Tone Mapping and Scotopic Image Synthesis

Scott McCloskey

COMP 646 - Computational Perception

November 27, 2003

1 Introduction

In a world with ubiquitous artificial lighting, we typically don't experience dimly-lit scenes for long enough to notice their unique appearance. Because the human visual system does not perform as well under low-lighting, we tend to avoid these situations. As such, most interest in low-light vision is motivated by scientific curiosity; most engineering solutions to problems in low-light resort to artificial methods (e.g. infra-red).

If, however, we hope to reproduce visual experiences accurately then it's necessary to understand the differences between scotopic (low-light) and photopic (bright light) vision. In particular, we'd like to be able to map radiance values from synthetic scenes to display intensities that mimic perception in low light. For real scenes, we'd like to define an image transformation to make digital images match the appearance of the world under low lighting. Both of these require a firm understanding of the human visual system's performance in scotopic conditions, particularly the performance of rod photoreceptors and the associated signal processing.

2 Spatial and Temporal Properties of Human Rod Vision in the Achromat

Since scotopic luminance is too low to generate a response from the cone system, perception is driven by the rods. In order to understand the performance of the human visual system under scotopic conditions, then, we need to understand the limits of the rod system.

In this paper, the authors present their findings about the spatial and temporal performance of rod-mediated vision. In particular, they measure the threshold frequencies (both spacial and temporal) above and below which change can be seen. Within this range, then, they measure the smallest change in frequency that can be discriminated.

One point is key in order to understand this paper in the context of our larger objective: the results in this paper were not obtained under scotopic conditions. The authors have chosen to test both the rod and cone systems under the lighting condition for which each has the best performance. The authors have determined that to be 180 scotopic trolands for the rod system and 2000 for the cone system.

The use of this measure is somewhat confounding. The similarly-named photopic troland is the amount of light incident on the eye's photoreceptors from a 1 $candela/m^2$ source viewed through a pupil of $1mm^2$. The conversion from photopic to scotopic trolands is given by [10], but depends on the light's wavelength. Because the conversion factor varies so widely, and because the authors don't specify the wavelength(s) of light used in their experiment, it's difficult to get a sense of what these numbers mean. As a gross approximation, the authors mention that 180 scotopic trolands is in the mesopic range, roughly 0.001 to 1 $candela/m^2$ [2]. Presumably 2000 scotopic trolands would be in the photopic range.

Normally, under mesopic lighting conditions, visual response is mediated by both the rod and cone systems. The authors are able to isolate the rod system because their subject (the co-author Knut Nordby) is an achromat - he sees with only his rods. Cited tests have indicated that his vision is otherwise normal, so the (implied) assumption is that the results of these tests reflect rod vision in the normal observer.

Their experimental setup is fairly straightforward: they display a sinusoidal grating, as outlined in the first equation in [3, page 389]. Because of the display's limit of 1 cycle/deg, viewing distance was changed in order to simulate higher spatial frequencies. A fixation point was provided, and the direction of variation was chosen to account for the achromat's involuntary (nystagmoid) eye movements.

Figure 1 of [3] shows the contrast sensitivity as a function of spatial frequency for both the normal and achromat viewer. As a side note, it seems that the other author has taken himself as a representative of normal observers, though they never state this explicitly. The sensitivity of the rod system is similar to that of the cone system for low frequencies, up to the rods' peak sensitivity around 0.5 cycles/deg. At higher spatial frequencies, the rod system has decreasing sensitivity until it's acuity limit between 6 and 7 cycles/deg.

Figures 2 and 3 show the minimum difference, expressed as a percentage, in spatial frequencies that can be discriminated. Again, for low frequencies the rod and cone systems show similar performance. Above 0.8 cycles/deg, however, the performance of the rod system degrades rapidly.

Figure 6 shows the contrast sensitivity of the rod and cone systems with respect to temporal frequency. The sensitivity of the rod system is lower then that of the cones for most of the range, and changes above 30 Hz can't be seen at all by the rod system. Figure 8 shows the smallest (percentage) change in temporal frequency that can be discriminated, given a 1 cycle/deg grating. The rod mechanism performs slightly worse then the cones over the range that it operates.

Much of the remainder of the paper is spent considering the implications of these results on the (then) current theory of the perceptual mechanism. Because this isn't immediately relevant to our objective, those implications will not be discussed here.

3 Tone Reproduction for Realistic Images

Classical methods in Computer Graphics such as radiosity and ray tracing have typically been concerned with the determination of the amount of light emitted or reflected from an object in a scene. While they can be extremely accurate in this regard, it is also necessary to create a mapping from the radiance values to some sort of display device. Most often this display device is a CRT monitor, which has a significantly lower range of luminances then were present in the original scene.

