Two Methods for Display of High Contrast Images

Jack E. Tumblin†   Jessica K. Hodgins†

Brian K. Guenter‡

†College of Computing and Graphics, Visualization and Usability Center

Georgia Institute of Technology Atlanta, GA 30332-0280

[ccsupjt|jkh]@cc.gatech.edu

‡Microsoft Research, Bellevue WA

July 16, 1998

1   Abstract

High contrast images are common in night scenes and other scenes that include dark shadows and bright light sources. These scenes are difficult to display because their contrasts greatly exceed the range of most display devices for images. As a result, the image contrasts are compressed or truncated, obscuring subtle textures and details. Humans view and understand high contrast scenes easily, "adapting" their visual response to avoid compression or truncation with no apparent loss of detail. By imitating some of these visual adaptation processes, we developed two methods for the improved display of high contrast images. The first builds a display image from several layers of lighting and surface properties. Only the lighting layers are compressed, drastically reducing contrast while preserving much of the image detail. This method is practical only for synthetic images where the layers can be retained from the rendering process. The second method interactively adjusts the displayed image to preserve local contrasts in a small "foveal" neighborhood. Unlike the first method, this technique is usable on any image and includes a new tone reproduction operator. Both methods use a sigmoid function for contrast compression. This function has no effect when applied to small signals but compresses large signals to fit within an asymptotic limit. We demonstrate the effectiveness of these approaches by comparing processed and unprocessed images.

2   Introduction

The ultimate goal of realistic image synthesis is to recreate the viewer's sensations of the original scene. This problem is a difficult one because the relationship between scene radiances and
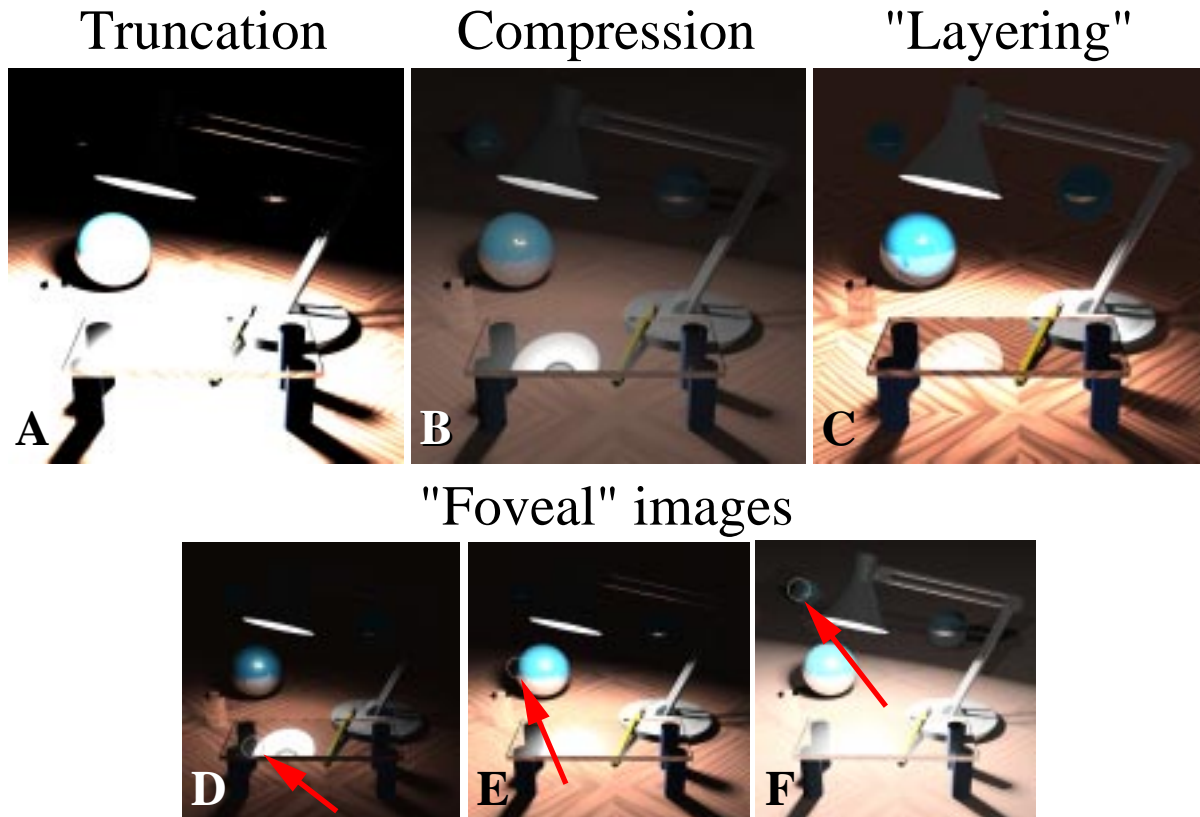
Figure 1: Contrasts are greater than 300,000:1 in the original scene. A) Truncation discards details in image shadows and highlights; B) contrast compression reveals shadows and highlights, but attenuates textures and fine details and lacks the appearance of high contrast; C) "layering" method preserves image details and provides the appearance of high contrast, D–F) the interactive "foveal" method adjusts the entire image for best display of a small circled region specified by the mouse cursor (at the end of the red arrows).

evoked visual sensations is poorly understood. Reproducing scene radiances or contrasts directly is often impossible because the range of the human visual system, from below 10^{-6} to about $10^{+8} cd/m^2$, dwarfs the output range of most displays [14]. Typical cathode-ray tube (CRT) display intensities, for example, are near $50 cd/m^2$, and the ratio between the largest and smallest pixel intensities is typically no more than 100:1.

The ratio between two chosen intensities or luminances is called "contrast" and we use only this narrow definition throughout the paper. Other common quantitative definitions are Weber contrast $\Delta L/L$ and Michelson contrast $(L_{max} - L_{min})/(L_{max} + L_{min})$, and these terms sometimes apply only to adjacent positions in an image or to locations separated by sharp boundaries. Some authors use contrast to name perceived quantities, such as "simultaneous contrast," others use it to describe neural signals, identify neural processes, or as an abstract, qualitative term for any important dissimilarity.

We have also imposed our own definitions for the terms "scene" and "image." A scene is the input to any picture capturing process and an image is the output. Waves crashing on cliffs on

Figure 2: Artist Chris Van Allsburg's drawing of a rocket launch at midnight [35] depicts an extremely high contrast scene in a low contrast image ($< 50 : 1$) without artifacts common to computer graphics such as truncation or "clipping," loss of fine detail, or halos around very dark or bright scene features. Both the streaks in the rocket exhaust (perhaps $10^{+8} cd/m^2$) and the shapes of clouds overhead lit only by starlight (typically $10^{-5} cd/m^2$) are clearly visible. From *THE MYSTERIES OF HARRIS BURDICK.* Copyright ©1984 Chris Van Allsburg. Reprinted by permission of Houghton-Mifflin Company. All rights reserved.

a winter night form a scene, and a painting of it is the image; my father tending a campfire is a scene, and a favorite old photograph of it is the image. A scene is only the purely objective and measurable part of the input; it includes shapes, textures, reflectances, and illuminants, but does not include subjective features such as "warmth." An image is objective and measurable; it is the reflectance of the paint or the radiances of a CRT's display screen.

Synthetic and real-world scenes often contain very high contrasts. For example, a scene with dark shadows, visible light sources, caustics or specular reflections is likely to contain contrasts as large as 100,000:1 or more. As a result, most displays with limited range force compression or truncation of image contrasts and may discard subtle but important textures and details of the original scene.

Limited display contrast is not an insurmountable problem because artists can produce

drawings of high contrast scenes using low contrast materials. They can, for example, convey convincing impressions of a rocket launch at midnight as shown in Figure 2 [35] or the cool shadows under a tree on a hot summer afternoon using only charcoal on paper, a method that provides contrasts of about 50:1. With better display methods, the 100:1 contrast range of CRT displays should be adequate.

The ease with which humans view high contrast scenes suggests that models of visual perception may help solve the problem of displaying high contrast images on a limited contrast display. This paper presents two simple methods inspired by the human visual system. In particular, humans form separate but simultaneous judgments of lighting and surface properties as if the scene were perceived in multiple layers [2]. The lighting layer contains most of the high contrasts while most of the image detail and texture is contained in the layers describing surface properties. The first method, therefore, compresses the lighting layers of an image and leaves the surface properties unchanged. The second method mimics the directional nature of visual adaptation. Because the human visual system adapts preferentially to available light in the direction of gaze, this method adjusts the entire image for best display of a small neighborhood around the viewer's center of attention.

The intent of both methods is to improve the accuracy and realism of displayed images. As shown in Figure 1A and B, using the typical methods of truncation or compression to display a high contrast scene can cause a severe loss of image details and textures. These details can be made visible by both our "layering" and the "foveal" methods, demonstrated in Figure 1C and D–F. The layering method statically reduces lighting contrasts and preserves reflectances; the foveal method dynamically readjusts the display for best reproduction of the contrasts in a small, targeted "foveal" neighborhood selected by the user's cursor movements.

The next section discusses the layering and gaze-directed adaptation of the human visual system in more detail; in Section 4 we review earlier models of adaptation used in computer graphics. Section 5 presents our implementation of the layering method in detail, followed by development of a revised tone reproduction operator in Section 6 used in a gaze-directed interactive foveal display program covered in Section 7. We conclude by analyzing the results of these two methods and discussing possible improvements.

3   Background

Psychology, psychophysics, photography, and computer graphics provide research results that are useful in addressing the problems inherent in displaying high contrast images. In this section we briefly summarize the studies of human visual perception that inspired the models in this paper and justify our assumptions and simplifications.

3.1 Layering Background

Humans see much more in an image than a simple map of intensities. The human visual system is adept at simultaneously extracting three-dimensional shapes, textures, reflectances, and other surface properties from a scene or image. The absolute magnitude of the intensities that define the image have little effect on these perceptions. For example, the surface properties of a child's wooden block are equally comprehensible when viewed in dim indoor lighting or by sunlight, though lighting changes may have increased scene intensities a hundredfold or more.

Based in part on the early computational vision work of Barrow and Tenenbaum [4], psychophysicists such as Arend, Gerbino, and Goldstein proposed that an image or scene is not viewed by the human visual system as a single entity but is decomposed into a stack of overlaid *intrinsic images* each of which describes a perceived scene quantity, such as illumination, reflectance, orientation, or distance [2]. Intrinsic image layers formally portray the ability to estimate multiple scene properties within an image, an ability well supported by examples and experimental evidence. Gilchrist [10], for example, placed a book with a bright red cover on the dashboard of his car on a sunny day. The book added a strong red reflection to his view of green objects seen through the windshield. Instead of a yellow color he saw both the red book reflection and the green objects simultaneously. In experiments by Adelson [1], Arend and Goldstein [2], Henneman [12], and Gilchrist, test subjects reported distinctly different but consistent perceptions of reflectance, illumination, transparency, image intensity, and other properties within viewed scenes. A recent book edited by Gilchrist [8] strongly supports this multidimensional or "layered" view of vision with cogent argument and extensive references.

This paper considers only six types of intrinsic image layers but many scenes contain more than one instance of each layer type. For a simple example, consider a photographic print lit by sunlight leaking through half-closed Venetian window blinds. The photographic print is illuminated by stripes of light and shadow from the blinds, but human viewers can sense this illumination layer without confusion. The reflectances of the photograph vary between zero and one, and these sensed values define a reflectance layer sensed without interference from the illumination. However, a viewer can also interpret the reflectance values as a complete image and decompose the reflectances of the photograph into another set of illumination, reflectance, and transparency values imputed for the photographed scene. The viewer's visual system recursively decomposes the photograph's reflectances into second set of intrinsic image layers.

Recursive decomposition is especially useful for intrinsic image layers of high contrast scenes that hold specular reflectances and transparency information, as in Gilchrist's example of a red book on a car dashboard. For a more complex case, consider a street scene near a modern office building constructed from panes of tinted, partially mirrored glass. Considering only reflectance, illumination, and transparency properties, a human observer may see and separately comprehend at least six intrinsic layers in the scene. Gazing at the glass on a sunny day reveals: 1) the diffuse reflectance of streaks and dirt on the glass surface. 2) the diffuse illumination of the streaks

and dirt by sunlight and light reflected from other buildings to form layer; 3) the tint of the transparency of the glass that forms a reflectance-like layer; 4) a faint building interior scene that illuminates the glass from behind; 5) the specular reflectance of the aluminized coating on the glass; and 6) the tinted mirror-image of the street scene that illuminates the specular coating. But now we have two opportunities for further recursion; both the building interior scene 4), and the mirrored street scene 6) may each be decomposed into another set of up to six layers. If these layers include transparencies or mirror-like specular reflections, more decompositions are possible.

Our visual system also seems to adjust its response as we direct our attention to various intrinsic layers. In the glass building example, the building interior seems dimmer and less distinct when closely inspecting the bright reflection of the street scene behind us, but the street scene's content fades and the building interior seems bright and clear when our attention is aimed inside. This change suggests the visual system may make separate visual adjustments to better assess the contents of each intrinsic image layer.

