

# The Statistics of Visual Representation

Daniel J. Jobson<sup>\*</sup>, Zia-ur Rahman<sup>‡</sup>, Glenn A. Woodell<sup>\*</sup>

<sup>\*</sup>NASA Langley Research Center, Hampton, Virginia 23681

<sup>‡</sup>College of William & Mary, Williamsburg, Virginia 23187

## ABSTRACT

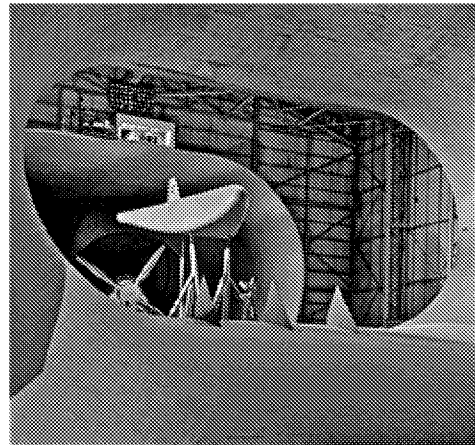
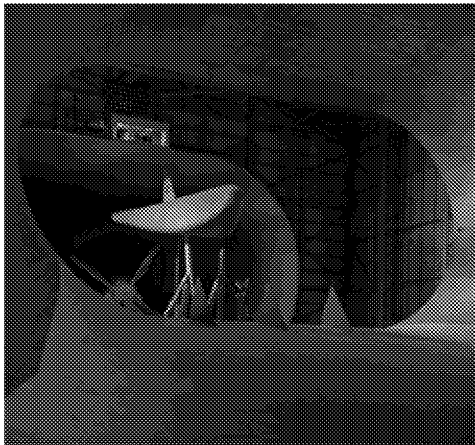
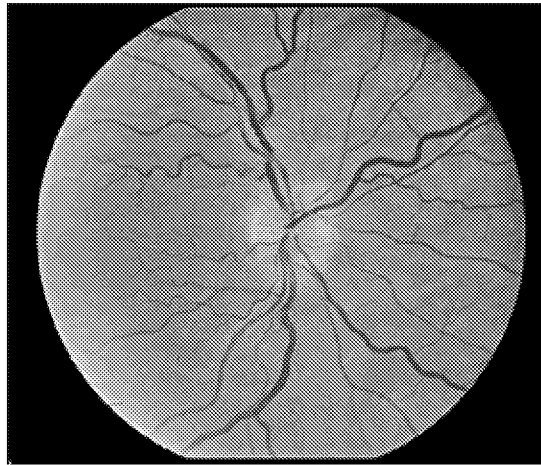
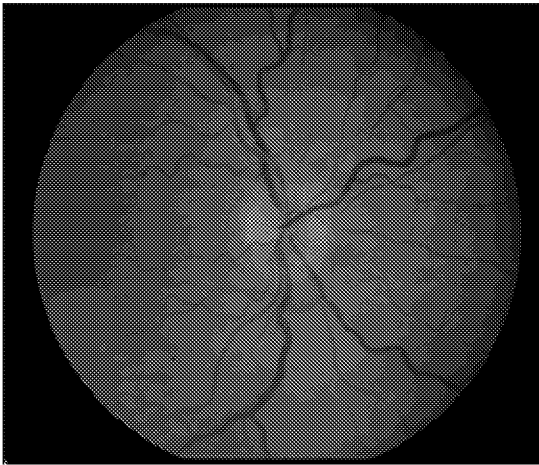
The experience of retinex image processing has prompted us to reconsider fundamental aspects of imaging and image processing. Foremost is the idea that a good visual representation requires a non-linear transformation of the recorded (approximately linear) image data. Further, this transformation appears to converge on a specific distribution. Here we investigate the connection between numerical and visual phenomena. Specifically the questions explored are: (1) Is there a well-defined consistent statistical character associated with good visual representations? (2) Does there exist an ideal visual image? And (3) what are its statistical properties?

## INTRODUCTION

The process of testing, developing, and extensively using the Multiscale Retinex with Color Restoration<sup>1-3</sup> (MSRCR) algorithm for image enhancement has brought forth several fundamental questions about the visual image. The MSRCR is a non-linear spatial and spectral transform that produces images that have a high degree of visual fidelity to the observed scene. In a previous paper,<sup>4,5</sup> we showed that the image of a scene formed using linear representation does not usually provide a good visual representation compared with the direct viewing of the scene. Given that a non-linear transform appears to be essential to the realization of good visual image rendition, we felt a need to further explore the connection between the numerical and the visual representations, i.e. between the numbers that are the digital image, and the visual image that they represent. With the MSRCR we felt we possessed an effective tool for large-scale experimentation and testing on highly diverse images (Figure 1). We asked questions such as: “Is there a statistical ideal visual image?” and “Do all good visual renderings share a convergent statistical character?” These questions, if answered in the affirmative, yield quantitative insights into visual phenomena and lay a general foundation for new definitions of absolute measures of visual quality, which can be used to automatically assess the quality of arbitrary images. Finally, these statistics point to hypotheses concerning the basic mathematical principles of visual representation, which define the general goal of image enhancement in a concise form.

