THREE METHODS OF DETAIL-PRESERVING CONTRAST REDUCTION FOR DISPLAYED IMAGES

A Thesis
Presented to
The Academic Faculty

by

(Jack) John Erwin Tumblin

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

College of Computing
Georgia Institute of Technology
December 1999

THREE METHODS OF DETAIL-PRESERVING CONTRAST REDUCTION FOR DISPLAYED IMAGES

Approved:

Greg Turk, Chair

Jessica Hodgins

Irfan Essa

Holly Rushmeier, IBM TJ Watson Research Center

Gary Meyer, University of Oregon

Date Approved

Acknowledgments

Research work presented in Chapters 5 and 6 was supported in part by NSF NYI Grant No. IRI-9457621, by funding from Mitsubishi Electric Research Laboratory and by a Packard Fellowship, each to Jessica Hodgins. This work also resulted in a paper published in ACM *Transactions On Graphics*, January 1999 issue [THG99] and two images from this paper (Figure 19 in Chapter 5) were chosen for the cover of the October 1998 issue. Early versions of the layering and foveal methods were developed during a 15-month extended student internship at Microsoft Research, graciously arranged by Brian Guenter. I also thank Houghton Mifflin Company for granting permission to use the copyrighted drawing by Chris Van Allsburg in Figure 7 for a small fee.

I am thankful for the insightful guidance, help and support I received from Greg Turk in the research that led to the LCIS solution. His deft, meticulously attentive and cheerful approach to difficult problems often leads to elegant solutions. I also thank Arno Schödl for thoughtful discussions on the continuous version of LCIS in Equation 40, and thank Huong Q. Dinh for critical and timely work on image alignment for Figure 35. Research work on the LCIS method presented in Chapter 7 was supported in part by NSF CAREER grant CCR-9703265 to Greg Turk, and resulted in a paper published in the SIGGRAPH'99 Conference Proceedings [TT99]. Figure 1 from this paper, shown as Figure 35 of this dissertation, was chosen as the Proceedings' title page image.

I am also very thankful to Jessica Hodgins for the help, encouragement and collaboration she provided in the tough times after my return from Microsoft. Her research group meetings re-connected me with the GVU Center, her wise counsel guided me back to productive research and collaboration with other faculty and her gentle insistence on rigorous good-humored scholarship sets a high standard for us all.

Contents

List of Figures

Summary

We are immersed in a world of high contrast scenes that we cannot directly reproduce in displayed images. Night scenes, sunny days and glaring reflections are filled with contrasts measured in ratios of thousands or millions to one, but we use display devices such as CRTs and printers with maximum contrasts measured in ratios of tens or hundreds to one. How can we reduce the large contrasts of a scene sufficiently for display yet still preserve the small contrasts of important scene details and textures made visible by local adaptation processes in human vision?

This dissertation argues that we should first separate the scene into "large features" and "fine details" and then construct the displayed image by combining compressed large features and preserved fine details. Most previous contrast-reducing methods either avoid this separation and suffer some loss of fine details, or perform separations based on linear bandpass filter decompositions that introduce halo-like artifacts in displayed images. Borrowing from computer vision, physiology and visual psychophysics, this dissertation presents three new display methods for high contrast scenes.

The *layering* method uses a new sigmoid-shaped function, similar to the response of film or retinal ganglia, to compress only the illumination components of a computer graphics rendering, preserving scene reflectances and transparencies as fine details. The *foveal* method interactively adjusts the displayed image for best reproduction in a small region centered around the user's direction of gaze. This "foveal" region is preserved as fine detail, and large peripheral features are compressed. The *LCIS* (Low Curvature Image Simplifier) method uses a variant of anisotropic diffusion to separate fine details from large features defined by scene boundaries and smooth shading. Only large, simple, but sharply-bounded scene features are compressed. Example images made from extremely high contrast scenes demonstrates how successfully each method captures visual appearance.

Chapter 1

Introduction and Thesis Statement

> **"Daylight is too easy.**
> **What I want is difficult–the atmosphere of lamps and moonlight."**
> *–Edgar Degas, circa 1895 [Deg98]*

Appearances are deceiving, especially when compared to light meter readings. Our mental assessments of scene contents are often markedly different from direct physical measurements of the light available to our eyes. The differences between appearances and measurements are largest at the outer extremes of human vision; in the scenes of lamps and moonlight that attracted Degas, or on a stroll through a dense forest pierced by brilliant shafts of sunlight, or while taking careful steps at night down a forest path lit only by starlight, or during a car ride down a dark country road as we shift our eyes between stargazing and the harsh glare of occasional oncoming car headlights. This dissertation addresses how we might better capture the appearance of such extremes in a displayed image.

This work imposes firm definitions for the terms *scene* and *image*. A scene is the input to any picture-making process and an image is the output. Waves crashing on cliffs on 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 quantities such as "warmth" that are the domain of talented artists like Degas. An image is also objective and measurable; it is the reflectance of the paint or the radiances of a CRT's display screen and not its visual appearance.

The mismatch between appearance and measurement pervades the entire range of vision and is not a simple one; our eyes are excellent as detectors of scene details but are careless and poor as light meters [RW69a]. For example, in a dark room, an ordinary 35mm slide projector with

an empty slide tray may appear to project a uniform white rectangle against a white screen, but measurements with a light meter will reveal that the center of the rectangle is about twice as intense as its corners. Once pointed out, this gross smooth shading is easy to see, as it is well above measured thresholds for detectable contrasts [HF86], but this factor-of-two variation is easily overlooked. However, any piece of lint or hair in the film gate immediately draws our attention. Why does the large, strong, smooth "hot-spot" at the rectangle's center receive much less attention than the tiny, dim, sharply bounded flaws caused by lint or hair? With a photographic slide in the projector's film gate the broad hot-spot is almost impossible to detect, but the hair or lint may be even more bothersome.

Nor does the absolute amount of light from a displayed image greatly affect its visual appearance. For example, a photographic print viewed by indoor office lighting does not appear remarkably different when viewed by direct sunlight outdoors, but the intensity of the displayed image may have increased by a factor of several hundred or more. Though we are aware that increased illumination has increased display intensity, we compensate accordingly; our assessments of the content of the photographed scene does not change.

In principle, the visual appearance of any scene, including Degas' "atmosphere of lamps and moonlight" can be exactly recreated by any display device that accurately reproduces the light that would reach our eyes from the original scene, but in practice this is almost never possible. No current photograph, paint, or electronic display can span all the viewable scene intensities, colors, movements and contrasts our eyes can accept, from a forest by starlight to the dazzling brilliance of a flash-bulb or a glinting snow drift on a sunny day. Instead, Degas' hard problem remains; how can the appearance of extremes of light and shadow be reproduced using only the tiny range of available display outputs?

Appearance-preserving transformations from scene to display, or *tone reproduction operators*, can solve Degas' problem and were first described in the computer graphics literature by Tumblin and Rushmeier [TR93] as illustrated in Figure 1. These operators rely on *observer models* that mathematically transform scene intensities into all the visual sensations experienced by a human observer viewing the scene, estimating the brain's own visual assessments. A tone reproduction

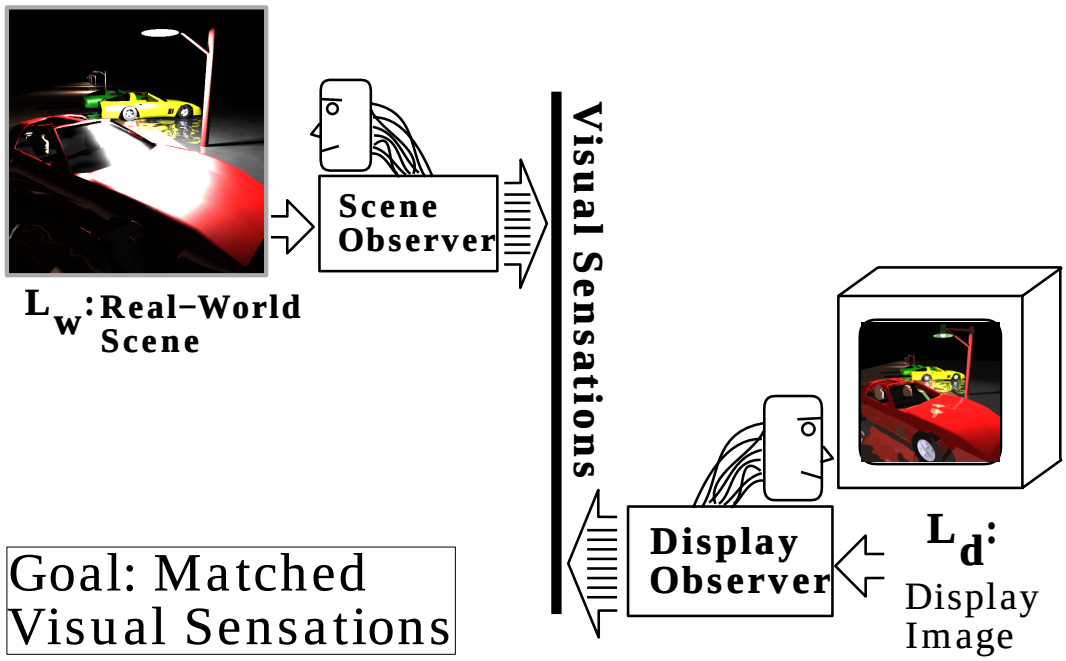

Figure 1: **Ideal Tone Reproduction Operator**
An ideal tone reproduction operator converts original real-world scene intensities L_w to display intensities L_d that evoke the same visual sensations.

operator matches the outputs of one observer model applied to the scene to the outputs of another observer model applied to the desired display image. Tumblin and Rushmeier offered a general framework for tone reproduction operators by concatenating a scene observer model with an inverse display observer model. When properly constructed such operators should guarantee the displayed image is veridical: it causes the display to exactly recreate the visual appearance of the original scene, showing no more and no less visual content than would be discernible if actually present to see the original scene.

Unfortunately, visual appearance is still quite mysterious, especially for high contrast scenes, making precise and verifiable tone reproduction operators difficult to construct and evaluate. *Appearance*, the ensemble of visual sensations evoked by a viewed image or scene, is not a simple one-to-one mapping from scene radiance to perceived radiance, but instead is the result of a complex combination of sensations and judgments, a set of well-formed mental estimates of scene illumination, reflectance, shapes, objects and positions, material properties and textures. Though all these quantities are directly measurable in the original scene, the mental estimates that make up visual

appearance are not. Measurements of these mental estimates by psychophysical experiments are indirect and difficult at best, and no complete, definitive and unified mathematical model of visual appearance exists.

Furthermore, an ideal tone reproduction operator would only guarantee correct visual appearance. For gifted artists such as Degas, conveying visual appearance is only a starting point; such artists capture even the emotions and impressions of the scene by an inspired orchestration of position, shape, color, light, shadow and texture and they may selectively reproduce, modify, suppress or exaggerate any component of visual appearance for better emotional effect. Degas did not write of just the appearance of lamps and moonlight, but of its "atmosphere." If we could somehow reach back one hundred years and give him an advanced electronic camera equipped with an ideal tone reproduction operator (a camera that does not yet exist), who could say what additional image adjustments he might want to make to better capture the atmosphere of his scenes? For his more ambitious intentions, I believe he might ask that we keep all the internal parts of the tone reproduction operator exposed and adjustable, permitting him to choose exactly what and how much of each measurable scene component is sensed by the display observer. Current tone reproduction operators in the literature fall far short of such idealized and powerful artistic aids.

Perhaps the most troublesome and paradoxical task of any basic tone reproduction operator is detail-preserving contrast reduction for high contrast scenes. Local adaptation, the ensemble of local sensitivity-adjusting mechanisms in the human visual system, reveals visible details almost everywhere in a viewed scene, even when embedded in scenes of astonishingly high contrast. Though most sensations of scene contents, such as reflectance, shape, color and movement can be directly evoked by the display outputs, large contrasts cannot. As shown in Figure 2, high contrasts must be drastically reduced for display, yet somehow the displayed image must retain the appearance of high contrast. At the same time it must also include all the low contrast details and textures revealed by local adaptation processes.

Most current scene-to-display mappings are poor at detail-preserving contrast reduction for three good reasons. First, available display contrasts are small and are easily overwhelmed by the scene contrasts, where contrast is the ratio between two measured light intensities. Newspaper

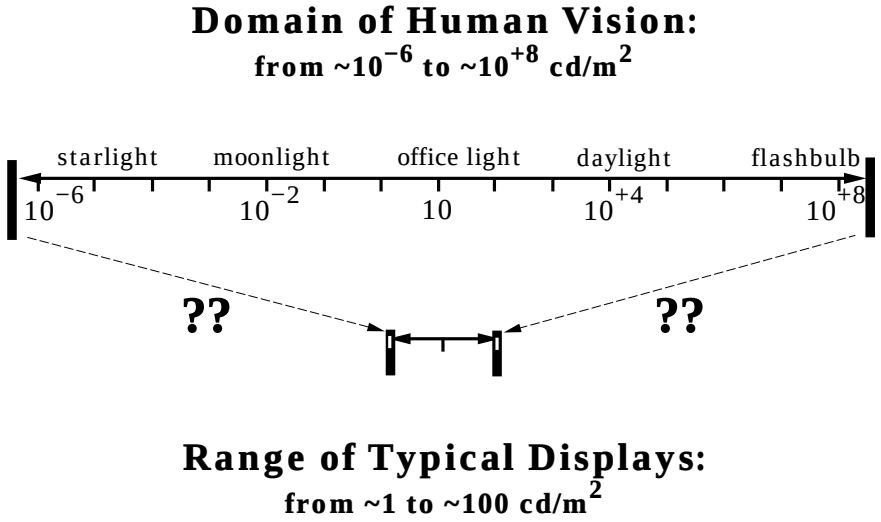

Figure 2: **Display Contrast Limitations**
The input range of human vision dwarfs the output range of most displays. A good tone reproduction operator should map the light intensities of the original scene to the display intensities yet faithfully preserve the appearance of the original scene.

photographs achieve a maximum contrast of about 30:1, typically CRT displays offer contrasts of no more than 100:1 and only the best photographic prints can provide contrasts as high as 1000:1 [Hun75]. However, scenes that include visible light sources, deep shadows and specular highlights can reach contrasts of 100,000:1 and well beyond. For example, contrasts in the foggy night scene depicted in Figure 35 (page 133) were greater than 100,000:1 and in the Stanford Church scenes of Figures 3 and 34 contrasts exceeded 250,000:1. Second, the simplest ways to adjust scene intensities for display will usually reduce or destroy important details and textures. Most commonly used adjustments are borrowed from photography and are given by

$$I_d = F\left(m \cdot I_s^{\gamma}\right) \tag{1}$$

where I_d and I_s are display and scene intensities in SI power units of candelas per square meter cd/m^2, m is a scale factor from film exposure, γ is contrast sensitivity and will compress contrasts for values of $\gamma < 1.0$ and function $F(I)$ limits I_d to fall between display minimum and maximum intensities. The simplest $F(I)$ truncates out-of-range intensities to the display limits, but this discards the fine details and textures in the scene's shadows, highlights, or both, depending on *exposure* or scale factor m. Compressing all scene contrasts uniformly by adjusting film gamma γ

may sufficiently compress the large contrasts for display, but it will also reduce smaller contrasts to invisibility, as shown in Figure 3. Choosing a better limiting function $F(I)$ such as the S-shaped response of film preserves details at mid-range intensities by gracefully compressing contrasts of only the scene highlights and shadows, but small details are badly attenuated at these extremes. Histogram-based mapping functions may preserve more of these details, but also cannot always prevent loss of details because overall scene contrasts greatly exceed display contrasts. Preserving contrasts in some selected ranges of scene intensity forces contrast compression in others.

Most existing computational models of early vision and local adaptation are poor predictors of high contrast appearance, offering a third reason why good detail-preserving contrast reduction has proved difficult. Much of the most rigorous literature of psychophysics concerns vanishingly small signal *threshold* measurements of the visual system, and computational models used in image processing and computer vision literature often apply them using linear bandpass decompositions of images such as filter banks, wavelets and image pyramids. Though these models often incorporate accurate small-signal models of receptive fields and neuronal responses to simple stimuli measured from the visual systems of higher primates, such measurements have proven to be poor predictors of visual response to complex stimuli or to even moderate strength signals such as displayed images [Gra89]. Because the human visual system is strongly and demonstrably nonlinear at almost every known stage, drastically extending linear models for use in high contrast scenes is an experimentally dubious method. Such models cause persistent ugly halo-like artifacts in displayed images, as illustrated in Figure 12 and discussed more extensively in Chapters 2 and 3.

My dissertation research is an attempt to find new scene-to-display mappings to help overcome vast mismatches between scene and display contrasts, yet avoid film limitations and halo artifacts. The three solutions presented here are meant to be practical for use in computer graphics and imaging applications and are consistent with current understanding of the human visual system. These new methods are capable of revealing scene details visible to human observers while avoiding halos or other objectionable artifacts, and are easily controllable for use as an expressive or artistic tool. For simplicity these methods do not address color appearance or other interesting aspects of vision such as temporal response, acuity, visual noise and intra-ocular scattering.

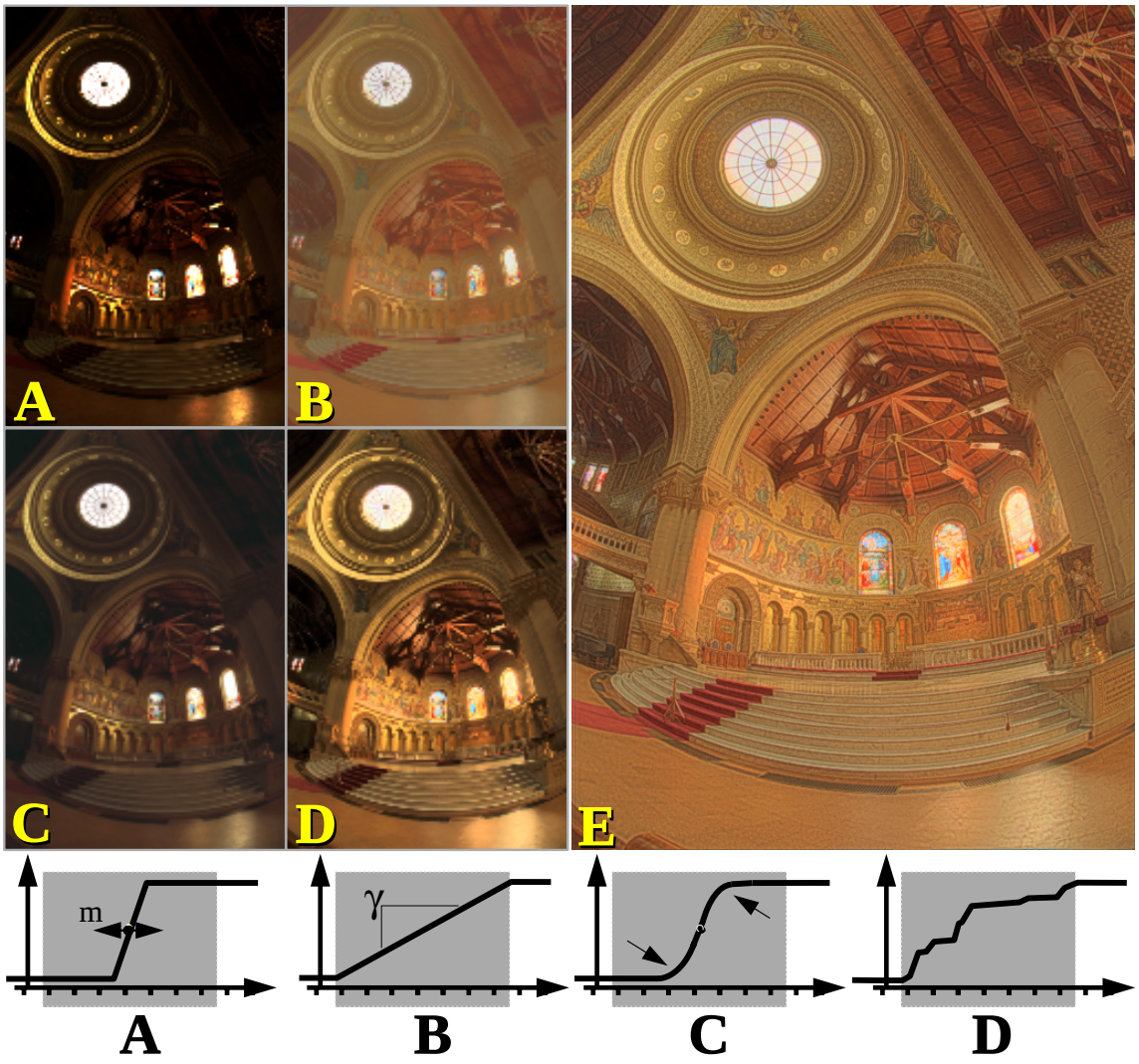

Figure 3: **Scene-to-Display Mapping Comparisons**
Contrasts are greater than 250,000:1 in the original scene captured by the radiance map method [DM97]. A) Truncation discards details in image shadows and highlights; B) gamma compression reveals shadows and highlights, but attenuates textures and fine details; C) Film-like response compresses details in highlights and shadows D) Ward-Larson's constrained histogram equalization [WLRP97] reduces gamma for least-used scene intensities, but E) the LCIS method of Chapter 7 can reveals more detail than is available to any global method by separating fine details from large features before contrast compression. Source image (radiance map) courtesy Paul Debevec, University of California at Berkeley.

Current methods for displaying the measurable contents of high contrast scenes do not adequately separate the viewed scene into the compressible and incompressible components of visual appearance, and contrast compression misapplied to incompressible scene parts leads to erroneous visual artifacts such as halos. Scene separation is the key insight of my thesis and led to the three solutions presented in this dissertation. To achieve dramatic contrast reduction, the compressible components must of course include all the large-amplitude components responsible for the largest contrasts; I will call these the *large features* of the scene. To preserve visible details, the incompressible components of the scene must include the small-amplitude details and textures gathered by local adaptation in the human visual system; I will call these the *fine details* of the scene. This dissertation is then a detailed argument in favor of:

THESIS:

Detail-preserving contrast reduction, necessary for visually accurate depictions of high contrast scenes on low-contrast displays, can be achieved with decompositions of scene intensities into compressible large features and preserved fine details.

In this dissertation, *large features* are high contrast, large area components of the scene, and *fine details* are low contrast, small area components that together describe the entire scene.

This thesis has two parts. First, I claim that detail-preserving contrast reduction is necessary, and second, that it yields well to new appearance-based decompositions of the scene. The bulk of the dissertation supports the second claim; the first only justifies the research done for the second, and is primarily supported by this introductory chapter and by reviewing current image display methods and psychophysics for high contrast scenes in Chapters 2 and 3. The first claim is also supported by comparing previous work presented in Chapter 4 with results presented for the foveal method in Chapter 6, the layering method in Chapter 5 and the LCIS method in Chapter 7. The second claim of the thesis is central, and after motivation by background and previous work chapters, is demonstrated with three working methods for detail-preserving contrast reduction, illustrated in Table 1.

The organization of the dissertation is straightforward. Background Chapters 2 and 3 discuss the problem of detail preserving contrast reduction in considerably more depth than the cursory

Method	Large Features	Fine Details
Layering	Illumination	Reflectance and Transparencies
Foveal	Scene Periphery	Small Region around Cursor
LCIS	Bounded Regions	Intra-region Details

Table 1: **Decompositions vs. Methods**

treatment already given. Chapter 2 surveys well-established methods for scene depiction and image reproduction, explains why these methods have been difficult to improve any further. The chapter ends with a brief summary of difficulties I discovered with multi-scale image intensity gradient estimates Chapter 3 reviews selected ideas from the computer vision and psychophysics literature that support the new approaches presented in this dissertation. Chapter 4 then uses this background material in a critical review of previously published work in the computer graphics literature on tone reproduction and adaptation models for display images.

Chapters 5– 7 each describe and demonstrate a new detail-preserving contrast reduction method developed according to my thesis statement. In the layering method of Chapter 5, the illumination components of a synthetically rendered high contrast scene are considered compressible large features, and scene reflectances are preserved in the display as fine details. For the foveal method of Chapter 6, an interactive display program captures the user's direction of gaze from mouse cursor positions and continually recomputes the best possible displayed image for a small neighborhood around the cursor, regarding the neighborhood contents as fine detail and the peripheral scene around it as compressible large features. The third and perhaps most general method uses "low curvature image simplifiers" or LCIS to separate fine details from large features as described in Chapter 7. Finally, Chapter 8 summarizes the most important accomplishments and applications of the three methods, outlines short-term and long-term improvements that may be possible, speculates on new applications of this work and discusses the difficulties of strong claims of correctness for the solutions presented here. The best confirmations for any detail-preserving contrast reduction method so far is how good or bad the resulting images look; there is not yet a sufficient understanding of human vision to permit objective measurement or predictions of the suitability of results for all possible images.

Chapter 2

General Background

In principle, accurate image-making is easy; just copy the scene to the display. In practice, no displayed image is an exact copy of the scene; it must include approximations and assumptions about visual perception and appearance, broad topics that are still deeply mysterious. Despite their length, this and the next chapter provide only a cursory overview of the vast background literature applicable to visual appearance, and for simplicity the chapters intentionally exclude studies of noise, color and motion, even though each of them are fundamental parts of visual appearance. Instead, this dissertation will only address achromatic detail-preserving contrast reduction and offers only luminance controls to achieve it. Luminance is a measure of radiant power weighted by human visual sensitivity at each wavelength (estimated as 1/threshold measurements), and its SI units of cd/m^2 are a useful expression of visible light intensity.

Contrasts are the central topic of this dissertation. More formally, *contrast* is the ratio between two chosen light intensities I_1/I_2 or luminances L_1/L_2, and I will use only this narrow definition for the term. Other common quantitative definitions include Weber contrast $\triangle L/L$ 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 even as an abstract qualitative term for any important dissimilarity, but contrast always refers to measurable luminance ratios throughout this document.