Several authors have shown that the perception of surface properties and their illuminants are largely independent, thus illumination layers rarely interfere with judgments of reflectance, and reflectance layers almost never disrupt the understanding of illumination, shadows, shapes, or transparency. An experiment by Gilchrist and Jacobsen [11] that is nicely summarized in [10] provides a striking example of this phenomenon. The experimenters arranged two small sealed rooms of equal dimensions and identical furniture layouts. A small aperture in the wall of each room provided a controlled view of the interior, and the room lights were placed outside the field of view. The experimenters painted all surfaces in one room, including the furniture, with a uniformly flat, non-glossy black paint; in the other room they used white paint. Because all surface reflectances in each room were diffuse and identical, any contrasts seen through the room apertures arose entirely from variations or edges in illumination. Gilchrist and Jacobsen adjusted the lamp intensities in each room so that light intensities seen through the viewing apertures were highest for the black-painted room and lowest for the white-painted room. Despite this unnatural ordering of intensities, test subjects who looked through the apertures immediately identified the black- and white-painted rooms, and 22 of 24 test subjects also perceived uniform reflectances within each room. These results hold even for simple radiosity renderings of such rooms as illustrated in Figure 3. Gilchrist and Jacobsen's experiment demonstrates that the visual system is attuned to detecting reflectances reliably and under widely varying illuminations, even without help from different reflectances in the scene. Such broad tolerance for lighting changes when making reflectance judgments suggests that the illumination layer of a viewed image or scene is less important and perhaps is sensed less critically than the reflectance layer.

Professional photographers routinely exploit our tolerance for changes in illuminants to produce better pictures. Most photographers use weak "fill lights" to brighten dark shadows while preserving the shadow boundaries and shadings. The resulting photograph reveals detailed reflectances of surfaces in the shadowed regions, but the added fill illumination is usually
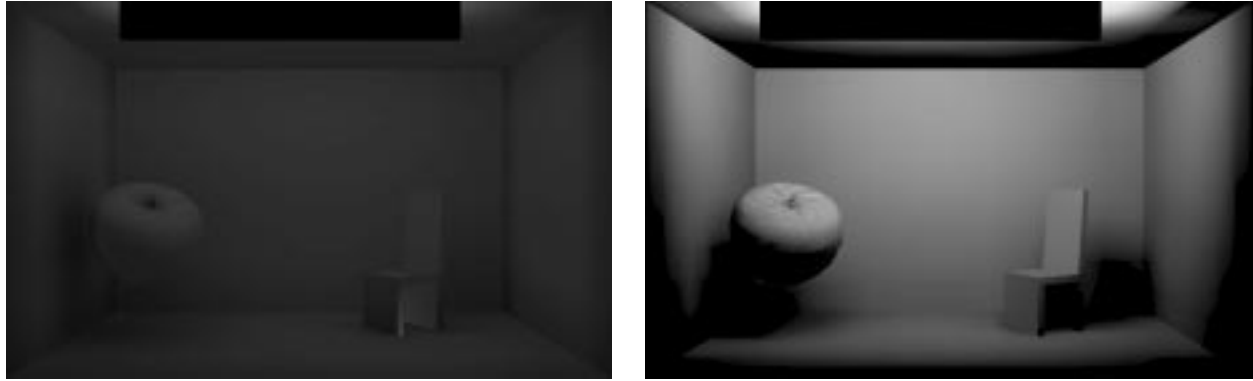
6

Figure 3: Progressive radiosity renderings of two matched rooms with constant reflectance everywhere (0.90 for the room on the left and 0.03 for the room on the right) illustrate that perceptions of reflectance are not easily disrupted by illumination. Despite the high peak pixel intensities in the image at the right due to strong lighting, the dimly lit image at the left retains the appearance of higher reflectance. Images rendered using HELIOS [3].

unnoticed.

Artists also seem to preserve scene reflectances far more diligently and accurately than scene illumination in images such as Figure 2. Assuming the original scene existed, the street surface would have been brilliantly illuminated by the rocket exhaust from the house, but the clouds overhead would have been lit only by starlight. The reflectance of the street was low, perhaps about 0.08, and the reflectance of the clouds was high, perhaps 0.70, but the difference in the strength of their illuminants was astronomical, probably as much as 10^{+7}:1. Van Allsburg's image reveals both the reflectance of the cloud tops and the fine surface texture of the street but assigns them very similar shades of gray; the huge illumination contrast is almost gone in the image, revealed more by context than by image intensities. Spectators at the scene of this house-launching could probably see all the street and cloud details Van Allsburg has drawn, but not simultaneously; they would have to gaze at each of them separately to let their eyes adjust to the huge changes in illumination. But Figure 2 combines all these separately sensed scene details together into one image, as if the scene's illumination contrasts were compressed to create the low contrast image.

Compressing only the illumination layers of an scene works well for low contrast displays because these layers contain most or all of the large scene contrasts. The illumination layers usually consist of smoothly varying light distributions with simple discontinuities at object or shadow boundaries; compressing or reducing them, therefore, is unlikely to obscure any noticeable low contrast scene details. Conversely, the reflectance layers contain most of the fine scene details and textures, and are not difficult to display because their contrasts are always small. Very low and very high diffuse reflectances, such as 0.01–0.04 for brushed black velvet and 0.93–0.97 for clean new snow [13] rarely form contrasts that exceed 100:1. This observation is certainly not new; homomorphic filtering methods used in image processing [21, 31] routinely exploit this property, and it was probably well known to those working to improve photographic film a

century earlier. The low contrasts of reflectance values are especially useful in computer graphics rendering because scene reflectances are usually known at each image pixel.

These experiments and observations lead us to suggest a method for constructing a low contrast image from a high contrast scene. First, split the scene into separate intrinsic image layers of illumination and reflectance values. Leave the reflectance layers unchanged, but compress the illumination layers; if more than one layer exists, equalize them so that none will unduly dominate the final display image, and their aggregate will not exceed the contrast range of the display device. Finally, combine the compressed illumination layers with the original reflectance layers to form a low contrast display image. To find a good method for the illumination compression and equalizing steps we again return to results from psychophysics.

Several experiments support the contention that the visual system has almost no direct sensation of light intensities but instead constructs estimates from delicate sensations of intensity changes. Experiments reported by Ripps and Weale [23] showed that test subjects asked to estimate light intensities routinely made errors of 30% or more. More recent experiments by Schubert and Gilchrist [25] show that human estimates of absolute intensity in a featureless, uniform visual field, or *ganzfeld*, are even less reliable. Test subjects viewed a *ganzfeld* made by controlled illumination of half a ping-pong ball placed over each eye. The illumination intensity changed extremely slowly at $0.045 \log_{10}$ units per minute. Because this rate-of-change was about one-tenth of the slowest perceivable rate measured in humans, their test subjects could report only the direct sensations of absolute intensity and were unable to integrate rate-of-change sensations. A three-fold increase or decrease in viewed intensity was required before test subjects could reliably report the direction of the gradual change. This experimental result suggests that human vision includes a very weak, insensitive response to absolute intensity, but constructs most perceptions of intensity from sensations of change. When researchers separately controlled the intensity of a small patch within the *ganzfeld*, test subjects easily detected the patch when the contrast between the patch and the *ganzfeld* was only a few percent. Test subjects were unable, however, to determine whether the intensity changed in the patch, the surroundings, or both. These experiments strongly support the hypothesis that intrinsic image layers must be constructed from sensations of change (perhaps contrast), rather than from absolute image intensities, and this "relational" approach to vision has strong support among some psychophysicists [8].

Measurements of neural signals supplied by each eye seem to support relational views, but also raise difficult questions about the mental construction of intrinsic image layers. Visual signals leave each eye through the optic nerve bundle, and every signaling fiber in this bundle is an axon (the output stem) of a retinal ganglion cell; these cells form the final neural layer and output of the retina. Each retinal ganglion cell responds only to light falling within a small area of the retina known as its receptive field, and receptive fields of nearby ganglia often overlap. By direct intracellular measurements, physiologists have established that a retinal ganglion cell responds primarily to *changes* in illumination across its receptive field, and its

output approximately encodes the contrast between light in a small "center" region and its surroundings. Each cell responds strongly to either increments or decrements of light in its central region, but not both; increment- and decrement-responding cells are called "ON-center" and "OFF-center" respectively [36]. This approximate encoding of locally measured contrasts indicates that "adaptation," the adjustment of visual sensitivity to fit the available light, is accomplished primarily within the retina.

However, the response of retinal ganglion cells to large local contrasts is bounded by gradual, asymptotic limits. Signals from retinal cells are difficult to measure, but experiments by Sakmann and Creutzfeldt (1969) and others (summarized in [36]) have shown ganglion firing rates in the cat approach a fixed upper limit as local contrasts exceed about 100:1, and their plots of firing rates revealed a family of smooth asymptotic curves. Retinal ganglion cells may directly encode the small contrasts (<100:1) caused by reflectance variations in a viewed scene, but the huge contrasts possible at illumination boundaries must drive both ON-center and OFF-center cells towards their asymptotic limits. Asymptotically limited signals from the eye might reasonably be expected to create asymptotically compressed mental assessments of the large scene contrasts that cause them, even if we do not know the methods or mechanisms used by the visual system to convert retinal signals to intrinsic image layers. A plausible assumption is that this perceived compression of large contrasts is symmetric, favoring neither the ON-center nor the OFF-center signals. The complementary response limits of ON-center and OFF-center cells plotted on log-log axes suggests that a sigmoid or "S-shaped" function can form a reasonable model of such perceptual compression. We will construct a sigmoid function in Section 5, and use this function in both the layering and the foveal display method.

Compressive sigmoid functions are also good descriptions of light reproduction by photographic film. Plotting display intensity versus scene intensity on log-log axes for commonly used film stocks shows a central linear region bounded by asymptotic lower and upper limits known respectively as the "knee" and "shoulder" of the film response. The knee compresses very dim scene features towards displayed black and very bright scene features are compressed towards white by the film's shoulder. This gradual leveling of film response to increasing contrast is more visually appealing than abrupt truncation because it helps preserve interesting details in the image highlights and shadows. Unlike film, the layering method presented in Section 5 applies the compressive sigmoid function only to the illumination layer of an image. Even at the extremes of scene illumination where the compression effect is strongest, the image details in the reflectance layers are unaffected and are still visible in the displayed image.

We suspect compressing the illumination layers is quite similar to what computer graphics animators do when they manually light a scene or write special purpose shaders to achieve a high contrast effect. For example a realistically rendered image of a jointed-arm desk lamp such as the one shown in Figure 1 will have radiances far different from the measurements of a real-world scene. The patterns of illumination and the distributions of light are similar, yet the contrasts are greatly reduced to avoid clipping by the display.

Taken together, these experiments and examples suggest that humans viewing scenes or images are far more aware of the content of the reflectance layers than they are of the absolute intensity in the illumination layers, and that an asymptotic "sigmoid" function is a plausible way to limit larger contrasts within a scene to form a display image. These observations form the basis for our first display method for high contrast images. In this method we capture intrinsic image layers during computer graphics rendering, compress the illumination layers with a sigmoid function, preserve the reflectance layers, and then combine layers to produce a display image as illustrated in Figure 1. In Section 5, we describe a practical implementation of this method and demonstrate its performance on test images.

3.2   Local Adaptation in the Foveal Region

Our second high contrast display method, the "foveal" display program, is inspired by eye movements and how they contribute to what we "see," that is, to our mental impressions of viewed scene content. The human eye is highly directional, adaptable, and nonuniform. Fine image detail and color information are detected almost exclusively in the "fovea," the two- to five-degree wide region of the retina centered at the direction of gaze, and both resolution and color sensing ability drops rapidly away from the center of this region [14]. To compensate for this lack of color and resolution throughout most of the visual field, the human eye makes quick jumps (*saccades*) or uses smooth eye movements (*glissades*) to examine interesting scene features. Somehow the new data gathered from eye movements are seamlessly assembled to form what we see; an inertially stable and uniform impression of the visual field, complete with color and fine detail everywhere. This mental impression is assembled without any conscious effort, and with very little awareness of the underlying eye movements and eye adjustments needed to create it, such as focusing and adaptation.

Adaptation is an ensemble of adjustments made by the human visual system in response to the amount of available light in a viewed scene. These adjustments include variations in pupil diameter, changes in concentrations of photopigment within the receptor cells of the retina, and light-dependent changes of neural processing in the retina and in interpretation by later stages of the visual system. The combined effect of these mechanisms allows us to read the lettering on the brilliant surface of an incandescent light bulb and to walk safely on a path lit only by starlight.

None of these adaptation mechanisms adjust instantly, and some reach equilibrium quite slowly. Photopigment concentrations in retinal rod receptors grow so slowly that most people need at least 45 minutes of darkness to reach the maximum sensitivity of "night vision," yet sensitivity and photopigment concentrations fall rapidly within tens of seconds of exposure to bright light. Pupil diameter changes also occur in seconds, but cannot change retinal illuminance by more than a factor of about 10:1. Other more significant adaptation processes due to neural interactions are generally much faster, most have a rapidly effective onset taking

only tens or hundreds of milliseconds, but some may take seconds to completely run their course. Adjustments for cone cells in the retina, which fill the fovea, are particularly fast; their multiplicative adaptation processes can be complete in as little as 50 milliseconds. See Spillmann and Werner [36] or other textbooks for a good summary of these processes. While temporal effects are important to a complete model of visual adaptation, we will ignore them in this paper for simplicity.

The huge input range of the human visual system is largely the result of adaptation processes. As summarized by Walraven and colleagues [36], several researchers have isolated the response of retinal photoreceptors from adaptation effects by measuring cell responses to very brief flashes of light. Their measurements indicate that without adjustment by adaptation processes, responses vary only in a narrow range of light intensities covering about two factors of ten, or 100:1. The light-sensing elements of many television cameras have a similar input range, and CRT image displays rarely exceed 100:1 contrast. This approximate match between photoreceptor and CRT contrast ranges raises an important question: could the low contrasts of CRT display images somehow convey the appearance of much higher contrast scenes by continually changing the image to mimic adaptation? Such an idea is not far-fetched; auto-exposure video cameras continually adjust sensitivity to match available light, and image sequences with transitions from dark to light surroundings are easy to understand. However, the foveal display program described in this paper attempts to model local and directional adaptation effects more closely.

