, and $*$ represents the convolution operator. The surround functions, F_k are given as:

$$F_k(x_1,x_2) = \kappa \exp[-(x_1^2 + x_2^2)/\sigma_k^2],$$

where σ_k are the scales that control the extent of the surround—smaller values of σ_k lead to narrower surrounds—and $\kappa = 1/\left(\sum_{x_1}\sum_{x_2} F(x_1,x_2)\right)$.

As mentioned in Section 1, we found that multiple surrounds were necessary in order to achieve a graceful balance between dynamic range compression and tonal rendition. The number of scales used for the MSR is, of course, application dependent. We have found empirically, however, that a combination of three scales representing narrow, medium, and wide surrounds is sufficient to provide both dynamic range compression and tonal rendition. Figure 1 shows the input image, the output of the MSR and the outputs when the different surround functions are applied to the original image. These are obtained by setting $k = 1$ and $W_k = 1.0$ in Equation 1. As is evident from Figure 1, none of the individual scales attains the goal that we are trying to achieve: visual realism. The narrow and medium surround cases are self-explanatory. The wide-surround case deserves some discussion because it is a "nice" output image. However, the lack of dynamic range obscures the features that were visible to the observer, hence it fails the test. The MSR processed image uses features from all three scales to provide simultaneous dynamic range and tonal rendition.

3. MSR WITH COLOR RESTORATION

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 therefore we can consider the desired color computation as a color restoration, which should produce good color

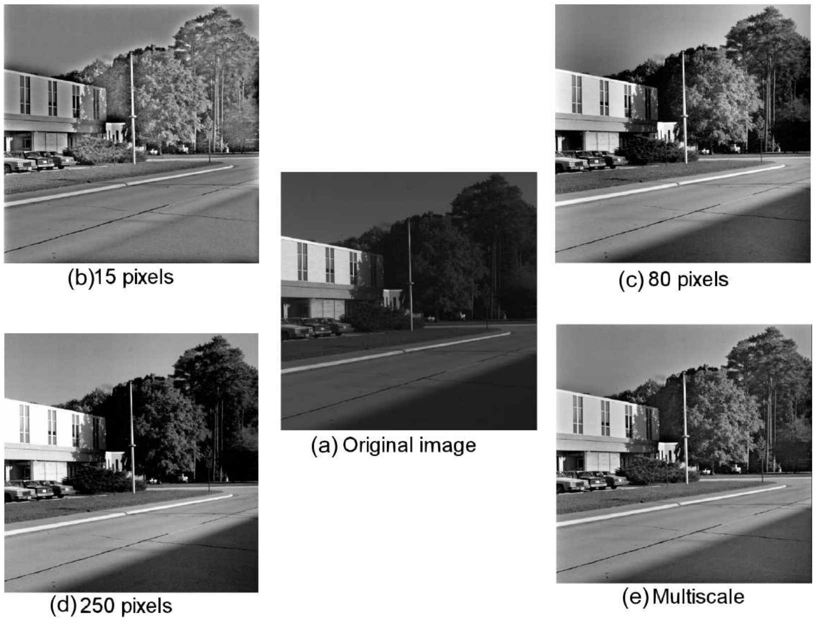

Figure 1. (a) The original input (b) Narrow surround (c) Medium surround (d) Wide surround (e) MSR output. The narrow-surround acts as a high-pass filter, capturing all the fine details in the image but at a severe loss of tonal information. The wide-surround captures all the fine tonal information but at the cost of dynamic range. The medium-surround captures both dynamic range and tonal information. The MSR is the average of the three renditions.

rendition for images with any degree of graying. 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.

We consider the foundations of colorimetry[17] even though it is often considered to be in direct opposition to color constancy models and is 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.[18] The reason for this choice is simply that it serves as a foundation for creating a relative color space and in doing this uses ratios that are less dependent on illuminant spectral distributions than raw spectrophotometry. We compute a color restoration factor, α based on the following transform:

$$\alpha_i(x_1, x_2) = f\left(I_i(x_1, x_2) / \sum_{n=1}^{N} I_n(x_1, x_2)\right), \tag{2}$$

where $\alpha_i(x_1, x_2)$ is the color restoration coefficient in the i^{th} spectral band, N is the number of spectral bands, I_i is the i^{th} spectral band in the input image, and $f(\,)$ is some mapping function. In a purely empirical fashion this was tested on several retinexed images to gain a sense of the visual impact. This proved to restore color rendition, encompassing both saturated and less saturated colors. Adding this to equation 1, the Multiscale Retinex With Color Restoration (MSRCR) is given by:

$$R_i(x_1, x_2) = \alpha_i(x_1, x_2) \sum_{k=1}^{K} W_k \left(\log\ I_i(x_1, x_2) - \log\left[F_k(x_1, x_2) * I_i(x_1, x_2)\right]\right). \tag{3}$$

The results of applying this transformation to the "monochrome" images are shown in Figure 2.