This dissertation uses the term *display* rather broadly. Though displays certainly include cathode-ray-tubes (CRTs) and photographic film, a display is any device that presents a viewable image. Methods used to make the image are unimportant; only the viewable distributions of image intensities are considered important here. Unusual items become displays by this broader

definition: highway billboards, ice sculptures, oil paintings and microscope eyepieces all convey viewable images, and even "Happy Birthday Bill!" written in frosting on a coconut cake is a valid display.

This chapter discusses common assumptions in existing image-making processes, reviews photographic and CRT image displays, and recounts some of the research results from psychology, psychophysics, neural physiology and computer vision that reveal some related properties of visual appearance.

2.1 Pernicious Image-Making Assumptions

Most image-making processes are terrifically "display-centric"; they gather, describe and manipulate images expressed in display system units instead of scene light measurements, and this bias helps makes scene-to-display mappings easy to overlook or misunderstand. For example, textbooks for drawing, sketching and painting such as [Mor50] use terms such as "tones" or "shades" for both the reflectances of the image and the scene intensities they represent. Photographic film in a camera maps scene intensities to film emulsion transparencies after chemical developing. Digital imaging systems routinely use arrays of integer RGB values to represent images, though frequently neither the intensity nor the color spectrum that each number represents is precisely or consistently defined. Worst of all, many computer graphics rendering methods implicitly assume RGB units are directly proportional to both scene and display intensities, and may regard the minimum and maximum intensities of both the scene and the display as perceived black and white, though often neither are measured nor displayed consistently – recall the "hot spot" at the center of a slide projector's output image. Even the term "image reproduction" implies the displayed image is a simple copy of the original scene.

Strong bias towards displays is understandable historically. Most of the oldest image-making disciplines such as sketching, pastel and watercolor renderings, pen-and-ink drawing and painting matured long before objective light measurements of a scene were available, and instead these methods rely on direct visual comparisons of the appearance of the scene and display. Because usually only the display materials are controllable, artists developed a deep understanding of how

to best exploit their materials to express scene appearance. Photography further reinforced the emphasis on displays. Though photography inherently measures scene lighting by the transparency of film emulsions, the scene-to-display mapping is determined mostly by chemistry and is difficult to control beyond film exposure and gamma (m and γ respectively in Equation 1). Instead, professional photographers often manipulate scene lighting, actually distorting the scene appearance for better display appearance, especially in high contrast scenes [Car95]. Such display-centric lighting is quite different from dramatic theatrical lighting where audiences directly view the lighted result, and on my first visits to film or television sets I was surprised at the flat, uniform appearance of a well-lit scene.

Existing analog television systems such as NTSC, PAL and SECAM [PG80] borrowed their fixed, standardized scene-to-display mappings from film and also use much of the same lighting equipment and terminology [Mil72]. Like film, analog television offers little more than exposure and gamma to control the scene-to-display mapping, and professional lighting for television and film share many important techniques for scene lighting adjustments that help achieve a pleasing display appearance. Also like film, analog television performs most of its scene-to-display mapping in the camera itself. The camera encodes scene intensities as voltage levels, and these are pre-distorted by inverse display gamma to better control the electron guns in display CRTs with minimal processing at the receiver [Hun75]. The RGB units of computer graphics originated as digital expressions of these CRT-controlling voltages, further cementing display units as the standard expression for images.

The bias towards displays in electronic imaging systems may help explain how several simplifying assumptions are so widely held that in many introductory texts [Car95, WW92, FvDFH96, GM98] they pass unnoticed as simple obvious truths that need no examination. This dissertation will challenge each of these assumptions:

- **Assumption 1: Linearity or Contrast Invariance**

  Accurately reproducing all scene contrasts will accurately reproduce its appearance. Equivalently, in Equation 1 if $F(x) = x$ and $\gamma = 1.0$, then a well-chosen scale factor m alone

is a sufficient tone reproduction operator for any scene and any display under any viewing conditions.

- **Assumption 2: Total Ordering**

  Scene-to-display mappings must be monotonic. Equivalently, choose any pair of scene location points (A, B); if the the scene intensities at these points obey $I_s(A) \geq I_s(B)$, then the corresponding display intensities must have the same ordering: $I_d(A) \geq I_d(B)$.

- **Assumption 3: More Contrast is Better**

  Higher displayed contrasts are closely tied to better display quality.

Though these assumptions may seem reasonable on first reading, each of them causes difficulties in reproducing high contrast scenes, and each relies on film-camera-like simplifications of much more complex processes in human vision. The next two sections review photographic film and CRT displays to better explain these film-like simplifications, followed by Section 2.4 offering research results that modify or confound each of these three assumptions and support the thesis of Chapter 1 page 8. Chapter 3 then offers selected background research work supporting each of the three detail-preserving contrast reduction methods presented in Chapters 5 – 7.

2.2 Photographic Image Reproduction

Though chemically complicated, photographic image reproduction is straightforward; scene intensities control the transparency of an emulsion, the chemically and optically active coating on photographic print paper and film [CHJ80]. Monochrome or "black-and-white" film in a camera records scene light intensities as the amount of emulsion transparency that appears after a series of chemical baths known collectively as developing. More scene light intensity causes less transparency in negative films or more transparency in positive or "reversal" films.[1] After developing, negative

[1] Despite their names, negative films swap black for white and reversal films keep whites and blacks as they are. Even 'reversal film' is a misnomer because only the developing processes are radically different; the film emulsions for negative and reversal films are very similar.

film images are usually copied to another negative emulsion for display, either the emulsion on another strip of film or the emulsion coating the front of a piece of photographic print paper.

Film display images are made by passing light through the display emulsion either once or twice. A photographic transparency such as a 35mm slide creates a displayed image by attenuating light from a source placed behind the emulsion, and the image is either viewed directly or projected through lenses onto a reflective screen. This display method can provide light that is more intense than the surroundings, reducing the influence of stray or ambient light on the displayed image. A photographic print display is illuminated by incident light from the viewer's side of the emulsion. Incident light passes through the emulsion, bounces off a reflective backing, usually white coated paper, and then passes through the emulsion again as it travels back towards the eye. Although two passes through the emulsion increases displayed image contrast, the ambient, scattered and reflected light in the emulsion or backing often limits photographic print contrasts, keeping them well below the contrasts available from transparency displays.[2] Image reproduction for color film is similar, except the film holds three light-sensitive emulsions separated by thin colored filtering layers to record different spectral components of the scene light on each emulsion. Color film developing then uses each emulsion to vary the opacity of a different colored dye to reproduce the contrasts of recorded spectral components.

Like several others that will follow Figure 4 plots a scene-to-display mapping on log-log axes that help reveal its effects on contrast. Here the horizontal axis expresses scene luminances, and the plotted curve maps each scene luminance to a corresponding display luminance on the vertical axis. Contrast between any two luminances L_1 and L_2 is given by $C = L_1/L_2$, and on a plot with logarithmic axes the distance between points at the L_1 and L_2 positions measures their contrast: $\log(C) = \log(L_1) - \log(L_2)$. Now suppose L_1 and L_2 are scene luminances that differ by an infinitesimal amount. The scene-to-display mapping converts each of them to a corresponding

[2] At the ACM SIGGRAPH Workshop on Rendering, Perception and Measurement (April 1999, Cornell University) Greg Ward-Larson demonstrated an experimental high contrast display he constructed. By strongly back-lighting just two stacked film transparencies and blocking external light, his device achieves a 5000:1 contrast ratio between its minimum and maximum displayed intensity. Theoretically these same high contrasts might be achieved by an ordinary photographic print if all ambient, scattered and reflected light could be eliminated.

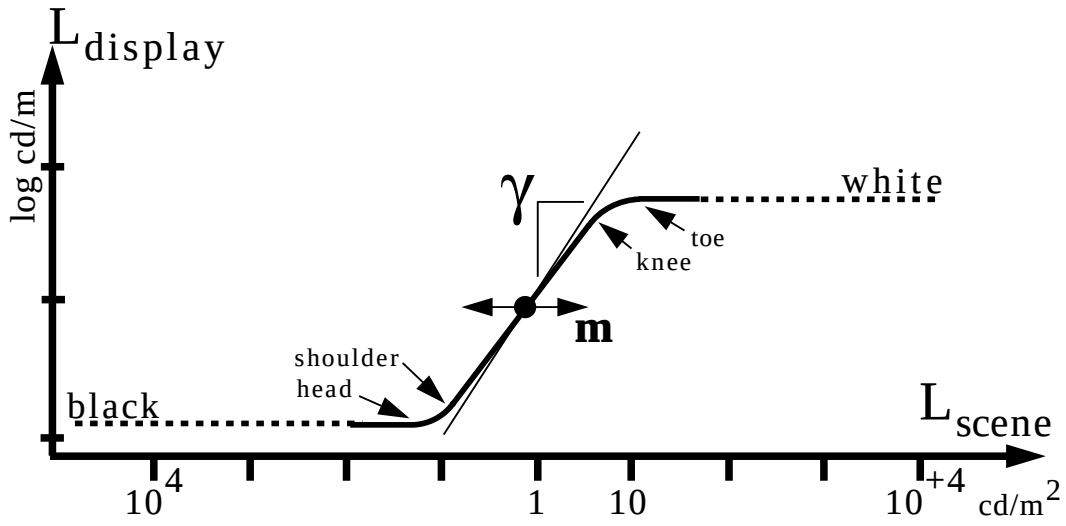

Figure 4: **Typical Photographic Scene-to-Display Mapping**
Photographic image reproduction transforms scene intensities to display intensities using a mapping function that forms a graceful S-shaped curve when plotted on log-log axes. As described by Equation 1, the slope of the nearly linear center region is γ, the horizontal offset of the entire curve is given by scale factor m and the shape of the curve's asymptotic approach to display minimum and maximum is set by limiting function $F(I)$. In film terminology, the nearly linear center region has a width known as the "latitude" measured between the "knee" at the upper limit and the "shoulder" at the lower limit. The outer asymptotic portions of the curve become constant at the "toe" on the upper end and the "head" at the lower end.

display luminance, and the distance separating these display luminances on a log-log plot is set by a derivative. If the derivative of the scene-to-display mapping at L_1 (or L_2) is less than or greater than 1.0 then the mapping reduces or expands their separating distance, and hence their contrast, on the display. Accordingly, a derivative of 1.0 exactly reproduces scene contrasts on the display, a slope of less than 1.0 compresses scene contrasts and a slope greater than 1.0 exaggerates them.

Figure 4 plots a typical scene-to-display mapping for a back-lit positive film transparency such as an ordinary 35mm slide; curves vary slightly for different film stocks. The mapping is a single monotonic curve that usually forms a graceful S-like or sigmoid shape, and this curve is well-described by Equation 1. The derivatives of this S-shaped curve are approximately constant in a broad, nearly linear center region, and fall gradually to zero at the curve's upper and lower extremes, gradually increasing contrast compression of the very bright scene highlights and darkest scene shadows to confine them to available display intensities. If the curve is inverted and the axes are re-labeled, such plots of scene-to-display mappings also describe the response of film to

light. These response curves are commonly found in film manufacturers' literature to describe photographic emulsions (for example, see [CHJ80] and websites for Kodak, Fuji and Agfa), and are known to photographers variously as the film's "characteristic curve,"[3] its "D-Log-E plot," or its "H-D plot," named after Hurter and Driffield who first published them in 1891 [Nel66a]. Hurter and Driffield devised a concise description of measured film response curves that has been borrowed by almost every imaging system since then. They parameterized the nearly linear center region of these plots with three terms; an exposure-like scale factor m describes its horizontal offset,[4] a γ term sets the slope or derivative and the third term that is the region's width, known as the film's latitude. Within this latitude the scene-to-display mapping is simply

$$I_d = m \cdot I_s^{\gamma} \tag{2}$$

where I_d and I_s are display and scene intensities in cd/m^2, m is an exposure-related scale factor that combines effects of the lens, aperture size, exposure time and the film speed or light sensitivity, and γ is the film's contrast sensitivity. Increasing or decreasing m scales the displayed image to make it appear brighter or darker. The γ value is more difficult to adjust because it is set primarily by the chemistry used in emulsion manufacturing and developing, but if $\gamma = 1.0$ the displayed image is just the scene intensities scaled by m, and reducing or increasing γ causes contrast compression or exaggeration as scene intensities are mapped to the display.

The nearly linear center region of characteristic curves of film has a width or latitude of somewhere between 1.0 to 3.0 decades (powers-of-ten or base-10 log units) corresponding to a range of

[3] Color film manufacturers measure a separate "sensitometric curve" for each spectral component.

[4] The m term is a simplification. Because Hurter and Driffield were characterizing film response and not the overall scene-to-display mapping, they separately measured film sensitivity or "film speed" by the symbol i, and also measured photon flux on the film, usually called exposure E, a product of scene intensity, lens aperture and exposure time. To avoid details of cameras and film, the single m term used throughout this dissertation as an "exposure-like scale factor" is the product of i and E terms in photography literature. This m term is used fairly consistently in computer graphics publications on tone reproduction; see [War94a], [Sch95], [FPSG96], [GTS+97], [PFFG98], [THG99] and others.

scene contrasts of somewhere between 10:1 and 1000:1.[5] The height of the curve determines the maximum achievable display contrast, and is also typically somewhere between 10:1 and 1000:1. In the photography literature the opacity of an emulsion is measured as its density D, the negative base-10 logarithm of its transparency; a density of $D = 3.0$ corresponds to emulsion transparency of 0.001. Like gamma and film speed, latitude and maximum density are primarily determined by film manufacturing and developing, though some photographers moderately adjust latitude and gamma by slightly pre-exposing or "flashing" the emulsion before use. Outside the film's approximately defined center region, beyond its latitude, the contrast sensitivity of the emulsion gradually falls to zero at extremely low or high intensity, vanishing into the display minimum or maximum; contrast sensitivity is the derivative of the characteristic curve, approximated by γ in its mid-range region. These outer regions of the film response curve are limited by the maximum possible opacity and transparency of the film, are named the "knee" and "shoulder" where they break from the linear center region, and are named the "toe" and "head" at their outermost limits where all contrast sensitivity vanishes. The names may seem inverted on Figure 4; only the highest scene intensities enter the knee-to-toe region, and the shoulder-to-head region captures only the deepest scene shadows, but the names originated in characteristic curves for negative film, traditionally plotted to sweep downwards with increasing scene intensity, placing the head and shoulders on the left, above the knee and toes to the right.

These outer film response regions are usually reserved for reproducing deep shadows and small specular highlights. If a large portion of the scene intensities are mapped outside the knee or shoulder then the display image is termed "under-exposed" or "over-exposed" and appears too dark or too light respectively. To capture these extremes in film response, Equation 1 adds the function $F(I)$ to Equation 2. The output-range-limiting function $F(I)$ models the asymptotic contrast-compression behavior of film in the knee-to-toe and the shoulder-to-head response regions, and the resulting expression describes the entire scene-to-display mapping for all possible scene intensities.

[5] Photographers often express latitude and contrast in units of "/f-stops," or base-2 logarithms; 8:1 contrast corresponds to 3 /f-stops, 16:1 is 4 /f-stops and so forth, but Kodak characteristic curves use base-10 log units.

In practice, virtually all photographic prints[6] and theatrical motion picture films are the result of at least one emulsion-to-emulsion copying operation; the film in the camera forms a negative image copied at least once by another negative emulsion for display. After developing, negative film from the camera is back-lit to project a negative image onto photographic print paper, or a negative release print film for motion pictures, and the copying emulsion is also processed as a negative to recreate the positive displayed image. If input intensities all fall within the film latitude on each emulsion, then the total scene-to-display mapping is a cascade through multiple instances of Equation 2, and the result can be rewritten in the same form. For example, suppose a photograph taken with negative film described by m_1 and γ_1 is carefully back-lit and projected onto negative photographic paper described by m_2 and γ_2. The scene-to-display mapping is then given by

$$\begin{aligned} I_d &= m_2 \cdot \left(m_1 \cdot {I_s}^{\gamma_1} \right)^{\gamma_2} \\ &= m_3 \cdot {I_s}^{\gamma_3} \end{aligned}$$

$$where:$$

$$m_3 = m_2 \cdot m_1^{\gamma_2} \tag{3}$$

$$\gamma_3 = \gamma_1 \cdot \gamma_2 \tag{4}$$

However, any scene intensities that fall outside the knee or shoulder of any of the film emulsions used will have their intensities and contrasts compressed. Repeated emulsion-to-emulsion copying of high contrast scenes can progressively distort intensities by film response compression beyond the knee and shoulder in the characteristic curve. Any deviation from constant gamma within the film latitude will also be amplified by emulsion-to-emulsion copying using the same film stock. Because motion picture films are often copied several times between the camera negative film and the release print, film manufacturers have designed films intended for copying known as "inter-negatives" and "inter-positives" with wide latitudes and complementary characteristic curves with nonlinearities balanced to cancel each other (for a quick summary, see `www.quantel.com/digfilm/inter.htm`).

[6] Polaroid(R) prints are a notable exception; the display print emulsion is exposed through the camera lens, then very fast reversal processing by paste-like chemicals begins as the print leaves the camera.

Proper selections for γ and m in photographic prints permits accurate reproduction of all scene contrasts within the film's latitude. As the human visual system senses small contrasts acutely over a wide range of scene intensities, accurately reproducing all scene contrasts might seem a good way to reproduce scene appearance. Surprisingly this is not true; people have a strong preference for exaggerated contrasts in photographic displays, and their preferences for reproduction of scene highlights depends strongly on the intensity of display surroundings. Under ordinary indoor viewing conditions, photographic prints made with $\gamma = 1.0$ and m chosen to stay within the film's latitude usually have an appearance that is far too dark and has too little perceived contrast.

As part of a career-long study of tone reproduction beginning with a widely influential paper published in 1920 (see [BB67b], [Nel66b]) Jones investigated viewer preferences for m, γ and characteristic curves. Jones and Nelson [JN42] and Jones and Condit[JC48],[JC49] photographed about 170 typical outdoor scenes and methodically sampled the space of parameters by making several thousand achromatic photographic prints. Aided by several other heroically patient observers, they compared and graded the visual appearance of these prints under typical indoor office viewing conditions (100 foot-candles illumination, where ideal paper with perfect 100% diffuse reflectance has a luminance of 33.80 cd/m^2 [RW69b]) and compiled "first choice" scene-to-display curves for several sets of scenes. On first reading their results are surprising; instead of preferring m values that preserve all scene contrasts equally, observers consistently chose prints where m was high enough to force scene highlights past the knee of the film's characteristic curve, compressing highlight details and lightening the rest of the displayed image. Prints where m kept all scene intensities within the film latitude were always assessed as "too dark." Preferred prints also had gamma values between 1.10 and 1.20, causing a moderate exaggeration of mid-range display contrasts. Any γ value less than about 1.10 was unanimously rejected as "too flat."

Unpublished experiments discussed by Nelson(see [Nel66b] pages 466–467, Figures 22.1-3) and confirmed in later work published by others show these odd viewer preferences vary as display intensities diverge from the intensities in the display's surroundings. The display image from a photographic print is an attenuated return of incident light and is almost never brighter than its surroundings. However, if the print is cropped to eliminate any white border and illuminated by

a light source that is aimed and masked to avoid lighting any of the print surroundings, then the displayed image can be made much more intense without affecting its surroundings. As the print intensity increases, viewers preferred less and less highlight compression, corresponding to lower m values that cause a darker print, a print that does not push scene highlights into the film's knee-to-toe region.

Nelson describes a similar and consistent dependence on display surroundings for back-lit transparencies, and supports them with reports of unpublished experiments. These experiments agree well with the more general and quantitative work of Bartleson and Brenneman [BB67b], [BB67a] who explain viewer preferences in display images as a matching of perceived intensities, also known as brightnesses. Bartleson and colleagues used back-lit and projected transparencies, and demonstrated that, as the display intensity rises above the intensity of its surroundings, viewers prefer less and less highlight compression, lowering the preferred m value to move scene highlights out of the knee-to-toe region of film response. They also showed that viewer's preferences for γ in the scene-to-display mapping depends strongly on display surroundings, varying from 1.5 for a projected transparency in dark surroundings down to gamma of 1.0 for strongly back-lit transparencies in ordinary office lighting. These gamma preferences were also measured independently by DeMarsh [DeM72] who reports that his measurements agree exactly with those made by S.B. Novick in studies for color television viewing.

Bartleson *et al.* [BB67b] suggest that viewers estimate a "reference white" in their surroundings, an estimate of the perceived luminance of a white diffuse reflector such as a piece of paper, and make scene reflectance judgments by comparing the displayed image with reference white. Photographic prints lit only by ordinary ambient light form display images that cannot exceed reference white found from the display and its surroundings, and choices of m for good contrast reproduction of highlights forces mid-range display intensities to values well below the viewer's reference white, leading to the "too dark" appearance reported in Jones' work. Both Nelson and Bartleson explain that viewers prefer to sacrifice highlight contrasts in favor of m values that more closely map reference white in the scene to reference white in the display surroundings. The same m value deemed "too dark" is acceptable if displayed highlight intensities can be raised above the purported

reference white, either by back-lighting a transparency or by carefully masked and aimed print illumination, but light leakage from either form of illumination can raise reference white and defeat the added lighting. If the display image is much brighter than its surroundings Bartleson suggests the displayed image itself supplies the reference white instead of the viewer's surroundings. Exactly how the visual system estimates reference white is unknown, as it seems to require a reliable way to factor display intensities into lightness (perceived reflectance) and brightness (perceived intensity) or illumination, perhaps by relying on a higher understanding of scene contents. Adelson has published excellent examples that show how suggesting 3D shapes or transparencies [Ade93] in a displayed image can cause viewers to assign a wide range of lightnesses to the same display intensities.

Photographic image reproduction offers only limited opportunities to adjust the scene-to-display mapping, usually constrained to only the scale factor m, the contrast sensitivity γ and the latitude. Film gracefully compresses contrasts outside of its latitude in its shoulder-to-head region for shadows and the knee-to-toe region for highlights, and such compression is an important and necessary part of appearance reproduction for photographic prints viewed under ordinary lighting. Viewer preferences for scene-to-display mappings in film are not intuitive; contrast exaggeration and compression are common occurrences, and the appearance of the displayed image depends on both the display intensities and its surroundings. Photographic images are usually displayed in only two forms, either as reflective prints viewed by ambient light or as projected transparencies viewed in a dark room, and the scene-to-display intensity mappings are radically different for each. Cathode ray tube (CRT) displays, perhaps the most common way to display computer graphics results, are an interesting mixture of these modes, viewed under a wide variety of emitted and ambient light amounts. The next section will review CRT displays, their outputs, and some of the studies of appearance and tone reproduction for television CRTs.

2.3 Ambient Light and CRT Displays

Cathode ray tubes (CRTs) are specialized high-voltage vacuum tubes for image display, and after over 70 years of continual improvement they are very nearly the only remaining class of high-vacuum

electron devices still widely manufactured.[7] The long narrow cylindrical neck at the back of a CRT holds a high-current electron gun that aims a steerable, tightly focused beam of electrons towards the back surface of a broad glass display screen at the front of the tube.[8] The back surface of the display screen is coated with a thin layer of rare-earth phosphor powders that emit visible light when bombarded with sufficiently energetic electrons. A yoke of magnetic coils around the narrow neck of the tube steers the electron beam in a scanning pattern or "raster" covering the screen surface, and the electron gun paints an image of phosphor light emissions made by varying the current of an electron beam.[9]

Cathode ray tube displays used in computer graphics and digital imaging borrow many design ideas and standards from analog television. In fact, many early microcomputers used television receivers or modified television monitors as displays for text or images, including products by Southwest Technical Products, Ohio Scientific, IMSAI, Northstar, Commodore, ATARI, Sinclair,

[7] Most of the rest are specialized tubes for high power broadcasting uses such as klystrons and traveling-wave tubes for microwaves and UHF TV, or triodes and pentodes for older radio and TV transmitters. Semiconductors are rapidly replacing even these devices. Lasers, spark gaps, neon tubes, fluorescent lamps, light bulbs, halogen lamps, sodium- and mercury-vapor lamps do not use a high vacuum.

[8] The neck of a color CRT has three electron guns and emits three nearly parallel beams. Unlike monochrome tubes, color CRTs also have a thin shadow mask, (a perforated plate or set of taut vertically aligned wires) near the back of the screen to restrict each gun and beam to one set of color phosphors. One gun illuminates only red phosphors, another hits only green phosphors and the third shines only on blue.

[9] Magnetically steered, rectangular raster-scanned CRTs are now reliable, cheap and by far the most plentiful, but CRTs allow other scanning schemes that are superior in some applications. For example, most oscilloscopes graph electronic signals by steering a constant-current beam across the screen, and deflect the beam electro-statically instead of magnetically using two pairs of thin plates that sandwich the beam and pull it to one side or the other with applied voltages. Some early computer graphics workstations such as the Tektronix 4104 and many air-traffic control radar displays steer the beam to draw an image with many sharply focused line segments known as strokes. Many flight simulators built to train airline and military pilots such as the Evans and Sutherland SPX or McDonnell-Douglas (now FlightSafety) VITAL systems use hybrid raster/calligraphic CRT displays that draw both raster and stroke features on the same color screen. The raster image shows 3D scenes, and the calligraphics draw runway lights and other bright points with brief but steady directed blasts of beam current. Evans and Sutherland has also experimented with calligraphics for special effects such as blowing snow and bright specular sun-glints.

Apple and many others.[10] Most analog television standards were designed for compatibility with motion picture film, minimal broadcast channel bandwidth and simple low cost receivers. To keep receivers simple, the standards require the broadcaster to transmit a signal that is already well suited for the display device. These requirements push the display-to-scene mapping task onto the broadcaster who can afford equipment to adjust it properly, but transmitting only the display-ready information in the television signal limits the adjustments the viewer can make to compensate for display surroundings [Hun75].