Adaptation has a strong local character because the human visual system adjusts separately at different locations within a viewed scene or image. These adjustments allow simultaneous sensing of texture and detail in both strongly shadowed and brightly lit regions. As a result, human vision almost never "clips" as a camera or display might. For example, trees silhouetted against a brilliant sunset may appear featureless black when photographed or rendered, but a human viewer will see leaf colors, bark textures, and other fine details of the tree if any of them subtends more than a few degrees of the visual field. Local adaptation allows us to recover the appearance of the tree within the scene.

Local adaptation depends strongly, but not entirely, on the image within the viewer's small, central fovea. For example, looking directly at the surface of an incandescent light bulb causes the remainder of the visual field to temporarily appear darker, indicating that the bright image on the fovea depressed perceived intensities everywhere. However, if the bulb is at least 20-30 degrees away from the direction of gaze, hand movements that reveal or block a view of the bulb have little or no effect on the apparent brightness of the rest of the scene. This foveal dominance of adaptation raises an interesting question; is local adaptation outside the fovea significant or necessary to the assembly of our mental impression of the scene?

For the foveal display program we claim the answer is no. A simple demonstration shows the human visual system can easily comprehend a scene from foveal intensities alone. Peer through a tube of paper rolled tightly enough to restrict your field of view to the fovea, a circle of between 2 and 5 degrees diameter centered at your direction of gaze. Next, get a friend to find an unfamiliar

picture and place it in front of the tube. By aiming the tube at various parts of the image you can easily understand the picture and build a detailed impression of its contents. Because peripheral scene intensities are not necessary for understanding the scene under these circumstances, local adaptation to them is also not required. Of course this demonstration is not conclusive because the paper tube responds instantly to hand movements and provides a fixed peripheral image; the foveal display program has much slower interactive response and the changing peripheral images could disrupt the scene appearance.

We also ignored the periphery for a more pragmatic reason; we do not know how to make display images that include effects of local adaptation without introducing "halo"-like artifacts around very dark or very bright image features, as discussed later in Section 4. Instead, we assume the effects of local adaptation on our mental impression of a scene can be adequately recreated by viewing uniformly processed images created from foveally dominated measurements of the scene.

Local adaptation is particularly useful when viewing high contrast scenes because small neighborhoods tend to be much more homogeneous than the entire image. Neighborhoods that include both shadowed and brilliantly lit features will have high contrast, but these regions are usually only a small fraction of the entire image. The problem of displaying high contrast images is largely a matter of handling these few particularly difficult neighborhoods appropriately.

We have applied these observations in the foveal display program, our second method for displaying high contrast images. The program is interactive; the user indicates a direction of gaze within the displayed image using the mouse cursor and the display program quickly computes and displays a new image best suited to the contrasts in the indicated region. Each new image is an attempt to display what the user's eyes would see in the scene after adapting to the new direction of gaze, and the program relies on the user's visual system to assemble the images into a consistent impression of the high contrast scene.

Because the display cannot reproduce all the original scene contrasts, out-of-range display values are asymptotically compressed towards black or white using the same "sigmoid" function devised for the layering method. We will describe this sigmoid function in Section 5, develop a new tone reproduction operator in Section 6, and finally give the implementation details of the foveal display program in Section 7.

4   Previous Computer Graphics Methods

Local control of sensitivity in the retina helps the human visual system comprehend high contrast scenes, and suggests that a position-dependent scale factor might reduce scene contrasts acceptably for a low contrast display. This approach converts the original scene or real-world intensities, L_w, to the displayed image intensities, L_d, using a position-dependent multiplying term $m(x, y)$:

$$L_d(x, y) = m(x, y) \cdot L_w(x, y). \tag{1}$$

Professional photographers use a similar technique to reduce contrasts in printed images. In this procedure, called "dodging and burning," the photographer moves an opaque hand-held mask to increase or decrease the exposure of the photographic paper around dim or bright portions of the image. However, unless the masks are moved skillfully, the adjacent areas of the image are over- or under-exposed, resulting in a dark or light "halo" effect around high contrast features.

Digital and electronic imitations of dodging and burning have shown similar weaknesses. The method proposed by Chiu *et al.* [7] used low pass filtering, defined by weighted averages of the neighborhood intensities, to construct a smoothly varying scale function that depends on image content. Their approach provides excellent results on smoothly shaded portions of an image; however, any small, bright feature in the image will cause strong attenuation of the neighboring pixels and surround the feature or high contrast edge with a noticeable dark band or halo. We believe the scaling function should change abruptly at the boundaries of high contrast features to avoid the halo effect, but we do not know how to construct a suitable scale function with this behavior.

A later paper by Schlick [24] reported problems with similar halo artifacts. Schlick used a first degree rational polynomial function to map high contrast scene luminances to display system values (e.g. RGB 0-255). This function works well when applied uniformly to each pixel of a high contrast scene, and is especially good for scenes containing strong highlights. Next, he made three attempts to mimic local adaptation by locally varying a mapping function parameter; one method caused halo artifacts, and his tests results indicated that the other two methods were inferior to the uniformly applied mapping function.

However, the uniformly applied function Schlick presents is quite elegant and practical. Users can find all parameters of the mapping function without photometric measurements of the display device, and can compute the mapping quickly because it does not require transcendental functions. The function preserves contrasts for dark image regions and asymptotically compresses image highlights sufficiently to avoid clipping on the display. Schlick's function inspired us to revise our sigmoid function in Section 5 for greater efficiency.

Tanaka and Ohnishi [32] noted that a mild form of halo artifacts have been used in paintings to identify and emphasize the presence of illumination edges. They created a locally varying scale factor from a Gaussian low pass filter to reduce image contrasts, and modeled their filters on the center-surround arrangement of retinal receptive fields. Their locally varying scale factor induces halo artifacts whose amplitude is proportional to local scene contrasts, but they claim the mild halos seen in their example images are desirable. Their method is simpler and faster to apply than that of Chiu et al. because it does not require repeated filtering of the out-of-range image remainders, but as a consequence Tanaka and Ohnishi's method cannot guarantee the output image will match the limited intensity or contrast range of the intended display device.

Jobson*et al.* [17, 22, 16], recently devised a full-color local scaling and contrast reduction method using a multiscale version of Land's "retinex" theory of color vision. Retinex theory estimates scene reflectances from the ratios of scene intensities to their local intensity averages.

Jobson, Rahman, and colleagues also use Gaussian low pass filtering to find local multiplying factors, making their method susceptible to halo artifacts. They divide each point in the image by its low pass filtered value, then take the logarithm of the result to form a reduced-contrast "single-scale retinex." To further reduce halo artifacts they construct a "multiscale retinex" from a weighted sum of three single-scale retinexes, each computed with different sized filter kernels, then apply scaling and offset constants to produce the display image. These and other constants (see table II, pg. 971 of Jobson [16]) give excellent results for the wide variety of 24-bit RGB images used to test their method, but it is unclear whether these robust results will extend to floating-point images whose maximum contrasts can greatly exceed 255:1, such as those used in Figure 1 or in the recent paper by Ward-Larson *et.al* [40].

While the multiscale retinex method does reduce halo artifacts, halo artifacts can grow with the logarithm of the maximum scene contrast, and no combination of weights and filter kernel sizes will eliminate them, as can be demonstrated by applying their method to a grayscale "step" image with value 0.001 on the left half side and 1.0 on the right. All multiscale retinexes will form a bright halo or overshoot on the right side whose width corresponds to the half-width of the largest filter kernel used. Retinexes also distort all scene contrasts by displaying the logarithm of locally scaled scene intensities. While the logarithm provides substantial contrast compression for high contrast scenes, it distorts even the moderate contrasts that could be precisely reproduced on a display device. Nonetheless, their results on example images are impressive and show promise for use where preservation of image detail is more important than perceived contrast fidelity, such as surveillance cameras, or in applications where parameters can be manually adjusted for best subjective effect, such as publications, still photography, or static video cameras.

With the exception of dodging and burning and the methods of Chiu, Jobson, Tanaka, and their colleagues, most imaging systems do not imitate local adaptation. Instead, almost all image synthesis, recording, and display processes use an implicit normalizing step to map the original scene intensities to the available display intensities without disturbing any scene contrasts that fall within the range of the display device. This normalizing consists of a single constant multiplier m:

$$L_d(x, y) = m \cdot L_w(x, y). \tag{2}$$

The multiplier is often ignored or explained as an imitation of global visual adaptation, but the exact value of m is the combined effect of several unrecorded adjustments to imaging equipment. For example, a film camera records scene intensities scaled by the lens aperture, exposure time, and film speed. A slide projector's displayed images are scaled by the strength of its light source. Computer graphic images are created and stored in RGB units made from normalized scene intensities, and gamma-corrected cathode-ray tube displays create intensities proportional to RGB units.

Image normalizing has two important properties; it preserves all reproducible scene contrasts and it discards the intensities of the original scene or image. Contrast, the ratio of any two intensities, is not changed if both intensities are scaled by the same multiplier. Normalizing

implicitly assumes that scaling does not change the appearance, as if all the perceptually important information were carried by the contrasts alone, but scaling display intensities can strongly affect a viewer's estimates of scene contrasts and intensities. While this scaling is not harmful for many well-lit images or scenes, discarding the original intensities can make two scenes with different illumination levels appear identical. Normalizing also fails to capture dramatic appearance changes at the extremes of lighting, such as gradual loss of color vision, changes in acuity, and changes in contrast sensitivity.

Tumblin and Rushmeier [34] tried to capture some of these light-dependent changes in appearance by describing a "tone reproduction operator," which was built from models of human vision, to convert scene intensities to display intensities. They offered an example operator based on the suprathreshold brightness measurements made by Stevens and Stevens [30, 28] who claimed that an elegant power-law relation exists between luminance L, adaptation luminance L_a, and perceived brightness B:

$$B = C \cdot \left(\frac{L}{L_a}\right)^{\gamma}. \tag{3}$$

These measurements, however, were gathered using "magnitude estimation," an experimental method that has not found universal acceptance among psychophysicists because results can vary strongly with context, because test subjects exhibit learning effects that make repeatable measurements difficult, and because these variations are not adequately explained by basic sensory processes [18]. More conventional methods measure only the detection thresholds for simple scene features. S. S. Stevens [29] argued that thresholds, though measured more reliably, are poor indicators of human response to large-scale or suprathreshold signals because measured thresholds depend on both the sensitivity and the background noise in neural processes. Insensitive visual mechanisms will have high measured thresholds, but high thresholds do not necessarily indicate low sensitivity. Very sensitive mechanisms may also have high thresholds if their response to weak signals must overcome strong background noise before detection. Stevens warned against misinterpreting threshold measurements as the inverse slope of human response curves (e.g. 1/threshold as "sensitivity") and vigorously objected to the practice of integrating threshold measurements to construct large-scale sensory response curves [29]. He attempted to measure directly the complete range of human vision, and the resulting power-law relation agrees reasonably well with the narrower results from more conventional threshold-finding experiments such as those by Blackwell [5]. More generally, extending threshold measurements to estimate suprathreshold performance is inadvisable for any nonlinear system such as the human vision. Nonlinear systems are not well described by the simple measures of sensitivity and impulse response that suffice for linear systems because no transfer function exists.

Tumblin and Rushmeier's tone reproduction operator used the results of Stevens and Stevens but exhibited several serious shortcomings, as shown in Figure 4. Images or scenes that approach total darkness processed with their method are displayed as anomalous middle gray images instead of black, and display contrasts for very bright images ($> 100 cd/m^2$) are unrealistically

exaggerated. Their method did not address the contrast limitations of displays and was presented in an awkward form that discouraged its use. In Section 6 we reformulate this method using less cumbersome notation and modify the operator to eliminate the anomalies with very dim and very bright images. The foveal display program described in Section 7 uses this revised tone reproduction operator.

Soon afterwards Ward [38] presented a much simpler approach to appearance modeling that also provided a better way to make dark scenes appear dark and bright scenes appear bright on the display. Ward observed that normalizing usually results in the appearance of moderate interior lighting when used to display any image computed using global illumination methods, regardless of the intensities of the original scene. He proposed using a light-dependent multiplying factor m to restore the appearances of different lighting conditions. The factor was built using contrast visibility data from Blackwell [33], data which showed that the smallest noticeable increase in luminance or "contrast threshold" of a small target on a uniform background grows nonlinearly as the amount of surrounding light increases. Ward chose his scale factor to match thresholds of the display to those of the original scene:

$$L_d = m \cdot L_w \tag{4}$$

where

L_d is the display luminance in cd/m^2,
L_w is the original scene or world luminance in cd/m^2, and
$m = \left[\frac{1.219 + L_{da}^{0.4}}{1.219 + L_{wa}^{0.4}} \right]^{2.5}$,

where

L_{da} is the display adaptation luminance, a mid-range display value
L_{wa} is the adaptation luminance for the real-world or scene, usually
$\log(L_{wa}) = mean\{\log(L_w)\}$.