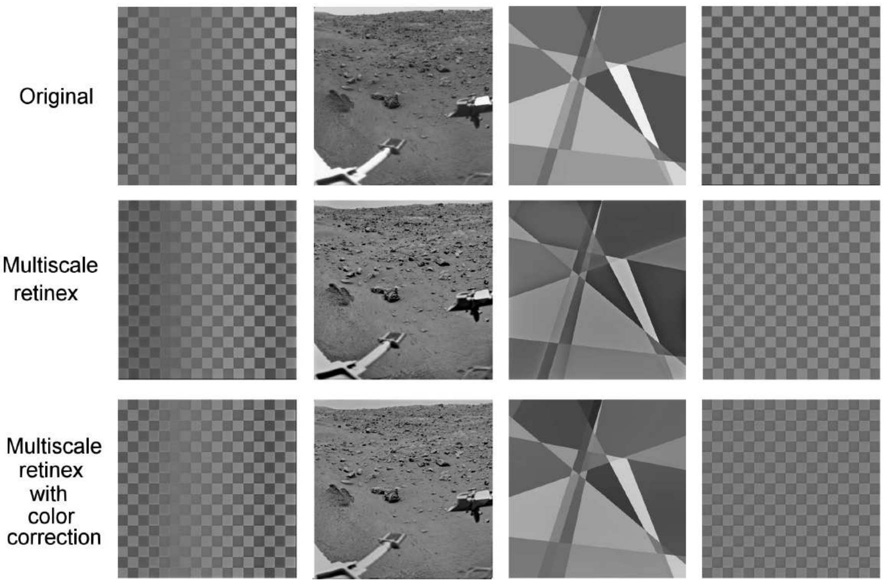

Figure 2. (Top row) Scenes that violate the gray-world assumption; (Middle row) the MSR output; note the graying of large areas of monochromes; (Bottom row) The MSRCR output; note that color constancy is diluted in order to achieve correct tonal rendition.

While we have called this additional color computation a "restoration" we have noticed in retrospect that depending upon the form of $f(\,)$, this can be considered as a spectral analog to the spatial operation of the retinex itself. If we use

$$\alpha_i(x_1, x_2) = \log\left(I_i(x_1, x_2) / \sum_{n=1}^{N} I_n(x_1, x_2)\right),$$

then the internal form of the Retinex process and the color restoration process is essentially the same. This mathematical and philosophical symmetry is intriguing since it suggests that there may be a unifying principle at work. Both computations are contextual in nature and highly relative and nonlinear. We can venture the speculation that the visual representation of wide ranging scenes must be a compressed mesh of contextual relationships even at the stage of 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.

4. THE MSRCR AND DIRECT VIEWING OF SCENES

Our work with the retinex[12, 13] led us away from the world of color and into the world of contrast/lightness perception of complex natural scenes. While the MSRCR synthesizes color constancy, dynamic range compression, and the enhancement of contrast and lightness, the emphasis here is on the latter. We have used the MSRCR on many tens of thousands of images and find that it brings the perception of dark zones in recorded images up in local lightness and contrast to the degree needed to mimic direct scene viewing. Only images with very modest dynamic ranges do not need such enhancement and for these the exposure must be very accurate to achieve a good visual representation. Wide ranging reflectance values in a scene, and certainly, strong lighting variations demand a rather strong enhancement to achieve anything like the visual realism of direct observation. The dynamic range compression of the retinex computation is the basis for the contrast/lightness enhancement. The generic character of the computation is the basis for using it as an automatic enhancement. A few examples of retinex enhancements will serve to convey the degree to which images need to be improved and provide a demonstration that the MSRCR does, in fact, perform this task with considerable agility (Figure 3) and without

Figure 3. Retinex examples to illustrate that the strength of the enhancement matches the degree of visual deficit in the original image. (a) Subtle enhancements