Fortunately, the response of cathode ray tube intensities to beam-strength controlling voltages closely resembles the response of photographic film to light, and a scene-to-display mapping method similar to film was used in television standards [DeM93]. Briefly, the emitted intensity of an ideal television CRT display can be approximated by:

$$I_{de} = V^{\gamma} \tag{5}$$

where I_{de} is the emitted display intensity, V is the normalized video voltage that varies between 0.0 and 1.0, and γ is a constant determined by the electron gun design and the circuitry that drives it; the display screen's phosphor material does not significantly affect γ [Poy98a]. The first NTSC standard specifies a television receiver CRT gamma of 2.8 ± 0.3, a rather narrow range for CRT manufacturing. To compensate for this consistent receiver characteristic and help reduce visible noise in displayed image shadows, the broadcast video signal is encoded with a compensating gamma of $1/2.2 = 0.45$; the resulting displayed image gamma is then $2.8/2.2 = 1.27$ and agrees well with the preferred display gamma values found for dim surroundings reported by Bartleson [BB67b], Nelson [Nel66b, JN42], DeMarsh [DeM72], Novick and others. The second NTSC meeting set standards for U. S. color television [DeM93] and also included careful specifications for TV receiver CRT color phosphors.

Computer graphics display systems using CRTs often take hidden liberties when converting digital RGB values to displayed images. In television receivers the V in Equation 5 comes from

[10] However, very few of these computers entirely complied with the rather finicky timing requirements of television standards, causing tremendous problems for videotaping and broadcasting.

broadcast signals already pre-distorted by a gamma of 1/2.2, but in most computer graphics CRT displays the V signal is the output of a digital-to-analog converter supplying V values that are directly proportional to computer-generated RGB values. These RGB values should also include the same compensating gamma of about 1/2.2 used routinely by television broadcasters, but this widely misunderstood "gamma correction" complicates color balancing and conversions [Gla95] is easy to overlook and is frequently omitted [Cat79, Poy93]. The display systems of some computers such as machines from Silicon Graphics, Sun Microsystems or Apple include hardware lookup tables and software support for gamma correction, but others including many PC-compatible video cards have no such compensation, causing image interchange and display problems for shared digital images that are still largely unresolved [Poy98b]. As display monitors have diverged from television designs in resolution and refresh rate, many no longer comply with NTSC gamma requirements and gamma values as low as 1.4 or as high as 3.5 have been reported (see [Gla95]). However, Poynton [Poy93] claims the actual range is much smaller and the wide-ranging numbers are mis-measurements caused by incorrect settings of black-level controls, knobs labeled "brightness" on many monitors. Lack of agreement between computers, monitors, measurements and image coding forces some users to measure and calibrate CRT displays for critical applications [Poy98b, Hal89a, Mey91].

Confusion and half-truths about gamma seem widespread in the computer graphics community, as several highly reputable computer graphics books describe gamma correction as a linearization of the display (see [Hal89b, SP94, CW93]; but [Gla95] found this oversight) and assume that V-to-display mapping gamma of 1.0 is best, in accordance with Assumption 1 on page 12. As discussed in the previous section, studies by Novick, DeMarsh [DeM72], Brenneman [BB67a] and others clearly show higher gamma values are preferred for all but the brightest displays and surroundings. I found Poynton's writings on this topic [Poy93, Poy98a, Poy98b] were best informed, most authoritative and offered hope for resolving the confusion.

In addition to gamma correction effects on display color, manufacturers have substantially improved color phosphors for CRTs [DeM93] since the NTSC color standards were set, and many high-output phosphors with improved spectral purity no longer conform to standard NTSC colorimetry, further aggravating computer color matching problems and the need for fast, accurate

measurements for computer graphics monitors [DeM77].[11] Though many standard computer graphics texts such as [FvDFH96, WW92] offer Equation 5 and explain how to adjust its gamma, most overlook the disruptive effects of ambient light, monitor brightness and contrast controls, and poor CRT high-voltage regulation. These effects can change display appearance as dramatically as gamma correction.

No display is ideal, but real-world CRTs are tremendously variable. Despite Equation 5, providing a CRT display with a video voltage of $V = 0$ obviously cannot create a displayed image of 0 cd/m^2. Ambient light in the displays surroundings give a CRT screen a dull gray-green appearance when the CRT is turned off and this same dim light intensity is added to the displayed image when the screen phosphors are glowing.[12] Excited phosphors emit light in all directions, and some light scatters and bounces within the display face and CRT interior to limit display contrasts even when viewing the CRT in complete darkness. More importantly, almost all CRT monitors and television receivers provide controls for brightness and contrast to improve display appearance for the current display surroundings and viewer preferences. These knobs are usually electronic adjustments for the offset and gain of the video amplifier (3 amplifiers for color) whose output controls the CRT's first grid voltage to regulate the electron beam current.[13] Unlike V in Equation 5, the first grid voltage has a strong negative bias, about -50 volts below the cathode, is not normalized and is easily driven past the electron gun's cutoff voltage. Beyond this cutoff, the beam current and I_{de} fall to zero. To model this beam cutoff behavior, define the function $Clip(x)$ as:

$$Clip(x) = \begin{cases} x & if\ x \geq 0, and \\ 0 & otherwise. \end{cases} \tag{6}$$

[11] Even the phosphors for television receivers have diverged from the NTSC standards. Some television monitors made by Tektronix Inc. and others intended for critical image evaluations include a front-panel switch to re-assert NTSC compliance for comparisons.

[12] Visual adaptation to bright phosphors make screen regions where beam current is zero appear black, but you can verify that these regions are still dull gray-green by rolling a piece of paper into a tube, closing one eye and viewing these regions by looking through the tube. Try turning off the CRT; the appearance of these so-called black regions will not change.

[13] This is a simplification; beam current depends on the difference between the cathode and the first grid voltages. Most CRT displays use fixed-offset video amplifiers to set the cathode voltage and let the brightness control adjust the voltage of the first grid.

Including beam cutoff, ambient light and the brightness and contrast controls in Equation 5 makes a more realistic but still flawed model of CRT response:

$$I_d = I_a R_s + I_{max} Clip\left(K \cdot (V + V_o)\right)^{\gamma} \tag{7}$$

where:

I_d is display image intensity,

I_a is ambient light on the display screen (in cd/m^2, not illuminant units) and

R_s is display screen reflectance.

I_{max}, a constant, is nominal maximum CRT emitted intensity and

K is display's contrast control setting. $K > 0$ and nominally $K = 1.0$.

V_o is black level or brightness control setting and nominally $V_o = 0.0$, but may be positive or negative.

γ is the CRT gamma as before, determined by electron gun design and the linearity of the video amplifier driving the first grid.

V is the display's input video signal limited to $0.0 \leq V \leq 1.0$ and usually found by normalizing values between 0 and 255 for 8-bit integer RGB units.

Log-log plots of curves from the V-to-display mapping of Equation 7 in Figure 5 look and behave like the bristles of a paint brush pushed against the bottom of a bucket of water. Part (A) shows the effects of only the brightness control with zero ambient light where the family of curves resemble splayed brush bristles. When brightness control V_o is positive and large, it prevents small video V values from achieving small display intensities but has little effect when V is large, and the lower, leftward parts of the bristle-like curves are splayed upwards. When V_o is zero Equation 7 is a scaled version of Equation 5, and on log-log axes the equation forms a straight line with slope given by γ; this is the central curve or bristle of the brush. When V_o is negative, it rapidly pulls its sum with V below zero to cutoff, splaying bristles downward and to the right. Now suppose the contrast control K is adjusted up and down; on log-log axes this scaling causes translation, moving all the bristle-like curves up and down as well. Parts (B) and (C) of the graph show the effect of adding ambient light in Equation 7; the additive light provides an asymptotic floor that

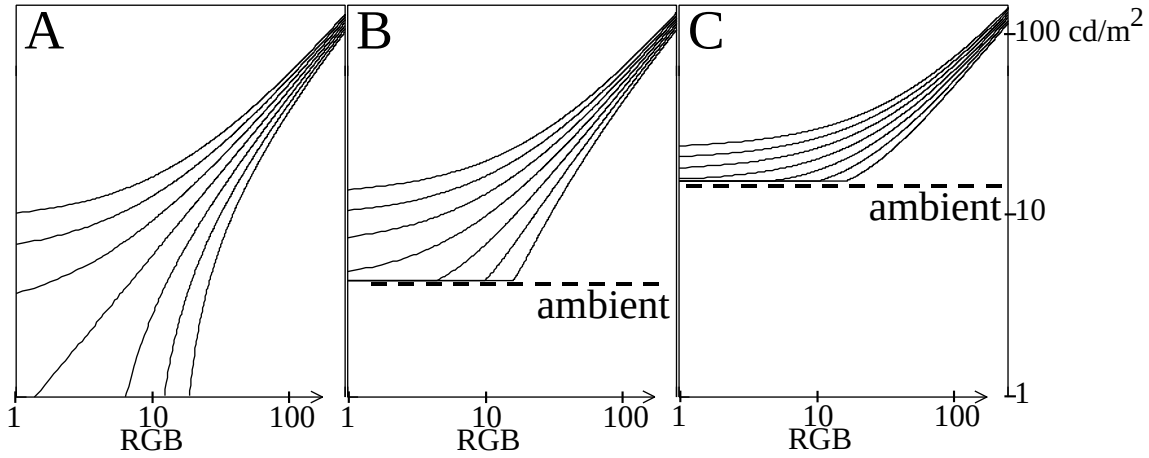

Figure 5: **CRT Response to Brightness Adjustments and Ambient Light Adjustments** Plots of Equation 7 on log-log axes show V-to-display intensity mapping:$K = 1.0$, $\gamma = 1.2$,$V_o = 0.1, ..., 0, ... - 0.1$, $I_{max} = 100$. (A) shows response with no ambient light: $V_o = 0$ response is a straight line, $V_o > 0$ curves the line upwards, $V_o < 0$ pulls the curve downward towards cutoff. In (B), dim ambient light ($I_a R_s = 2.0\ cd/m^2$) bends all curves upwards, linearizing the $V_o = -0.1$ curve above the beam cutoff value for V. In (C), strong ambient light ($I_a R_s = 10\ cd/m^2$) bends all curves upwards, drastically compressing contrasts for small video values V.

distorts the brush bristles. As the emitted light intensity approaches and falls below the intensity of the reflected ambient light, the display intensity changes less and less; each curve can venture no lower on the graph than the minimum or floor value set by the ambient term. If a high brightness value V_o already keeps a bristle above the ambient light level, then adding or deleting the floor term $I_a R_s$ makes little difference, but for smaller V_o the ambient light lifts and splays the curve leftward, just as a brightness adjustment might. But ambient light affects all the bristles and even the $V_o = 0$ curve is now lifted and pushed leftwards asymptotically; instead one of the bristles that was previously splayed rightwards due to negative V_o now closely approximates the straight line of Equation 5. As the plot demonstrates, negative V_o can act to cancel the effect of ambient light on the displayed image, but does so at the expense of driving smaller video signal values V below CRT cutoff. Happily, much of this loss can be remedied by increasing the contrast control K to raise all the bristles upwards. Though this changes the shape of all the bristles near the ambient light level, one of the leftward (negative V_o) bristles will again form a straight line resembling Equation 5.

Adjusting the CRT display controls to maintain this straight-line, constant-contrast sensitivity curve would seem to be the best strategy according to Assumption 1 on page 12, and predicts a

preference for lower brightness settings with higher ambient light. Instead, the opposite is true; even when contrast settings are not changed, with increased ambient light, I personally prefer to adjust brightness upwards, reducing the contrast sensitivity but avoiding cutoff of shadow details in displayed images; informal tests on several displays in the Graphics, Visualization and Usability Lab here at Georgia Tech suggest this is a common preference. As noted earlier, much more formal tests indicate viewers prefer reduced television display gamma with increasing ambient light; about 1.5 for television viewed in darkness, but the preference drops to about 1.2 in dim interior lighting and to 1.0 for bright surrounds. DeMarsh [DeM72] observed that adding ambient light itself reduces effective displayed gamma, but adjusting brightness upwards reduces gamma still further. As shown in Figure 5, increasing the brightness control also further reduces the derivative of the graph for small values of V, further reducing the displayed contrasts in shadows.

However, even this model of CRTs is probably inadequate. This expression for video gain and offset shown here is simple but circuit implementations often have some coupling between contrast and brightness controls; adjusting one affects the other. Displayed images on CRTs also have other secondary effects that more significantly disrupt the simple model of Equation 5. Keeping the electron beam tightly focused, properly deflected (and beams converged for color monitors) and at the desired beam strength everywhere in a CRT is a difficult engineering feat that is imperfectly achieved in even the most sophisticated monitors. For example, a high-current electron beam for a bright display also requires much more power for deflection and focusing; imperfect high-voltage supply regulation causes very bright images to grow and blur on-screen. A blurred electron beam has a lower peak electron flux, causing scanlines that are both broader and dimmer, and because monitor gamma is greater than 1.0 this dimming is exaggerated. The video amplifiers and/or the electron guns can also limit output at excessively high contrast settings, causing contrast compression in the brightest portions of the video screen. Electron beams are more difficult to focus at strongly oblique angles, and larger and better-quality monitors use "dynamic focus" circuitry to adjust focusing anode voltages with deflection angle to improve focus at the corners of a CRT. Similar angular effects on CRT spot size often cause a "hot spot" at the center of a CRT display similar

to those found on 35mm slide projectors, but the smoothly peaked intensity is rarely noticeable without photometer measurements.

All these limitations demonstrate that CRT displays are highly variable, and few assumptions about the V-to-display mapping are safe on uncalibrated CRTs. We can reasonably expect that ambient light and preferences for brightness control settings may reduce the contrast range available for small V values, and over-driven electron guns or video amplifiers may cause similar contrast compression for large V, producing a slightly S-shaped mapping. We cannot rely on a fixed γ at display mid-range values because it is affected by the computer choice, the display, ambient light and the brightness and contrast control settings. Far from being a linear scaling of RGB values to screen intensities, a CRT bears closer resemblance to an unknown film stock when used as a digital image display.

2.4 Challenging the Assumptions

The review of film and CRT displays in the previous two sections has already exposed serious weaknesses in the three assumptions presented at the beginning of this Section (page 12). This section presents additional evidence against them by showing why human vision is profoundly different from photography. Grouped assumption by assumption, this evidence will help reveal why each is erroneous or incomplete and will suggest more promising alternatives for contrast reduction.

Film and television cameras offer an attractive model for human vision that is simple, appealing and wrong. Superficial comparisons between photography and vision show strong similarities, and may erroneously suggest that humans see scenes in very much the same way that cameras photograph them. Many telling parallels exist. Both eyes and cameras sharply focus scene intensities onto a thin, light-sensitive membrane, suggesting the human retina acts as a living, light-sensitive emulsion to sense and encode scene intensities. Most film emulsions [CHJ80], television camera tubes (leddicons, saticons, etc.) and human retinal photoreceptors [WECH+90] each have a usable range of only two or three decades of contrast before reaching limits in highlights and shadows. Eyes and cameras both use several sensitivity-adjusting mechanisms to adapt to the available light in viewed scenes. Both cameras and eyes adjust an entrance aperture or iris to smaller diameters for

increasing amounts of light. In low light human eyes grow more sensitive and respond more slowly with lowered resolution and faintly visible noise; similarly, photographers use longer exposure times and more sensitive, more grainy film stocks, and videographers boost television sensor sensitivity but suffer ghosting, slow response, and noticeably increased grain-like video noise. Film records scene intensities as film transparency, television and computer graphics records intensities as video signal voltages or digital codes, and humans can describe a scene as a map of darkness and light, as if the human visual system also directly senses and records scaled scene intensities as a camera would.

Within its latitude, television and film have nearly constant contrast sensitivity at all scene intensities, or equivalently, the response to light is linear when plotted on log-log axes as in Figure 4. Similarly, a naive extrapolation of Weber's Law suggests that perceived intensity versus measured intensity is also linear on log-log plots, but if any simple curves exist at all they are unlikely to be linear; measurements on photographic transparencies suggest hooked curves [BB67a] that depend on the viewer's amount of adaptation.

Weber's Law was first presented in a massive 1860 book *Elements of Psychophysics* by Gustav Fechner that helped establish and legitimize the field of psychophysics, the quantitative measurement of sensations [Gol96]. In addition to presenting three fundamental measurement methods, Fechner described the work of physiologist Ernst Weber with a simple mathematical "law" named after him (some call it the Weber-Fechner fraction). Using comparisons between pairs of hand weights, Weber measured people's ability to sense weight differences. He measured difference thresholds, or just noticeable differences (JNDs), and found that for all but the very lightest weights,[14] JNDs are very nearly a fixed fraction of the weight itself [Gol96]. Other experiments show similar fixed fractions for other senses including hearing, smell, taste and vision. Fechner interpreted thresholds as indicators of sensitivity, and integrated threshold measurements to construct functions that directly mapped stimulus strength to sensation strength. Plotted on log-log axes, such

[14] weights light enough to be difficult to detect in your hand, such as a butterfly wing

assemblages of threshold measurements form a single straight line, suggesting constant contrast sensitivity. Exactly 100 years after Fechner's massive publication, Stevens [Ste61] offered a fascinating challenge to this idea (see a brief discussion on page 69).

However, the measurements themselves say nothing about our ability to sense *changes* in stimuli, either changes over time or spatial position. Weber and Fechner were concerned with independently sensed stimuli, but neural interconnections among neighboring neural pathways in the visual system interferes with this independence.[15] Television and film cameras sense light independently at different image locations, but the eye does not. Weber's test subjects held weights in each hand, but measurements of visual difference thresholds predicted by Weber's Law (see [Wan95], [PFFG98] or [Poy93] for discussions and [Bla46] or [Bla81] for exhaustive measurements), are usually concerned with the minimum noticeable contrast between a small, round, sharply bounded target region and a large uniform background. For only these very limited conditions, Weber's Law gives a very good approximation over about 8 decades of background intensities spanning 10^{-2} to 10^{+6} cd/m^2. Matching human visual response to film models requires us to badly misinterpret Weber's Law as a general-purpose measure of human contrast sensitivity or "visual gamma."

Though appealing, the eye-as-camera model fails badly at every step. Both eyes and cameras use a thin light-sensing membrane, but film emulsions simply record intensities while the retina performs far more extensive signal processing, leading one author to the book title "The Retina: An Approachable Part of the Brain" [Dow87]. Each eye's retina contains about 125 million light-sensitive photoreceptor cells, but conveys their content with only about 1 million optic nerve fibers arranged in an orderly cable-like bundle leaving the back of the eye. Each fiber is the output tendril (the axon) of a retinal ganglion cell that fires in response to some *change* in local retinal illuminance. Its firing signals a sensed change within the spatial, temporal or spectral content of the scene.

Though eyes and cameras both have multiple sensitivity-adjusting mechanisms, they use them very differently. Most of a camera's broad input range is achieved by wide variations in the chosen exposure value m, made almost entirely by changes in lens aperture and exposure time rather than

[15] for example, local adaptation and lateral inhibition in the retina: see [WECH+90]

changes in film sensitivity.[16] Conversely, most human sensitivity regulation occurs in the retina– the eye's iris cannot change retinal illumination by more than about 7:1, and some authors think its purpose is to optimize depth-of-field for the amount of light available [HF86] rather than adjust sensitivity.

Contrast sensitivity or γ for film and television is reasonably steady for all mid-range scene intensities and is, of course, independent of scene content, but in the human visual system the relationship between measured scene contrast and perceived contrast is not straightforward. Even for small signals, Weber's Law does not generalize well. Constant contrast sensitivity holds true only for small differences between moderate-sized, simple, sharply bounded targets on a uniform background and little else. Difference thresholds are not proportional to background intensities for targets that are too small, (where the square-root-like response known as the DeVries-Rose law applies [SEC83]), targets that are too large (where response follows cube-root-like curves) and targets with complicated shapes (such as Gaussian-windowed sinusoids known as Gabor patches [Gra89, Geo79]). Attempts to merge threshold measurements from sinusoidal and disk-like stimuli into a single mathematical model to reliably predict thresholds for both has also proven cumbersome and difficult [HG92]. Nor are thresholds straightforward for combinations of simple targets such as sinusoids of different frequencies; one stimulus can obscure or "mask" another, nonlinearly raising its detection thresholds [Gra89] and attempts to measure contrast sensitivity above thresholds yield other curves, some of them contradictory [BB67a, Ste61, Ste61, Geo79], and none following simple linear relationships. Seemingly minor flaws in otherwise uniform background luminance can also greatly disrupt thresholds for targets [Gra89], as can the viewer's conscious memory and the expected appearance of the target [DG81]. Reputable texts such as [Gra89, Kau86, SEC83, Cor70, Sch86, Wan95] and signal detection theory [MC91] give stern warnings that threshold measurements are not valid predictors of response to stimuli even mildly above threshold. Though extrapolations may give plausible answers, these answers have little or no experimental validity and are impossible to verify except individually for each stimulus. To apply threshold measurements to high contrast scenes is perhaps

[16] Fuji Fujichrome Velvia slide film at ASA50 and Kodak Royal Gold MAX film at ASA 800 were the slowest and fastest 35mm camera film commonly available at a large nearby photo store, but the latter is only 16 times more light sensitive than the former.

the worst possible abuse of data, as it takes measurements of the weakest, tiniest visual response and stretches them all the way across the harshest extremes of vision.

The light measuring mechanisms of film and television image sensors are extremely uniform in sensitivity and resolution, but the retina is wildly nonuniform; it adjusts sensitivity locally and has most of its color-sensing and detail-resolving ability concentrated in a tiny central region known as the "fovea." Moving along the retina away from the fovea, the type (rod or cone) and density of light-sensitive photoreceptor cells falls extremely rapidly [HF86], even if plotted on logarithmic axes, while local neural interconnections between retinal cells grow broader and more elaborate. Near the center of the fovea a single photoreceptor's output may dominate the response of a retinal ganglion cell and its single outgoing optic nerve fiber, but in far peripheral regions the firing rates of retinal ganglia are each influenced by thousands of neighborhood photoreceptors. [Dow87].

Because film and television sensors integrate intensities over exposure time, photographic a moving scene will blur the recorded image. Human retinas primarily sense changes in intensity, and directly sense and signal the local scene velocities [Gra89]. Moving scenes that cause blurring in cameras do not cause equivalent visual sensations of blur, and such sensations are fleeting and easily ignored. Humans frequently move their image sensors; eye movements help us track moving objects in viewed scenes for foveal examinations of scene features. Conversely, film or TV sensors are kept stationary within cameras, and professional camera operators keep most camera movements small. Large camera movements such as tracking shots are kept simple and smooth to avoid nauseating the display viewer.

Human eye movements are far more than just a tracking convenience; we are blind without them. Our eyes continually tremble, jump and glide to wipe the retina across viewed scene intensities, and even when viewing static scenes these movements are essential. Even a fixed stare at a stationary scene includes small ocular tremors or nystagmus, a faint high frequency shaking caused by the eye's positioning muscles that continually perturbs the scene falling on the retina, moving a distance of a few photoreceptor diameters measured at the fovea. If this movement is canceled by anesthetizing eye muscles or by a precise optical apparatus, viewers report a curious form of temporary blindness; the scene entirely disappears and viewer has no sensation of light or dark, as if asleep [Gil90]. These

and related experiments strongly support claims by Gilchrist and others that visual appearance is a reconstruction of scene contents built from sensations of change in retinal illumination (see [Gil94], pages 1–34).

Television or film photography analogies for human vision may help explain the broad acceptance of the three assumptions presented in Section 2.1 on page 12. If eyes behaved as biological film cameras then tone reproduction is easy. With film-like retinas in our eyes the tone reproduction operator of Figure 1 collapses to the simple film copying rules given by Equations 3 and 4. Furthermore, if human contrast sensitivity is describable by γ in Equation 1 and is approximately constant, as Weber's Law seems to suggest, then γ for the scene observer and inverse display observer in Figure 1 cancels, and indeed any tone reproduction operator will consist of only Equation 1 where $\gamma = 1.0$, as stated in Assumption 1(Linearity or Contrast Invariance). Assumption 2 (Total Ordering) would then follow by inspection because Equation 1 is monotonically increasing, and Assumption 3 (More Contrast is Better) seems reasonable because more available display contrast reduces the amount of contrast compression necessary for scene highlights and shadows. Other evidence against these assumptions is also strong:

Assumption 1: Linearity or Contrast Invariance

Film-like models of vision are common and often implicit in photography literature, but most publications do not agree with Assumption 1; accurately reproducing the contrasts of a scene is not always the best way to reproduce its appearance. In a hugely influential 1920 paper, L. A. Jones implicitly used film-like vision models to construct tone reproduction operators for film and printing. His method plots the desired characteristic curve of an operator using a 4-quadrant graphical method that neatly depicts the transformation from scene to film to display. Jones' method is still widely used [Hun75, DeM72], and may be the origin of the term tone reproduction. However, as discussed in Section 2.2 Jones' tone reproduction operators were not devised from psychophysical data such as Weber's Law, but by careful testing of user preferences. His work and later studies by Bartleson, Brenneman, DeMarsh, Novick and others discussed in Section 2.2 showed viewers prefer some contrast exaggeration depending on display surroundings, strongly suggesting different values

for the "visual gamma" of the scene and display observers in Figure 1 and a dependence of both γ and m on display or scene surroundings. Largely unknown or overlooked in the computer graphics literature, these results have not displaced Assumption 1 as they should.

Assumption 2: Total Ordering

Film analogies for vision make the total ordering assumption seem utterly straightforward and obvious, but again, appearances are deceiving; higher measured intensity does not always correspond to higher perceived intensity. Several simple illusions shown in Figure 6 reveal that the human visual system sometimes resorts to a partial ordering when converting scene intensities to scene appearance. In part (A) and (B) of the figure, we see large-area intensity variations where none exist, variations induced by differences in surrounding intensities. Such mismatches between measured and perceived intensity and perception are difficult to explain with an ordinary nonlinear sensor such as photographic film; the effect of surrounding intensities clearly reveals neighborhood or adjacency effects on appearance, but without a well-defined neighborhood size. The Craik-Cornsweet-O'Brien (CCOB) illusion of part (C) is even more difficult; consider measured and perceived intensities along a horizontal scanline through the center of the image. If perceived intensity is computed as some combination of neighborhood intensities, then the perceived intensities on either side of the black-white discontinuity should form a better and better match with increasing distance from the central discontinuity. Instead, the perceived mismatch between the gray interior and gray exterior of the figure stays large and does not improve.