Because Ward's method scaled image intensities by a constant factor m, it did not change scene contrasts for display. Although his method provided visually pleasing results on many images, some published night scenes computed with his method seem to show lowered contrast [38]. This contrast reduction may be due to the loss of linearity commonly found at the smallest output values of many displays. The lower bounds on the display luminance L_d shown in Figure 4 are usually set by light from the display surroundings and cause all scene intensities below about $10^{-2} cd/m^2$ to appear as featureless black on the display. Figure 4 also shows that the scale factor m maps all adaptation luminance values L_{wa} above about $100 cd/m^2$ to almost the same display value L_d. Such choices for m effectively normalize scene luminances; boosting
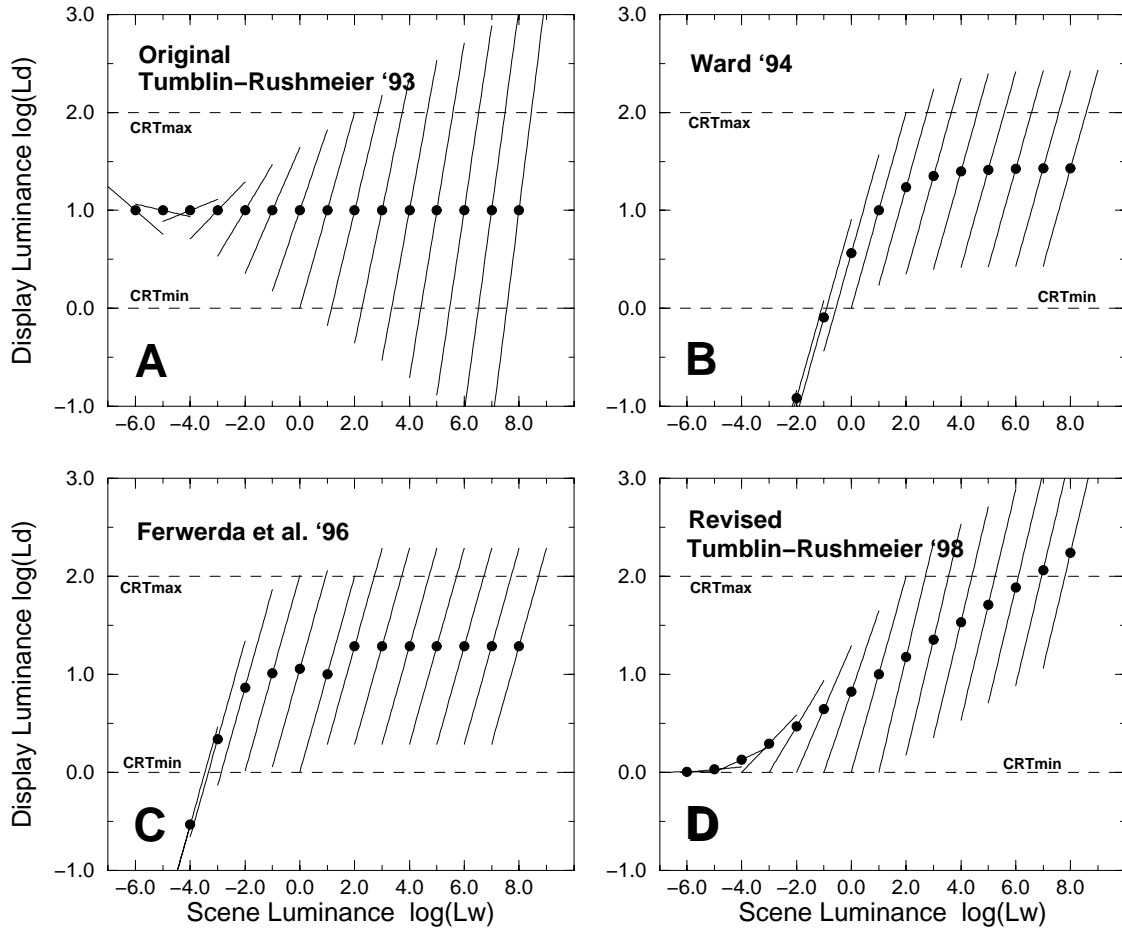
16

Figure 4: Log-log plots show input-to-output mappings of the four tone reproduction operators discussed in Sections 4 and 6. Each plotted line shows display luminance L_d versus scene luminance L_w for a one adaptation value L_{wa}. Adaptation values cover the entire range of human vision in factor-of-ten steps from 10^{-6} to $10^{+8} cd/m^2$ and are marked by a dot where $L_w = L_{wa}$. Each curve shows L_d computed from L_w values between $0.1L_{wa}$ and $10.0L_{wa}$. Output luminances of most CRT displays fall within the 1 and $100 cd/m^2$ limits marked by dashed lines. A) Tumblin and Rushmeier's operator [34] uses Equation 3 and reduces the slope of each curve for smaller L_{wa} to reduce displayed contrasts. Extremely dark scenes are displayed as medium gray with reversed contrasts, and extremely bright scenes exaggerate contrasts unrealistically. B) Ward's tone operator [38] of Equation 4 never causes contrast reversals and always maps dark scenes to dark display images, but maps to black all scene luminances L_w below about $0.01 cd/m^2$, and almost normalizes scenes with $L_{wa} > 100 cd/m^2$. C) Ferwerda et al. [9] extended the dark response of Ward's method, but display luminance is not a monotonically increasing function of L_{wa} near $1 cd/m^2$. D) The revised Tumblin-Rushmeier operator of Equation 17 reduces displayed contrasts for very dark scenes, preventing contrast reversals and exaggerations. Increasing L_{wa} values map to monotonically increasing display intensities to better map dark scenes to dark displays and bright scenes to bright displays.

the illumination intensities in a bright scene by a factor of 10 will produce nearly identical display images.

Ferwerda and colleagues later offered an extended appearance model for adaptation that successfully captured several of its most important visual effects [9]. By modeling the gradual transition from cone-mediated daylight vision to rod-mediated night vision, their method depicted reduced luminance sensitivity, color sensitivity, and spatial resolution with decreased light. Like Ward, they converted original scene or image intensities L_w to display intensities L_d with a multiplicative scale factor m, but they determined their m values from a smooth blending of increment threshold data for both rods and cones in the retina, as shown in Figure 4. Their inclusion of threshold data for rod-mediated vision extended the usable range of their operator down to about $10^{-4} cd/m^2$, which is much closer to the absolute threshold of vision. They included both a spatial-filtering step and a color-controlling step to simulate the reduced acuity and loss of color sensitivity of night vision. They also provided a simple method to mimic the time course of adaptation for both dark-to-light and light-to-dark transitions. As with Ward's method, their m choices act chiefly as a normalizer for all scenes with L_{da} above about $100 cd/m^2$, and does not modify image contrasts for display, though the Gaussian filter used in the resolution-controlling step will attenuate small high contrast features in the image.

More recently Ward and colleagues published a new and impressively comprehensive tone reproduction operator based on iterative histogram adjustment and spatial filtering processes. Their operator reduces high scene contrasts to match display abilities, and also ensures that contrasts that exceed human visibility thresholds in the scene will remain visible on the display. They model some foveally dominated local adaptation effects, yet completely avoid halo artifacts or other forms of local gradient reversals, and include new locally adapted models of glare, color sensitivity, and acuity similar to those used by Ferwerda *et al.* [9]. Their example images are quite beautiful and convincing, and their method appears straightforward to implement.

However, the underlying method of histogram adjustment is troublesome for three reasons. First, the method has no position dependence; a pixel at the center of the image is equally affected by intensities of distant and nearby pixels. Second, the method enforces a monotonically increasing mapping from scene intensity to display intensity. Artistic renderings such as Figure 2 routinely violate this restriction because differently illuminated regions of the image, such as the sky, the house, and the street, are rendered using overlapping intensity ranges to achieve greater display contrast. Third, the histogram adjustment method can occasionally reduce large scene contrasts inconsistently. Spans of scene intensities held by large numbers of pixels are probably the most important parts of the scene, and are rightfully assigned larger portions of the display range. However, contrasts with more unusual scene pixels can be distorted. For example, choose two scene pixels that form a large contrast and appear in an empty or sparsely populated region of the scene's histogram. In the displayed image, the contrast between this pixel pair may be reduced to nearly the threshold of visibility. Now choose another pair of scene pixels whose contrast is one third that of the first pair and are located in a densely populated region

of the scene's histogram. In the displayed image the contrast of this pixel pair may be nearly unchanged, leading to a curious reversal; the small scene contrast would be displayed as much larger than the large scene contrast.

A few other computer graphics researchers have modeled the appearance of extremely bright, high contrast scene features by adding halos, streaks, and blooming effects to create the appearance of intensities well beyond the abilities of the display. Nakamae *et al.* [20] proposed that the star-like streaks seen around bright lights at night are partly due to diffraction by eyelashes and pupils, and they presented a method to calculate these streaks in RGB units, implicitly normalizing them for display. Later Spencer, Shirley, and others [27] presented an extensive summary of the optical causes and visual effects of glare and modeled their appearance by using several adjustable low pass filters on the intensities of the original scene. Small, extremely bright light sources that cover only a few pixels, such as street lights at night or the sun leaking through a thicket of trees, are expanded into large, faintly colored, glare-like image features that have a convincing and realistic appearance.

Despite progress in modeling the light-dependent changes in appearance that occur over the entire range of human vision, few methods offer the substantial contrast reduction needed to display these images without truncation or halo artifacts. The method of Tumblin and Rushmeier reduces display contrasts somewhat to match the eye's lowered contrast sensitivity in night vision, but their method also increases contrasts for scenes brighter than the display, thereby making truncation problems more severe. Ward's first method does not change image contrasts, nor does the method of Ferwerda and colleagues except as a byproduct of their acuity-limiting Gaussian filter. Linear filters used by Nakamae to model diffraction and by Spencer to model intra-ocular scattering may incidentally reduce contrasts of small features, but as noted by Spencer and colleagues, a need remains for a perceptually valid method to reduce scene contrasts.

Sections 5 and 7 present implementation details of two simple display methods for high contrast images that both use a simple sigmoid function to asymptotically limit display contrasts. In the "layering" method of Section 5 the sigmoid function reduces contrasts of only the illumination layers for display. The interactive "foveal" method described in Section 7 uses the revised tone reproduction operator presented in Section 6 to map scene intensities to display intensities, then applies the sigmoid function to reassign out-of-range pixel intensities to the available display limits.

5   Implementation and Evaluation of Layering

As discussed in Section 3.1, the human visual system extracts several simultaneous perceptions from an image. The layering method of contrast reduction relies on this ability to perceive multiple illuminants and surface properties separately and on the natural tolerance for changes in illumination routinely exploited by photographers. Our layering method uses standard computer graphics rendering methods to create separate layer images representing high contrast scenes.

Deep shadow:
0.4 cd/m^2 –>(50, 30, 23)

Lightbulb:
175,000 cd/m^2 –>(255, 255, 255)

Shroud reflection:
40,000 cd/m^2 –>(240, 240, 240)
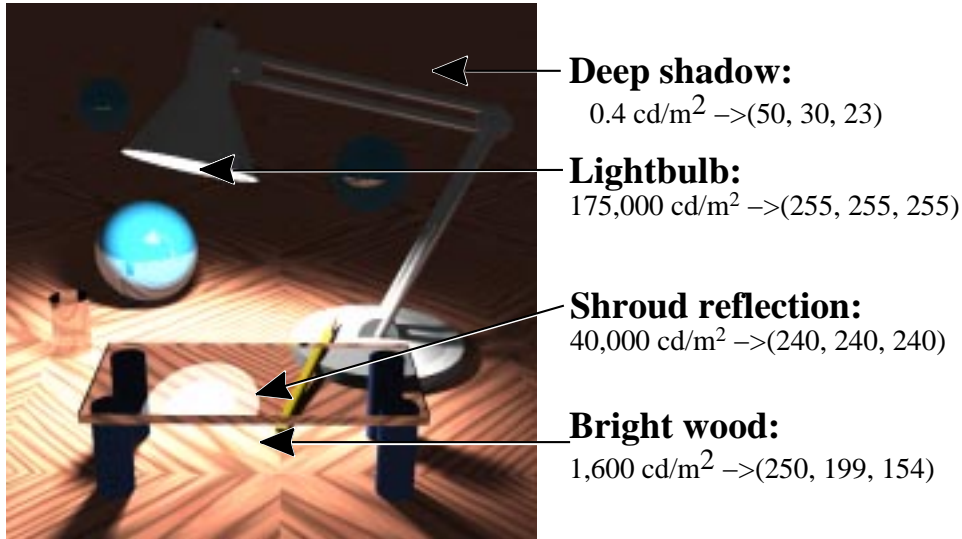
Bright wood:
1,600 cd/m^2 –>(250, 199, 154)

Figure 5: Applying the layering method to a high contrast scene ($> 300,000 : 1$) reduces display intensities to nominal pixel values without discarding fine textures and image details.

These layer images are floating-point images that describe scene properties. To construct a display image such as Figure 5 from these layers, we compress the illumination layers to reduce their contrast and then combine them with the reflectance layers.

We use six different types of intrinsic image layers grouped together in three pairs to represent the high contrast scene or image; an example of each is shown in Figure 6. The first pair describes diffuse reflectance and illumination in the scene, the second pair describes specular reflectance and illumination, and the third pair describes transparency and illumination. The original scene is then expressed in layers by:

$$Scene(x, y) = K_d(x, y)I_d(x, y) + K_s(x, y)I_s(x, y) + K_t(x, y)I_t(x, y) \qquad (5)$$

where for all (x, y) points in the image,

K values form reflectance layers ($0 \leq K \leq 1$), and
I values form illumination layers (cd/m^2).

The diffuse illumination layer, I_d, describes the amount of light received from all directions at each visible surface point in the scene but excludes all light subject to mirror-like reflections from the specular direction $\hat{S}$. Given the surface normal vector $\hat{N}$ of unit length and a unit-length vector $\hat{E}$ pointing from the surface to the eye or camera, the vector $\hat{S}$ is mirror-like; it is coplanar with $\hat{N}$ and $\hat{E}$, points outwards from the surface, and forms an equal but opposite angle with $\hat{N}$ such that $\hat{N} \times \hat{E} = -\hat{N} \times \hat{S}$. The diffuse reflectance layer, $K_d(x, y)$, is the fraction of the diffuse illumination, $I_d(x, y)$, that is reflected towards the eye.

