Scene Normalization Mechanisms in Humans

John J. McCann
Polaroid Corporation, Cambridge, Massachusetts

Abstract

Using a variety of complex displays these experiments compare human image processing with photographic, hybrid and electronic imaging systems. These experiments measure human responsiveness to changes in averages and maxima, as well as test for evidence of long-distance spatial interactions. The result of this comparison is that imaging systems would have improved color and tone-scale performance if they mimicked human vision.

Introduction

There are two important traditions in modeling human color vision: Colorimetry and Color Constancy. Colorimetry studies the quanta catch at a point that controls color matches; it is inherently a single pixel process. Color Constancy studies color sensations, that is, how colors look —red or green, etc. It is inherently a field process, comparing each pixel with all the others.

When reproducing the colors of a scene, if the color gamuts of the original and the reproduction media are the same, then a point by point Colorimetry calculation can match all the pixels and the reproduction matches the original. When the gamuts, the levels of radiances, or the color of the illuminants are different, then Colorimetry can match only a fraction of the pixels. If one matches all the pixels as closely as one can, then the reproduction does not look like the original. Instead, one must introduce normalization mechanisms to get a good reproduction. This paper discusses experiments studying the nature of human normalization mechanisms. If one can predict how humans will process a scene, then we can make a reproduction system that generates what we see, regardless of the radiances in the scene and gamut limitations.

All imaging systems normalize. Human vision responds to a 10 log unit range of radiances. The radiance from snow on a sunny day at the top of a mountain is 10^{10} greater than the threshold of vision. Despite this remarkable range of response at the receptor level, the neuron signals that transmit vision to the brain have a dynamic range of only 100 to 1.

Photographic film can record images over nearly the entire range of human sensitivity. However, there is a similar limitation in that a photographic print can only display a dynamic range of 30 to 1—limited by surface reflection. Photographic and electronic imaging systems all use a normalization mechanism for both exposure and color balance. With the exception of professional photographers with special equipment, most imaging systems and especially computer imaging systems use a variant of an average of the field of view to normalize the image. It is common thinking that human vision acts just like a camera. The experiments below are in conflict with that assumption.

Tests for a Dependence on Averages

McCann[1] changed the average radiance by introducing highly colored surround papers. He used the most saturated colored papers available, replacing 75% of each paper with the new gray-world-changing paper. The new paper surrounded each Mondrian paper on all sides. The intent was to make a major change of the average radiance and measure a large appearance change. The color shifts were small, on the order of one chip in the Munsell book. Next, he repeated the Color Mondrian experiment, but this time holding the "gray-world" average constant. The original experiment varied the illumination. In this, new surround papers were found that exactly compensated for the illumination shift. Observers matched papers in the Mondrian and reported the same color appearances with the constant "gray-world" averages as they did with different averages. Both sets of experiments show that there is little to no change in appearance caused by major adjustments to average radiances.

Tests for a Dependence on Maxima

Land and McCann[2] described a model for lightness that normalized to the maximum in each waveband. Recent experiments test this hypothesis directly.[3] Experiments using the "destroy the match" technique support the hypothesis that the normalization process for color constancy in humans consists of three independent normalizations to the maxima in each of the long-, middle- and short-wave cone quanta catch. These experiments construct arrays of papers that have the same relative reflectances, but different absolute values for each waveband. The experimenter chose two illuminants such that two different sets of reflectances with two different illuminants sent identical stimuli to the eye. Since both displays were in fact identical, they looked the same. The experimenter then introduced a wide variety of new identical patches to both displays. The introduction of any paper which sent to the eye a higher quanta catch for any type of cone destroyed the match. Papers that sent to the eye less light than those already in the display had no effect on the match between corresponding areas. These results support the theory that the process controlling color constancy normalizes to the maxima in the field of view. Furthermore, they confirm that humans normalize not just to the white, but independently normalize to each waveband. A new, higher radiance white introduces three new maxima, a *yellow, magenta or cyan* introduce two and a *red, green or blue* introduce one new maxima. Regardless, any new maximum changes color appearances of other areas.

Tests for a Dependence on Spatial Parameters

The human normalization process works over very wide angles, but the mechanism is not strictly global. Beginning with a dim square of light on a darker surround, the square appears a light gray. When surrounded by a much brighter area, the square appears black. Separating the brighter area from the square causes the lightness of the square to increase a little. After a separation of 1.25 degrees the square looks nearly the same dark gray, supporting the idea of long-distance spatial interactions.