## THE INITIAL HYPOTHESIS AND ITS MODIFICATION

As a starting point, we explored the idea that good visual representations seem to be based upon some combination of high regional visual lightness and contrast. To compute the regional parameters, we divide the image into non-overlapping blocks that are 50×50 pixels. For each block, a mean,  $I$ , and a standard deviation,  $\sigma_f$ , are computed. A first approach was to postulate that for visually good rendition the contrast×lightness product should be above a minimum value, with the additional constraint that each component cannot fall below an absolute minimum value (Figure 2). This regional scale is sufficiently granular to capture the visual sense of regional contrast. Both the contrast and the lightness can be measured in terms of the regional parameters. The overall lightness is measured by the image mean,  $\mu = \bar{I}$ , which is also the ensemble measure for regional lightness. The overall contrast,  $\bar{\sigma}_f$ , is measured by taking the mean of regional standard deviations,  $\sigma_f$ , and it provides a gross measure of the regional contrast variations. The global standard deviation of the image did not relate, except very weakly, to the overall visual sense of contrast. Image frame sizes ranged from 512×512 to 1024×1024 pixels. The coupling of the constraints of minimum contrast-lightness product with minimum contrast and lightness as separate entities defines the zone in Figure 2 labeled “visual good”. Further, this figure suggests that there may exist a contour of much higher contrast-lightness, which can be considered a “visual ideal”.