The specular illumination layer, I_s, gives the amount of light subject to mirror-like reflections towards the eye. Specular illumination is received at each surface point along the direction of $\hat{S}$,
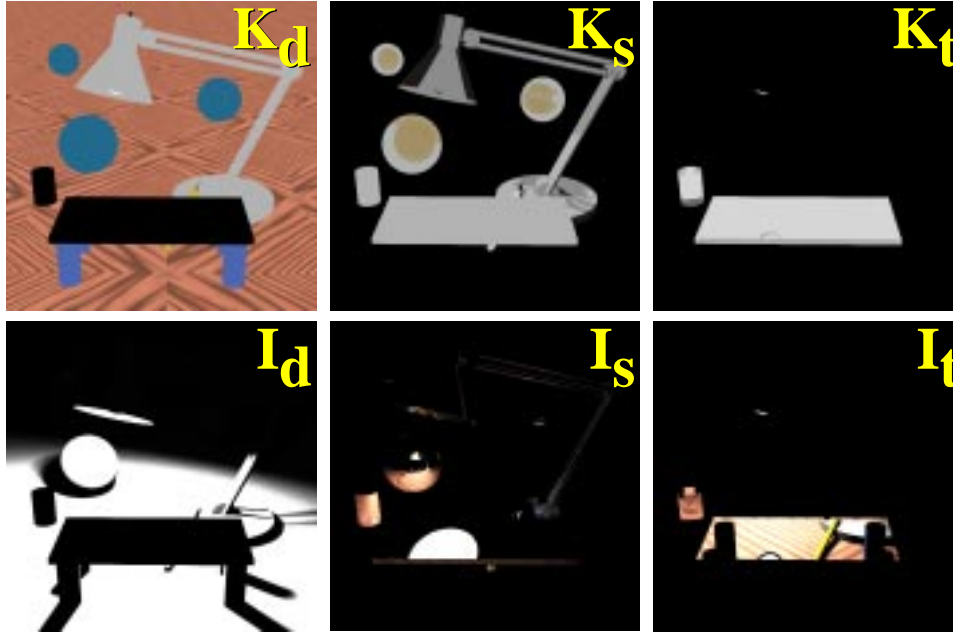
Figure 6: Layer images: the top row shows diffuse reflectance K_d, specular reflectance K_s, and transparency K_t. These three images have rich, complex detail but low contrast. The bottom row shows corresponding layer images for diffuse illumination I_d, specular illumination I_s, and transparent illumination I_t. These images contain few details but extremely high contrasts.

so the diffuse and specular illumination layers, I_s and I_d, together represent the total irradiance of all visible surfaces. The specular reflectance layer, K_s, is the fraction of the specular illumination, I_s, that is reflected toward the eye.

The transparent illumination, I_t, is somewhat unconventional because it describes the light intensity *behind* transparent objects and measures only the irradiance components in directions that follow an unobstructed path through the transparent object and towards the eye. Refraction at the surfaces or interiors of transparent objects may bend these paths; as a result the transparent illumination layer image may contain lens-like distortions as shown in the transparent cylinder in Figure 6. The transparency layer, K_t, describes the fraction of I_t transmitted through the transparent object to the eye. To define a directly visible light-emitting surface, set I_t to the surface emittance and K_t to 1.0.

As discussed earlier in Section 3.1, the human visual system appears capable of recursive decomposition and separate adaptation to some scene layers. The layering method restricts recursive decomposition to the specular and transparent illumination layers, I_s and I_t. To perform the decomposition, we assume the visual system may treat each of these layers as a new scene, and this new scene may itself be decomposed into new set of its own diffuse, specular, and transparent layers, as diagrammed in Figure 7. Each node in the figure represents a scene and contains all the layers needed to represent it, and the root node describes the entire scene. In some cases the I_s or I_t layer of a scene is recursively decomposed; it is replaced by an edge leading to a new scene with its own set of layers. For example, in the layered image shown in
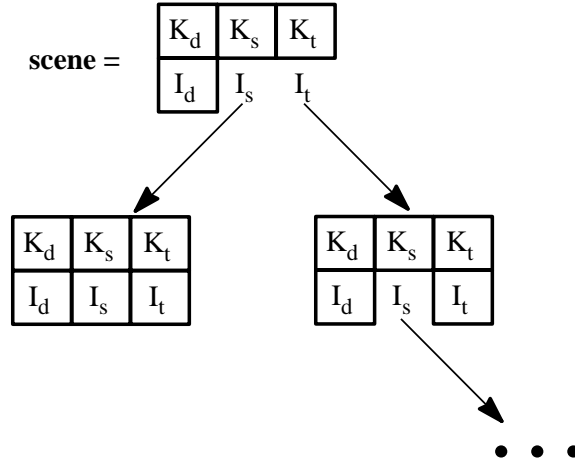
21

Figure 7: Recursive decomposition of layer images: both specular illumination and transparent illumination layers form comprehensible images that may be decomposed into more layers. Complex images or scenes with multiple transparencies and reflections form a tree of layer images.

Figure 5 and diagrammed in Figure 7, the uppermost specular illumination layer I_s is replaced by a new scene reflected in the glass tabletop. Recursive decomposition replaces this I_s layer with an edge to a new node with six child layers, as shown in the graph. Of these, the diffuse reflectance and illumination layers K_d and I_d include the interior of the lampshade and the white reflective shroud around the light bulb; the specular reflectance and illumination layers K_s and I_s include the highlights on the shroud and on the light bulb, and the transparency and transparency illumination layers K_t and I_t include the light bulb's frosted glass envelope and the illumination behind it. Returning to the root of the graph, the transparency illumination layer I_t is the new scene seen through the glass tabletop. In the graph, this layer is replaced by an edge to a new node containing six child layers. The child layers for diffuse reflectance include the wooden floor seen through the tabletop. The transparency and transparency illumination child layers are zero-valued because no transparent objects exist behind the glass tabletop, but the specular reflectance and illumination layers K_s and I_s are not empty because they contain a sliver of the lightbulb surface and glints from the pencil and lamp base. Further decomposition is possible on I_s as shown by the ellipsis in the graph.

In the layering method we assume the human visual system separately compresses and adapts to each illumination layer to help reduce large perceived contrasts due to mismatched illumination. In the scene in Figure 5, separate adaptations to the diffuse and specular illumination layers I_d and I_s permit us to see both the dimly lit wooden floor texture in the background and the shape of the light bulb reflected in the glass table top; compressing contrasts within the diffuse illumination layer I_d ensures the wood texture is also visible in the foreground. To merge these diverse layers we estimate a separate "adaptation luminance value" L_{wa} at the middle of the range of each layer's luminances, compress the contrasts of each illumination layer around the central L_{wa} to emulate the effects of local adaptation, scale layer intensities to match

adaptation values L_{wa} to the display, and then combine all layers to form the displayed image.

We concur with Schlick [24] that tone reproduction operators should probably be achromatic and therefore depends only on the luminance of the layer. We estimate the luminance for each layer using a quick binary fraction approximation of the Y luminance signal of the NTSC television standard [15]:

$$L = L(x, y) = \frac{5}{16}R(x, y) + \frac{9}{16}G(x, y) + \frac{2}{16}B(x, y). \tag{6}$$

where R, G, and B are color spectral components expressed in cd/m^2, and L is the luminance or gray-scale value of the scene or image. We then express each R,G,B color component as a fraction of the pixel's luminance value for fast conversion of display luminance to display RGB values. For example, a pixel where $RGB = (7, 11, 13)cd/m^2$ yields $L = 10cd/m^2$, and $(R/L, G/L, B/L) = (0.7, 1.1, 1.3)$. For each illumination layer, we define the adaptation luminance value L_{wa} as the mean luminance of all non-zero pixels L_w measured on a logarithmic scale:

$$\log(L_{wa}) = mean\{\log(L_w)\}. \tag{7}$$

This logarithmic scale directly corresponds to contrasts: given any two luminance values, L_1 and L_2, the distance between them on a logarithmic scale $\log(L_2) - \log(L_1)$ is the logarithm of their contrast, $\log(L_2/L_1)$. Accordingly, the mean of $\log(L)$ is the centroid of contrasts within a layer, and hence is a plausible "midrange value" for contrast compression.

Perhaps the simplest method for compressing contrasts of an illumination layer is to scale its values around L_{wa} on log-log axes with the scaling constant γ to form compressed layer image L_c:

$$\log(L_c(x, y)) = \log(L_{wa}) + \gamma(log(L) - \log(L_{wa})) \tag{8}$$

or equivalently, $L_c = L_{wa}\left(\frac{L}{L_{wa}}\right)^{\gamma}$, where $0 < \gamma < 1$, and compression increases as γ approaches zero. The contrast compression term is named γ because it is analogous to the power-law response of an idealized CRT display given by $L = V^{\gamma}$, where L is normalized display intensity (L_{out}/L_{max}), and V is normalized video input signal (V_{in}/V_{max}) [15].

We found γ compression unacceptable because it compresses both large and small contrasts equally. Illumination layers often contain small areas of high contrast, such as specular highlights or directly visible light sources, and large areas of low contrast, such as the gradual illumination changes across interior walls. A γ value small enough to make the high contrasts displayable often makes the low contrasts invisible. A better compression function should be

- *progressive:* to compress large contrasts more severely than small contrasts,

- *monotonic:* to guarantee that small luminances remain smaller than large luminances,

- *symmetric:* to affect very dark and very light regions equally,

- *asymptotic:* to compress an infinite scene range to a finite display range,

- *minimal:* to compress scene contrasts just enough to fit display contrasts and no more, and

- *adjustable:* to suit viewer preferences.

Many functions satisfy these goals. After examining the function proposed by Schlick [24] we adopted a similar first degree rational polynomial that forms a "sigmoid," or S-shaped curve when plotted on log-log axes:

$$sig(x) = \left(\frac{x^g + (\frac{1}{k})}{x^g + k} \right) \cdot D \tag{9}$$

where:

x is the normalized scene, found by dividing scene by adaptation luminance: L/L_{wa},
$sig()$ is normalized display luminance, $0 < sig() \leq 1$,
k^2 is the maximum achievable output contrast; $sig(\infty) = D$ and $sig(0) = D/k^2$,
D is an output scaling constant to map maximum scene luminance to maximum display luminance,
g is the gamma (γ) setting parameter, where γ is the slope of the curve at $x = 1$ when plotted on log-log axes:

$$\gamma = g \cdot \frac{(k-1)}{(k+1)}. \tag{10}$$

The k, D and g parameters adjust the shape and size of the $sig()$ function response curve, but are awkward to specify directly. Instead, we find their values from the limits of the desired mapping between scene luminances and display luminances. As shown in Figure 8, these limits form a rectangular "limit box" around a portion of the $sig()$ function curve. The width of the limit box is set by x_{max} and x_{min}, the maximum and minimum normalized scene luminances respectively, and the height is given by C, the amount of display contrast used.

The limit box provides an intuitive way to specify any desired $sig()$ function. Choosing values for x_{min}, x_{max}, and C along with L_{wa} provides enough information to uniquely specify a $sig()$ function that sweeps across the limit box from its lower left to its upper right corner. This $sig()$ function applies just enough contrast compression to map x_{min} and x_{max} to display minimum and maximum, yet stays symmetric about L_{wa}, even if L_{wa} is off-center within the limit box. If L_{wa} is closer to x_{max} than x_{min} then scene luminances near x_{min} will be compressed more severely than those around x_{max}. Solutions for k and g exist for any limit box where $x_{max}/x_{min} > C$; if $x_{max}/x_{min} < C$ then no $sig()$ function is needed because all scene contrasts can be directly reproduced on the display without compression. If the gamma setting parameter g is held constant, the γ of the resulting curve varies smoothly as the limit box changes size and shape, and increasing g smoothly increases the γ of the curve. The γ value grows from zero as C rises above zero, and if $g = 1.0$ then γ asymptotically approaches 1.0 as C increases. We found an analytic expression for k using limit box terms and g by writing the equation
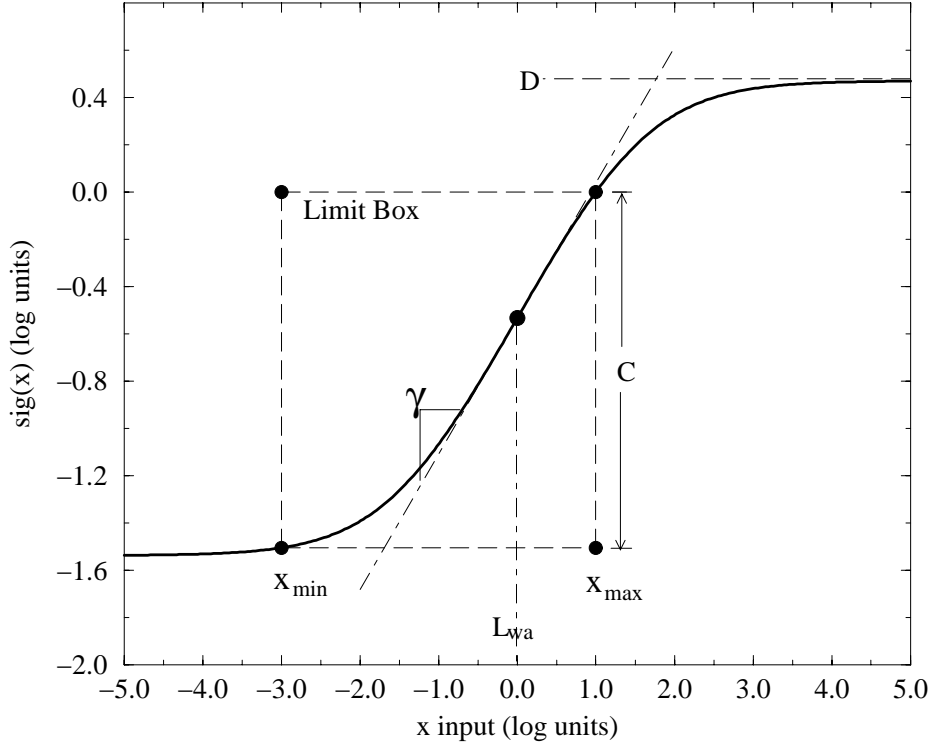
Figure 8: Both the layering and foveal methods use the $sig()$ function of Equation 9 to reduce high scene contrasts to fit within the abilities of the display. Users set $sig()$ function parameters by specifying a "limit box" for the mapping from scene to display luminances. The limit box is bounded by upper and lower scene luminance limits x_{max} and x_{min} and by display contrast limit C. The $sig()$ function has an inflection point and is symmetric about $x = 1$ (or $\log(x) = 0$) where slope is given by γ.