McCann and Savoy[4] measured the appearance of a patch (3 mL) with and without a much brighter surround (1000 mL). The experiment measured the appearance of the square test spot by matching. The surrounds were varied in circumference, that is 1, 2, 3 and 4 sides. The separations between the test area and the 1000 mL surrounds were 1.25, 2.5, 5.0 and 7.5 degrees. Without the 1000 mL surround the test patch appeared the same as a standard lightness patch of 7.7. Table 1 lists the matching lightnesses for each separation and number of sides in the surround.

Table 1. Matching lightness for a 2.5 degree, 3mL test area on a 2mL background

	1.25 degrees	2.5 degrees	5 degrees	7,.5 degrees
1 Side	5.7	5.6	5.6	5.5
2 Sides	4.5	4.9	5.0	5.1
3 Sides	3.9	4.3	4.2	4.4
4 Sides	3.6	3.9	4.3	4.6
Without 1000 mL				7.7

Figure 1 shows the data from Table 1 in three dimensions. The surface shows a dependence on both separation and the number of sides of surround. The presence of the 1000 mL surround takes the test patch to a maximum lightness of 5.7 with maximum separation and one side. With 4 sides and 1.25 degree separation the minimum lightness of 3.6 was measured. Intermediate surrounds give intermediate results. The total range is 2.1 on scale of 9.0 in lightness.

The above results show two features of the human normalization process. First, the absence of the 1000 mL surround is markedly different from its presence anywhere in the field of view. Without the 1000 mL surround the observers chose lightness 7.7. With it 7.5 degrees away they chose 5.5 with one side and 4.3 with four sides. Second, the influence of the test patch varies with the number of sides and the separation. The test patch separated from the surround by 7.5 degrees is 4.6, while separated by 1.25 degrees it is 3.6.

These experiments demonstrate two essential properties of the human normalization process. First, the normalization process is a long-distance spatial interaction mechanism. Vision is a spatial process, not a single pixel based process. In this case, the appearances of test patches were modified by surrounds 7.5 degrees away. Second, these long-distance interactions are not globally uniform, rather they show a small, but significant, influence of spatial parameters.

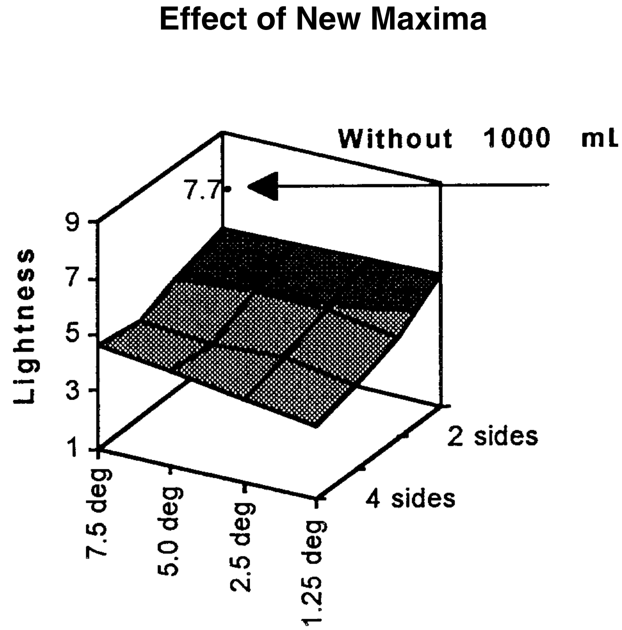

Figure 1. Graph of the data in Table 1

The Role of Scattered Light

The experiments that demonstrate the strongest influence of long-distance neural interactions on the retina are those that use contrast phenomena to change appearance. Introducing a much brighter surround makes a test square look much darker. The much brighter surround will substantially increase the quanta catch of the test square due to scattered light in the eye; in the above experiment the radiance at the center of the square on the retina increased from 3 mL to 30 mL. The paradox is that the much brighter surround makes the test area ten times brighter and darkens its appearance from light gray to black. This demonstrates that the decrease in contrast between objects from scatter is counteracted by an increase in contrast from neural mechanisms.

The two mechanisms, *scattered light*- an optical mechanism—and *long-distance contrast interactions*—a neural mechanism- tend to cancel each other out. The influence of these neural interactions would be much greater if the human eye had no scatter. As well, it can be argued that these neural interactions not only helped the visual system cope with unpredictable illumination, but also generated a better synthetic image of the world, less distorted by scatter in the eye.