**Figure 1. Examples of original and optimized images**

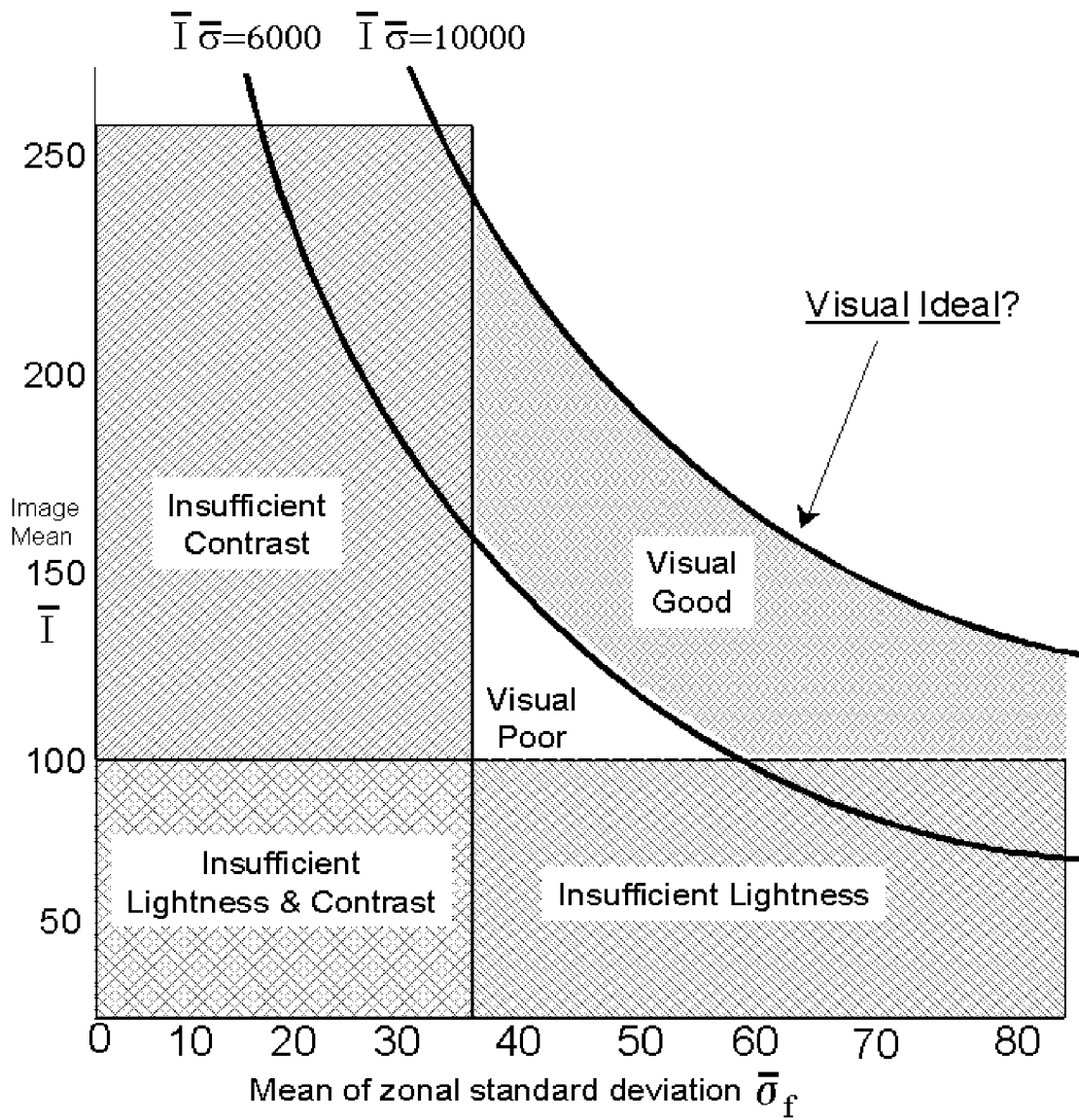


Figure 2. Initial hypothesis; contrast and lightness

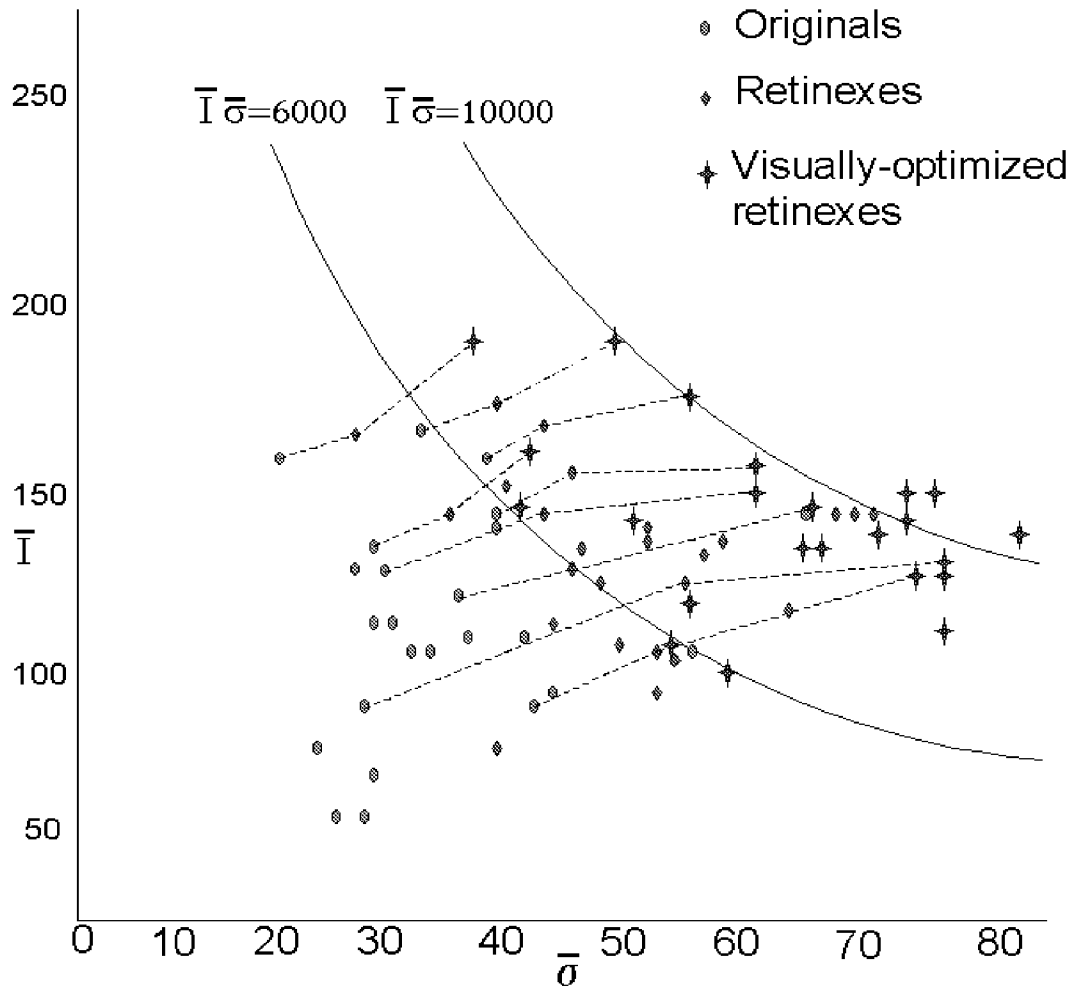
To test this hypothesis, we performed some preliminary experiments. The first was to visually optimize a small sample of images using the MSRCR and any other more conventional processing, such as contrast stretch and sharpening. Even this small data set demonstrates that the initial hypothesis is not entirely satisfactory. Though the data exhibit a trend of clustering about a fairly stable mean with quite variable values for the standard deviation, these did not follow any of the particular contours for minimum contrast-lightness product.