These partial-order illusions are much easier to explain if we regard visual appearance not as a nonlinear transducer of intensity, but instead as an imperfect reconstruction of intensity built from sensations of change, such as local estimates of intensity gradients. Over-estimating large gradients or contrasts at patch boundaries in Figure 6 (A) and (B) would exaggerate intensity differences along the borders of the patches. Small gradients are difficult to measure over small neighborhoods in an image, suggesting that in part (A) the zero-gradient patch interior is indistinguishable from the tiny gradients necessary to plausibly "fill-in" between the exaggerated intensity estimates at the patch borders. In the CCOB illusion of Part (C), outer boundaries suggest zero gradients at

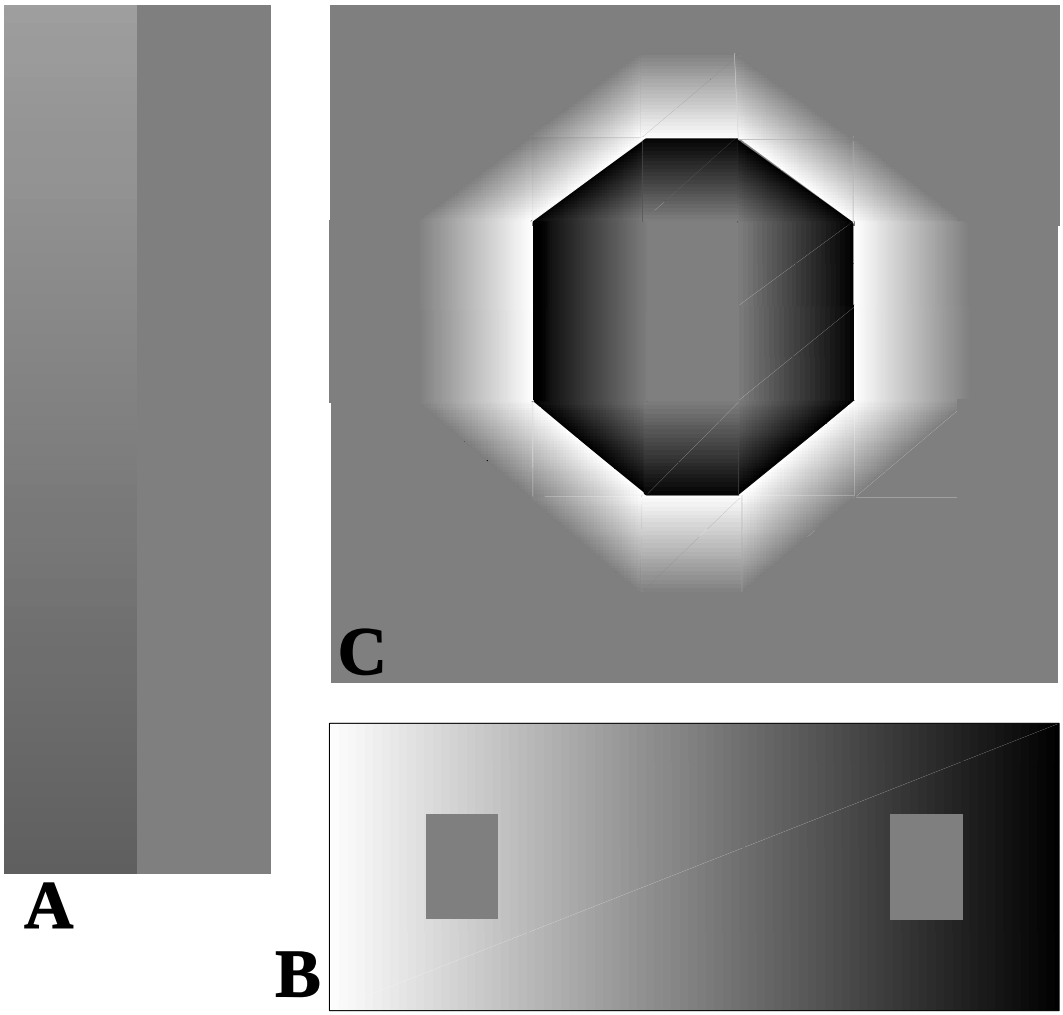

Figure 6: **Failure of Assumption 2: Total Ordering**
Each of these illusions demonstrate a failure of the human visual system to impose a total ordering from measured intensity to perceived intensity; instead, the we see a partial ordering that seems to preserve local relationships across boundaries of large features in the display, but permits incorrect gradient and intensity estimates. In (A), only the left strip has nonzero intensity gradient, and despite its appearance the right gray strip is NOT shaded. In the "simultaneous contrast" illusion (B) both gray squares are precisely the same intensity. The Craik-Cornsweet-O'Brien (CCOB) illusion in (C) induces the gray region at the center to appear darker than the gray at the outer edges of the figure by gradually reducing gradients on either side of a step-like boundary.

the left-most and right-most extremes and the visual system may extend this zero-valued estimate into regions of small but nonzero gradients nearer the central discontinuity, underestimating them to reconstruct different perceived intensities on left and right.

The CCOB illusion allows some exploration of the locality of this partial ordering. Not surprisingly, reducing the width of the central discontinuity eventually weakens and then eliminates the illusion; in the limit the discontinuity appears only as a thin black line adjacent to a thin white line on a uniform grey background; clearly the illusion depends on a minimum neighborhood size. Kingdom and Moulden [KM88] showed that the CCOB illusion is also weakly cascadable; multiple copies of Figure 6 part (C) placed end-to-end strengthens the perceived mismatch, but the difference is not additive and is limited by outer boundaries with its surroundings. They tried eliminating these outer boundaries by curving the central discontinuity and extending it to form a closed circle, and found concentric illusions have a measurably stronger cascading effect, perhaps because the closed circle eliminates a boundary that permits comparisons to the same background intensities. Both results suggest that the visual system must fall back to a partial ordering only for large, smooth and sharply bounded features that keep non-adjacent intensities well separated, and can reasonably maintain total ordering over small neighborhoods of fine features.

Visual estimates of scene intensities depend on more than just the measurable boundaries and gradients of scene intensities; even inferred boundaries and 3-D shapes are a strong influence on visual appearance. Kanisza, Gerbino and their colleagues (see [Kan90] for a quick summary) offer a family of simple illusions to illustrate that the human visual system contains low-level mechanisms to infer the location of obscured scene boundaries. Some authors suggest these sensations are computed by so-called "end-stopping cells" in V1 of the striate cortex, cells that respond strongly to ends of line segments used as test stimuli in Hubel and Weisel's extensive mapping of cells in the visual cortex of macaque monkeys. Another elegant example by Adelson [Ade93] also shows that inferring 3D shape within a displayed image strongly influences mental estimates of scene intensities.

The existence of such partial ordering in visual appearance suggests a new way to reduce scene contrasts for display. First, separate the scene into multiple components that obey the partial

Figure 7: **"The House on Maple Street** -*It was a perfect Lift-off."*–Chris Van Allsburg
This low contrast (<50:1) image [VA84] preserves high contrast appearance and scene details. Both the rocket exhaust streaks (perhaps 10^{+8} cd/m^2) and starlit clouds overhead (typically 10^{-5} cd/m^2) are clearly visible. From *THE MYSTERIES OF HARRIS BURDICK*.

ordering. Without a thorough understanding of visual appearance these components are not rigorously defined, but they might be scene regions such as the sky, the house, the street and and the rocket exhaust as in Figure 7, or they may be scene factors such as illumination and reflectance as in the layering method of Chapter 5, or some other plausible decomposition. Next, apply a separate scene-to-display mapping function for each component, perhaps Equation 1, choosing each mapping to best preserve the details of each component, obey the partial ordering of intensities with other displayed components, and to further reduce the overall display contrasts. Consider the these multiple mapping functions in reverse; every display intensity now represents multiple scene component intensities, and each might be different. Multiple mapping functions offers a simple way to re-use display intensities and break the detail-destroying one-to-one relationship between scene and display intensities imposed by film and television methods. Such orderings are in fact already in widely used by skilled artists, either by the "coarse-to-fine" drawing method discussed in Section 3.3 and shown in Figure 7, or by photographers using the darkroom technique known as "dodging and burning" discussed by [CHS+93] and in Section 4.1.

However, the existence of partial orderings in human vision is also troubling; no definitive tone reproduction operator for high contrast scenes can exist without a thorough understanding of the mechanisms underlying this curious aspect of visual appearance.

Assumption 3: More Contrast is Better

For most imaging systems, higher displayed contrast is considered higher quality. For example, advertisements for photographic print papers tout "rich, deep blacks" and "vivid colors" caused by high contrasts between spectral components. Low contrast photographic prints are often dismissed as "muddy" and less desirable [Sch92], and recent announcements for new TFT-LCD flat-panel displays such as the Super High Aperture displays from Sharp Corporation (see `www.sharp.co.jp/sc/library/lcd_e/s2_5_1e.htm`). advertise contrasts as high as 300:1, even though ambient light falling on the display panel in typical office lighting is very likely to limit achievable contrasts to much less than this lofty figure. Studies of viewer preferences by Novick and DeMarsh suggest television viewers crave contrast. At the 1964-65 Worlds' Fair, Bartleson and

Brenneman [BB67b] created a memorable and widely acclaimed photographic display with large, high resolution back-lit photographic transparencies placed within a 'light trap' in a dark room that maximized contrast and prevented "spill" light from illuminating surroundings. With no other visible surroundings, viewers adapted only to the displayed high contrast image and the result was breathtakingly vivid and appealing (I remember it!). Perhaps some of the visceral immersion effects of head-mounted displays are due to similar effects of contrast and optical isolation.

Though artists and illustrators may have created a virtue from necessity, high contrast displays are much less consistently valuable in more traditional imaging methods. For example, the deep shadows and deft highlights in a Rembrandt oil painting adds drama unavailable to all but the best photographers, even though contrasts of an oil painting often cannot match the contrast abilities of a good photographic print. Conversely, the best watercolors seem to intentionally avoid high contrasts [CAS+97], instead expressing appearance and atmosphere with subtle, low contrast washes, shadings and textures, artfully understating features that might appear harsh or distracting in the original scene. Though "realistic" methods for detail-preserving contrast reduction might exploit all the available display contrast, watercolors and other pleasing renderings suggest that artists may want the ability to adjust and manipulate the amount of contrast reduction and detail present in a displayed image. Oddly, some artists have made paper-cut images with no reflectance contrasts at all–the entire image is depicted by shapes cut into clean, unmarked, stacked sheets of paper or fiberboard; all visible contrasts are due only to self-shadowing, paper texture and cut edges. Odder still, skilled artists can convey or suggest all scene contrasts and shadings in a low-contrast line drawing that depicts only carefully chosen scene boundaries, with no measurable shadings whatsoever in the displayed image!

By abandoning these three assumptions we can re-examine our picture-making processes for digital images. Digital imaging systems are no longer bound by the chemically fixed scene-to-display mappings of film. Nor are they constrained to fixed signal-to-screen mappings of broadcast television receivers. Instead, digital imaging systems can split apart picture-making into independent steps: scene capture, image display and scene-to-display mapping. The mapping can be done by any computable algorithm, and this dissertation presents just three new alternatives to film-like methods

for detail-preserving contrast reduction. Clearly the literature of vision, psychophysics, drawing and illustration contain the raw materials for many more interesting rendering methods.

Chapter 3

Specific Background

Building on the general background material of the previous chapter, this chapter will review additional material that supports the three new methods for detail-preserving contrast reduction presented in Chapters 5– 7.

3.1 Layering Method 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 hundred-fold or more.

Based in part on the early computational vision work of Barrow and Tenenbaum [BT78], 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 [Are94]. 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 [Gil90], 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 (probably grass, trees or shrubbery) 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 [Ade93], Arend

and Goldstein [Are94], Henneman [Hen35], 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 [Gil94] strongly supports this multi-dimensional or "layered" view of vision with cogent argument and extensive references.

The layering method presented in Chapter 5 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 estimated 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,
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 [GJ84] that is nicely summarized in [Gil90] 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 8. With simpler stimuli, Arend and Goldstein found judgments of reflectance are extremely consistent under widely varying illumination intensity [JG87] and gradient [JG90]. These experiments demonstrate that the visual system is attuned to detecting reflectances reliably and under widely varying illuminations,

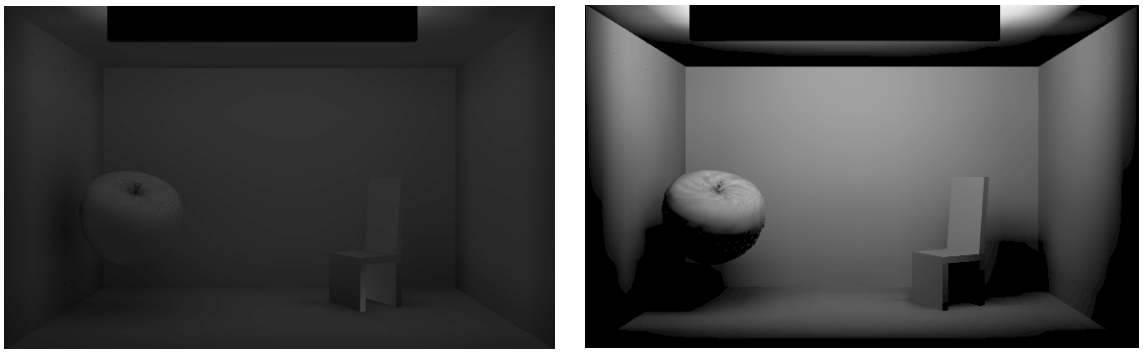

Figure 8: **Illumination can't hide reflectance**
Progressive radiosity renderings of two matched rooms with constant reflectance everywhere (0.90 for the room on the left and 0.03 for the room on the right) illustrate that perceptions of reflectance are not easily disrupted by illumination. Despite 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 [Ash94].

even without help from different reflectances in a complex 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 fine details in the shadowed regions, but the added fill illumination is usually unnoticed. Careful lighting can also severely mislead viewer estimates of scene illumination. To create the appearance of night for motion picture photography, cinematographers may use near-daylight illumination levels to make a recordable scene, but secondary effects convince the audience to see an extremely dim night scene. Such "day for night" scenes may include blue-tinted camera filters to darken long-wavelength colors, very dark low-angle shadows that hide scene details, hidden light sources behind major characters or objects to create strong silhouette boundaries, and camera settings to cause intentional under-exposure to push much of the scene into the "toe" of film response.

Other artists also seem to preserve scene reflectances far more diligently and accurately than scene illumination in images such as Figure 7. Assuming such a fanciful 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 (typical for asphalt roadways [HWS71]), 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 7 applies a partial ordering to combine 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, a simple set of large scene features; 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 [HWS71] rarely form contrasts that exceed $100:1$. This observation is certainly not new; homomorphic filtering methods used in image processing [OSS68, Sto72] 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 to the method presented in Chapter 5 for constructing a detailed but 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 using separate film-like mapping functions. If more than one layer exists, equalize each mapping 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 I 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 [RW69a] showed that test subjects asked to estimate light intensities routinely made errors of 30% or more. More recent experiments by Schubert and Gilchrist [SG92] 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 [Gil94].

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 greatly 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 [WECH$^+$90]. 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 extremely difficult to measure, but experiments by Sakmann and Creutzfeldt (1969) and others (summarized in [WECH$^+$90]) have shown that ganglion firing rates in the cat approach a fixed upper limit as local contrasts exceed about 100 : 1, and their plots of firing rates revealed a family of smooth asymptotic curves. Retinal ganglion cells may directly encode the small contrasts ($< 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 film-like sigmoid or S-shaped function can form a reasonable model of such perceptual compression. I will construct a sigmoid function in Chapter 5, and use this function in both the layering and the foveal display method.

Compressive sigmoid functions are also fair descriptions of light reproduction by photographic film, and the similarity of film and retinal ganglion response may help explain the innately appealing

and realistic appearance of photographed scenes. As already discussed in Section 2.2, the scene-to-display mapping of film gradually loses scene details at the outer limits of film response. Unlike film, the layering method presented in Chapter 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.

I 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 15 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 partially ordered display image. These observations form the basis for the first display method for high contrast images. In this method, I 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 15 and 19. In Chapter 5, I describe a practical implementation of this method and demonstrate its performance on test images.

3.2 Foveal Method Background

The second method for detail-preserving contrast reduction, 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,[1] 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 [HF86]. 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 7:1. 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 and much of the macula, are particularly fast; their multiplicative adaptation processes can be complete in as little as 50 milliseconds. See Spillmann and Werner [WECH+90]

[1] Or perhaps more accurately, in the macula.

or other textbooks for a good summary of these processes. While temporal effects are important to a complete model of visual adaptation, I will ignore them for simplicity.

The huge input range of the human visual system is largely the result of retinal adaptation processes. As summarized by Walraven *et al.* [WECH$^+$90], 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 here 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 I 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.

I also ignored the periphery for a more pragmatic reason; I did not yet 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 3.1 and Chapter 4, though later work led to just such a method presented in Chapter 7. Instead, I 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.

I have applied these observations in the foveal display program, the 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 film-like sigmoid function devised for the layering method. I will describe this sigmoid function in Chapter 5, develop a new tone reproduction operator used by the foveal program in Appendix A and finally give the implementation details of the foveal display program in Chapter 6.

3.3 LCIS Method Background

The third detail-preserving contrast reduction method is inspired by the use of boundaries and shadings by skilled artists and illustrators. When drawing or painting, many artists capture the visual appearance of a viewed scene using a coarse-to-fine drawing process that implicitly defines a hierarchy of boundaries, details and shadings. Many begin with a sketch of large, important scene features and then gradually add finer, more subtle details. An artist's simplest initial sketches usually contain sharply defined boundaries separated by large, smoothly shaded regions. These first boundaries outline the largest, highest-contrast and most important scene features. The artist then refines the work by adding more shadings and boundaries, also sharply defined, to build up fine details that fill out the visually empty regions between earlier large boundaries, gradually building up a complete image filled with detail everywhere [Mor50].

This method works particularly well for high contrast scenes because it permits separate contrast adjustments at every stage of increasing detail and refinement. An artist drastically compresses contrasts of large features drawn first, then adds the fine details and textures with little or no attenuation to ensure they are visible in the final image. Using this artistic boundary hierarchy also allows the artist to easily emphasize or mute the boundaries of objects, controlling their prominence

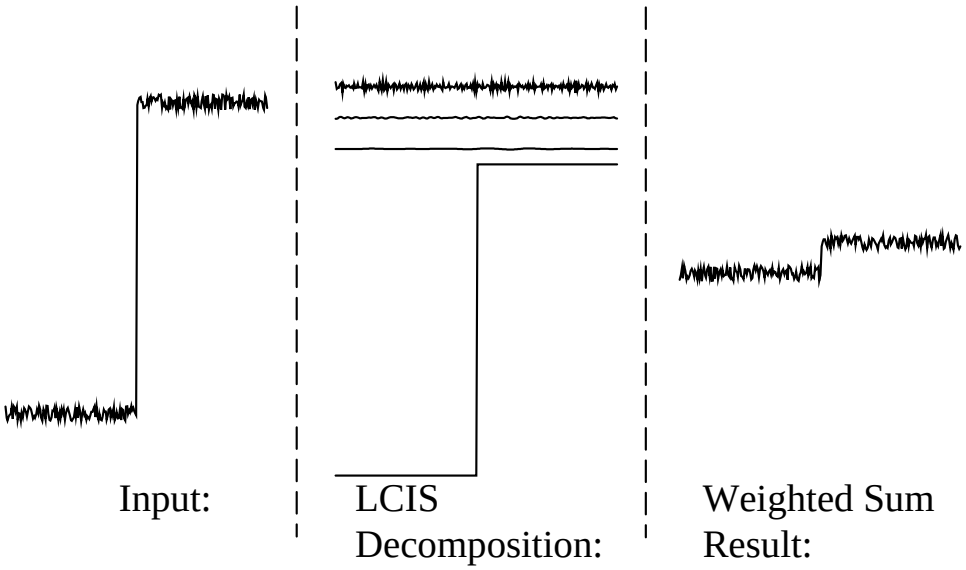

Figure 9: **LCIS on a Scanline**
Applied to a scanline from a high contrast scene, an LCIS hierarchy separates large features and fine details as an artist might. Compressing only its simplest features reduces contrasts but preserves details.

in the image and directing the viewer's attention to important regions of the scene. The low-curvature image simplifier or LCIS method described in Chapter 7 uses modified diffusion equations to form a mathematical imitation of this artist's method.

An artist's progressive refinement of an image into boundaries and shadings is quite different from popular and widely used image decompositions based on linear bandpass filters, such as steerable-filter pyramids [SFAH92], multi-rate filter banks and wavelets [AH92, SDS96]. Instead of a hierarchy of sinusoids, the artist uses a hierarchy of boundaries and shadings. For example, consider a simple high contrast scene made from two adjacent sheets of rough-textured construction paper. A black-colored sheet on the left is dimly but uniformly lit, but a white sheet on the right is illuminated by a strong white light source sharply masked to fall only on the white paper. To an artist, the scene has only one strong boundary and one faint texture everywhere, as shown in the scanline plots of Figure 9 (created by LCIS), but to a linear filter decomposition this is a rich, broad-band scene, as in Figure 10. At its largest scale, the linear filter hierarchy is a blurred wash from black to white showing only that the left and right intensities differ greatly. Every successively finer level contains a strong, zero-mean, ripple-like "detail" that sharpens and narrows the transition

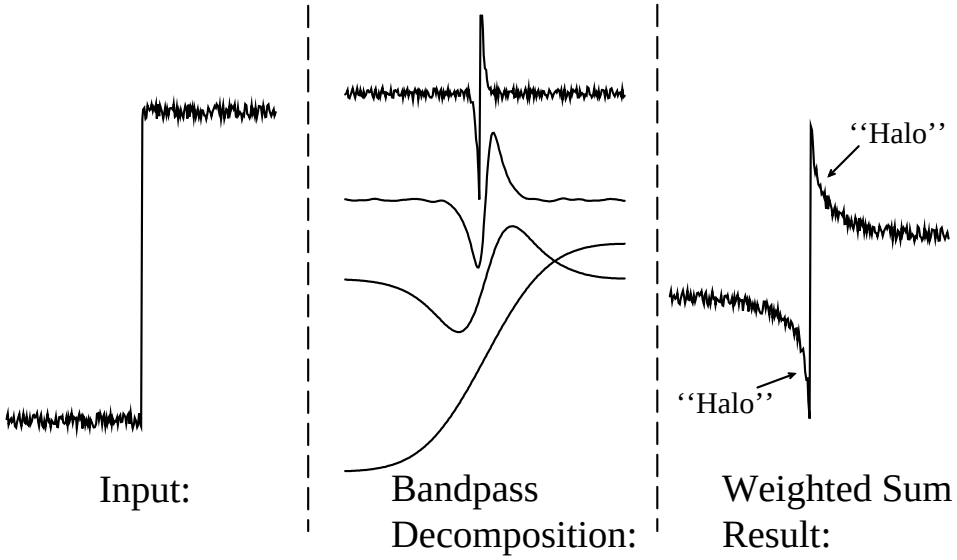

Figure 10: **Linear Bandpass Filtering on a Scanline**
A linear filter hierarchy does not adequately separate fine details from large features. Compressing only the low-frequency components to reduce contrasts causes halo artifacts.

from black to white, as if each finer level were improving the focus of a camera. At the finest levels these focus-like details of the black-to-white transition overwhelm the much weaker variations from the paper texture.

From an artists point of view, linear filter hierarchies fail to completely separate the large features of a scene from its small details and textures. Big strong ripple-like pieces escape from the scene's large, high-contrast features and corrupt the fine details of the cardboard texture. The artists' method of compressing only the coarsest levels of a hierarchy fails with linear filters because it allows the edges of large features to escape attenuation by mixing with fine details such as the paper texture. The resulting display image, as shown in Figure 12, suffers from artifacts known variously as "halos" [Sch95], "overshoot-undershoot" [AH92], or "gradient reversals" [WLRP97].

The scene decomposition into boundaries and shadings commonly used by artists is well suited for detail-preserving contrast reduction, but a linear bandpass filter decomposition is not. Typical frequency-selective bandpass decompositions do not distinguish between sharp boundaries of large features and the small details between them as an artist might. Instead, linear bandpass filters segregate the scene components according to spatial scale alone with no regard to signal amplitude;

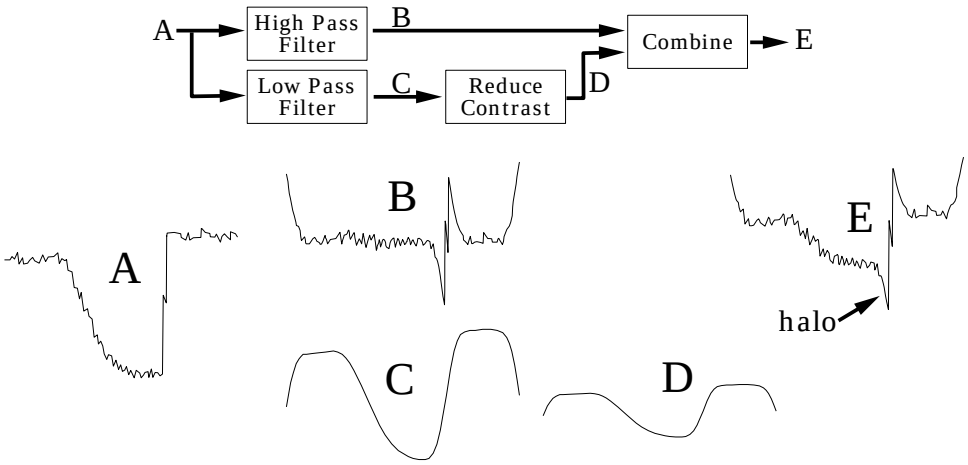

Figure 11: **Halo-Forming Process**
Local adaptation models based on bandpass decompositions suffer from halo artifacts. Separating input signal A into high frequency components B and low frequency components C does not completely separate small details and textures from large, high contrast features as desired; B contains large spike-like features. Attenuating C to make D and combining it with B passes on these spikes to the output image E, causing halo artifacts.

it selects sinusoidal components rather than boundaries, as shown in Figure 11. From a scanline of a high contrast scene (A), the fine detail signal (B) contains large spike-like features at the locations of sharp high contrast scene boundaries. These spikes are caused by the uniform smoothing process applied to (A) to make the large feature signal (C): low-pass filters smooth away local variations without regard to amplitude or large feature boundaries. Reducing the contrast of (C) may provide a displayable large feature signal (D), but combining it with the fine detail signal (B) creates a display image (E) with large spikes that appear as halo-like artifacts around high contrast boundaries in the original scene. Though the method does perform detail-preserving contrast reduction, it permits a halo artifact that can easily overwhelm the details and destroy the appearance of the displayed scene, as demonstrated by Figure 12.