$C = sig(x_{max}/L_{wa})/sig(x_{min}/L_{wa})$ and solving for k:

$$k(x_{max}, x_{min}, L_{wa}, C, g) = \frac{1}{2L_{wa}^g(x_{max}^g - C \cdot x_{min}^g)}(B_p + \sqrt{B_n^2 + C \cdot A^2}) \qquad (11)$$

where

$A = 2L_{wa}^g(x_{max}^g - x_{min}^g),$
$B_p = ((x_{max} \cdot x_{min})^g + L_{wa}^{2g})(C - 1),$ and
$B_n = ((x_{max} \cdot x_{min})^g - L_{wa}^{2g})(C - 1).$

To maintain normalized display output to ensure $sig() = 1$ when scene luminance L_w reaches x_{max} in Equation 9 let:

$$D = \frac{(\frac{x_{max}}{L_{wa}})^g + k}{(\frac{x_{max}}{L_{wa}})^g + \frac{1}{k}}. \qquad (12)$$

In Section 6 we will specify a desired gamma γ_d for the $sig()$ function curve. Though we have no analytic solution, finding the value of g that produces a gamma value of γ_d is a simple root finding problem. On log-log axes, adjusting the g parameter is equivalent to scaling the sigmoid curve function $sig()$ about its inflection point at $x = 1$, shrinking or stretching the entire curve to modify γ, its slope at $x = 1$. To find a $sig()$ curve that both fits the limits box and has the desired gamma γ_d we must find the g to satisfy Equation 10 where k is given by Equation 11. The equation is well behaved and converges quickly with conventional root finding methods.

The layering method applies the $sig()$ function to each illumination layer using nominal values of $x_{max} = max(L)$, $x_{min} = min(L)$, $\gamma = 1.0$ and $C = \sqrt{C}_{max}$. The value of C is an ad-hoc choice, made to consume only half of the contrast range of the display (when plotted on log-log axes, as in Figure 8) and allow room for additional contrast from reflectance layers. Choosing C by more rigorous methods may improve layering results. Choosing D using Equation 12 normalizes the output of $sig()$ function for easy conversion to display units; maximum scene luminance x_{max} causes $sig()$ output of 1.0, and x_{min} produces an output of $1/C$. We scale $sig()$ outputs by a constant equal to maximum display luminance L_{dmax} to convert to photometric units (cd/m^2).

Finally, all layers are combined to construct a reduced contrast display image. The compressed and scaled illumination layers are each converted from luminance images back to RGB images by multiplying them with their corresponding $(R/L, G/L, B/L)$ images. If any luminance layer was compressed with a γ other than 1.0 we apply this same contrast sensitivity change to the color ratio images as $(R/L)^\gamma, (G/L)^\gamma, (B/L)^\gamma$. Then the compressed illumination layers I_d, I_s, and I_t are multiplied by their associated reflectances K_d, K_s, and K_t, progressively collapsing the tree of layer images diagrammed in Figure 7 from the leaves upwards to the root node to form the output image. Because the tree describes a sequence of nested multiplies and adds, our implementation computes a composite K image for each of the illumination layers, multiplies each of them with their corresponding I images, and sums the result. We compute the composite K images by traversing the layer image tree in depth first order so multiplications of common parent K layers are performed only once. Composite K images provide considerable savings for the image of Figure 9 because the tree held 36 layer images, including 7 illumination layer images at the leaves of the tree. Computing the 7 composite K images also allowed us to experiment with different types and amounts of compression for illumination layers and quickly revise the output image.

Figures 1, 5, and 9 show the results of layering applied to extremely high contrast images, and clearly reveal deep shadow detail and the brilliant surfaces of light sources. In Figure 1, the illumination layer compression reveals the wood grain of the floor in both the brightly lit and deeply shadowed areas and exhibits gradual shading on the interior of the lamp housing. In Figure 9, layering reveals the driver's face in the red car and details of both the street light and the wet street surface reflecting the car's glossy paint in the foreground.

Capturing layer images is straightforward with many synthetic image renderers, particularly those with shading models that use explicit diffuse and specular terms such as those by Gouraud,
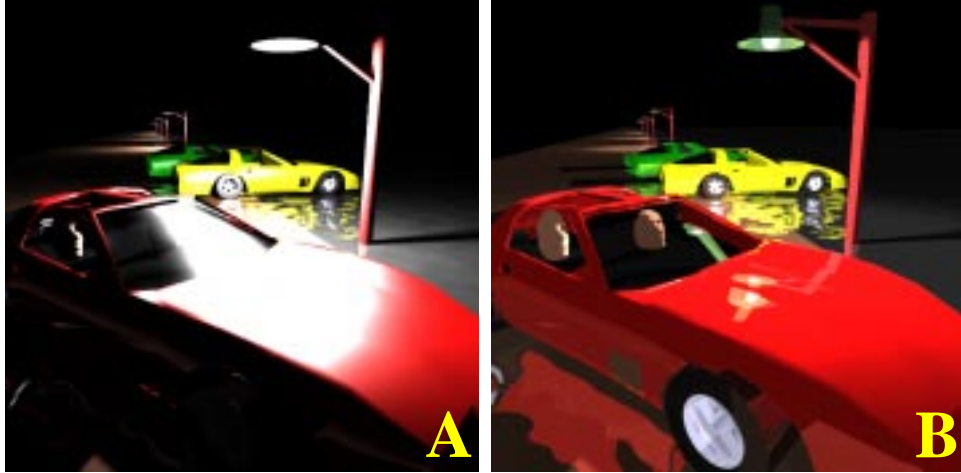
Figure 9: The original high contrast ($> 300,000 : 1$) image A loses both highlight and shadow detail to clipping; in image B the layering method maintains the appearance of high contrast while revealing the driver's disembodied face behind the windshield and the reflection of the streetlight in the car hood.

Phong, or Cook-Torrance. The layering method is not a new form of rendering, but does require the renderer to save computed values for illumination, reflectance, and transparency. To capture the layer images shown here we modified the shader functions of a commercially available ray tracer [41] to record K and I values in separate floating-point image files. Auxiliary programs then compressed and combined the layer files to form the final images shown here. Though the layering method was intended as an automatic technique that needs no external parameters, we found that a wide range of x_{max}, x_{min}, C, D and γ_d values for the compression function produced pleasing display images. Adjusting these parameters provides a convenient and intuitive way to interactively change the emphasis of various image components according to their importance in a scene.

Capturing layer images might be more difficult in renderers that compute global illumination solutions or use bidirectional reflectance distribution functions (BRDFs) to describe surfaces, such as RADIANCE [39], HELIOS [3], or the commercial software products offered by LightScape. These renderers explicitly compute illumination both from light sources and from inter-reflections with other surfaces. They may also include complex angular dependencies in surface reflectances. We expect that the six image layers defined here can be captured from the intermediate results of such renderers, but we have not attempted to do so.

Images from the layering method sometimes contain subtle aliasing artifacts. The nonlinear compression of illumination layer images will cause some error at pixels that contain large illumination boundaries, because the result of compressing the pixel's single value is not the same as compressing two or more illuminants within the pixel independently, then combining them. We suspect these errors can be greatly reduced either by retaining subpixel geometry information or by careful use of a transparency or "alpha" value computed at each pixel to help

represent such boundaries.

Our experience with layering also suggests that in addition to a "layered" image decomposition, the visual system may further segregate images into regions or objects. For example, the layering method grouped together the dim specular reflections of the blue chrome spheres in the background of Figure 5 with the dazzling reflections of the glass tabletop, and adjusting the compression function on I_s affected the appearance of both. Computer graphics has a tremendous advantage over traditional photography and other image capture methods because the 3-D scene description used to render an image is likely to contain an adequate partitioning of the objects as well.

6   Revised Tone Reproduction Operator

The foveal method of Section 7 depends on an improved tone reproduction operator which we will now construct by revising and updating the method of Tumblin and Rushmeier [34]. We begin by building on the work of Ferwerda and Ward. They set display luminance L_d from scene or world luminances L_w using $L_d = m \cdot L_w$, but we include a (γ) term to adjust contrasts just as m adjusts intensities:

$$L_d = m \cdot \left(\frac{L_w}{L_{wa}}\right)^{\gamma} \tag{13}$$

As before, m is a scale factor based on a model of human visual adaptation, and γ is the change in human contrast sensitivity between the original scene and the displayed image. Unlike the scale factor m, the γ term affects small and large luminances differently; display contrasts increase as the scene contrasts increase, but grow more slowly when the γ term is smaller.

We find γ from Tumblin and Rushmeier's original tone reproduction operator [34], restated here with less awkward notation and with corrections for the anomalies in very dim and very bright images. Their operator was based on work by Stevens [30, 28], who claimed that a viewer fully adapted to a uniform background luminance L_a viewing a test patch of luminance L will experience a brightness sensation B, related to L by a power law. Rewriting the claim in SI units gives:

$$B = C_0 \left(\frac{L}{L_a}\right)^{\gamma} \tag{14}$$

where

L is luminance in cd/m^2,

B is brightness in *brils*; a fully dark-adapted viewer senses one *bril* when viewing a patch of 1 *micro-lambert* intensity,

L_a is adaptation luminance in cd/m^2,

$C_0 = 0.3698$, a constant due to measurement units,

28

γ is contrast sensitivity, an exponent that depends on adaptation luminance L_a.

In Stevens and Stevens, the contrast sensitivity term γ falls linearly with decreasing $\log(L_a)$, passes through zero at $L_a = 2.3 \cdot 10^{-5} cd/m^2$, and is negative for smaller values. We have modified their γ term in three ways. As shown in equation 18 below, γ is limited to no more than 2.655 to match measurements that indicate human contrast sensitivity stops increasing above about $100 cd/m^2$ [26]. We also limited its minimum value to zero to prevent negative contrasts in extremely dim images and added a constant offset of $2.3 \cdot 10^{-5}$ to L_a so that contrast sensitivity approaches zero asymptotically near the threshold of vision.

Tumblin and Rushmeier used two instances of Equation 14 to convert scene luminances to display luminances. One instance computes the perceived brightnesses of the display B_d, and the other finds the perceived brightnesses B_w of the "real world" or scene. B_d is set equal to B_w to make the perceived quantities match:

$$B_w = C_0 \left(\frac{L_w}{L_{wa}} \right)^{\gamma_w} = B_d = C_0 \left(\frac{L_d}{L_{da}} \right)^{\gamma_d} . \tag{15}$$

Solving for L_d in terms of L_w and L_{wa}:

$$L_d = L_{da} \left(\frac{L_w}{L_{wa}} \right)^{\left(\frac{\gamma_w}{\gamma_d} \right)} . \tag{16}$$

The result has the same form as Equation 13, except the C_0 terms cancel and reveal an anomaly; all mid-range scene luminances map to mid-range display luminances near L_{da}, therefore the display appears a uniform gray in dim scenes where contrast sensitivity is low. We remove this anomaly by appending a new scale factor term $m(L_{wa})$.

Our $m(L_{wa})$ function is built from a simple conjecture about visual appearance; we claim as the scene adaptation luminance value L_{wa} increases from starlight to the threshold of eye damage, the corresponding display luminances should grow steadily from display minimum to maximum. We choose $m(L_{wa})$ to vary according the same log-linear expression Stevens used to find contrast sensitivity γ in Equation 18, forming an almost straight line (or a straight series of dots in Figure 4D) when plotted on log-log axes. For L_{wa} values below $100 cd/m^2$, changes in m match changes in contrast sensitivity and cause scene luminances of $L_{wa}/\sqrt{C_{max}}$ to map precisely to the minimum display luminance. Above $100 cd/m^2$, reaching minimum display luminance requires scene luminances further below L_{wa}. The revised tone reproduction operator is given by:

$$L_d = m(L_{wa}) \cdot L_{da} \cdot \left(\frac{L_w}{L_{wa}} \right)^{\left(\frac{\gamma_w}{\gamma_d} \right)} \tag{17}$$

where

L_{da} is the display adaptation luminance, typically between 10–$30 cd/m^2$,
L_{wa} is scene adaptation luminance, found from scene luminances L_w using:

29

$$log(L_{wa}) = mean\{log(L_w + 2.3 \cdot 10^{-5} cd/m^2)\},$$

γ_d is $\gamma(L_{da})$, and γ_w is $\gamma(L_{wa})$, Stevens' contrast sensitivity for a human adapted to the display and the scene respectively. Find these γ values using:

$$\gamma(L_a) = \begin{cases} 2.655 & \text{for } L_a > 100 cd/m^2 \\ 1.855 + 0.4 \log_{10}(L_a + 2.3 \cdot 10^{-5}) & \text{otherwise,} \end{cases} \tag{18}$$

$m(L_{wa})$ is the adaptation-dependent scaling term to prevent anomalous gray night images:

$$m(L_{wa}) = \left(\sqrt{C_{max}}\right)^{(\gamma_{wd}-1)} \tag{19}$$

where

C_{max} is the maximum available display contrast (30 to 100 typical),
and $\gamma_{wd} = \left(\frac{\gamma_w}{1.855 + 0.4 \log(L_{da})}\right)$.