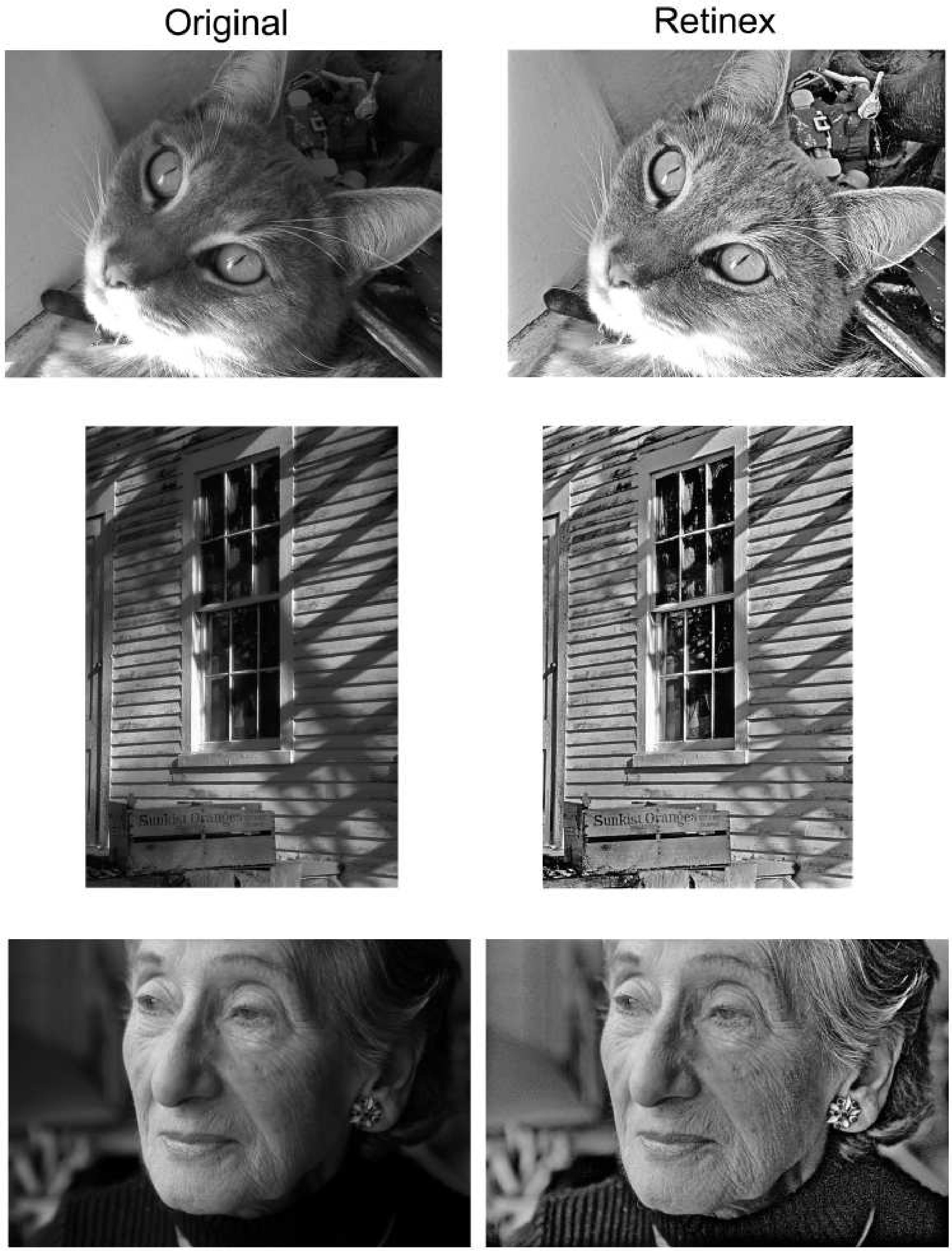

Figure 3: Continued: (b) Moderate enhancements

Figure 3: Continued: (c) Strong enhancements

human intervention. These examples highlight a major facet of retinex performance: intrinsically, the degree of automatic enhancement matches the degree of visual deficit in the original acquired image.