For the second experiment we used a larger number of samples (24 images), but otherwise the experiment was identical to the first. The image samples were selected to be as diverse as possible so that the results would be as general as possible. While the MSRCR performs a visually dramatic transformation in most images, the output image can sometimes be further visually optimized, especially by the application of a sharpening filter. This can be expected as a result of vagaries in pre-MSRCR image pre-conditioning, and the blur introduced in the original images by the optics of the image acquisition devices. While the MSRCR is robust with respect to image pre-conditioning, it cannot be completely immune.<sup>6</sup> The post-MSRCR fine-tuning was confined to modest adjustments in brightness and contrast, and sharpening. The results (Figure 3) are shown in stages to make clear the migration (selected points connected by dotted lines) of the data points from the original image data to the MSRCR values then to the visually optimized final destinations. In general, the primary migration is to higher contrast values with relatively smaller increases in lightness. This confirms and quantifies the visual judgment that most images need contrast improvement to be better visual renderings. The visually optimized outputs do converge to a range of approximately 40-80 for global mean of regional standard deviation and global means of 100-200. Again the data do not follow any specific contours for minimum contrast-lightness, but rather appear to be gravitating to a box (Figure 4). So we revise our initial hypothesis accordingly.

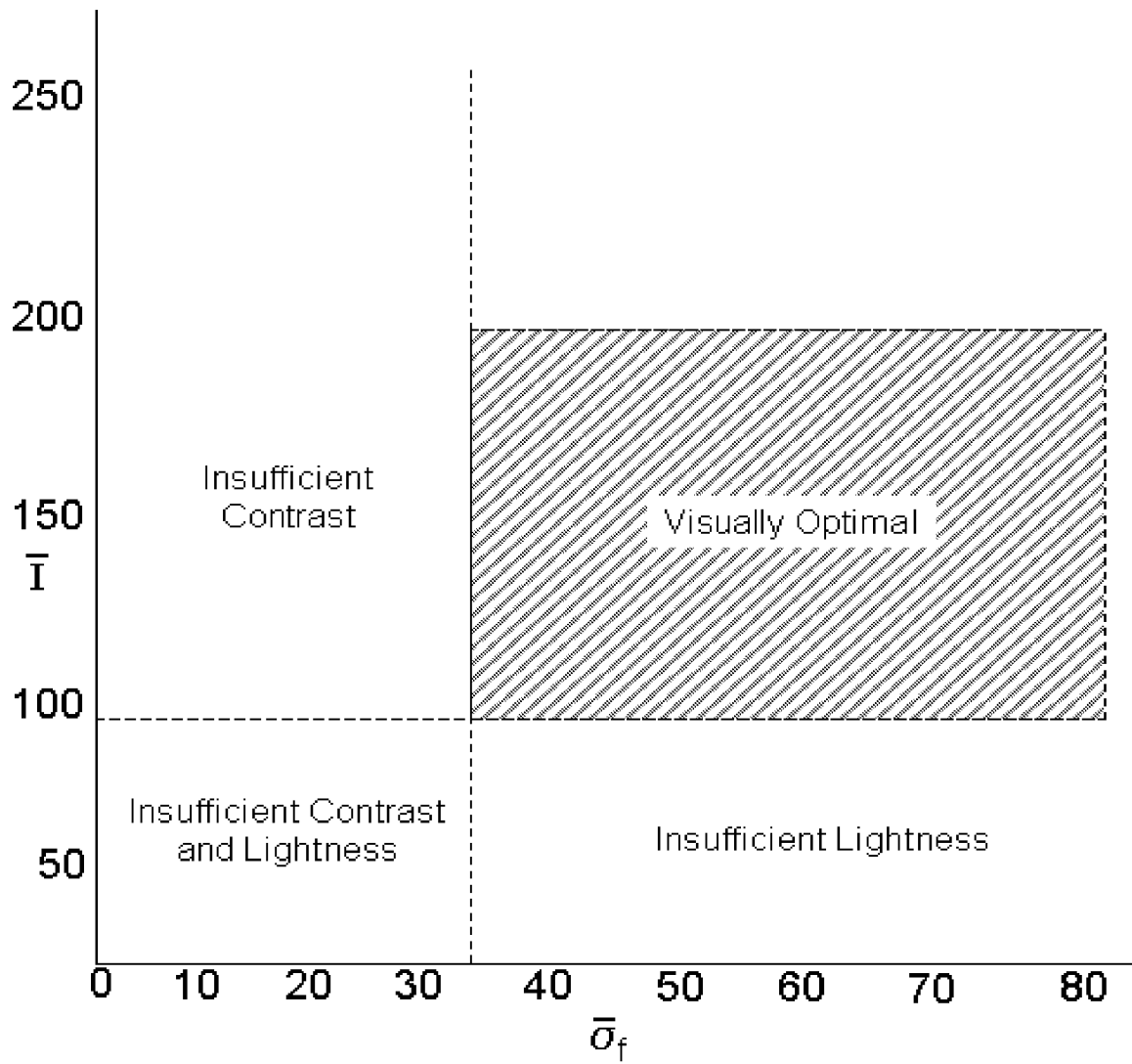
There is a sense of increasing visual quality within this box from left to right. To that extent, we can say that the extreme right edge of the box could be regarded as an ideal, but not an ideal that can be realized by all images. Rather it is an ideal that can be approached by some enhanced images. The extreme left of the box is problematic in the following way. The data points here are associated with feature impoverished images—those with large stretches of uniform space—e.g., a small object against a large blank background. Therefore the placement of the left boundary requires a semantic decision and demands a judgment be made about the minimum amount of feature information that an image must contain in order to be considered to be visually good. At the extreme we can certainly agree that the null image cannot be visually good, so at some point we should be able to say an image is intrinsically bad if there is just too much blank space. For a photographer this would correspond to needing to zoom in and have the subject fill more of the blank space in order to have a satisfactory picture. Perhaps in a more informational sense, blank space in images conveys little semantic information except about the relative size of this nullity. This seems intuitively less informative than the world of features, which conveys information about objects and textures.