The m term steadily increases display brightnesses as the scene adaptation luminance L_{wa} increases towards the upper limits of vision, as shown in Figure 4. We will apply the operator defined by Equations 17, 18, and 19 in Section 7.

7   Implementation of the Foveal Display Program

As discussed in Section 3.2, the foveal display program evokes the visual sensations of a high contrast scene by computing new displayed images in response to the user's eye movements. The program regards the mouse cursor position as the user's direction of gaze in the scene, and considers a small circular region around the mouse cursor as the user's "foveal neighborhood," the portion of the scene currently viewed by the user's fovea. Users may adjust the diameter of the program's foveal neighborhood to match personal preferences and nominally subtend 2–5 degrees in the original scene. In response to mouse clicks on the image, the program computes a new image as it might appear after foveally dominated adaptation, with intensity and contrast of the displayed image determined by the tone reproduction operator presented in the previous section. Any out-of-range display intensities are asymptotically compressed towards display black or white by the $sig()$ function of Section 5 to help preserve details and textures in image shadows and highlights.

The foveal display program works in four steps. First, in response to a mouse click the program finds the position and diameter of the foveal neighborhood and briefly displays a thin circle enclosing it, as shown in Figure 10. Second, the program computes the foveal adaptation luminance value L_{wa} from scene luminances in the circled neighborhood using a precomputed image pyramid. Third, the program uses L_{wa} in the tone reproduction operator described in
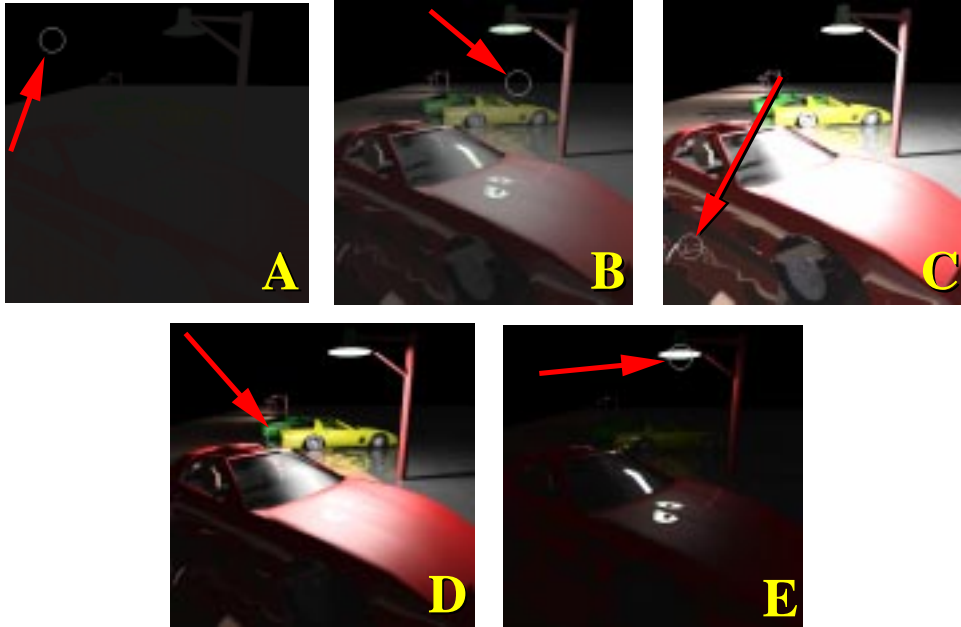
Figure 10: Example display images from the interactive foveal display program: each red arrow points to the foveal neighborhood, a circled region around the user's direction of gaze. A) Placing the circled region on the night sky causes adaptation luminance L_{wa} near the absolute threshold of vision where contrast sensitivity is extremely low causing the anomalously dim street light. B) Adapting to distant terrain lit by moonlight greatly increases contrast sensitivity, causing contrast compression in the foreground. C) Directing attention to the deep car shadows in the foreground reveals reflected details of the wet ground below the car. D) Adapting to the distant yellow car shows its reflection on wet pavement. E) Gazing at the bulb of the streetlight causes extremely high adaptation luminance, darkening the rest of the scene. Images A) and E) illustrate the limitations of globally applied adaptation models; local adaptation in the human visual system ensures that the street light and car have a bright appearance for any direction of gaze in the scene.

Equation 17 to find the desired display image luminances at each pixel. Finally, it applies the asymptotic compression function $sig()$ to find displayed luminance values without truncating image highlights and details in the foveal region.

The foveal program must update the displayed image rapidly because the program relies on the user to remember and assemble a sequence of images into a coherent impression of the high contrast scene. For quick response, the program uses an image pyramid [6] of log scene luminances to find L_{wa} values in constant time for any foveal neighborhood diameter, and we recommend Ward's 32-bit per pixel RGBE format [37] to store and manipulate high contrast scene values, though we used 32-bit floating point values for each color component in our test program to ensure accuracy. Our program, written in Visual C++ 5.0 running under WindowsNT 4.0 on a 90-Mhz Pentium machine with 48MB of memory achieves a 4 Hz display update rate on a 256x256 pixel image, and was not optimized for speed.

Without image pyramids, computing foveal adaptation luminance L_{wa} can be slow for large diameter foveal neighborhoods. Foveal L_{wa} is a localized form of the global adaptation

value computed by Equation 7, where each foveal neighborhood pixel's contribution to L_{wa} is smoothly weighted by distance to favor pixels nearest the cursor. The weighting function has an approximately Gaussian shape and is applied to the logarithm of neighborhood pixel luminances; their weighted sum is $\log(L_{wa})$. The time required to compute L_{wa} directly from scene pixels grows linearly with the number of neighborhood pixels.

Using image pyramids allows computation of L_{wa} in constant time, but the method is more easily explained by a simpler problem. Suppose the foveal neighborhood diameter $diam$ is limited to one of two fixed widths, either W or $2W$ pixels. We may choose to precompute two images to store L_{wa} for each pixel; one image for diameter W named $lev0$, another for diameter $2W$ named $lev1$. To find L_{wa} quickly use $diam$ to select an image and use the cursor position to select a pixel. However, the $lev1$ image is much smoother than $lev0$ because it was computed with a foveal neighborhood four times larger; we can reasonably approximate all $lev1$ values by interpolating them from a much smaller precomputed image. For example, we may decide to reduce the size of the original $lev1$ image by a factor of four using $(2,2)$ decimation by discarding every other scanline and every other pixel on the remaining scanlines. To find L_{wa} for any value of $diam$ between W and $2W$ we can approximate L_{wa} by first finding its value in $lev0$ and $lev1$ and then interpolating between them according to the ratio of W and $diam$. Image pyramids use similar ideas.

The image pyramid is a stack of images built by recursive low-pass filtering and $(2,2)$ decimating. The base level of the pyramid holds the pixel-by-pixel logarithm of the input image luminance, and each successively higher pyramid level is a smoother, smaller version of the previous level, ending with a single pixel image at the top level. The program builds each pyramid level using Burt and Adelson's separable 5-tap Gaussian-like filter kernel $[0.05, 0.25, 0.40, 0.25, 0.05]$. This filter is approximately circularly symmetric and does not introduce the half-pixel offsets found in the more widely used MIP-map pyramid filters, yet it is very fast to compute because it is symmetric and applied separately in the horizontal and vertical directions. Building a complete pyramid for an $N \times N$ image requires only $4/3N^2$ storage locations, $2N^2$ multiplies and $8/3N^2$ adds.

Sampling an image pyramid level at the cursor position is equivalent to finding a weighted sum of input image pixels in a neighborhood around the cursor. The neighborhood's diameter doubles with each successively higher pyramid level, as shown in Figure 11. To approximate a continuously variable neighborhood size, the program linearly interpolates between samples in two adjacent pyramid levels, using the logarithmic relation between neighborhood diameter and pyramid level. The pyramid levels are numbered sequentially, with $lev = 0$ as the base, and $lev = levmax$ for the tip image. The pyramid base is a copy of the $\log(L_w)$ image; therefore, each pixel in the $lev = 0$ image can be regarded as the input image averaged over a local neighborhood with a diameter of $diam = 1$ pixel. The spacing between pixels doubles in each successively higher pyramid level when measured against the pyramid base, so that $diam = 2^{lev}$ pixels. To approximate L_{wa} at the cursor position for a neighborhood diameter that is a power of two,

One sample of image pyramid at level L ⟷ **SAME** **Weighted average in neighborhood of diameter 2^L**
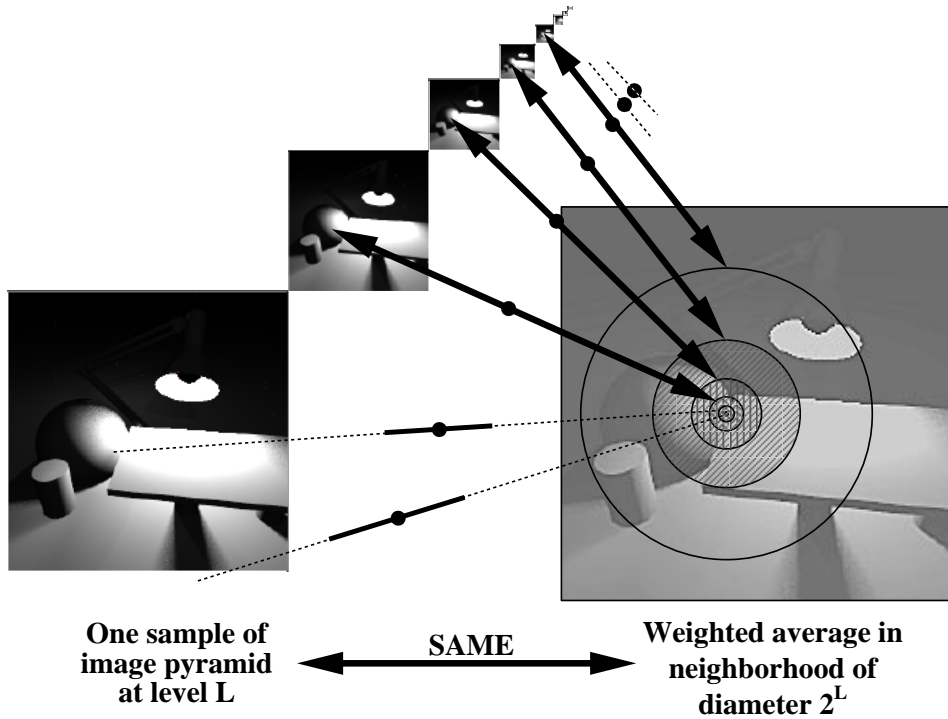
Figure 11: Image pyramids allow the foveal display program to find $\log(L_{wa})$, a weighted sum of neighborhood pixel values, in constant time for any neighborhood diameter.

the program finds the cursor's value within the pyramid level selected by $lev = \log_2(diam)$. For neighborhood diameters that are not a power of two, lev is split into integer and fractional parts, $levInt$ and $levFrac$ respectively. The program finds the cursor-position value at both level $levInt$ and at level $levInt + 1$, linearly interpolates between these two values using $levFrac$, and converts the interpolated value from log units to the luminance value L_{wa}.

The locally measured adaptation luminance, L_{wa}, determines how the foveal display program will convert original scene luminances, L_w, to display luminances, L_d. Though the operator in Equation 17 can adjust the contrast of the displayed image by changing γ_w, it cannot guarantee that the computed display luminances L_d are within the range of the display device. To avoid clipping, the foveal program combines the $sig()$ function of Equation 9 and Figure 8 with the tone reproduction operator of Equations 17–12 to compute the compressed display luminance:

$$L_d = sig(x, x_{max}, x_{min}, L_{wa}, C, D_{fov}, \gamma), \tag{20}$$

where

$x = \frac{L_w}{L_{wa}},$

$x_{max}, x_{min} = max(L_w), min(L_w)$ measured over the entire scene,

$C = C_{max}$, the maximum contrast available from the display,

$D_{fov} = D \cdot m(L_{wa})$ using D found in Equation 12 and m given by Equation 19, and

33

$\gamma = \left(\frac{\gamma_w}{\gamma_d}\right)$ found in Equation 18.

The x inputs to the $sig()$ function are the original scene contrasts as measured against the adaptation luminance value L_{wa}. The x_{max}, x_{min}, and C parameters are constants that ensure the $sig()$ function can accept all scene luminances without truncation, and the D value maps scene L_{wa} values to display luminances according to the revised tone reproduction operator of Equation 17. The γ term adjusts display contrasts to match contrast sensitivity of a human observer viewing the original scene.

Interactive viewing of high contrast images with the foveal display program resembles the familiar behavior of an automatic exposure camera, and Figures 1 and 10 show typical screen images. The foveal program reproduces all displayable scene contrasts in the small circled neighborhood around the cursor, but other regions that form high contrasts when compared to the adaptation luminance L_{wa} are compressed towards display black or white and temporarily lose low contrast details. As the cursor or the L_{wa} value moves nearer to these obscured regions, their details emerge in the displayed image.

8   Discussion

Both the foveal and layering methods display images of high contrast scenes while preserving details commonly lost to truncation or scaling in conventional methods of image display. Both methods are supported by results from the psychophysical literature, are straightforward to implement, and are not computationally expensive. The foveal method can be applied to high contrast images from any source, but the layering method is useful only for synthetically generated images. The layering results can be displayed statically and are suitable for printed images, but the foveal method is interactive, and requires a computer display to convey the impression of high contrast.