During the course of developing this computation we were led to reexamining some of the most basic ideas about the imaging process and found that some no longer appeared tenable. Assuming that the goal of imaging is to produce a good visual representation that is comparable to the direct observation of the scene—or to provide good visibility for non-visual images such as those formed with thermal Infrared—we had to discard the idea that imaging is a replication process that produces minimal distortion of measured signals or radiometry. Instead, we had to accept the idea that *imaging is a process of profound transformation that intrinsically involves nonlinear spatial processing*. This shift arises entirely from considering the image as a visual entity and the evident visual shortcomings of the linear representation of image data (Figure 4). In general the linear representation is not a good visual representation. This is consistent with the conclusion of a study of the data handling and processing for color negative film scanning.[19] Tuijn describes the correction for all transfer functions so that the image data is linear, and then explains that this is often visually inadequate—weak in contrast and color. In order to explore this further, we displayed known linear data taken with a Nikon D1 camera in linear mode on linearized color computer monitor (gamma correction of 1.6). For a wide array of images, the displayed image is too dark (Fig 4), and the retinex enhancement (also shown for comparison) was required to produce a good visual representation. The linear representation can approach a good visual rendering for a very restricted class of scenes—those with diffuse illumination and restricted ranges of reflectances (or those where white surfaces which can be saturated). Even so, for this cooperative class a substantial degree of contrast stretching (gain/offset) is required to achieve a good visual display/print.

While image data can be quite arbitrary in a statistical sense—the histograms of images vary widely—we observe that the retinexed data were not. As noted in a previous paper,[13] histograms of MSRCR processed images tend toward a characteristic Gaussian-like shape. More recently we have studied regional means (visual lightness) and standard deviations (visual contrast) and found that they tend to converge on consistent global aggregates. This implies that a good visual representation can be associated with well-defined statistical measures for visual quality. In scientific terms, this implies the existence of a canonical visual image as a statistical practical ideal. Such a defined ideal can then serve as the basis for the automatic assessment of visual quality. While this work is still underway, we can show some preliminary results which are encouraging. By following the general idea that the retinex brings regional means and standard deviations up to higher values and that these approach an ideal goal, we have constructed tentative visual measures and performed some testing. The measures were set empirically on a small diverse test image set and then were applied to a broad array of images of all sorts. Figure 5(a) shows a sample of the automatic visual quality assessment by classification into one of three classes—poor, good, excellent. The classification scheme is based upon the map shown in Figure 5(b).

While more study and development is necessary, the early results do support the idea of a canonical visual image with well defined statistical properties. Further, the investigation indicates that the MSRCR is a valuable tool for research purposes—in this case, to define a new statistical measure of visual quality.

5. A HYPOTHETICAL DETERMINISTIC DEFINITION OF VISUAL INFORMATION

While the retinex experience provides new avenues for the study for statistical image processing, it also suggests deterministic pathways as well. The generic character of the retinex computation suggests that some new quantitative definition of visual information may be possible. A deterministic definition would contrast with previous statistical ones based upon information theory.[20, 21] Specifically the MSRCR is approximately performing a log of the ratio of each pixel in each spectral band to both spatial and spectral averages. The suppression of spatial and spectral lighting variations is achieved at the expense of accepting a significant degree of context dependency. Simply put, the MSRCR appears to mimic human perception in producing color and lightness that are influenced by the visual setting in which they occur. The exchange of spatial and spectral lighting dependencies for spatio-spectral context effects appears to be a very basic element of human vision and the MSRCR computation. While we do not have a clear definition of information in a semantic sense, or visual information as some subset of all information, the idea that information is context-relationships is appealing.