The issue of convergence for the visually optimized rendition versus original image data requires more experimentation with an even larger more diverse sample. Results exhibit two primary trends summarized schematically in Figure 5(b). Figure 5(a) (for ~100 images) shows the clustering of actual data points. These data support the idea that the visually optimized representations compared to original data do converge in two senses: (1) mean values cluster and do so reasonably tightly around an average of about 165 whereas original image distributions exhibit mean values that scatter rather more evenly across a wider range, and (2) the frame average of regional standard deviations for the visually optimized images all shift to significantly higher values, but do not necessarily converge to any particular value. Further, these same frame averages do shift above a minimum of about 35. Figure 5(b) summarizes these trends for a still larger set of images (~300). So we conclude that these data support the idea that there are distinctive statistical characteristics for good visual representation and that the distinctiveness is sufficiently strong that it can serve as a partial basis for defining new visual measures that automatically assess visual quality. The partial overlap in the two classes in Figure 5 indicates that these two parameters alone are not completely distinguishing. Overall the visually good representation possesses a mean of ~165 and a frame average standard deviation above 35-40. These large samples support the modified hypothesis of Figure 4. A remnant of the initial contour hypothesis (Figure 2) appears possible in Figure 5(a) and is more definite in the actual plots summarized in Figure 5(b). However this statistical tendency is largely overwhelmed by the confines of the box in Figure 4. At the most it appears to be a secondary effect in the statistics of visual representation.

When images are displayed on monitors, their intensity profile is typically modified using the gamma-transformation given by:  $I_o(x,y) = [I_i(x,y)]^{1/\gamma}$ , where  $I_i(x,y)$  is the input value, and  $I_o(x,y)$  is the modified value. A value of  $\gamma = 1$  is the linear transform. In order to gauge our results against a linear baseline for the original image data, we determined that most digital images are super-linear and should be corrected to approximate linearity by gamma transforming the processed image using  $\gamma = 0.63$ . While this has negligible impact on standard deviation values, it does adjust the mean downward from about 165 to about 128. The implications of this are discussed next.



**Figure 3. Emerging statistical trends in visual optimization**



**Figure 4. Modified hypothesis**

## **IMPLICATIONS: DEFINING UNDERLYING MATHEMATICAL PRINCIPLES**

These data coupled with visual examinations of large numbers of retinex visual optimizations led to the definition of two mathematical principles that are being followed with the observed trends and results. If we view the digital image as a 3-dimensional box (Figure 6), the visually optimal representation appears to jointly satisfy two mathematical conditions for this box. These are:

1. For any spatial scale ranging from near local to near global, visual optimization centers the distributions of regional means near the mid-level of the box (128 for the 8-bit image) and spreads the signal excursions (as quantified by mean of regional standard deviations) out to fill the box as much as possible.
2. Visual optimization spatially minimizes any over- or under-shoots of the box. This is a statement that both clipping to zero and saturation are spatially limited to small zones of infrequent occurrence.