Pixel Processing vs. Field Processing

As described at the beginning of the paper, Colorimetry is inherently a pixel based model. Its input is the radiant flux at a pixel; there are no terms for other pixels in the image. With such a limitation single "pixel processes" are limited to proportion scaling. The same scaling has to be applied to all pixels that have the same input value. The problem is, as seen in the above experiments, human vision uses nonlinear, spatially dependent scaling. The ratio-product-reset model [2,5] is a field model that requires as input the radiant flux for all pixels in the field of view. This process is intended to be highly nonlinear and should not be modified to loose that property.[6]

This departure from proportional scaling described above is essential for a model that attempts to mimic human vision. Reflectances in the world are limited by surface properties to a range of roughly 30 to 1. Common real life scenes have been measured to have 1000 to 1 radiance ranges.[7] Humans see a perfectly normal image under these circumstances, despite the fact that the optic nerve has a 100 to 1 dynamic range. A rigid global normalization would maintain the dynamic range of the scene: 1000 to 1 in and 1000 to 1 out. A spatially dependent normalization process has a unique benefit in that it achieves a highly nonlinear dynamic range compression: 100 to 1 in and less than 100 to 1 out.

This can be seen in an analysis of McCann and Savoy data used in Figure 1. Figure 2 shows the values of the calculated radiance on the retina. This process begins with the array of radiance for the entire display and calculates the amount of light scattered from the 1000 mL surround into the center of each test square.

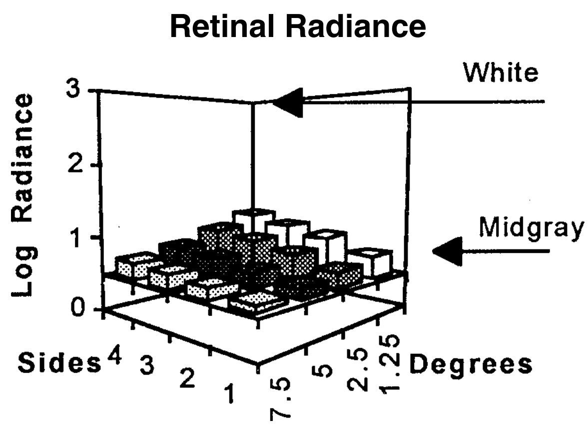

Figure 2. Calculated radiances on the retina with intraocular scatter. The sixteen test areas were described above in Figure 1. The radiance values were calculated by Allan Heff using a two dimensional array of radiances of the entire displays and the technique described by Stiehl, McCann and Savoy.[8]

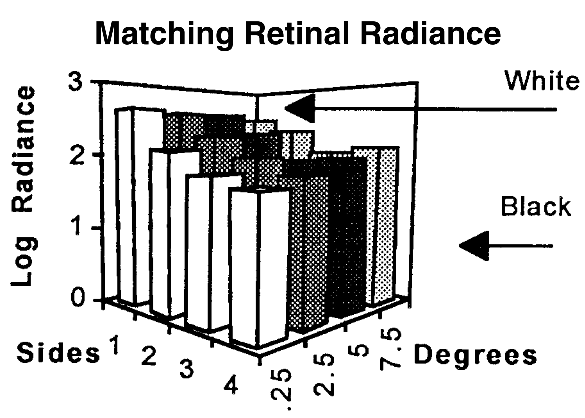

Figure 3. Calculated radiances on the retina with intraocular scatter. The sixteen test areas were chosen by observers to match the areas displayed above in Fig. 2. The radiance values were calculated by Alan Stiehl using a two dimensional array of radiances of the entire displays and the technique described by Stiehl, McCann and Savoy.[8] *The orientation of this plot has been rotated 180 degrees from that of Figure 2 for better visualization of the data.*

Next consider the radiances at the center of the test patches in the standard display that matched each of the McCann and Savoy test patches.

All 16 test areas have a radiance at the display of 3 mL or 0.48 log mL. That value is shown as the baseline in Figure 2. The effect of scatter in the eye is to increase the radiances of all patches. The lowest value of log radiance after scatter is 0.56; the highest value is 1.15. In these displays white (lightness 9.0) is 3.0 and middle gray (lightness 5.0) is 0.68. In this frame of reference, half of the lightness scale is represented by (3.00 - 0.68) 2.32 units log radiance.

The range of log radiances with scatter for the 16 matches in the standard display where each patch is surrounded by white is 1.66 to 2.58. The major benefit of the transformation from Figure 2 to Figure 3 is compression of the dynamic range from white to black. In Figure 3, as the labels show, 2 log units of radiance represent the full range of white to black. In Figure 2 half the range was represented by 2.32 log units radiance.