Though both methods were intended for display use, the layering method also shows some promise as a scene lighting tool. Our experience with layering has shown that interactively adjusting L_{wa} and the $sig()$ parameters x_{min}, x_{max}, C, D and γ for each illumination layer while viewing the combined result provides an intuitive way to produce images with attractive lighting. Layering seems especially well-suited to interactive lighting design tools such as those by Kawai, Painter, and Cohen [19] and may help to reduce the tedious cycles of adjusting lights and re-rendering.

Creating the appearance of a high contrast scene is often difficult with conventional lighting methods. For example, the first image in Figure 12 shows the layering result, and images A–C show the best approximations to the layering result that we could achieve by changing the intensities of the two light sources. Increasing the ambient illumination revealed the floor in the background, but no intensity value for the light bulb appeared correct. Reducing the light bulb
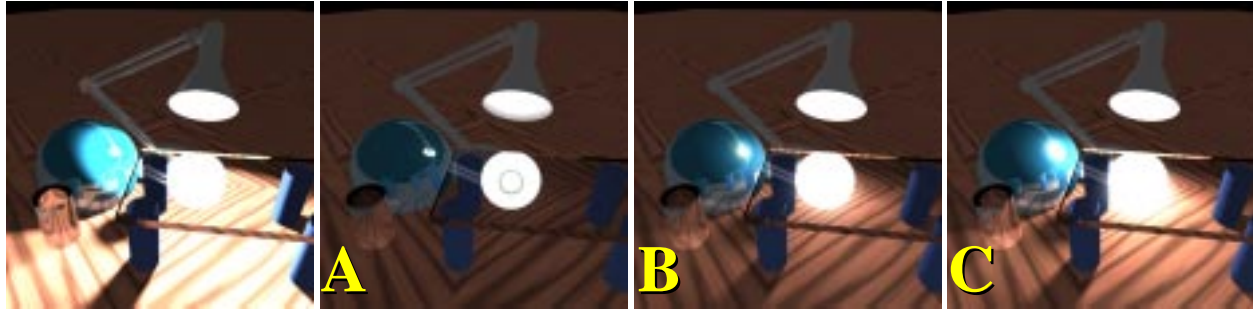
Figure 12: The layering result shown in the first image is difficult to achieve by re-lighting and re-rendering the original scene, as shown in images A-C. The floor and horizon in the background are easily revealed by increased ambient light, but we found no acceptable intensity for the light bulb. Reducing intensity to reveal the bulb's reflection in the glass tabletop eliminates table leg shadows (A), but increasing it enough to show dramatic leg shadows in (C) causes nearby wood textures to clip to white. The intermediate choice (B) still lacks the appearance of high contrast provided by the layering result.

intensity enough to detect its shape reflected in the glass tabletop caused the strong shadows from the table legs to disappear (A), but increasing its intensity enough to deepen the shadows caused the wooden floor texture beneath the glass tabletop to clip to white (C). As a compromise, in image B we chose the highest light bulb intensity that would avoid clipping the wooden floor texture. Despite our efforts, this image lacks an appearance of extremely high contrast and would need additional skillfully placed light sources to resemble the layering result shown in the leftmost image.

The layering and foveal methods are preliminary steps towards perceptually valid display methods for high contrast scenes, and these methods capture just a few of the many directly measurable response properties of human vision. We have not addressed many important adaptation properties, some of them already examined by other authors, such as local adaptation in both foveal and peripheral vision, temporal effects, the subtractive adaptation process that gradually augments contrast sensitivity and aids color lightness constancy, visual masking effects, and color adaptation. Both methods make global changes to images to prevent image artifacts; we intentionally avoided local variations of intensity scale factor m, contrast scale factor γ, and asymptotic compression function $sig()$. The human visual system makes such local nonlinear adjustments routinely with very few apparent artifacts, but we do not know how to model this process mathematically.

Some images produced by the foveal display program illustrate the need for a better model of local adaptation. In Figure 10, selecting a circled neighborhood in the night sky will choose an extremely low adaptation luminance near the absolute threshold of vision, where contrast sensitivity approaches zero. When this "foveal" adaptation is applied to the entire image, even the street light is reduced a dim gray shape instead of the brilliant white appearance our eyes would see.

The layering and foveal methods could be extended to include other previously published

visual effects models as well. For example, layering is well suited for use with the visual glare, diffraction, and scattering models of Spencer *et al.* [27], and the foveal method could include the wide-ranging models for color, acuity, and the time course of adaptation developed by Ferwerda and colleagues [9]. High speed implementations of the foveal method might lead to interesting experiments using eye-tracking equipment or head-mounted displays in which the displayed image actively guides adaptation of the user's visual system. Combinations of the foveal and layering methods may also be possible, where the user's direction of gaze assigns attention to layer images according to their content adjusts their *sig*() function parameters for emphasis.

Currently the most accurate and appealing low contrast renditions of high contrast scenes are made by highly skilled artists. Ansel Adams's rich photographic prints capture the dramatic appearance of western landscapes, and several intriguing books of charcoal drawings by Chris Van Allsburg [35] (Figure 2) show astonishingly high contrast scenes with fine detail visible everywhere. These works set a high standard for realistic image display that is far beyond the current capabilities of computer graphics. Studying texts for drawing and photography may provide important guidance for future high contrast display methods.

Acknowledgments

References

[1] E. H. Adelson. Perceptual organization and the judgment of brightness. *Science*, 262:2042–2044, 24 December 1993.

[2] L. Arend. Surface colors, illumination, and surface geometry: Intrinsic-image models of human color perception. In A. L. Gilchrist, editor, *Lightness, Brightness, and Transparency*, pages 159–214. Lawrence Erlbaum, Hillsdale, NJ, 1994.

[3] I. Ashdown. *Radiosity: A Programmer's Perspective*. John Wiley and Sons, New York, NY, 1994.

[4] H. G. Barrow and J. Tenenbaum. Recovering intrinsic scene characteristics from images. In A. R. Hanson and E. M. Riseman, editors, *Computer Vision Systems*, pages 3–26. Academic Press, San Diego, CA, 1978.

[5] H. R. Blackwell. Contrast thresholds of the human eye. *Journal of the Optical Society of America*, 36(11), November 1946.

[6] P. J. Burt and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4):217–236, October 1983.

[7] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface '93*, pages 245–254, May 1993.

[8] A. L. edited by Gilchrist. *Lightness, Brightness, and Transparency*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.

[9] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual adaptation for realistic image synthesis. In *SIGGRAPH '96 Conference Proceedings*, Annual Conference Series, pages 249–258, New York, NY, August 1996. ACM SIGGRAPH, Addison Wesley. Held in New Orleans, Louisiana, 4–9 August 1996.

[10] A. L. Gilchrist. The perception of surface blacks and whites. In I. Rock, editor, *The Perceptual World*, pages 63–78. W. H. Freeman, New York, NY, 1990.

[11] A. L. Gilchrist and A. Jacobsen. Perception of lightness and illumination in a world of one reflectance. *Perception*, 13:5–19, 1984.

[12] R. H. Henneman. A photometric study of the perception of object color. *Archives of Psychology*, 179:5–89, 1935.

[13] C. D. Hodgman, R. C. Weast, and S. M. Selby, editors. *Handbook of Chemistry and Physics*, pages 2954–2955. Chemical Rubber Publishing Co., Cleveland, OH, 42nd edition, 1971.

[14] D. C. Hood and M. A. Finkelstein. Chapter 5: Sensitivity to light. In K. R. Boff, L. Kaufman, and J. P. Thomas, editors, *Handbook of Perception and Human Performance*, volume Volume I: Sensory Processes and Perception, pages 5–3 – 5–6. John Wiley and Sons, New York, NY, 1986.

[15] R. W. G. Hunt. *The Reproduction of Colour*. Kings Langley:Fountain Press, 3rd edition, 1975. (5th edition is now available).

[16] D. J. Jobson, Z. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, July 1997.

[17] D. J. Jobson, Z. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *IEEE Transactions on Image Processing*, 6(3):451–462, March 1997.

[18] L. Kaufman. Section I: Theory and methods, Overview:10. Scaling. In K. R. Boff, L. Kaufman, and J. P. Thomas, editors, *Handbook of Perception and Human Performance. Volume I: Sensory Processes and Perception*, pages 1.55–1.66. John Wiley and Sons, New York, NY, 1986.

[19] J. K. Kawai, J. S. Painter, and M. F. Cohen. Radioptimization - goal based rendering. In *SIGGRAPH '93 Conference Proceedings*, Annual Conference Series, pages 147–154, New York, NY, August 1993. ACM SIGGRAPH, Addison Wesley. Held in Anaheim, California, 1–6 August 1993.

[20] E. Nakamae, K. Kaneda, T. Okamoto, and T. Nishita. A lighting model aiming at drive simulators. In *SIGGRAPH '90 Conference Proceedings*, Annual Conference Series, pages 395–404, New York, NY, August 1990. ACM SIGGRAPH, Addison Wesley. Held in Dallas, Texas 6–10 August 1990.

[21] A. V. Oppenheim, R. W Schafer, and T. G. Stockham. Nonlinear filtering of multiplied and convolved signals. *Proceedings of the IEEE*, 56(8):1264–1291, August 1968.

[22] Z. Rahman, D. J. Jobson, and G. A. Woodell. Multi-scale retinex for color image enhancement. In *Proceedings, International Conference on Image Processing*, volume 3, pages 1003–1006, June 1996. Held in Lausanne, Switzerland 16-19 September 1996.

[23] H. Ripps and R. A. Weale. The visual stimulus. In H. Davson and L. T. Graham Jr., editors, *The Eye*, volume 2A: Visual function in Man: Part I: On Human Vision, pages 43–99. Academic Press, New York, NY, 2nd edition, 1969.

[24] C. Schlick. Quantization techniques for visualization of high dynamic range pictures. In G. Sakas, P. Shirley, and S. Mueller, editors, *Photorealistic Rendering Techniques*, Proceedings of the 5th Eurographics Rendering Workshop 13-15 June 1994, pages 7–20, Berlin, 1995. Springer Verlag.

[25] J. Schubert and A. L. Gilchrist. Relative luminance is not derived from absolute luminance. *Investigative Opthalmology and Visual Science*, 33(4):1258, April 1992.

[26] R. Shapley and C. Enroth-Cugell. Chapter 9: Visual adaptation and retinal gain controls. In N. N. Osborne and G. J. Chader, editors, *Progress in Retinal Research*, volume 3, pages 263–346, see p. 283. Pergamon Press, Oxford, 1983.

[27] G. Spencer, P. Shirley, K. Zimmerman, and D. P. Greenberg. Physically-based glare effects for digital images. In *SIGGRAPH '95 Conference Proceedings*, Annual Conference Series, pages 325–334, New York, NY, August 1995. ACM SIGGRAPH, Addison Wesley. Held in Los Angeles CA. 6–11 August 1995.

[28] J. C. Stevens and S. S. Stevens. Brightness function: Effects of adaptation. *Journal of the Optical Society of America*, 53(3):375–385, March 1963.

[29] S. S. Stevens. To honor Fechner and repeal his law. *Science*, 133(12):80–86, January 13 1961.

[30] S. S. Stevens and J. C. Stevens. Brightness function: Parametric effects of adaptation and contrast. *Journal of the Optical Society of America*, 50(11):1139, November 1960. Program of the 1960 Annual Meeting.

[31] T. G. Stockham. Image processing in the context of a visual model. *Proceedings of the IEEE*, 60(7):828–842, July 1972.

[32] T. Tanaka and N. Ohnishi. Painting-like image emphasis based on human vision systems. In D. Fellner and L Szirmay-Kalos, editors, *EUROGRAPHICS '97*, volume 16(3), pages C253–C260. The Eurographics Association, 1997.

[33] CIE Publication 19.21 Technical Committee 3.1. *An Analytic Model for Describing the Influence of Lighting Parameters upon Visual Performance, Volume I: Technical Foundations*. Commission Internationale De L'Eclairage(CIE), 1981.

[34] J. Tumblin and H. Rushmeier. Tone reproduction for computer generated images. *IEEE Computer Graphics and Applications*, 13(6):42–48, November 1993.

[35] C. Van Allsburg. *The Mysteries of Harris Burdick*. Houghton Mifflin, New York, NY, 1984.

[36] J. Walraven, C. Enroth-Cugell, D. C. Hood, D. A. I. MacLeod, and J. L. Schnapf. Chapter 5: The control of visual sensitivity: receptoral and postreceptoral processes. In L. Spillmann and J. S. Werner, editors, *Visual Perception: The Neurophysiological Foundations*, pages 53–101. Academic Press, San Diego, CA, 1990.

[37] G. Ward. Real pixels. In James Arvo, editor, *Graphics Gems II*, chapter II.5, pages 80–85. AP Professional, Cambridge MA, 1991.

[38] G. Ward. A contrast-based scalefactor for luminance display. In P.S. Heckbert, editor, *Graphics Gems IV*, chapter VII.2, pages 415–421. AP Professional, Cambridge MA, 1994.

[39] G. Ward. The RADIANCE Lighting simulation and rendering system. In *SIGGRAPH '94 Conference Proceedings*, Annual Conference Series, pages 459–472, New York, NY, July 1994. ACM SIGGRAPH, Addison Wesley. Held in Orlando FL 24–29 July 1994.

[40] G. Ward Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, October-December 1997.

[41] C. Watkins and S. Coy. *Photo-Realism and Ray Tracing in C.* M & T Press, 1992. Also available in *Essential Books on Graphics Programming.* CD-ROM compilation, Dr. Dobb's Journal, Miller Freeman 1995.