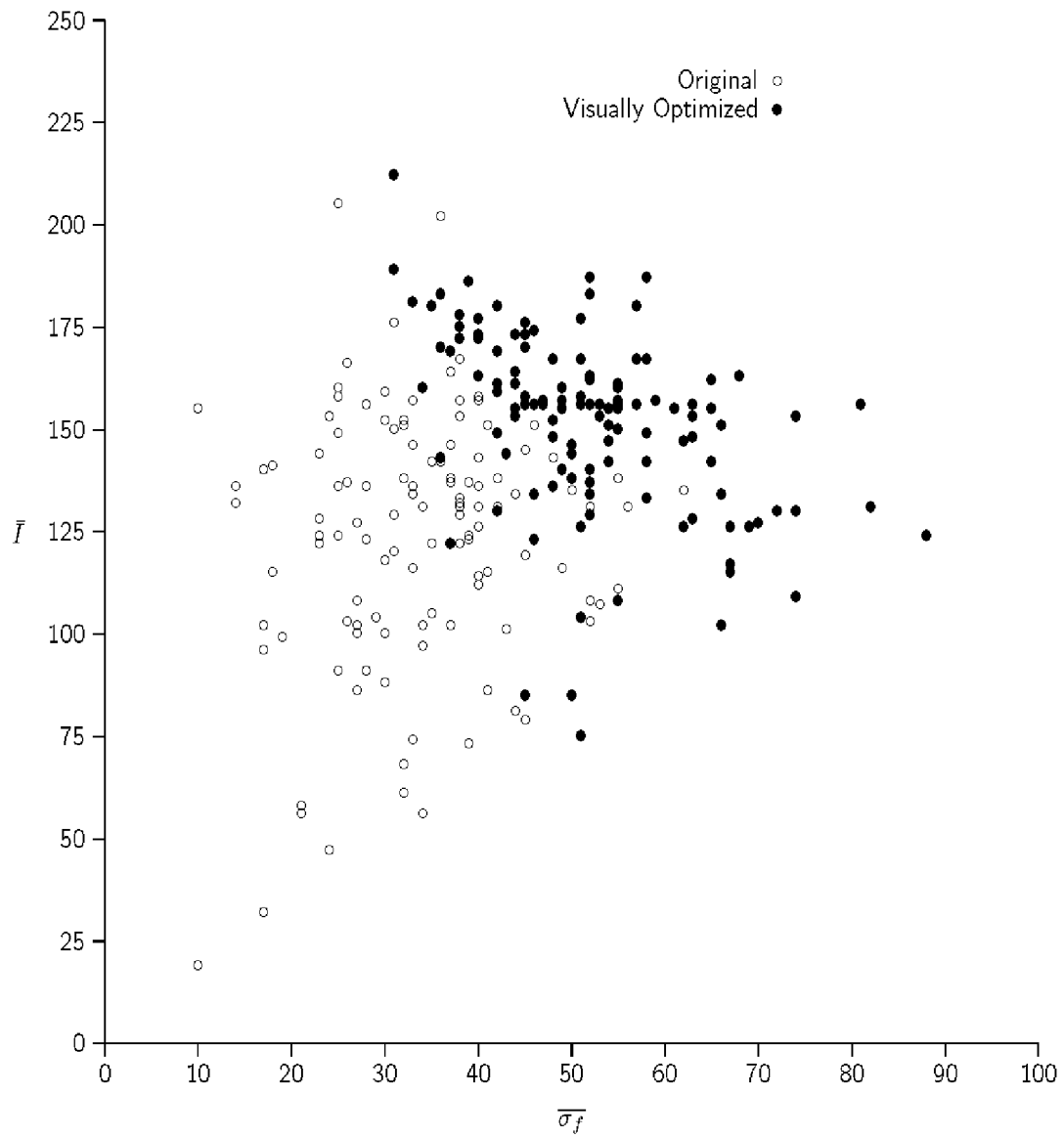
Since the data presented here are all for one spatial scale (the 50×50 pixel region), these two mathematical principles are postulated as working hypotheses concerning the underpinnings of the visual optimization process. Scale changes will not affect the statement about optimization forcing the mean to the midpoint of the dynamic range, but clearly can affect the statement about standard deviation. The statement regarding under- and over-shoots is not scale dependent and appears to be general as long as the original image data prior to optimization is not strongly clipped or saturated over large spatial zones.

These principles suggest that the visually optimal, in more vernacular terms, is centering data on the middle of the box, and spreading the contrast out vertically in the box to the maximum extent possible while minimizing excursions outside the box. This certainly seems to be a pre-conditioning of image signals to most efficiently occupy the box space.

## **DISCUSSION: THE IDEAL VISUAL REPRESENTATION**

The mainstream of the data presented here is associated with optimization to the point of producing a good visual representation. But it is interesting to consider what might be an ideal visual representation. In preceding discussion we noted that feature-impooverished images are debatable cases, and that there seems to be a minimum of feature occurrence necessary just to achieve a “good” visual image. At the opposite extreme, an ideal visual representation fundamentally needs to be feature-rich and the optimization needs to achieve a strong sense of visual contrast approaching that found in the graphics world of illustration. Clearly this is not possible, as already noted, for all images. So only a restricted class of images are even candidates for an optimization that approaches some ideal. In numerical terms, we can see that the mean value of about 165 should not be affected by “good” versus “ideal,” but that the “ideal” will exhibit much larger values of contrast (standard deviation) in the range of 60-90. Images, which can be enhanced to this level, are ones for which there is a high degree of reflectance diversity in the scene, rich feature densities, and successful retinex dynamic range compression for scenes, which have strong lighting variations. An extreme case, which does not have the visual sense of being “ideal” is the printed text image. While text images do have high standard deviations (~90), they do not represent natural scenes, and can be compressed to binary data (not needing 8 or more bits). Further we think that the act of reading is far more of a local raster scanning process than the more global visual sense of comprehending pictures. The visual judgment applied to pictures is therefore not likely to be involved in the reading of text.