The naive solution to the mapping problem is to use the brightest and darkest points in the image as the reference white and black, respectively. As Tumblin and Rushmeier demonstrate in the sidebar in [9], however, this doesn't make sense in extreme lighting. In extremely dark situations, detail is lost due to the lesser response of rod photoreceptors. In extremely bright situations, detail is lost as the cone responses become saturated.

The main point of this paper is the proposal that tone reproduction operators be developed to account for these effects. The objective of such an operator should be to determine values that will drive a given display device to create an image that an observer would judge to match the appearance of the original scene. Given that the authors are primarily concerned with synthesized scenes, the notion of an original scene is somewhat arbitrary.

In figure 2, the authors outline the form of one such operator. Given an image representing the luminance of the scene, we first model the response of a real world observer to get a notion of the appearance to that person. We then perform the inverse operation, taking the appearance and determining what luminance values need be emitted by the CRT in order for the display observer to have the same sensation. Note that we get different luminance values then we started with, as the display and real scene observers will have different states of adaptation. Knowing the luminances needed to get a matching appearance, we determine the display values that will generate them.

This type of operator requires that we know quite a bit. First of all, it assumes that we can determine the real world observer's state of adaptation. More imposing is the requirement that we determine the state of the display observer's adaptation. This is unknowable, given that we have no control over where the user puts his or her monitor, so it is necessary to make some assumptions here. Finally, it assumes that the monitor itself is well characterized.

The authors go on to develop their own tone reproduction operator based on the examples of film and television systems, with an interest toward preserving perceived brightness. For the sake of brevity, I will not address the details of this, as we will see richer operators in [5] and [2].

4 A Model of Visual Adaptation for Realistic Image Synthesis

In this paper, Ferwerda, et. al. develop a tone reproduction operator to account for threshold sensitivity, color sensitivity, and acuity based on luminance adaptation.

They start by outlining the physiological mechanisms that implement visual adaptation. These mechanisms are the dilation and contraction of the pupil, the rod and cone systems, bleaching and regeneration of pigments, and neural adaptation.

- With respect to the pupil, the mechanism is obvious; the range of pupil sizes accounts for about a log unit change in light incident on the photoreceptors.
- The rod and cone photoreceptors allow for adaptation by providing specialized hardware for different ranges of luminance.
- When intensity is high, the rate of regeneration of photopigments allows for adaptation. These pigments are not regenerated as quickly as they would be used, so the sensitivity is reduced.
- When intensities are high, the neural processes of the retina and ganglion adjust to allow for further adaptation. The base level of nerve firing is adjusted to allow for a greater range of response.

The authors go on to cite psychophysical results for acuity, threshold sensitivity, and color sensitivity as a function of luminance. See the paper for these results.

Assuming that we can estimate the adaptation levels of the display and real-world observers, they formalize a method of finding the display luminance that is needed to preserve threshold sensitivity and (indirectly) color sensitivity. In order to simulate the loss of acuity in low-light, a Gaussian filter is designed whose band-pass region preserves only those frequencies that would be visible. We will see a refinement of this notion in [8].

5 A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes

In this paper Larson, Rushmeier and Piatko build on [2] to develop a tone reproduction operator that accounts for acuity, color and threshold sensitivity in a manner that makes better use of the display dynamic range.

In particular, they use the idea that luminance adaptation happens locally in a neighborhood of approximately 1^o of visual angle. This notion has motivated previous work in local tone-mapping operators. One problem with such operators, however, is that their output may not be globally consistent. That is, in an area of constantly decreasing intensity, local operators might result in an image with unnatural inflections.

In order to account for this, the authors develop a model that is simultaneously local and global. They build a histogram of local adaptation luminances, and attempt to find a global mapping from them.

They take, as a starting point, the tone mapping that we'd get from classical histogram equalization - ranges of input luminances that occur frequently are mapped to wider ranges of output luminances. One of the well documented problems with equalization happens when the image is largely constant (imagine the much-ballyhooed "broad side of a barn" that I can't hit with a golf ball). In such images, the dynamic range is stretched too much, resulting in an amplification of noise and an exaggeration of local differences.

In order to account for this tendency, the authors introduce a constraint that caps the slope of the tone mapping function to that of the naive linear mapping. Later on they further restrict the slope to be no greater then the ratio of the threshold detection of the display observer to the same value of the real world observer. They use the threshold detection data from [2].

Beyond the tone-mapping, the authors account for acuity on a local level. Using the local adaptation estimates that were used to generate the starting histogram, the local acuity is determined for each neighborhood. Based on that value, the area is filtered so that more detail is displayed in area of high luminance then in areas of low light.