Instead, LCIS offers a new kind of image decomposition that more closely follows artistic methods for rendering high contrast scenes. The decomposition separates an image into an orderly hierarchy of boundaries and shadings. I argue that large features and fine textures are better distinguished by the strengths of their boundaries and shadings than by the strengths of their sinusoidal components. In Chapter 7, I present a new image decomposition that is a hierarchy

Figure 12: **Halo Effects in Display Images**
Using linear bandpass filter decompositions for detail-preserving contrast reduction frequently leads to strong halo artifacts around very bright or very dark sharply bounded objects. Though many scene details are preserved, in the left image note the dark ring around the lightbulb and the light outlines around the dark table legs, and in the right image note similar strong halos at the edge of the bed, at the edge of the wall mirror and above and below the top edge of the chairs near the window. Figure 29 demonstrates how the LCIS method avoids halos in these same scenes.

of *boundaries* instead of a hierarchy of frequencies, using a controllable detail-removing process I call low curvature image simplifiers (LCIS). Inspired by anisotropic diffusion, LCIS is a progressive detail-removing process that drives all image curvatures (the second partial derivatives of intensity) towards zero or infinity, automatically segmenting the image height field into piecewise smooth regions of nearly uniform gradient, bounded by infinitely sharp gradient discontinuities.

3.3.1 Boundaries and Shading

Though no firm quantitative model of high contrast visual appearance exists, there are many tantalizing hints that boundaries and shadings must play a central role. Influential theories of human and computer vision rely heavily on boundaries, such as David Marr's "primal sketch" [Mar82] which proposes we first detect and complete scene boundaries to form a simple sketch of its contents, then use additional information from shading, context, memory, and reasoning to form firm convictions about the materials, objects and lighting of the viewed scene. Classic work by Kanisza [Kan90] on the appearance of illusory contours suggests human vision has sophisticated low-level mechanisms

that allow us to rapidly find, infer and complete hidden boundaries in complex scenes. Robust edge or boundary detection and image segmentation has long been a basic topic of most computer vision texts (see [HS93, Nal93, Can86]), and the computer vision and image understanding literature holds many innovative techniques to achieve the low-level goals edge detection, completion and interpretation. High-level vision approaches such as "shape from shading" and "model-based vision" also rely heavily on estimates of scene boundaries and gradients or shading, and suggests these low-level components play an important role in human vision as well.

Boundaries are a simple and efficient way to mark the position of the most important changes in a viewed scene, including changes from foreground to background, changes from shadow to light, or changes from one object to another, and this efficiency may help reduce the size and complexity of visual processing tasks. Mathematical simulations by Daugman [Dau90] (also see Zetzsche [ZS87]) strongly support this view; boundary-like representations in a linear-filter-based model of the striate cortex reduced entropy of 8-bit source images from $S = 7.57$ (of a possible 8.0) to 2.55, a reduction of 5 binary log units without discarding any image content. Artists intuitively exploit this efficiency: skillful line drawings demonstrate that boundaries alone are often sufficient to express scene contents and may even suggest scene intensities and gradients. Conversely, the lack of distinct boundaries can make judgments of small gradients difficult, as shown in Figure 6, and missing or obscured boundaries can induce misjudgments of scene intensities. Much of the biological machinery in the early stages of human vision are also well suited to find and track important boundaries and gradients within complex moving scenes, and few computer vision methods are as effective as the human visual system at these tasks [Gei89, Gei84].

Unlike film and many computational models of vision, the retina of the eye continually moves against the viewed scene, as mentioned earlier on page 33 converting even stationary boundaries to flickering illumination of retinal photoreceptors, and these movements are crucial to visual appearance. Numerous studies suggest the optic nerve bundle leaving the eye is primarily signaling *changes* in retinal illumination, and not its absolute light level, such as [SG92, GJ84, Gra89, Gil94, Dow87]. In static scenes, these retinal illumination changes are only produced by boundaries, abrupt shadings, scene changes, or eye movements. In addition to quick jumps (saccades) between points of

interest and smooth tracking movements (glissades) across a scene, our eyes tremble slightly at all times (nystagmus), and this trembling is essential to vision. Canceling ocular tremor with specialized optical apparatus that stabilizes scene intensities on the retina causes temporary blindness–in a few seconds the viewed scene fades to an indeterminate nothingness, neither black nor white nor colorless, like the blindness of sleep or long-closed eyelids, but vision returns immediately with any ocular disturbance or normal eye movements.

In related experiments by Krauskopf, Yarbus and others (see Gilchrists' fascinating short summary [Gil90]) retinal stabilization of one or more boundaries in a scene can drastically change a test subject's assessment of intensity, reflectance and color. Krauskopf found that stabilizing the boundary of a large green disk against a red background caused test subjects to report that the background invaded and completely filled in the green region to uniform red. Yarbus [Yar67] found that stabilizing a large region boundary can brighten a small colored patch embedded within it, even though the boundary is far away and not connected to the patch by any intervening boundaries. These and other experiments lead some vision researchers to suspect that lightness and brightness estimates are assembled by some sort of selective integration process, an adding up of changes measured across a selected set of sensed scene boundaries and gradients to estimate reflectance and illumination. Whittle and Challands [WC69] and Land and McCann's "retinex" theory [LM71] (also reprinted in [LM88]) were early tests and applications of this integration idea, and an extensive treatise edited by Gilchrist [Gil94] supports and explores it thoroughly.

Reliance on boundaries for visual appearance may also help explain why our eyes so easily accept or forgive poorly adjusted display devices. Though precisely adjusting the brightness, contrast and gamma of a CRT display often dramatically improves the appearance of its images, most users easily ignore all but the most severe misadjustments if the most important boundaries and shadings are still visible.

However, the boundary integration process for images is ill-defined and both biologically and mathematically troublesome. Mathematically, how should boundaries be defined for more complex images? How should boundaries represent regions of high gradients? Biologically, how is appearance constructed from signals of change alone? Image processing texts such as [GW87] usually regard

the higher derivatives of an image such as gradients and curvatures as pixel constructions, difference signals that are usually too badly corrupted by noise to permit easy estimates of scene boundaries. If higher derivatives are inherently noisier and less trustworthy than direct measurements of pixel intensity, why would the visual system express the scene in this way, using encodings of change in retinal illumination? How does it use them without massive errors; more specifically, how does the human visual system find a stable and consistent estimate of scene contents without the instability and ambiguity that plagues straightforward schemes for vector field integration?

3.3.2 Bandpass Filters and Boundaries

Numerous authors have drawn analogies between the shape of receptive fields, the spatial maps of the inhibitory and excitatory effects of light on neuronal firing rates and the impulse response of linear filters. Further, the extremely regular, hypercolumn structure of the first layers of the visual cortex (V1) bears a strong resemblance to linear bandpass image decompositions such as image pyramids, steerable filters and pyramids [SFAH92] multi-rate filter banks and wavelets [AH92, SDS96]. I thought perhaps these bandpass decompositions might form a reasonable starting point for a model of the early neural processes in the visual system where many adaptation mechanisms have been identified, but was sorely disappointed.

Laplacian image pyramids and related multiscale bandpass image decompositions have several favorable properties for modeling local adaptation. Bandpass decompositions split an input image into multiple output images according to spatial frequency content. Fine image details appear solely in the high-frequency images, and the separation assigns broad, slowly varying components to the low frequency images. The decomposition into multiple images rather than just a fine/coarse pair offers better control of image detail and offers the ability to vary the amount of compression among various larger image features. Pyramid methods compute pixels in the low-frequency images by efficiently finding a weighted sum of local neighborhood pixels in the original image, and might plausibly serve as estimates of the amount of local adaptation at each pixel. Image pyramids are also precisely reversible; the original input image can be exactly reconstructed from its image pyramid.

Many researchers have presented variations on image pyramids for modeling visual perception processes, such as Blommaert [BM90], Cannon and FullenKamp [CF91], Giesler [Gei89], Haig [Hai93], Pelli [Pel90, Pel91] and Teo and Heeger [TH94, Hee87], Watson [Wat87b, WAJA89, Wat87a] and Zetzsche and Caelli [ZC89] and many more. These image-pyramid models imitate some aspects of the massively parallel signal processing channels of the early visual system, and allowed researchers to test models predicting their ensemble behavior, especially for vanishingly small signals known as threshold measurements. However, accurate models proved difficult to construct (for example, see [DuB92] or [HG92] because the visual system can selectively combine or "pool" the responses of many sensory channels to greatly improve its detection ability for many weak but complex signals (See [Gra89] Part III). Interactions between channels are believed to be even stronger for high contrast stimuli [Gra89], but the nature of these interactions are not known and are difficult to measure.

Accordingly, bandpass decomposition models must be regarded with some suspicion and used with caution. Still, such models are plausible because they form a reasonable first-order approximation of neuronal responses in early visual processing, and are in agreement with the "multiple parallel mechanisms" model of visual thresholds advocated by Donald Hood, Norma Graham and others [Gra89]. However, few of these bandpass models were intended for suprathreshold stimuli, and few if any were used to directly predict visual appearance (perhaps work by Peli [Pel91]). Instead, most were attempts to predict visibility thresholds of complex stimuli, such as a camouflaged military tank hiding in a forest or the effects of texture on visibility of polygonal facets [FPSG97]. Most do not include explicit models of visual adaptation, and most use input images already encoded in display units (RGB,0-255).

My early attempts to model local adaptation to high contrast scenes were extensions to the Laplacian image pyramids of Burt and Adelson [BA83]. Though the more recent work on steerable pyramids by Simoncelli *et al.* [SFAH92] adds directional selectivity and removes aliasing artifacts, I chose the Burt pyramids for simplicity, and intended to substitute steerable pyramids later. The generalized pyramid scheme I used is a straightforward revision to Laplacian image pyramids that permits inclusion of arbitrary functions for smoothing and detail extraction. Generalized pyramids

allowed me to experiment with intensity-dependent contrast sensitivity functions, nonlinear contrast responses such as those measured by Georgeson [Geo79] and later applied by Pattanaik *et al.* [PFFG98], saturation functions to selectively limit contrasts and various encodings of bandpass details, such a ratiometric measures.

My experiments with bandpass image decompositions were disappointing and I eventually abandoned them. I found halo artifacts were inescapable in every variant I tried. Despite a wide variety of smoothing functions including Volterra filters (neighborhood polynomials) and median filters, despite many different detail-encoding functions including add/subtract and rational polynomials, some form of halo artifacts appeared in every test image. I eventually concluded that halo artifacts are inherent in bandpass decompositions because sharply bounded high contrast features will have strong components at *all* spatial frequencies. For example, the step-like discontinuity in Figure 11 is a large image feature that should be compressed; its large spatial extent causes a strong signal in the low-frequency version of the image. However, the discontinuity also appears as a large negative and positive spike in the high frequency images along with the fine details and texture. By appearing with the fine detail, these spikes avoid the compression intended for the large-scale step feature that caused them, and remain behind to cause a halo in the output image.

A second look at the receptive fields in the visual cortex offers another linear interpretation of their purpose; instead of bandpass decompositions, perhaps these cells offer intensity gradient vector estimates. Though center-surround receptive fields of retinal ganglia send bandpass-like signals towards the brain to indicate local contrasts, simple cells in the visual cortex (V1, after layer 4) use these relayed signals to construct multiscale directionally selective responses. These directional receptive fields are widely interpreted as local, directional bandpass filters of about 1 octave, as suggested by threshold measurements for sinusoidal grating patches and summarized by Graham [Gra89], but some suggest they might also be interpreted as a contributor to local estimates of boundaries or intensity gradient vector instead. This gradient interpretation is closer to Marr's primal sketch idea [Mar82], in which early visual processes help partition the viewed scene using intensity edges and estimates of object boundaries, and is also compatible with junction- and

corner-detector interpretations of the response of cortical complex cells. An image pyramid that is holding gradient vector estimates might also avoid the halo problems of bandpass decompositions.

Conversely, the idea of cortical gradient estimators raises several practical difficulties. First, such encodings might exaggerate perceived visual noise. Second, how could the visual system reliably integrate these gradient estimates to construct a stable mental impression of scene appearance? Neural signals from the eye are both noisy and suffer from saturation effects at high contrast boundaries. Further, integration across the entire visual field seems inconsistent with physiology; connections between visual neurons are short and local until the later, least-understood sections of the visual cortex. Vector integration processes are also notoriously unstable; as I recall a deft comment from Jim Kajiya in a 1994 discussion on this topic,

> "I just can't believe my brain performs numerical integration to see things; it never crashes like my computer does!"

However, I found I could address all of these difficulties with a multiscale estimate of image gradient vectors I called the "forward difference" or FD pyramid. Instead of encoding the difference between a pixel and a weighted sum of its neighbors, I encoded changes in forward differences on multiple scales. Using a generalized pyramid, the base of the pyramid stored the forward differences in the x and y direction for each pixel in the original image. The smoothing function used to construct the next higher level of the pyramid finds the weighted sum of neighborhood forward differences and decimates the result, as in a bandpass pyramid, but the detail-extracting function is almost ratiometric;

$$\vec{detail} = \frac{\vec{in}}{1 + ||\vec{bk}||}, \tag{8}$$

where:

$\vec{detail}$ is the vector-valued signal encoding local changes in gradient,

$\vec{in}$ is the local gradient estimate found by forward difference,

$||\vec{bk}||$ is the magnitude of the neighborhood gradient estimate.

When both $i\vec{n}$ and $b\vec{k}$ vectors are large, the detail signal is the forward difference at a pixel approximately normalized by the average forward difference of its neighbors. When the neighborhood forward differences are small or zero, the detail signal is directly proportional to the forward differences. As a result, the amplitude of large-scale step-wise discontinuities are expressed in the coarse, or low spatial frequency layers of the FD pyramid, and the exact position of large step feature is captured in the nonzero portions of the fine or high spatial frequency layers. Though the fine FD images still hold fine details of the large image features such as the step discontinuity, it is encoded ratiometrically; compressing the coarse FD image will scale all the associated large feature gradient estimates on all scales, yet leave fine details untouched in regions of low coarse gradient.

The FD pyramid, once constructed, can also be integrated to form an image by purely local operations in accordance with physiology. Begin by converting the pyramid from ratiometric encoding to ordinary multiscale forward differences. Next, starting at the coarsest pyramid image, which contains 4 pixels, use the forward differences to perform path integrals from the image midpoint to find pixel values. At each successively finer pyramid level, compute new and revised pixel values from local path integrals around previously defined pixels. As multiple paths exist between pixel locations from a previous pyramid level and new pixel locations, each new pixel value estimate can be computed from path integrals weighted by their length in the image plane, improving their accuracy and reducing the effect of any uncorrelated noise that might be present in the forward difference signals. Thus forward difference pyramids should offer resistance to noise, halo-free contrast reduction, entirely local, coarse-to-fine integration for image reconstruction and numerical stability.

I implemented FD pyramids and found that for simple 1D test signals such as Figure 11 they performed as expected. However, experiments on more complex, realistic images revealed serious problems with vector conservatism. Any form of integration to convert a gradient vector valued image to a scalar valued image requires that the vector field is self-consistent, or conservative; the sum of all vectors (or in my case, forward differences) on any closed path must equal zero, or mathematically, the curl of the vector field must be zero everywhere. In a conservative vector field, traveling in a direction perpendicular to the local gradient vector will always form a closed path corresponding to a constant-intensity contour in the image. Non-conservative vector fields are

ambiguous about the image they represent; some paths are not closed. Unfortunately, compressing the large-scale components of the forward difference pyramid often caused non-conservatism. Despite many attempts, I was unable to find an acceptable way to restore or preserve conservatism. Least-squares fitting is one plausible solution, and other authors such as [FDB92] offer more localized redistribution of error to correct conservatism problems, but I found no methods that did not re-introduce halo artifacts.

My attempts to use bandpass and forward-difference pyramids to model local adaptation and high contrast appearance were disappointing, but raised interesting questions. How does the human visual system avoid the vector conservatism problems I encountered? How is visual appearance constructed from signals that encode local intensity changes? How are these changes integrated to build and maintain stable and consistent estimates of scene appearance from the noisy encodings of local change sent by retinal ganglia? Surely the biological signals are noisier, less uniformly compressed and less consistent than the contents of my forward difference pyramids. Why don't these biological errors and inconsistencies cause artifacts or instabilities in the visual appearance of high contrast scenes?

Chapter 4

Previous Computer Graphics Methods

To augment the existing 80-year span of literature on tone reproduction for film, television and printing, the computer graphics community has much more recently addressed the problem of accurately capturing the appearance of high contrast scenes on low contrast displays. Because advances in psychology, psychophysics and neurophysiology have improved general understanding of the human visual system, many of these papers use models of local and global adaptation mechanisms to devise new scene-to-display mapping functions, often combined with the results and, unfortunately, the assumptions from film methods. This chapter briefly reviews the papers I found that were most closely related to the new methods I present in the chapters that follow.

4.1 Local 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 $m(x, y)$:

$$L_d(x, y) = m(x, y) \cdot L_w(x, y). \tag{9}$$

Simple image-based localized gain controls are already used in the "unsharp masking" process used to restore apparent sharpness to photographs reprinted by halftone processes [Hun75]. Though mechanically elaborate, unsharp masking reduces the contrast of low frequency components, and may be regarded as a weak halo-forming process. However, unsharp masking is performed only on display images and does not offer any significant contrast reduction as was required for Figures 35,

29 and 34. The tiny amounts of contrast reduction it offers are sufficient to enhance perceived sharpness, but are not enough to significantly disrupt displayed gradients near exaggerated boundaries.

Professional photographers use a related technique to locally reduce contrasts photographic print 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, as illustrated in Figure 12 in the previous chapter.

Digital and electronic imitations of dodging and burning have shown similar weaknesses. The method proposed by Chiu *et al.* [CHS+93] 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. I believe the scaling function should change abruptly at the boundaries of high contrast features to avoid the halo effect, but constructing a suitable scale function with this behavior has proved difficult.

A later paper by Schlick [Sch95] 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 me to revise the sigmoid function described in Chapter 5 for greater efficiency.

Tanaka and Ohnishi [TO97] 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, Rahman and colleagues [RJW96, JRW97a, JRW97b], 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, based on a line- or contour-integration process that has some support in the psychophysics literature and earlier computational models of vision [CG84, GT88, Gro90]. To implement a process similar to this line-integration mechanism without incurring conservatism problems, Jobson, Rahman and colleagues also use Gaussian low pass filtering to find local multiplying factors, but this choice makes 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 "multi-scale 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 [JRW97a]) 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 15 or in the recent paper by Ward-Larson *et al.* [WLRP97].

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.

Building on a rigorous paper on perceptual image metrics [FPSG97], in 1998 Pattanaik and colleagues [PFFG98] presented an impressively thorough tone reproduction operator that performed contrast reduction using an intricate model of local adaptation assembled from extensive psychophysical data, including acuity, chromatic adaptation, contrast matching and many measured nonlinearities of vision. However, contrast reduction is chiefly due to attenuation in a linear filter hierarchy; despite many admirable qualities, their method is still susceptible to strong halo components.

4.2 Global Methods

With the exception of dodging and burning and the methods of Chiu, Jobson, Tanaka, Pattanaik and their colleagues, most imaging systems do not imitate local adaptation and thus are not bedeviled by halo artifacts. 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 and corresponds to film exposure value m discussed in

$$L_d(x,y) = m \cdot L_w(x,y). \tag{10}$$

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.

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 accurately by the contrasts alone, but scaling display intensities can strongly affect a viewer's estimates of scene 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 [TR93] tried to capture some of these light-dependent changes in appearance by introducing a tone reproduction operator 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 [SS60, SS63] who claimed that an elegant power-law relation exists between luminance L, adaptation luminance L_a (corresponding to the reference white value discussed by Brenneman [BB67b, Nel66b]) and summarized in Chapter 2, and perceived brightness B:

$$B = C \cdot \left(\frac{L}{L_a}\right)^{\gamma}. \tag{11}$$

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 [Kau86].

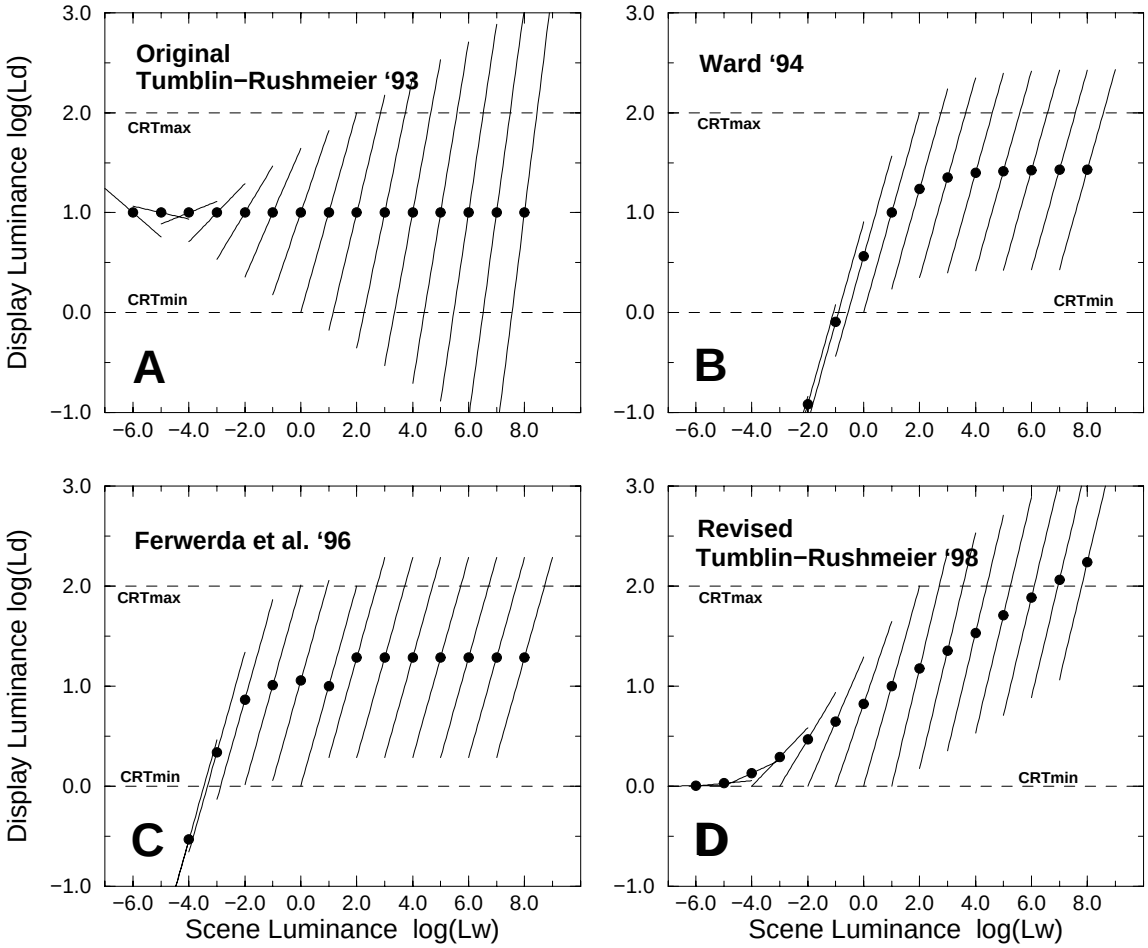

Figure 13: **Tone Reproduction Comparisons**

Log-log plots show input-to-output mappings of the four globally applied tone reproduction operators discussed in this chapter. Each plotted line shows display luminance L_d versus scene luminance L_w for 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, each 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 [TR93] uses Equation 11, and reduced line slope for smaller L_{wa} reduces displayed contrasts. Very dark scenes are displayed as medium gray with reversed contrasts, and very bright scenes exaggerate contrasts unrealistically.
- **(B)** Ward's tone operator [War94a] of Equation 12 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.* [FPSG96] 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 Appendix A, Equation 55 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.

More conventional methods measure only the detection thresholds for simple scene features. Stevens [Ste61] 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 [Ste61]. He attempted to directly measure the complete range of human vision, and the resulting power-law relation agrees reasonably well with both the film-model results of Jones [Nel66a], Brenneman [BB67a] and others discussed in the previous chapter, and the narrower results from more conventional threshold-finding experiments such as those by Blackwell [Bla46].

Tumblin and Rushmeier's tone reproduction operator used the results of Stevens and Stevens but exhibited several serious shortcomings, as shown in Figure 13. 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 Appendix A, I re-formulate 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 Chapter 6 uses this revised tone reproduction operator.

Soon afterwards Ward [War94a] presented a much simpler approach to appearance modeling for computer graphics 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 [Bla81], 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:

$$L_d = m \cdot L_w \tag{12}$$

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 [War94a]. This contrast reduction may be due to the loss of contrast reproduction ability commonly found at the smallest output values of many displays and discussed in Chapter 2. The lower bounds on the display luminance L_d shown in Figure 13 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 13 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 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 [FPSG96]. 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 13. 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 choice of value for m acts 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.* [FPSG96]. 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 7 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.* [NKON90] 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 [SSZG95] 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.