The “ideal” should however be associated with a near-perfect sense of clarity and sharp features so sharpness is important component of “ideal” which we do not specifically address here. We did however make frequent use of post-retinex sharpening to reach the visual optimizations. These considerations of “ideal” cannot be related to aesthetics, where often the diffuse or impressionistic or murky are the most beautiful and may be “ideal” in that strictly aesthetic sense.



**Figure 5(a). Large sample of original and optimized images**



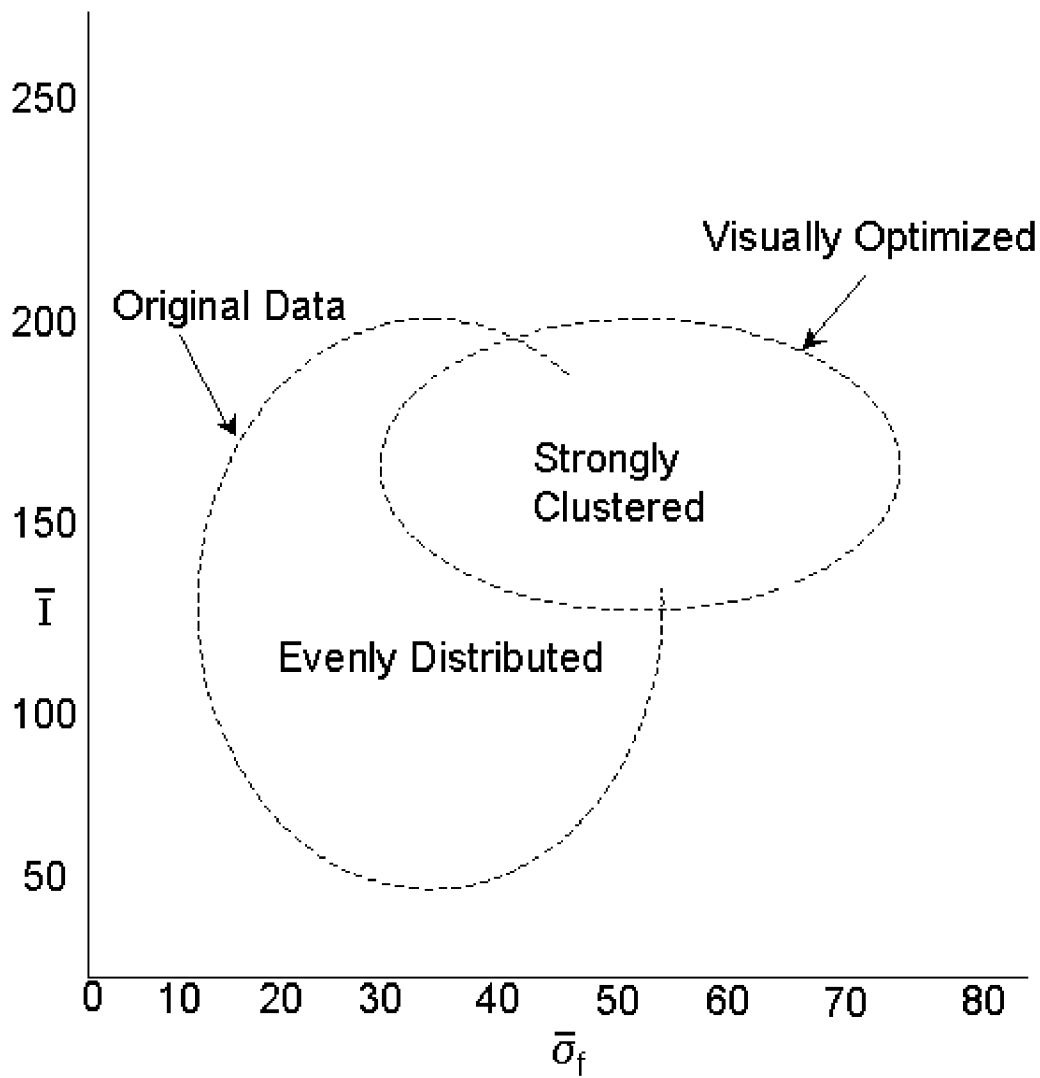
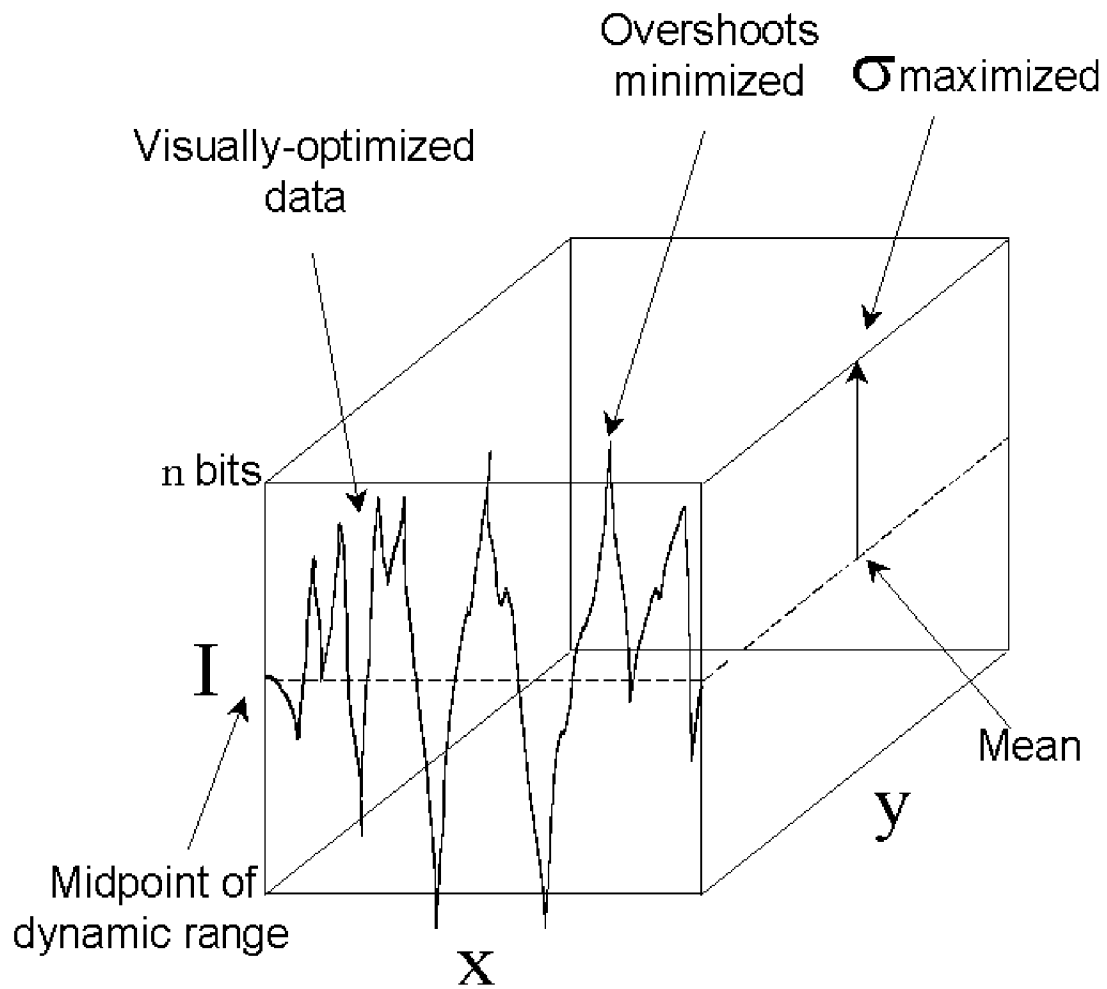