Figure 4: Visual inadequacy of the linear representation

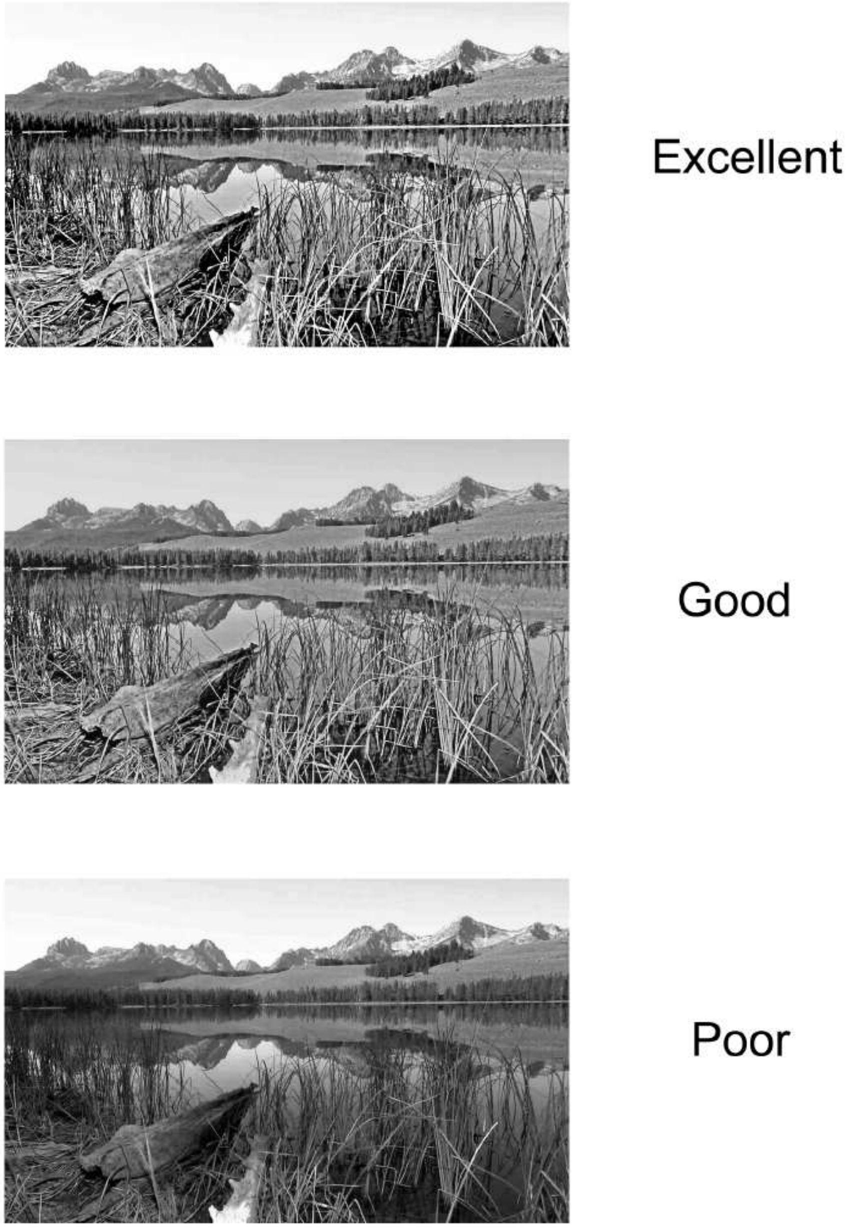

Figure 5: Preliminary performance of Visual Measures for automating visual assessment (a) Global classes

The additional factor of a log function suggests a compactness which may be leading in the direction of symbolic representation—the symbol being the ultimate conciseness and carrier of meaning.

The establishment of context relationships is central to at least the senses of vision and hearing. Music seems to be based upon pitch relationships with certain ratios producing consonance or dissonance in varying degrees. Speech recognition must contend with the difficulties of speaker variations, the interdependencies of phonemes, and all manner of extraneous variations in loudness, temporal rates, degrees of clarity, and the like. For vision, the awesome task of transforming the signals of vision into the sense of vision must succeed in extracting information in the presence of all manner of extraneous variations as well as find some very concise ultimately symbolic representation. Context must be a critical element of vision information as it is in speech and music

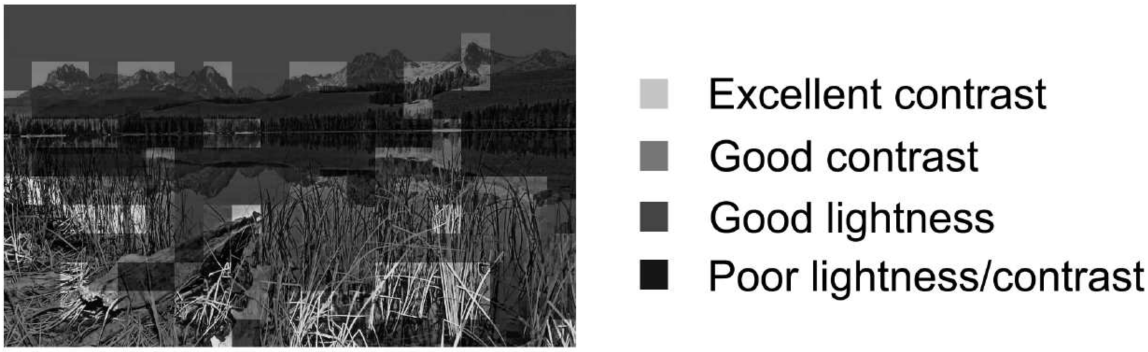

Figure 5: Continued: (b) Visual map showing regional classes

where isolated acoustical events become perceived as a fluid temporal mesh of meaningful words or melody, harmony, and rhythm. Signals are not meaningful in isolation and for vision such contextual relationships as edge connectedness, textural uniformity, and color reflectances differences seem fundamental to building a some sort of "visual information". Perhaps the retinex transformation moves one step in this direction by reducing extraneous variations, increasing spatial and spectral differences, and providing a foundation for a structure of relatedness which with subsequent processing can become symbolic.