Several tone reproduction operators such as those by Ward [WLRP97], Ferwerda [FPSG96] and their colleagues make extensive use of psychophysical contrast threshold measurements to model the appearance of suprathreshold images, including images of extremely high contrast. I dislike this practice and avoid it in my work for three reasons. First (as discussed earlier on page 31), standard texts such as [Kau86, Gra89] and numerous well known experiments warn that each threshold measurement applies only to a specific stimulus on an extremely uniform background, and is very susceptible to disruptions. Second, the effects of these disruptions are not simple. In some cases, detection thresholds for two or more combined or pooled stimuli are much smaller than the thresholds for either one measured separately, yet the presence of one stimulus can obscure or mask another, greatly increasing its detection thresholds. Vision is inherently nonlinear; responses to large input signals are not proportional to the measured responses to small input signals, especially at the extremes of contrast where saturation effects in the retina limit the ability to detect additional contrast. Third, the practice of interpreting thresholds as measurements of visual sensitivity, especially suprathreshold sensitivity, is demonstrably incorrect. For example, the strongly peaked "Contrast Sensitivity Function" made by plotting contrast sensitivity thresholds for sinusoidal patch stimuli versus spatial frequency is often mistaken for the "frequency response" of the human visual system. As shown by Georgeson [Geo79], this function rapidly flattens and loses its frequency selectivity as stimulus contrast exceeds 10%, (though even this claim is disputed: see Biondini and DeMattiello [BD85]).

Despite substantial progress, I still consider the tone reproduction problem troublesome and largely unsolved. The mismatch between the contrast sensing abilities of our eyes and the contrast reproducing abilities of our displays is enormous. By far, this mismatch is the largest disparity that must be overcome by tone reproduction, but it is here that published solutions are weakest.

I regard appearance preserving contrast reduction as the central problem of tone reproduction because all other light dependent effects, such as loss of color, acuity, noise processes and temporal response, are directly dependent on the solution used for contrast reduction. Current displays can emulate these other visual effects by directly reproducing them; only the high contrasts of a scene are unreachable. Accordingly, precise modeling of visibility thresholds is premature and is easily

invalidated by later work; these are details that deserve close attention only after the appearance of the large scene contrasts have been captured.

Each of the current methods are compromises between sufficient reductions of large contrasts and preservation of the small, low contrast details that decorate them. Some make the trade-off explicit, such as the compressive function of Schlick [Sch95] or the histogram adjustments of Ward-Larson and colleagues [WLRP97]; others make implicit compromises, such as Tanaka and Ohnishi [TO97] or Nakamae [NKON90], still others ignored contrast reduction entirely, such as Tumblin and Rushmeier [TR93], Ward [War94a] and Ferwerda [FPSG96]. Previous attempts to split the scene into large compressible features and small incompressible details by modeling the local adaptation processes of human vision has caused halo-like artifacts in various forms, such as the halos mentioned by Chiu and Shirley [CHS+93], by Tanaka and Ohnishi [TO97], by Schlick [Sch95], and suppressed but not eliminated in the multiscale retinexes of Rahman, Jobson and colleagues [RJW96, JRW97b, JRW97a]. Instead, I am searching for methods to cleanly separate large-scale compressible features from small scale preservable details, with each component represented in ways that allow compression without introducing visible artifacts. The next three chapters each present a new method to address this persistent problem of appearance-preserving contrast reduction.

Chapter 5

Layering Method

Persistent difficulties with locally varying contrast reduction using image pyramids led me to look for a new approach that might be more suitable for simpler globally applied compression functions. Inspired by the "relational" approach to lightness and brightness perception advocated by Gilchrist and others [Gil94], the layering method uses computer graphics renderings from a modified ray tracer to decompose scenes into reflectance, transparency and illumination components.

The layering method assumes the human visual system constructs separate but simultaneous mental estimates of several scene properties at once, but senses scene reflectances far more carefully than scene illumination. This assumption is plausible because small differences in surface reflectance are easily detectable under almost any illumination, no matter how uneven or colorful. Conversely, changes to illumination that preserve boundaries and shadings are easily ignored.

The layering method supports the thesis of Chapter 1 by example; it separates the input scene into large features and fine details, and performs detail-preserving contrast reduction by compressing only the large features of the scene and combining them with preserved fine details. For the layering method, large features of a scene are defined as the illumination components or "layers" of the scene, and the fine details are the reflectance and transparency layers. To find these large features, this chapter presents simple modifications to collect intermediate results during computer graphics scene rendering to record contents of the scene layers defined below.

Figures 14 and 15 demonstrates the effectiveness of the layering method by comparisons using a floating-point 2D scene intensity map made by a modified ray-tracer. In Part (A), display truncation or clipping destroys all details in scene highlights and shadows. In (B), adjusting the gamma value γ in Equation 1 reduces scene contrasts sufficiently to reveal the contents of shadows and highlights, but small contrasts are lost to excessive attenuation, giving the scene a dull washed-out appearance.

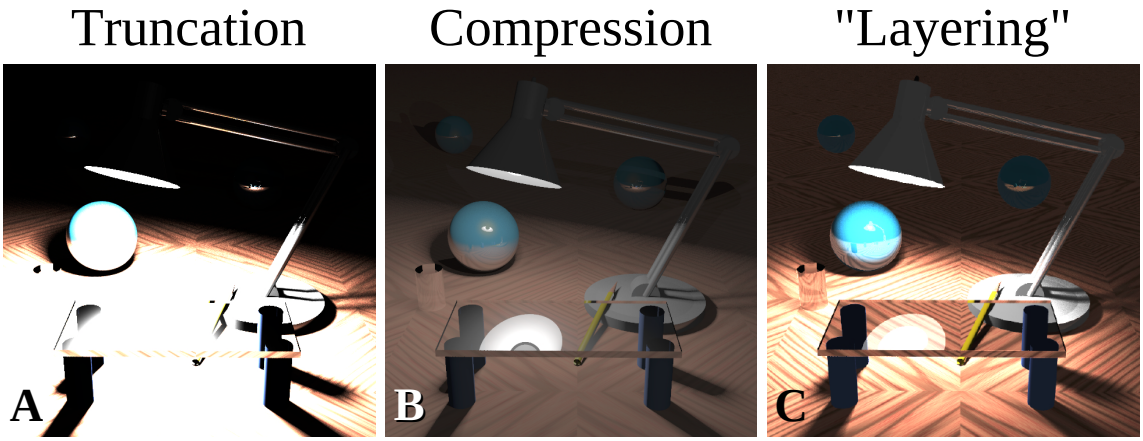

Figure 14: **Layering Comparisons**
Contrasts are greater than 300,000:1 in this synthetic scene. In (A), truncation discards details in image shadows and highlights; In (B), contrast compression by adjusting γ (Equation 1) reveals shadows and highlights, but lacks the appearance of high contrast and attenuates textures and details; in (C) the layering method preserves image details and maintains more of the appearance of a high contrast.

In part C) the layering method has reduced only the illumination contrasts for display and has preserved the much smaller contrasts due to reflectance and transparency. Note that wood-grain texture is visible everywhere on the floor, both in the deeply shadowed background and through the brilliant specular reflection of the lampshade's interior, as reflected in the toy-sized glass-topped table. These effects are achieved by separate processing of scene layers.

5.1 Layer Separations

The layering method uses six different types of intrinsic image layers grouped together in three pairs to represent the original high contrast scene; an example of each is shown in Figure 16. 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:

$$Scene(x,y) = K_d(x,y)I_d(x,y) + K_s(x,y)I_s(x,y) + K_t(x,y)I_t(x,y) \tag{13}$$

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

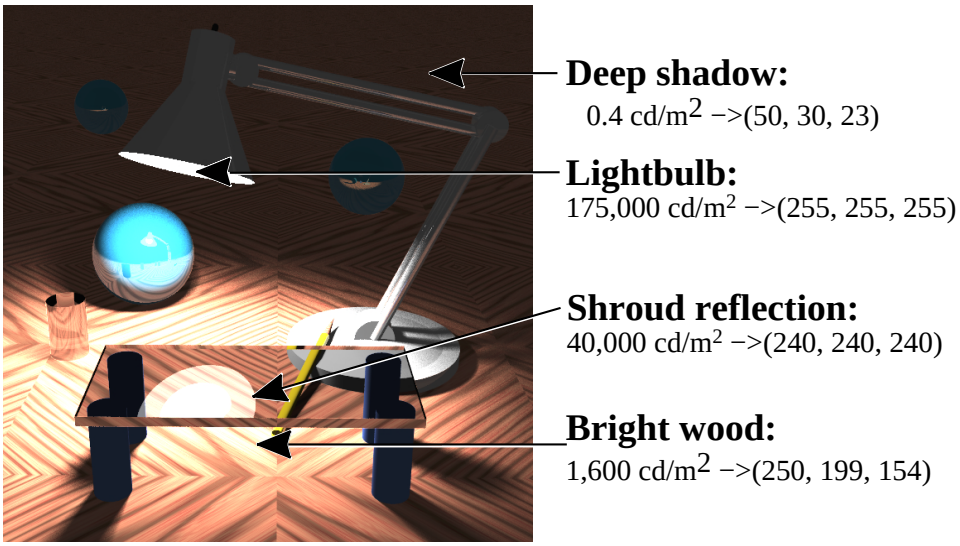

Figure 15: **Layering Results I**
Applying the layering method to a high contrast scene ($> 300,000 : 1$) reduces display intensities to nominal pixel values without discarding fine textures and image details.

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 co-planar 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 $\hat{S}$, 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.

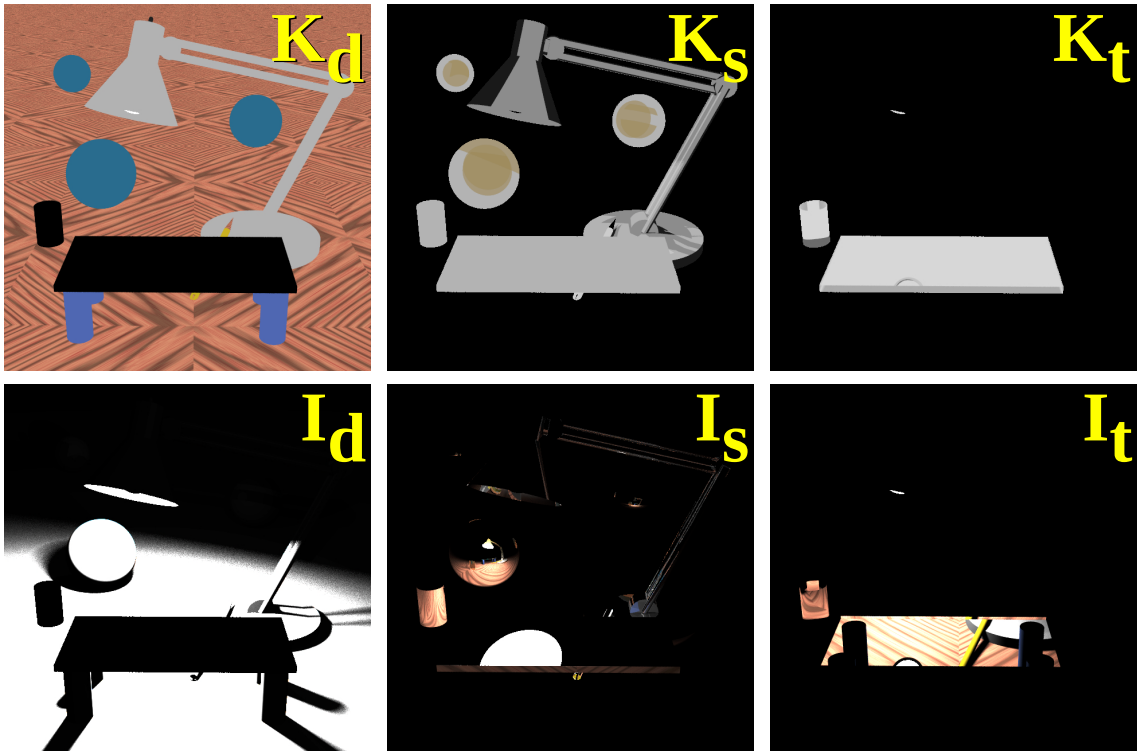

Figure 16: **Layer Images for Figure 15**
The top row shows diffuse reflectance K_d, specular reflectance K_s and transparency K_t. These three images have rich, complex detail but low contrast. The bottom row shows corresponding layer images for diffuse illumination I_d, specular illumination I_s and transparent illumination I_t. These images contain few details but extremely high contrasts.

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 16. 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 Chapter 3 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

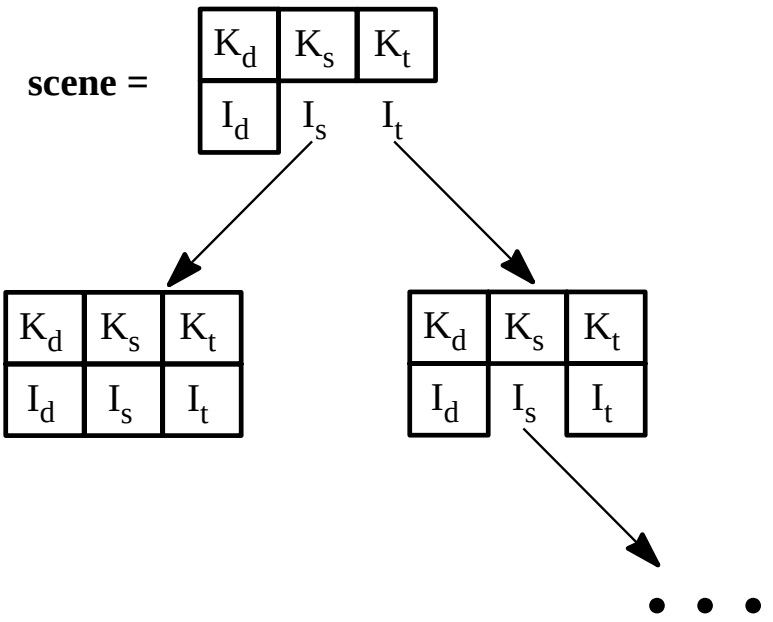

Figure 17: **Layer Recursion**
Recursive decomposition of layer images: both specular illumination and transparent illumination layers form comprehensible images that may be decomposed into more layers. Complex images or scenes with multiple transparencies and reflections form a tree of layer images.

the decomposition, I 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 17. 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 Figure 15 and diagrammed in Figure 17, 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.

5.2 Layer Compression

In the layering method, I 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 15, 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, I 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.

I concur with Schlick [Sch95] that tone reproduction operators should probably be achromatic and therefore depend only on the luminance of the layer. I estimate the luminance for each layer using a quick binary fraction approximation of the Y luminance signal of the NTSC television standard [Hun75]:

$$L = L(x,y) = \frac{5}{16}R(x,y) + \frac{9}{16}G(x,y) + \frac{2}{16}B(x,y). \tag{14}$$

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. I 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 $L = 10\ cd/m^2$, and $(R/L, G/L, B/L) = (0.7, 1.1, 1.3)$. For each illumination layer, I define the adaptation luminance value L_{wa} as the mean luminance of

all non-zero pixels L_w measured on a logarithmic scale:

$$\log(L_{wa}) = mean\{\log(L_w)\}. \tag{15}$$

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 mid-range 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:

$$\log(L_c(x,y)) = \log(L_{wa}) + \gamma(log(L) - \log(L_{wa})) \tag{16}$$

or equivalently, $L_c = L_{wa}\left(\frac{L}{L_{wa}}\right)^{\gamma}$, where $0 < \gamma < 1$, and compression increases as γ approaches zero. See Sections 2.2 and 2.3 for an extensive discussion of γ.

I 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.

5.3 Sigmoid Compressive Function

Many functions satisfy these goals. After examining the function proposed by Schlick [Sch95], I adopted a similar first degree rational polynomial that forms a sigmoid or S-shaped curve when plotted on log-log axes:

$$sig(x) = \left(\frac{x^g + (\frac{1}{k})}{x^g + k} \right) \cdot D \tag{17}$$

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:

$$\gamma = g \cdot \frac{(k-1)}{(k+1)}. \tag{18}$$

The k, D and g parameters adjust the shape and size of the $sig()$ function response curve, but are awkward to specify directly. Instead, I find their values from the limits of the desired mapping between scene luminances and display luminances. As shown in Figure 18, 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 $x_{max}/x_{min} > C$;

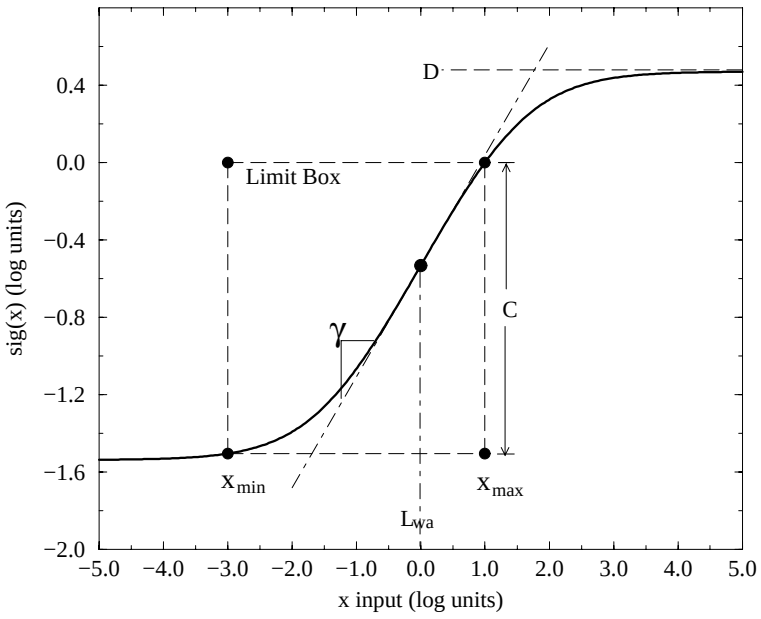

Figure 18: **Sigmoid Function $sig(x)$**
Both the layering and foveal methods use the $sig()$ function of Equation 17 to reduce high scene contrasts to fit within the abilities of the display. Users set $sig()$ function parameters by specifying a limit box for the mapping from scene to display luminances. 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.

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. I found an analytic expression for k using limit box terms and g by writing the equation $C = sig(x_{max}/L_{wa})/sig(x_{min}/L_{wa})$ and solving for k:

$$k(x_{max}, x_{min}, L_{wa}, C, g) = \frac{1}{2L_{wa}^{g}(x_{max}^{g} - C \cdot x_{min}^{g})}(B_p + \sqrt{B_n^2 + C \cdot A^2}) \tag{19}$$

where

$$A = 2L_{wa}^{g}(x_{max}^{g} - x_{min}^{g}),$$

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

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

To maintain normalized display output to ensure $sig() = 1$ when scene luminance L_w reaches x_{max} in Equation 17 let:

$$D = \frac{(\frac{x_{max}}{L_{wa}})^{g} + k}{(\frac{x_{max}}{L_{wa}})^{g} + \frac{1}{k}}. \tag{20}$$

In Appendix A, I will specify a desired gamma γ_d for the $sig()$ function curve. Though I 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 I must find the g to satisfy Equation 18 where k is given by Equation 19. The equation is well behaved and converges quickly with conventional root finding methods. For simplicity I used the bisection method.

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 18) and allow room for additional contrast from reflectance layers. Choosing C by more rigorous methods may improve layering results. Choosing D using Equation 20 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 $1/C$. I 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, I 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

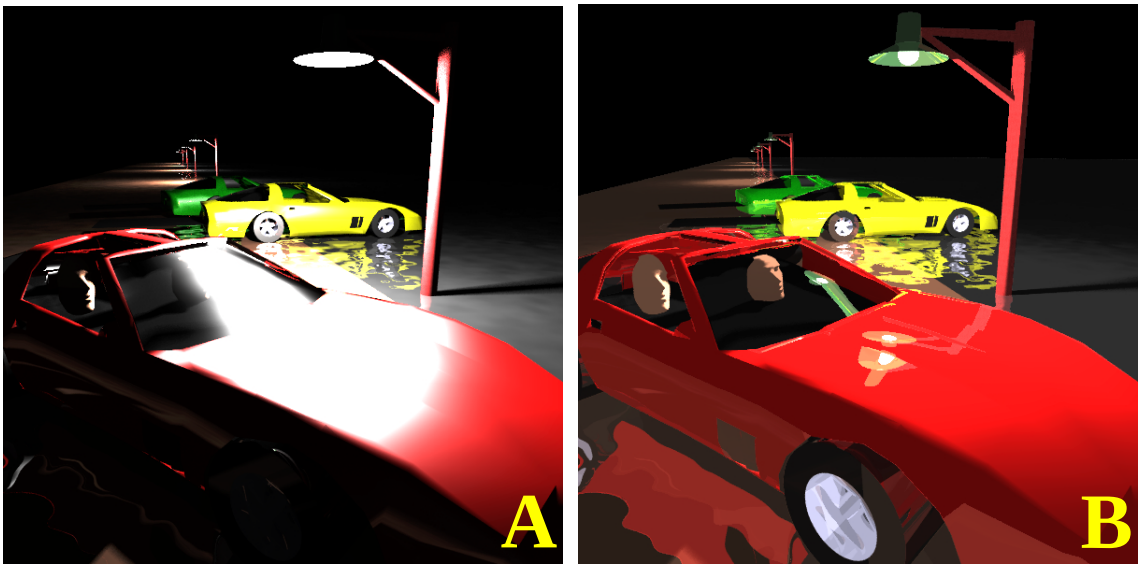

Figure 19: **Layering Results II**
The original high contrast ($> 300,000 : 1$) image A loses both highlight and shadow detail to clipping; in image B the layering method maintains the appearance of high contrast while revealing the driver's disembodied face behind the windshield and the reflection of the streetlight in the car hood.

are multiplied by their associated reflectances K_d, K_s and K_t, progressively collapsing the tree of layer images diagrammed in Figure 17 from the leaves upwards to the root node to form the output image. Because the tree describes a sequence of nested multiplies and adds, my 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. I 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 19 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.

5.4 Results

Figures 15 and 19 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 15, 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 19, 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, 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 I modified the shader functions of a commercially available ray tracer [WC92] 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, I 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 [War94b], HELIOS [Ash94], or the commercial software products offered by LightScape. These renderers explicitly compute illumination both from light sources and from interreflections with other surfaces. They may also include complex angular dependencies in surface reflectances. I expect that the six image layers defined here can be captured from the intermediate results of such renderers, but I 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. I 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.

Though 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 [KPC93] 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 20 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 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.

My 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 15 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.

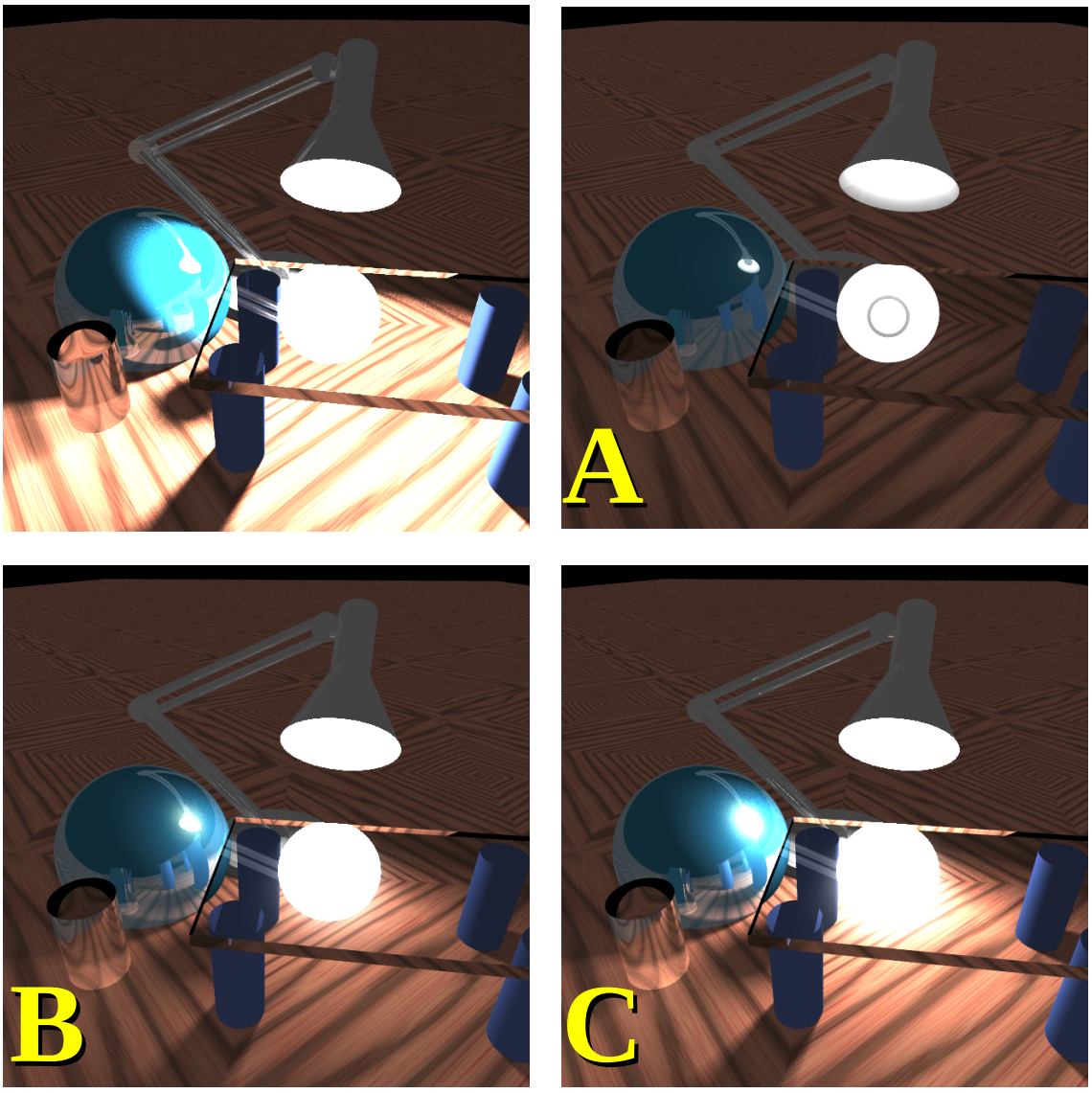

Figure 20: **Layering Differs from Relighting**
The layering result shown in the first image is difficult to achieve by re-lighting and re-rendering the original scene, as shown in images A-C. The floor and horizon in the background are easily revealed by increased ambient light, but we found no acceptable 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.