They develop a similar method for local color reproduction by transforming input RGB images to CIE XYZ space. Lacking a standard transformation, the authors develop a scotopic luminance coordinate that is the best (least squares sense) fit of the color patches of a *Macbeth ColorChecker Chart*^{TM}. They don't explicitly state it, but the fit is presumably based on the CIE Standard Scotopic Observer sensitivity (see [4]).

Based on the local adaptation level, then, the luminance is taken to be the scotopic or photopic luminance, or - if the luminance is in the mesopic range - a linear combination thereof. Color is ramped accordingly.

The key point of this method is that tone, sharpness, and color can be be altered based on local conditions to best use the limited output range. This is possible because, as the authors state, the human visual system is less sensitive to low-frequency then high-frequency changes.

6 A Spatial Post-Processing Algorithm for Images of Night Scenes

The main contribution of this paper is the development of a new (global) filtering scheme intended to better represent the experience of scotopic vision. Based on the experiences of Knut Nordby as outlined in [6], the intent is that the filtering operation should maintain the sharpness of edges.

Because this material was covered in my presentation, this space will be used for further comments on the paper.

- First of all, Nordby's testimony that his is a visual world where things "appear... well-focused" ([6]) should be taken with a grain of salt. What, exactly, does Mr. Nordby have as a reference?
- Secondly, the parameters of their method require manual intervention. The σ_{blur} value used in their filtering scheme could be derived in a principled manner from an estimate of the real world observer's adaptation. This would require some information about the visual angle subtended by the scene which, in the case of a photographic image, could be derived from information about the lens and sensor.

  Unlike the σ_{blur} value, however, the γ_{edge} value (used to preserve edge sharpness) seems arbitrary. The authors ominously state that "a value of 1.25 works well for *a wide range of naturally occurring images*" ([8, page 3], emphasis mine).
- In their examples section, the authors assume that input images have been previously processed in order to reduce brightness, etc. If such processing has already been done then the image can't be in a standard RGB space for which the transformation to scotopic luminance is defined.
- The authors state that a blue hue should be given to scotopic image, but don't justify this. It seems that, if one color needs to be chosen, blue is the best choice - see [1] - but we'd hope for a principled method to determine the magnitude and hue of the perceived color shift.

7 Conclusion

The literature reviewed here, to a large extent, represents the development of tone mapping operators to account for luminance adaptation. There is certainly room for future work, though. The edge-preserving filtering scheme of [8] could be applied locally as in [5]. Other methods, as in [7], could be used to preserve edges while low-pass filtering. Additional psychophysical experiments could lead to a more principled choice of σ for the addition of noise to scotopically-viewed images. Experiments to determine color sensitivities at different adaptation levels might confirm the use of a linear ramp in the mesopic range, or suggest an alternate ramping.

On the personal level, the work in [8] provided enough information for me to get images that are roughly consistent with my memories. While this is sufficient for those of us interested in aesthetics, a significant amount of work is necessary to make these methods acceptable in a scientific sense.

References

[1] S. Buck. *What is the Hue of Rod Vision?* Color Research & Application, 26 Suppl., S57-S59, 2001

[2] J. A. Ferwerda, S. Pattanaik, P. Shirley, D.P. Greenberg. *A Model of Visual Adaptation for Realistic Image Synthesis.* Proc. ACM SIGGRAPH '96, pp. 249-258, 1996.

[3] R. F. Hess, K. Nordby. *Spatial and Temporal Properties of Human Rod Vision in the Achromat.* Journal of Physiology, Vol. 371, Issue 1, pp. 387-406, 1986.

[4] R. W. G. Hunt. *The Reproduction of Colour* England: Fountain Press. 1995.

[5] G. W. Larson, H. Rushmeier, C. Piatko. *A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes.* IEEE Transactions on Visualization and Computer Graphics, Vol. 3, No. 4, pp. 291-306, 1997.

[6] K. Nordby. *Vision in a Complete Achromat: A Personal Account.* http://bpeyes.com/achromat.htm

[7] P. Perona, J. Malik. *Scale-Space and Edge Detection Using Anisotropic Diffusion.* IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 12, No. 7, pp. 629-639, 1990.

[8] W. B. Thompson, P. Shirley, J. A. Ferwerda. *A Spatial Post-Processing Algorithm for Images of Night Scenes.* Journal of Graphics Tools, Vol. 7, No. 1, pp. 1-12, 2002.

[9] J. Tumblin, H. Rushmeier. *Tone Reproduction for Realistic Images.* IEEE Computer Graphics and Applications, Vol. 13, No. 6, pp. 42-48, November 1993.

[10] Photopic to Scotopic Troland Conversion, http://retina.anatomy.upenn.edu/~lance/modelmath/photscottd.html