Observers chose each of the corresponding points as matches. They look identical. That means that somewhere beyond the retina these two very different sets of input data are processed to have the identical output values. Considering the limit in dynamic range of the neurons in the nervous system, it is of considerable advantage to vision to have a normalization mechanism that at the same time reduces the range of radiances sensed by the retina. Again, this cannot be achieved by any form of proportional scaling; this has to be a nonlinear calculation that creates a new synthetic representation from the quanta caught at the retina

A good example of the dynamic nonlinear nature of this process can be seen in the above data. In Figure 2 the lowest value of log radiance after scatter is 0.56; the highest value is 1.15. This is a range of 0.59 log units. The maximum radiance corresponds to 4 sides and the minimum separation. In Figure 3 the range of log radiances after scatter for the 16 matches in the standard display where each patch is surrounded by white is 1.66 to 2.58. This is a range of 0.92 log units. The interesting fact is that now the maximum value corresponds to 1 side and 7.5 degrees separation. The reversal is the familiar phenomenon called simultaneous contrast. Simultaneous contrast has canceled 0.59 log units of scatter and added 0.92 log units of contrast.

The ideal imaging system is one that captures the image as the human retina does. Thus it can accomplish color matching by mimicking the spectral sensitivities of human cones. As long as the scene is perfectly uniform in illumination, the model can stop here.[9] For all other processes beyond painting reproduction, non uniform illumination makes the range of inputs exceed the range of print media outputs. Normalization and dynamic range compression are required. The ideal reproduction system is one that calculates the color appearance and writes those values on film or other media.[7, 10]

Conclusions

All imaging systems have a normalization process. Humans have one that has the following properties:

- Unlike automatic mechanical devices, humans are highly insensitive to the avg. of the field of view.

- Humans are responsive to maxima, in each wave-band.
- Experiments show that humans use long-distance neural interactions across the field of view.
- Experiments show that these long-range interactions are not a linear scaling process, but are spatially dependent.
- Experiments show a combined normalization and dynamic range compression system.
- Scattered light is a significant factor in analyzing experimental data.
- Scattered light and neural interactions tend to cancel each other.

In summary, human vision normalizes to the maxima, but in such a way that it exhibits a dependence on spatial properties of the image. These findings show that the lack of linear, global scaling provides human image processing with the highly desirable feature of dynamic range compression.

References

1. J. J. McCann, Psychophysical Experiments in Search of Adaptation and the Gray World, *IS&T 47th Annual Meeting Proceedings*, **2**, May 1994, pp. 397-400; (see page 14, this publication).
2. E. H. Land and J. J. McCann, Lightness and Retinex Theory, *J. Opt. Soc. Am.*, **61**: 1-11 (1971).
3. McCann, J. J., Rules for color constancy, *Ophthal. Physiol. Opt.*, **12,** 175-177 (1992). J. J. McCann, Color Constancy: Small overall and large local changes, *Proc. SPIE*, **1666**, 1992, pp. 310-321.
4. J. J. McCann and R. L. Savoy, Measurements of light ness: Dependence of the position of a white in the field of view, *Proc. SPIE*, **1453**, 402-411 (1991).
5. J. J. McCann, S. McKee and T. Taylor, Quantitative studies in Retinex theory: A comparison between theoretical predictions and observer responses to "Color Mondrian" experiments, *Vision Res.*, **16**: 445-458 (1976). J. Frankle and J. J. McCann, Method and apparatus of lightness imaging, U. S. Patent 4,384,336 (1983).
6. D. H. Brainard and B. A. Wandell, Analysis of Retinex theory of color vision, *J. Opt. Soc. Am. A* **3**: 1651-1661 (1986). This reference modified the model intent to find a numeric estimation of a highly nonlinear image dependent operation. See J. J. McCann, The role of nonlinear operations in modeling human color sensations, *Proc. SPIE,* **1077**: 355-363, (1989).
7. J. J. McCann, Calculated color sensations applied to image reproduction, *Proc. SPIE*, **901**: 205-214, (1988).
8. W. A. Stiehl, J. J. McCann and R. L. Savoy, "Influence of intraocular scattered light on lightness-scaling experiments, *J. Opt. Soc. Am.* **73**: 1143-1148 (1983).
9. J. J. McCann, Color sensations in complex images, *IS&T/SID Color Imaging Conference Proceedings* **1**, November, 1993, pp.16-23 (see page 56, this publication).
10. J. J. McCann, The application of color vision models to color and tone reproductions, *Proc. Japan Hardcopy '88*, 1988, pp. 196-199.

published previously in the IS&T 1994 Color Imaging Conference Proceedings, page 5