Chapter 6

Foveal Method

The foveal method exploits the directional nature of human vision and adaptation. Human eyes are highly directional, adjustable and nonuniform, and sense fine detail and color almost exclusively in the fovea, a 2-5 degree wide region of the retina centered at the direction of gaze. Eyes continually adapt as a viewer's gaze lands on various important scene details, adjusting to the available light to improve sensing ability and avoid saturation. Despite these changes, the human visual system somehow assembles signals from the eye into a stable mental impression of surroundings that are uniformly detailed and colorful. Though photoreceptors in the retina continually adapt to the amount of light in a small local neighborhood, our mental impression of a high contrast scene is dominated by adaptation in the fovea. For example, gazing directly at a bare lightbulb will cause the surrounding room to appear darker temporarily, but hiding and revealing the same bulb with your hand while gazing in a direction 30 degrees away from the bulb has little effect on the room's appearance.

Like the layering method, the foveal method supports the thesis of Chapter 1 (page 8) by example; it separates the input scene into large features and fine details, and performs detail-preserving contrast reduction by compressing only the large features of the scene and combining them with preserved fine details. For the foveal method, large features of a scene are the peripheral contents of viewed scene, and fine details are defined as the contents of the small foveal neighborhood of the scene in the current direction of gaze. This small central region dominates adaptation in human vision, and this same region determines the moment-by-moment choice of display parameters for the entire scene.

6.1 Methods

The foveal method uses an interactive computer display program to imitate adaptation in the users direction of gaze. The display program presents a changeable displayed image to the user, and the user indicates each new gaze direction on the image by mouse clicks. With each mouse click, the display program computes a new image of the scene as it might appear to a viewer adapted to the amount of light in a small neighborhood that approximates the fovea around the mouse cursor. In addition, the display program applies a tone reproduction operator to control display contrasts and intensities according to a scene viewer's sensing ability. The foveal method then assumes the display program user can then mentally assemble the sequence of displayed images into a consistent impression of the original high contrast scene despite the limited abilities of the display.

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 to 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 5.3 to help preserve details and textures in image shadows and highlights.

6.2 Implementation

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 21. Second, the program computes the foveal adaptation luminance value L_{wa} from scene luminances in the circled neighborhood using a pre-computed image pyramid. Third, the

program uses L_{wa} in the tone reproduction operator described in Appendix A Equation 55 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 [BA83] 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 [War91] 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. The foveal display program was written in Visual C++ 5.0 and was not optimized for speed. Running under WindowsNT 4.0 on a 90-Mhz Pentium machine with 48MB of memory, it achieves a 4 Hz display update rate on a 256x256 pixel image.

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 15, 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 starting with a simpler task. Suppose the foveal neighborhood diameter $diam$ is limited to one of two fixed widths, either W or $2W$ pixels. We may choose to pre-compute 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 pre-computed image. For example, we may decide to reduce the size of

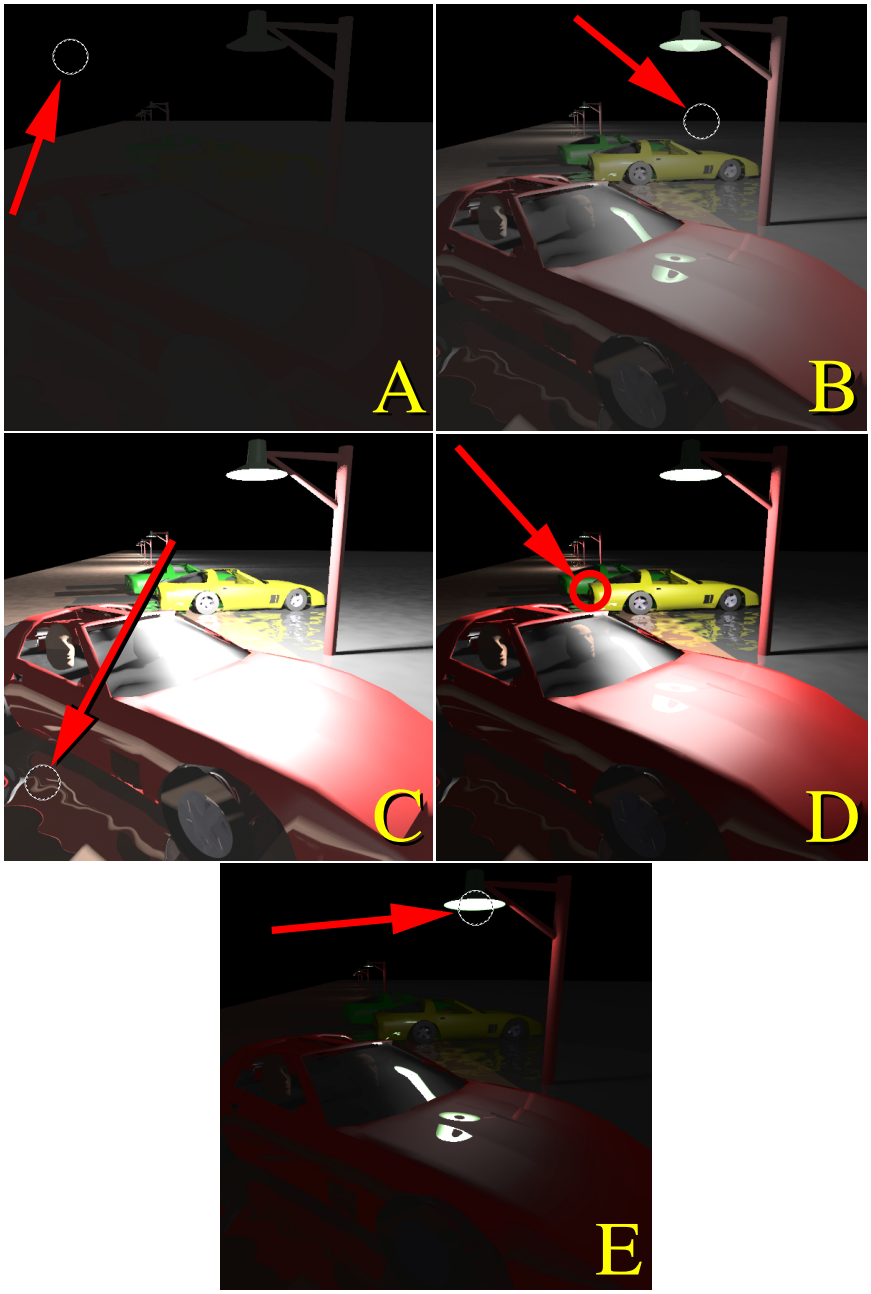

Figure 21: **Foveal Method Results**
Red arrows points to the foveal neighborhood around user's direction of gaze. (A) Adaptation luminance L_{wa} for night sky causes extremely low contrast sensitivity, and an anomalously dim street light. (B) Distant moonlit terrain increases contrast sensitivity and reveals cars. (C) Gazing into nearby shadows reveals puddle reflections below the car. (D) Adjusting to the distant yellow car shows its reflection on wet pavement. (E) The streetlamp causes extremely high adaptation luminance, darkening the rest of the scene. Images (A) and (E) reveal limitations of globally-applied adaptation models; human vision adapts locally, ensuring that both the street light and car have a bright appearance for any direction of gaze.

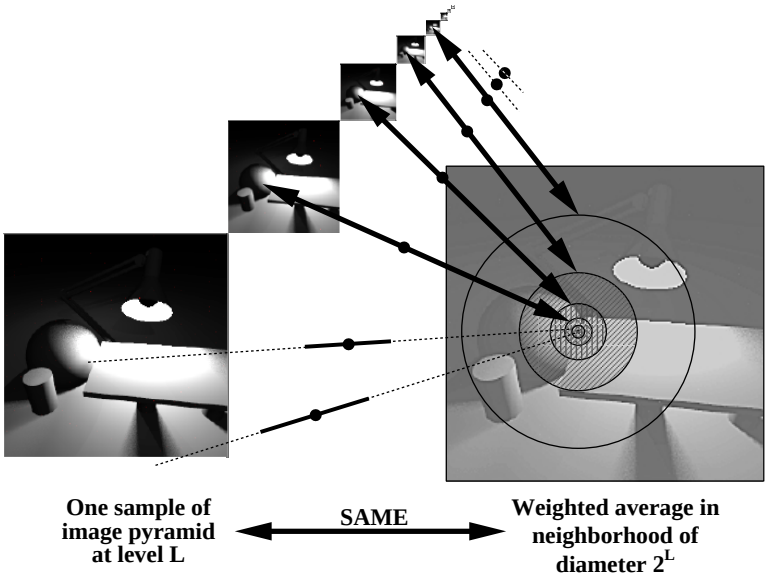

Figure 22: **Image Pyramid for $\log(L_{wa})$**
Image pyramids allow the foveal display program to find $\log(L_{wa})$, a weighted sum of neighborhood pixel values, in constant time for any neighborhood diameter.

the 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 22. 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, 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 55 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 Equations 17– 20 and Figure 18 with the tone reproduction operator of Appendix A Equations 55– 57 to compute the compressed display luminance:

$$L_d = sig(x, x_{max}, x_{min}, L_{wa}, C, D_{fov}, \gamma), \tag{21}$$

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 20 and m given by Equation 57 and

$\gamma = (\frac{\gamma_w}{\gamma_d})$ found in Equation 56.

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 55. The γ term adjusts display contrasts to match contrast sensitivity of a human observer viewing the original scene.

6.3 Results

Interactive viewing of high contrast images with the foveal display program resembles the familiar behavior of an automatic exposure camera, and Figure 21 shows 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.

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.

Some images produced by the foveal display program illustrate the need for a better model of local adaptation. In Figure 21, 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.* [SSZG95], and the foveal method could include the wide-ranging models for color, acuity and the time-course of adaptation developed by Ferwerda *et al.* [FPSG96]. 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.

Chapter 7

LCIS Method

The low curvature image simplifier (LCIS) method presented in this chapter attempts to mathematically imitate a commonly used artistic method to depict high contrast scenes and preserve important scene textures and details. As discussed in Chapter 3 Section 3.3, skilled artists produce very appealing renditions 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 7 [VA84] 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 for any scene.

LCIS is probably the most versatile of the three detail-preserving contrast reduction methods presented in this thesis because it accepts input from any scene source, either synthetic renderings as in Figure 24, or real-world radiance maps such as in Figures 23 and 34, and produces images suitable for almost any display, either static or interactive. Often real-world radiance maps do not supply sufficient scene information for use with the layering method of Chapter 5, and static displays such as printers, film and television are not well suited for the interactive foveal display method of Chapter 6.

Like the foveal and layering method, the LCIS method also supports the thesis of Chapter 1 (page 8) by example; it separates the input scene into large features and fine details, and performs detail-preserving contrast reduction by compressing only the large features of the scene and combining them with preserved fine details. For the LCIS method, large features of a scene are defined as large, simple, low-curvature regions separated by sharp, ridge-like boundaries. To find these large features this chapter presents an "LCIS operator," a diffusion-like process that finds and sharpens major boundaries in the scene by a shock-forming process, and also smoothes away

Figure 23: **Gamma Compression and LCIS applied to a Real-World Scene**
A radiance map [DM97] of a tree and streetlight on a foggy night captured contrasts >100,000:1, (as shown by factor-of-ten scaling in thumbnail images). LCIS method result (right) preserves fine details impossible to capture in a single photograph and lost in the gamma compression version (left), including long, dramatic fog streaks, asphalt texture and tree details in highlight and shadow.

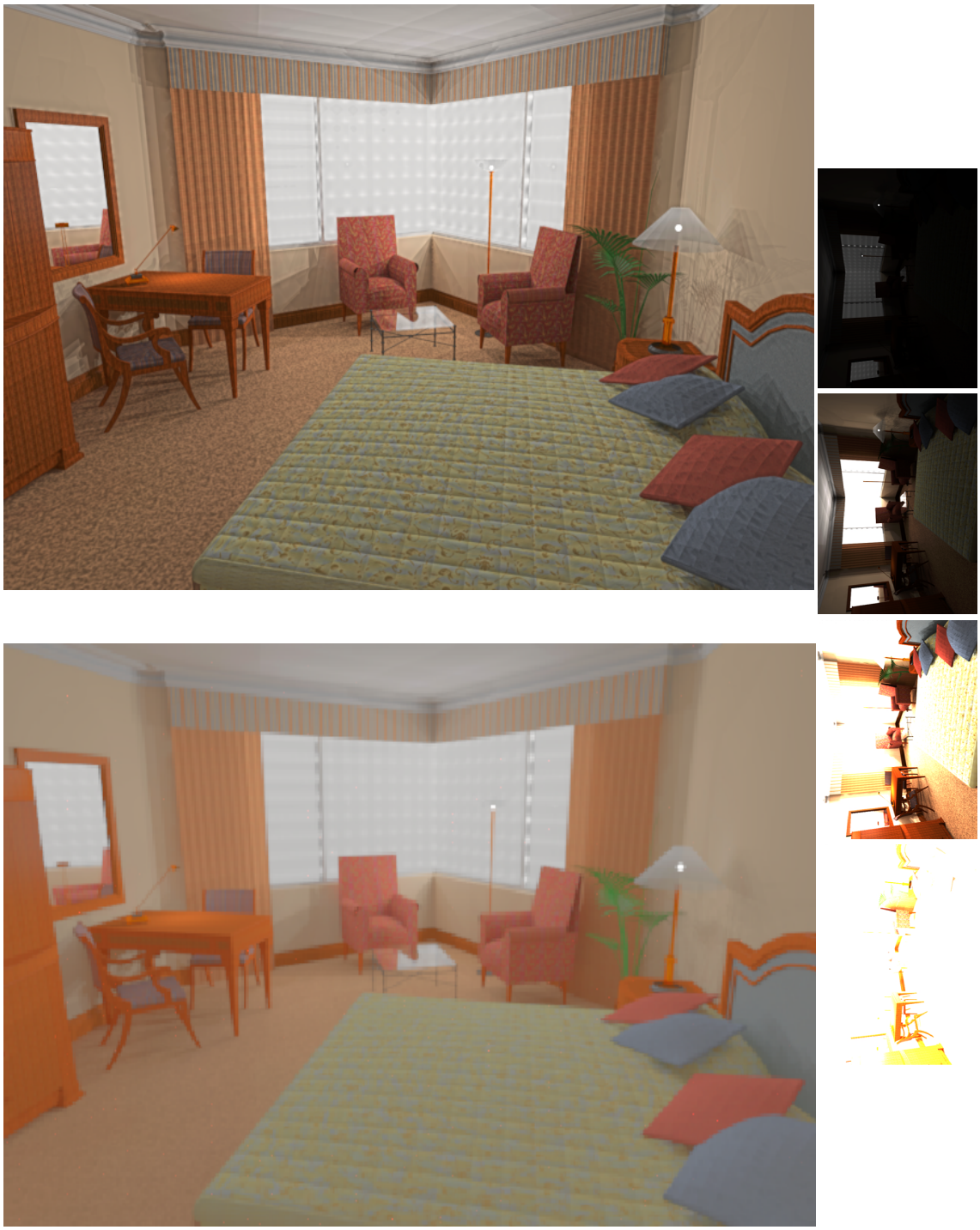

Figure 24: **Gamma Compression and LCIS applied to a Synthetic Scene**
Gamma Compression (bottom) loses details kept by LCIS (top) in this high contrast hotel room scene ($> 70,000 : 1$). Note the floor lamps and complex wall shadows. (Source Image: Proposed Burswood Hotel Suite Refurbishment (1995). Interior Design-The Marsh Partnership, Perth, Australia. Computer Simulation-Lighting Images, Perth, Australia. Copyright © 1995 Simon Crone.)

the details between these shocks by a relaxation process that gradually minimizes curvature. Fine details are defined as the difference between the original scene and its large features; these are the complex, high-curvature small-amplitude variations smoothed away by the LCIS operator.

As discussed in Chapter 3 Section 3.3, when drawing or painting, many artists capture visual appearance with a coarse-to-fine sequence of boundaries and shadings. Many begin with a sketch of large, important scene features and then gradually add finer, more subtle details. Initial sketches hold sharply defined boundaries around large, smoothly shaded regions for the largest, highest contrast and most important scene features. The artist then adds more shadings and boundaries to build up fine details, to fill in the visually empty regions and to capture rich detail everywhere. Such a procedure implicitly defines a hierarchical or coarse-to-fine scene decomposition, and we will use the LCIS operator to construct a similar decomposition for detail-preserving contrast reduction. This LCIS hierarchy is based on scene boundaries and shadings, an approach markedly different from the bandpass linear filter decompositions that may suffer from halo artifacts illustrated in Figure 12 and discussed in Chapter 3 Section 3.3. The LCIS method produces no such halos.

7.1 Boundaries, Shocks and Smoothing

The central importance of boundaries and shadings in both vision and artistic renderings (see Chapter 3 Section 3.3) suggests a new image decomposition method. To an artist, shadings usually refer to regions of nearly uniform intensity gradient. Because the gradients change smoothly and gradually with position, the region has low curvature; the second derivatives of scene intensity with respect to image position are small. An image made entirely of low curvature regions has region boundaries defined by gradient discontinuities, and these may include both ridge-like and step-like features, but only ridge-like primitives are necessary, as a step may be regarded as two adjacent ridge-like features.

The intensities and locations of these ridge-like boundaries alone are sufficient to construct a novel form of simplified image by interpolating between the boundaries with a curvature-minimizing process. The result has an interesting physical analogy; imagine image boundary intensities as a

height field made from a frame of thin wires. Dipping the wires in soapy water forms minimal-curvature bubble membranes between the wires, and interpolates intensities in the empty regions between the wire-frame boundaries of the simplified image. Now suppose the wire frame is refined further by adding more wires and bending the existing wires to more closely follow the smaller details of the original scene. With more wire and more careful bending, the soapy wire-frame becomes an increasingly accurate representation of the original scene, and with enough wires and bending eventually (in the limit) the entire scene can be represented exactly. Such an artist-like coarse-to-fine hierarchy can be constructed algorithmically if we can create a well-behaved method to find these boundaries and smooth away their intervening details.

Anisotropic diffusion has shown great success as a boundary-finding intra-region smoothing method and has gathered widespread attention in the image processing literature. Similar methods have been given many names such as variable conductance diffusion, nonlinear diffusion, geometry-limited diffusion and edge-affected diffusion. Anisotropic diffusion methods triggered a tremendous surge of new ideas and published solutions for image segmentation, restoration and enhancement. A search of the INSPEC publication database yielded over 300 citations of the early paper by Perona and Malik [PM90]. Though they published work on this idea as early as 1987 [PM87], their lucid analysis in the later journal paper drew widespread attention and is sometimes cited as the origination of anisotropic diffusion. Mathematically, it is a gradual, time-dependent evolution of an image towards a piecewise-constant approximation, as shown in Figure 27. The change in image intensity over time is determined by the same class of partial differential equations (PDEs) that govern diffusion of heat or other fluids through solids.

All diffusion behaviors are determined by just two factors: a *motive force* vector field $\boldsymbol{F}(x, y)$ and a *conductance* scalar field $C(x, y)$. The motive force pushes the fluid, and the conductance controls how easily the fluid may move within the material in response to the motive force. Together, motive force and conductance cause flux $\boldsymbol{\Phi}(x, y)$, the moment-by-moment flow of fluid through the material. The behavior is simplest if conductance is kept a constant everywhere ($C(x, y) = C_0$), and is known as isotropic (e.g. uniform) diffusion, but diffusion behavior is more interesting when conductance changes as the image evolves over time, as in anisotropic diffusion and LCIS.

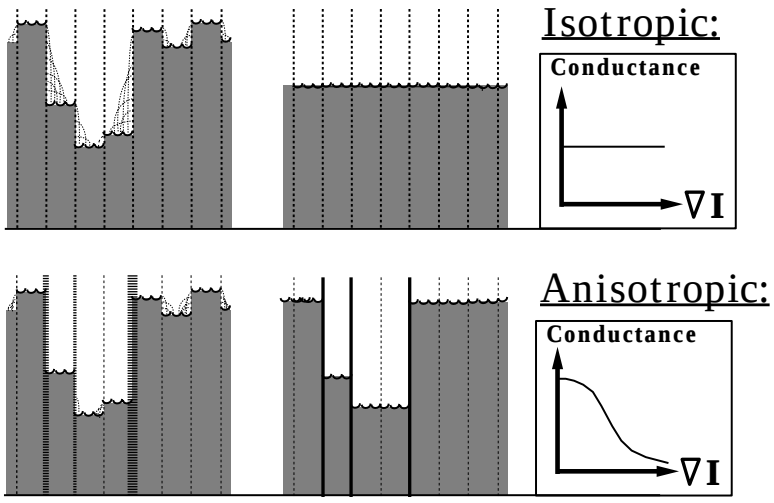

Figure 25: **Diffusion in a Soggy Egg Crate**
Diffusion equations cause changes to pixel values that correspond to water-level changes in a soggy egg-crate. In the egg-crate, motive forces from gravity push water through cardboard walls, causing flows that smooth away water level variations. The smoothing rate is set by the conductance or water permeability of the cardboard, and higher conductance permits faster smoothing. In isotropic diffusion, conductance is constant everywhere, uniformly smoothing away all water level differences. In anisotropic diffusion the conductance varies inversely with gradient; high gradients sharpen rapidly and become step-like boundaries or "shocks" but small details and low gradients between these shocks are rapidly smoothed towards a piecewise constant result.

Diffusion equations cause images to change in ways that are easy to understand using a *soggy egg-crate* analogy, a thought experiment where water seeps through cardboard. Imagine a huge mesh of interlocking cardboard strips arranged in egg-crate fashion to form a grid of many identical four-sided cells as illustrated in Figure 25. Let each cell represent an image pixel, and instantaneously fill each grid cell with water to a height proportional to pixel intensity. As time passes, water slowly diffuses through the cardboard cell walls and tries to equalize water levels in adjacent cells. Diffusion equations govern the gradual changes in these water levels, and when applied to images these diffusion equations change pixel values in the same way.

Water levels change in the soggy egg-crate in response to fluid flow between cells, water movement pushed along by a motive force and permitted by the conductance or water permeability of the cardboard. More formally, the water velocity or flux vector $\mathbf{\Phi}$ at every point is the product of the motive force vector $\boldsymbol{F}$ and the conductance scalar C. Gravity provides the motive force in

the soggy egg-crate analogy by attempting to push water "downhill," pushing water in the negative gradient direction $-\nabla I$ to minimize water level differences. Where the image gradient is low, adjacent egg-crate cells contain almost the same amount of water, and motive force between these cells is low. High image gradients correspond to large differences in adjacent water levels, causing a large motive force in the maximally downhill direction given by the negated gradient.[1] Flux is then written:

$$\mathbf{\Phi} = -C\nabla I. \tag{22}$$

High flows do not always lead to rapid changes in the water levels of a cell, because a rapid inflow may be matched by an equally rapid outflow, as might happen in the middle of a large area with constant high gradient. Instead, cell water levels change only in response to differences in local water flow. In the limit as cell size shrinks to zero, these flow differences are given by the divergence of the negative flow vector $\nabla \cdot -\mathbf{\Phi}$. As a result, diffusion changes an image over time according to:

$$I_t = \nabla \cdot (-\mathbf{\Phi}) = \nabla \cdot (C\nabla I) \tag{23}$$

Where:

t is time,

x, y are spatial coordinates of the image,

C is the conductance scalar,

$\mathbf{\Phi}$ is flux, the fluid velocity vector.

(Subscripts denote partial derivatives such that I_t is $(\partial/\partial t)I(x, y, t)$, the time rate of change in pixel intensity or water level, I_x is $(\partial/\partial x)I(x, y, t)$, I_{xx} is $(\partial^2/\partial x^2)I(x, y, t)$ and so forth.)

Isotropic diffusion is the special case where conductance of cardboard cell walls is the constant value C_0 everywhere. Because conductance is constant it no longer affects divergence, reducing

[1] The egg-crate analogy is slightly deceptive here; motive force between two adjacent cells depends only the cell wall area that is wetted on only one side: motive force is not dependent on pressure differences, which is much greater at the bottom of the cells than at the top. This area-only dependence acts as a forward-difference estimate of image gradient; pressure dependence would cause a motive force that is quadratically related to image gradient.

equation 23 to $I_t = C_0(I_{xx} + I_{yy})$:

$$\begin{aligned} I_t &= \boldsymbol{\nabla} \cdot (C_0 \boldsymbol{\nabla} I) && (24) \\ &= C_0 \nabla^2 I. && (25) \\ &= C_0(I_{xx} + I_{yy}). && (26) \end{aligned}$$

The behavior of the isotropically diffused water levels over time also corresponds exactly to repeated convolution of the image with a Gaussian filter kernel. Given unlimited time, such smoothing removes all differences between pixels and gradually moves all water levels and pixel intensities towards the mean image value. Accordingly, isotropic diffusion acts as a uniform, non-directional smoothing operation with no regard for image boundaries. The conductance constant C_0 determines how fast this smoothing occurs.

In anisotropic diffusion the conductance depends on the image, and both the image and the conductance evolve over time in more interesting ways. In their seminal 1990 paper, Perona and Malik [PM90] noted that conductance controls the rate of local image smoothing, and proposed that conductance should vary inversely with a local "edginess" estimate to find, preserve and sharpen image edges. This edginess value is a measure of the likelihood that a point is near an edge or boundary. Low conductance at likely edge locations and high conductances elsewhere preserves "edgy" features, yet rapidly smoothes away the details and textures between them, and simple edginess estimates work well. They used gradient magnitude scalar $\|\boldsymbol{\nabla} I\|$ and offered two inverse functions to find variable conductance $C(x, y, t)$. Thus anisotropic diffusion is:

$$\begin{aligned} I_t &= \boldsymbol{\nabla} \cdot (C(x, y, t) \boldsymbol{\nabla} I) && (27) \\ &= \frac{\partial}{\partial x}(C I_x) + \frac{\partial}{\partial y}(C I_y) && (28) \\ &= C(x, y, t) \nabla^2 I + (\boldsymbol{\nabla} C(x, y, t) \cdot \boldsymbol{\nabla} I), && (29) \end{aligned}$$

where

$$\begin{aligned} C(x, y, t) &= g(\|\nabla I\|) && (30) \\ &= g\left(\sqrt{I_x^2 + I_y^2}\right). && (31) \end{aligned}$$

Perona and Malik [PM90] offer two choices to find conductance from edginess:

Version I:

$$g(m) = e^{-(\frac{m}{K})^2} \tag{32}$$

Version II:

$$g(m) = \frac{1}{1 + (\frac{m}{K})^2.} \tag{33}$$

where K is the "conductance threshold" for edginess estimate m. Conductance reaches its midpoint value of 0.5 when $m = K$ and approaches zero as $m \to \infty$.