Figure 5(b). Large sample - overall trends in optimization



**Figure 6. Underlying mathematical principles of optimization**

## CONCLUSIONS

Guided by the extensive experience of enhancing images using retinex methods, we find that good visual representations require a non-linear spatial and spectral transform of raw digital image data, which results in consistent statistical trends. These trends provide a new quantitative understanding of the goals of image processing for visual rendition and a partial foundation for constructing visual measures for automatically assessing the quality of visual representation. In general, visually optimized images are more tightly clustered about a single mean value and have much higher standard deviations. Further the results support the idea that visual optimization centers the data mean on the mid-point of the image dynamic range and spreads the signal excursions out across the dynamic range to a maximal extent while at the same time limiting any over- and under-shoots spatially. This overall trend relates to most efficiently occupying the data space with the actual image data. In general visually optimized images are improved in terms of both regional lightness and contrast with the latter being the most strongly affected.

## REFERENCES

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* **6**, pp. 965-976, July 1997.
2. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and Performance of a Center/Surround Retinex," *IEEE Transactions on Image Processing* **6**, pp. 451-462, March 1997.
3. Z. Rahman, D. J. 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, 1997.
4. D. J. 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.
5. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Retinex processing for automatic image enhancement," in *Human Vision and Electronic Imaging VII*, B. E. Rogowitz and T. N. Pappas, eds., Proc. SPIE 4662, 2002.
6. Z. Rahman, D. J. Jobson, and G. A. Woodell, "Resiliency of the Multiscale Retinex Image Enhancement Algorithm," in *Proceedings of the IS&T Sixth Annual Color Conference: Color Science, Systems, and Applications*, pp. 129-134, IS&T, 1998.