6. CONCLUSIONS

The visual image remains an enigma full of surprises, some of which we have encountered in our experiences with retinex image processing. Though we do not understand the intricacies which allow the human vision to encompass very wide dynamic ranges, and provide color constancy, we have developed an approach that seems to mimic these behaviors. Because of this, our thinking about the imaging process has changed in basic ways:

1. Imaging should be considered as a process of transformation rather than replication with minimal distortion.
2. The statistical convergence of MSRCR image enhancements to a histogram which closely matches Gaussian distributions, leads us to postulate the existence of a canonical visual image with consistent statistical aggregate characteristics. Further, these can be used to construct entirely new visual measures which can be the basis for the automatic assessment of visual quality of arbitrary images by the computer.
3. A new deterministic definition of visual information emerges from the computational form of the retinex—namely that visual information is in some sense the log of spatial and spectral context relationships within the image.

A computation like the MSRCR appears to have two very useful properties simultaneously: a diminishment in the dependence of the appearance of the image on extraneous variables such as spatial and spectral lighting, and the construction of compact context relationships. The former is inherently useful because it can lead to better image classifications, and the latter because it shows very clearly that the appearance of a color is dependent not only on the spectral characteristics of a pixel, but also its surround. Together, these properties may be able to provide a basis for bringing more advanced levels of visual intelligence into computing.

Acknowledgments

Dr. Rahman's work was supported with the NASA co-operative agreement NCC-1-01030.

REFERENCES

1. E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci.* **83**, pp. 3078–3080, 1986.
2. E. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Nat. Acad. Sci.* **80**, pp. 5163–5169, 1983.
3. E. Land, "Recent advances in retinex theory," *Vision Research* **26**(1), pp. 7–21, 1986.
4. A. C. Hurlbert, *The Computation of Color*. PhD thesis, Massachusetts Institute of Technology, September 1989.
5. A. C. Hurlbert, "Formal connections between lightness algorithms," *Journal of the Optical Society of America A* **3**, pp. 1684–1693, 1986.
6. A. C. Hurlbert and T. Poggio, "Synthesizing a color algorithm from examples," *Science* **239**, pp. 482–485, 1988.
7. A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," *IEEE Transactions on Neural Networks* **2**, pp. 237–247, March 1991.
8. 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, Morgan Kaufmann, San Mateo, CA, 1991.
9. 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.
10. 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.
11. 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.
12. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. on Image Processing* **6**, pp. 451–462, March 1997.
13. 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* **6**, pp. 965–976, July 1997.
14. Z. Rahman, G. A. Woodell, and D. Jobson, "A comparison of the multiscale retinex with other image enhancement techniques," in *Proceedings of the IS&T 50th Anniversary Conference*, pp. 426–431, IS&T, 1997.
15. Z. Rahman, D. Jobson, and G. A. Woodell, "Resliency of the multiscale retinex image enhancement algorithm," in *Proceedings of the IS&T Sixth color Imaging Conference: Color Science, Systems, and Applications*, pp. 129–134, IS&T, 1998.
16. D. Jobson, Z. Rahman, and G. A. Woodell, "Spatial aspect of color and scientific implications of retinex image processing," in *Visual Information Processing X*, S. K. Park, Z. Rahman, and R. A. Schowengerdt, eds., pp. 117–128, Proc. SPIE 4388, 2001.
17. W. D. Wright, *The Measurement of Colour*, Hilger and Watts, London, second ed., 1958.
18. P. Lennie and M. D. D'Zmura, "Mechanisms of color vision," in *CRC Critical Reviews of Neurobiology*, vol. 3, pp. 333–400, 1988.
19. C. Tuijn, "Scanning color negatives," in *Proceedings of the IS&T Fourth color Imaging Conference: Color Science, Systems, and Applications*, pp. 33–38, IS&T, 1996.
20. F. O. Huck, C. L. Fales, and Z. Rahman, "Information theory of visual communication," *Philosophical Transactions of the Royal Society of London A* **354**, pp. 2193–2248, Oct. 1996.
21. F. O. Huck, C. L. Fales, and Z. Rahman, *Visual Communication: An Information Theory Approach*, Kluwer Academic Publishers, Boston, MA, 1997.