Anisotropic diffusion is especially interesting because both its edge-preserving and its smoothing abilities are self-reinforcing, as illustrated in Figure 26. Small-gradient regions have high conductance, allowing easy fluid flow that further reduces gradients. Large-gradient regions have low conductance, discouraging flow as if forming a weak barrier. However, higher conductances of its surroundings let fluid erode and steepen the already large gradients. Water seeps inwards towards the uphill side of the barrier, and fluid quickly drains away from the downhill side, making the large gradient region narrower and steeper, strengthening its barrier effect. The region quickly evolves into a step-wise discontinuity with infinite gradient and zero conductance known as a "shock." As a result, anisotropic diffusion transforms an image into a piecewise constant approximation with step-like discontinuities in regions of high "edginess."

More importantly, the self-reinforcing behaviors of anisotropic diffusion improve its performance as a boundary finder. Gradient magnitudes much larger than the gradient threshold K in Equations 32 and 33 will consistently form shocks, but the boundary/not-boundary decision is not a simple threshold testing process. Image behavior at points where gradient magnitude is near K is strongly influenced by image surroundings; gradients less than K may still form shocks if another shock is forming nearby, and small, isolated fine details with gradients greater than K are still smoothed away. Thus anisotropic diffusion finds boundaries according to both their gradients and their surroundings, sharpens the boundaries to create shocks and smoothes away all textures and

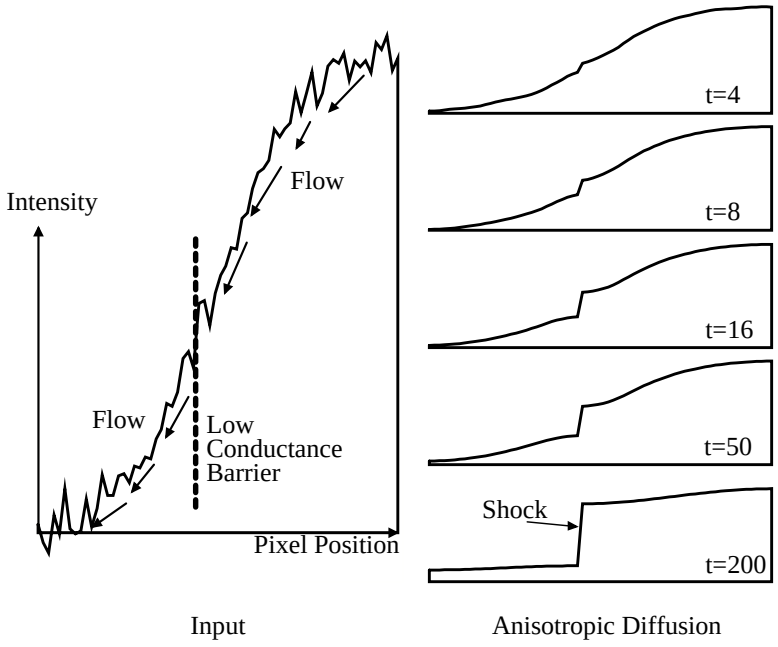

Figure 26: **Shocks in Anisotropic Diffusion**
Anisotropic diffusion rapidly forms step discontinuities or shocks in high gradient regions.

details between them. In this way anisotropic diffusion describes a robust, well-behaved boundary-finding method where boundaries are defined by image gradients and their surrounding neighborhoods. Anisotropic diffusion offers a nonlinear detail removal method that identifies and sharpens boundaries while smoothing away all details and textures between them.

Though numerically stable and guaranteed to converge to a piecewise-constant solution as $t \to \infty$, Nitzberg and Shiota [NS92], You *et al.* [YXKT95] and others have shown that anisotropic diffusion is ill-posed; infinitesimal changes in input can cause very large changes in output due to the shock-forming process. Shocks usually form at local gradient maxima and follow image boundaries closely, but this is not always true. Regions of high, approximately uniform gradient may develop shocks anywhere within the region. Instead of a single large, centrally placed shock, anisotropic diffusion may develop multiple shocks placed seemingly at random, causing "stairsteps" in the region as shown in Figure 26 and explored by Whitaker and Pizer [WP93]. Several authors have proposed methods to remove stairstepping, including [WP93, LKG94, Act98], but most force

tradeoffs between resistance to multiple steps and ability to capture small crenelations in scene boundaries.

Despite the attractive edge-preserving and intra-region smoothing properties of anisotropic diffusion, its piecewise constant output is a poor choice as the coarse image output of a simplifier function. Many high contrast images contain large smooth high gradient regions that should be regarded as a single coarse image feature, such as a large region of a bare wall lit only by a small lamp, as in Figure 29. The corresponding fine image in this region should contain only the wall texture. Though anisotropic diffusion will quickly smooth away the wall texture in a few iterations, no arrangement of shocks can reasonably approximate the coarse features of this region. Stairstepping can exacerbate the problem, causing multiple coarse image contours that do not correspond to any significant input image features.

7.2 LCIS Equations

Inspired by anisotropic diffusion, I have created a related set of PDEs that capture its self-reinforced smoothing and shock-forming behavior, but is driven by higher derivatives of the image intensities $I(x, y)$. Instead of evolving an image towards a piecewise constant approximation by driving all gradients towards zero or infinity, these equations smooth and sharpen an image towards a piecewise linear approximation by driving all curvatures towards zero or infinity. Because boundary conditions usually prevent a zero curvature solution, I call this operator a low curvature image simplifier (LCIS).

As with anisotropic diffusion, LCIS equations describe fluid flow, but both the motive force pushing the fluid and the variable conductances permitting flow are computed differently. The motive force of anisotropic diffusion is the negative gradient $-\nabla I$, but LCIS pushes fluids to encourage uniform gradients; it pushes outwards from intensity peaks or ridges with negative curvature and inwards towards pits or hollows with positive curvature. Therefore the LCIS motive force vector should follow positive derivatives of curvature, but these form a tensor with no obvious single direction. Instead, I define the motive force vector using simpler directional derivatives and vector integration. To evaluate the motive force at image point $I(x_0, y_0)$, I first define a 1-D line L through

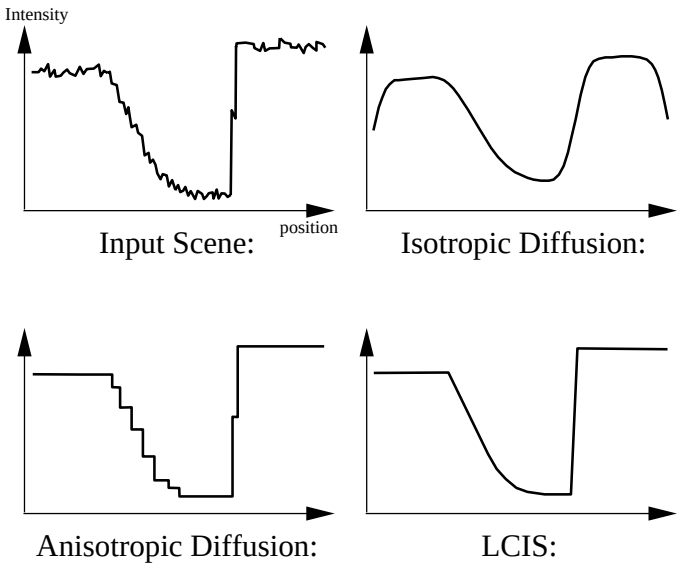

Figure 27: **Diffusion Comparison**
Isotropic diffusion uniformly smoothes the entire image; anisotropic diffusion forms step-like shocks at persistent high gradients and smoothes away all intensity variations between them, sometimes forming "stairsteps" [WP93]; but LCIS forms ridge-like shocks at persistent high curvatures and smoothes away the gradient variations between them.

the point with orientation θ and signed distance parameter α along line L. We evaluate the image I along line L and find its third derivative $I_{\alpha\alpha\alpha}$ as a measure of curvature change in the direction given by θ:

$$A = x_0 + \alpha \cos\theta, \tag{34}$$

$$B = y_0 + \alpha \sin\theta. \tag{35}$$

$$\begin{aligned} I_\alpha(A,B) &= I_x(A,B)\cos\theta + I_y(A,B)\sin\theta \\ &= \left(\left(\frac{\partial}{\partial x}\right)\cdot\cos\theta + \left(\frac{\partial}{\partial y}\right)\cdot\sin\theta\right) I. \end{aligned} \tag{36}$$

$$I_{\alpha\alpha\alpha} = \left(\left(\frac{\partial}{\partial x}\right)\cdot\cos\theta + \left(\frac{\partial}{\partial y}\right)\cdot\sin\theta\right)^3 I, \tag{37}$$

where (A,B) are (x,y) coordinates of line $L(\alpha)$, I_α is the directional derivative of I along line L and $I_{\alpha\alpha\alpha}$ is the third derivative of I along line L.

If $I_{\alpha\alpha\alpha} > 0$, then the curvature $I_{\alpha\alpha}$ is increasing along line L as it passes through point (x_0, y_0); flow in that direction would help equalize curvatures on either side of the point and reduce $I_{\alpha\alpha\alpha}$. Accordingly, I let $I_{\alpha\alpha\alpha}$ define the strength of an infinitesimal motive force vector along line L and I

sum up these tiny forces for all orientations θ to find the force vector's x and y components, labeled "East" and "North" to avoid confusion with partial derivatives, and given by: $\boldsymbol{F} = (f_E, f_N)$;

$$f_E = \frac{1}{\pi}\int_0^{2\pi} I_{\alpha\alpha\alpha}\cos\theta d\theta = \frac{3}{4}(I_{xxx} + I_{yyx}), \tag{38}$$

$$f_N = \frac{1}{\pi}\int_0^{2\pi} I_{\alpha\alpha\alpha}\sin\theta d\theta = \frac{3}{4}(I_{xxy} + I_{yyy}). \tag{39}$$

For LCIS conductance, I use Equation 33 from anisotropic diffusion, but now the m argument is given by a new edginess estimate. As the desired ridge-like shocks have infinite curvature, I construct m from a non-directional measure of curvature magnitude: $m^2 = 0.5(I_{xx}^2 + I_{yy}^2) + I_{xy}^2$. The low curvature image simplifier is then defined as:

$$I_t(x, y, t) = \boldsymbol{\nabla} \cdot (C(x, y, t)\boldsymbol{F}(x, y, t)) \tag{40}$$

where:

> $\boldsymbol{F}(x, y, t)$ is the motive force vector computed from partial derivatives and given by $\boldsymbol{F} = (f_E, f_N) = \frac{3}{4}(I_{xxx} + I_{yyx}, I_{xxy} + I_{yyy})$.
>
> The conductance $C(x, y, t) = g(m)$ is computed from the edginess-to-conductance Equation 33 borrowed from anisotropic diffusion, but instead uses a curvature-based edginess m given by:
>
> $$m^2 = 0.5(I_{xx}^2 + I_{yy}^2) + I_{xy}^2.$$

Low curvature image simplifiers (LCIS) share several important properties with anisotropic diffusion. Equation 40 is adiabatic; intensity is neither created nor destroyed. LCIS meets the goals set forth by Perona and Malik [PM90]; varying its conductance threshold K defines a continuous scale space that exhibits causality, immediate localization and piecewise smoothing. Conductance is inversely linked with the motive force, causing rapid shock formation at image boundaries, and smoothing between boundaries is self-reinforcing, though asymptotic. Unlike anisotropic diffusion, LCIS shocks are discontinuities in gradient instead of intensity; they form ridge-like features that appear step-like when adjacent. Just as large high gradient regions can cause multiple shocks or stairstepping in anisotropic diffusion results, LCIS can also form multiple shocks unpredictably in large regions of uniform high curvature. However, high curvature regions tend to be smaller due to

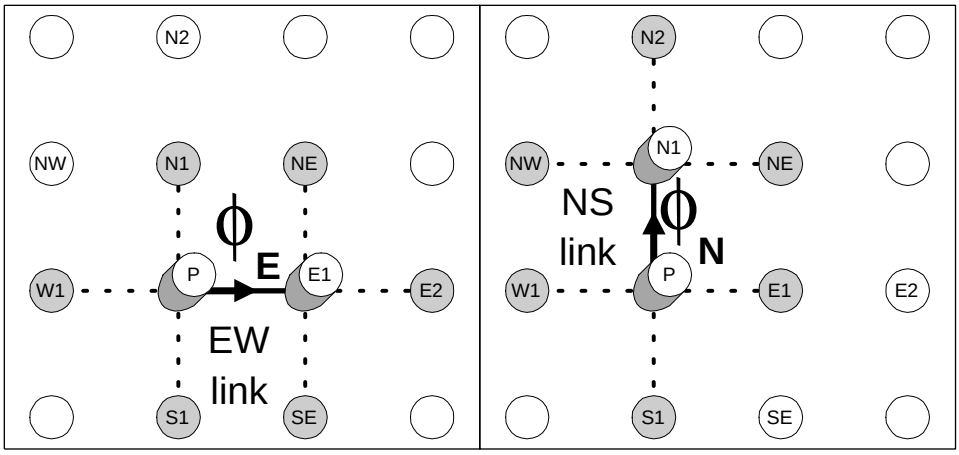

Figure 28: **Discrete Implementation of LCIS**
My implementation of LCIS transfers intensity through "links" between pixels. On each timestep, flux ϕ_E,ϕ_N flows through EW and NS links respectively, computed from the gray-shaded pixels.

the larger intensity range they require, and I have found that multiple ridge-like shocks are far less visually apparent than stairsteps. Finally, both anisotropic diffusion and LCIS are formally defined for continuously variable (x, y). Any practical implementations must use discrete approximations.

7.3 LCIS Implementation

A pixel-based LCIS approximation is straightforward to implement. I use explicit integration with a fixed timestep to find $I(t)$, compute a new image on each timestep, assume constant conductance and flux during each timestep, and compute flux only between 4-connected neighboring pixels. The computation is entirely local; each new image is computed only from pixels in the previous timestep's image, and each new pixel is computed from a fixed set of neighboring pixels in the previous image.

Continuing with the fluid flow analogy, imagine that each pixel is a tank holding a fluid volume equal to the pixel intensity. Each pixel reservoir is tied to each of its four-connected neighbors through separate pipelines or "links." As shown in Figure 28, pixel P has links to pixels $E1$, $N1$, $W1$ and $S1$. During a fixed timestep T, I transfer a fluid quantity flux ϕ through a link, decreasing the source pixel intensity and increasing the destination. On each timestep, I compute flux for each link and then adjust the current image by the flux amounts to create a new output image.

The left and right sides of Figure 28 show the two types of links. The drawing on the left shows an EW link connecting pixels P and $E1$. The flux ϕ_E that flows through this EW link is computed from the values of eight neighborhood pixels shown in gray and connected by dotted lines. The sign of ϕ determines flow direction: $\phi_E > 0$ flows in the $+x$ direction, which lowers the intensity of pixel P and increases pixel $E1$ by the same amount. Similarly, link NS connects pixels P and $N1$, and positive flux ϕ_N flows in $+y$ direction to diminish P and increase $N1$.

Motive force through each link is found from forward-difference estimates of third partial derivatives of the image at each link center. For the EW link between pixels P and $E1$:

$$I_{xxx} = (E2 - W1) + 3(P - E1), \tag{41}$$

$$I_{yyx} = (NE - N1) + (SE - S1) + 2(P - E1), \tag{42}$$

and for the NS link between P to $N1$,

$$I_{yyy} = (N2 - S1) + 3(P - N1), \tag{43}$$

$$I_{xxy} = (NE - E1) + (NW - W1) + 2(P - N1). \tag{44}$$

Conductance of each link is found from forward difference estimates of second partial derivatives. I define:

$$P_{xx} = E1 + W1 - 2P, \qquad P_{yy} = N1 + S1 - 2P,$$

$$E_{xx} = E2 + P - 2E1, \qquad E_{yy} = NE + SE - 2E1,$$

$$N_{xx} = NE + NW - 2N1, \quad N_{yy} = N2 + P - 2N1,$$

and

$$N_{xy} = (NE - E1) - (N1 - P),$$

$$S_{xy} = (E1 - SE) - (P - S1),$$

$$W_{xy} = (N1 - P) - (NW - W1).$$

The square of the edginess estimate m is:

$$m^2_{EW} = (P^2_{xx} + P^2_{yy} + E^2_{xx} + E^2_{yy})/4 +$$

$$(N^2_{xy} + S^2_{xy})/2 \text{ for EW links and} \tag{45}$$

$$m_{NS}^2 = (P_{xx}^2 + P_{yy}^2 + N_{xx}^2 + N_{yy}^2)/4 +$$
$$(W_{xy}^2 + N_{xy}^2)/2 \text{ for NS links.} \quad (46)$$

Assuming constant curvatures and flow rates during each timestep, the flux through each link is the product of timestep length, motive force and conductance, given by $\phi_E = TF_EC_E$ and $\phi_N = TF_NC_N$. I recommend a timestep of $T \leq 1/32$ for stability. F_E and F_N are the motive forces driving EW and NS flux, given by $F_E = (I_{xxx} + I_{yyx})$ and $F_N = (I_{xxy} + I_{yyy})$. Conductances through NS and EW links are computed from Equation 33

$$C_E = \frac{1}{1 + \left(\frac{m_{EW} D_E}{K}\right)^2} \quad (47)$$

and

$$C_N = \frac{1}{1 + \left(\frac{m_{NS} D_N}{K}\right)^2}. \quad (48)$$

where m_{EW}, m_{NS} are edginess estimates from Equations 45 and 46, D_E, D_N are leakfix multipliers explained below, initialized to 1.0.

Estimating image derivatives with adjacent pixel differences causes "leakage" problems. Shocks in continuous images form perfectly impermeable boundaries to prevent any fluid flow across them. Though discrete images also form shocks, neither the gradients nor the curvature estimates reach infinity due to the fixed, finite spacing between pixels, allowing small fluid flows or leaks across boundaries that should be impermeable. Small leaks over many timesteps gradually erode the image boundaries and eventually destroy them all. Though several papers (*e.g.* [LC94]) offer strategies for stopping the time evolution before boundary erosion is too large, any chosen stopping time is a compromise between adequate intra-region smoothing and minimal leakage. Instead, I devised a simple leakage fix that works quite well for both discrete LCIS and anisotropic diffusion.

The leakage fix is a single self-adjusting "leakfix multiplier" value D_E or D_N stored for each EW or NS link respectively and used in Equations 47 and 48. I noticed in Equation 33 that shock forming drives all the edginess estimates m rapidly away from the conductance threshold K in the earliest timesteps. Edginess estimates at boundaries are boosted towards infinity by shock formation, and self-reinforced smoothing drives all other m below K and towards zero. To identify and

prevent leakage as an image evolves, the LCIS implementation continually compares m against K to find links that cross image boundaries and should hold shocks, and I adjust D_E or D_N of these links to amplify their edginess estimates m and drive conductance towards zero. Leakfix multipliers grow exponentially with time in links where m is consistently larger than K, but settles rapidly back towards 1.0 if edginess falls below K. In my implementation, initially $D_E = 1.0$ and $D_N = 1.0$ for all pixels, then for each timestep:

$$D_E = \begin{cases} D_E(1.0 + m_{EW}) & \text{if } (m_{EW} > K), \\ 0.8D_E + 0.2 & \text{otherwise, and} \end{cases} \tag{49}$$

$$D_N = \begin{cases} D_N(1.0 + m_{NS}) & \text{if } (m_{NS} > K), \\ 0.8D_N + 0.2 & \text{otherwise.} \end{cases} \tag{50}$$

The leakage fix also provides a convenient marker for boundaries; I label any link with a leakfix multiplier greater than 10.0 as a "boundary link" that may cross a ridge-like boundary shock in the image. Even though conductance drops to zero at LCIS shocks, my analysis of continuous LCIS showed ridge-like shocks should not evolve into step-like shocks during intra-region smoothing. To prevent this divergence in the discrete implementation, I also mark the pixels on either end of a boundary link as "boundary pixels" and stop all subsequent flux into or out of these pixels; see Appendix B for source code. With this simple two-part fix I have not encountered any noticeable problems with leakage or boundary erosion.

7.4 Contrast Reduction: LCIS Hierarchy

Discrete LCIS mimics the artist's drawing process in reverse; it selectively removes details from a scene to leave only smoothly shaded regions separated by sharp boundaries. I can easily recover the removed details by subtracting the LCIS-smoothed image from the original scene, and then follow the artists scheme for detail-preserving contrast reduction: I strongly compress the contrasts of the simplified image, then add in the details with little or no compression, as shown in Figure 30. The boundary-preserving smoothing behavior avoids formation of strong halo artifacts common to

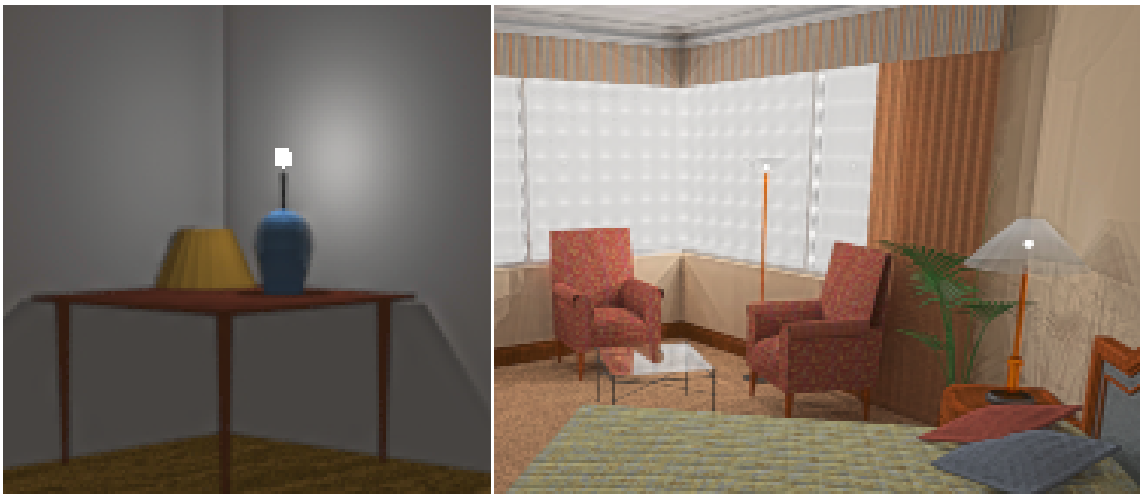

Figure 29: **LCIS Method avoids Halos**
The LCIS method avoids halo artifacts that frequently corrupt results of linear filter-based contrast reduction methods as shown earlier in Figure 12.

linear filter methods, as demonstrated by comparing Figure 29 made by the LCIS method with the same scenes processed by linear filtering in Figure 12.

The LCIS hierarchy shown in Figure 31 shows the expandable multiscale form used to make Figures 29, 34 and 35. Just as an artist may create an image by progressive refinement, the LCIS hierarchy extracts preserved scene details with a progressive set of LCIS settings. The LCIS method first converts the scene to its base-10 logarithm so that pixel differences directly correspond to contrasts (intensity ratios). Color processing here is rudimentary: the method applies LCIS only to scene luminances and reconstructs color outputs using color ratios as suggested by Schlick [Sch95]. Next, the LCIS hierarchy makes a set of progressively simpler images by applying LCIS with progressively larger K values starting from zero: when $K = 0$, the LCIS operator has no effect on the input image. Values for the church scene in Figure 34 are: $K_1 = 0.06$, $K_2 = 0.10$ and $K_3 = 0.16$. LCIS with the largest K value makes the simplest or *base* image, and the LCIS hierarchy extracts a graduated set of detail images $(det_0, det_1, ...)$ by subtraction, with the finest details in det_0. Next, it compresses their contrasts by scaling (for the church image in Figure 34 I chose $w_{0,1,2,3} = 1.0, 0.8, 0.4, 0.16$, $w_{color} = w_3$) and then adds and exponentiates to find display intensities. My test-bed software allows interactive adjustment of the parameters K, w and the number of LCIS timesteps, and by trial and error I found 100 to 1000 timesteps and K from 0.02

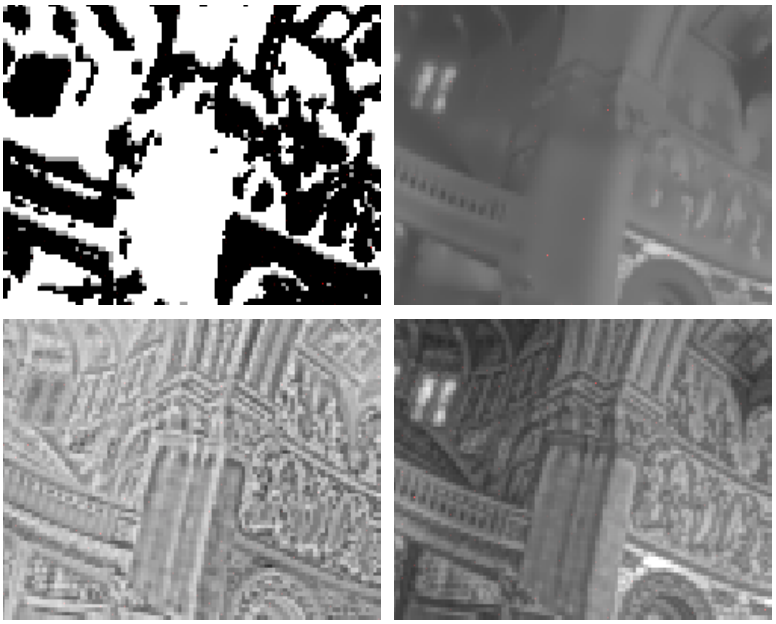

Figure 30: **Images from an LCIS Hierarchy**
Images from an LCIS hierarchy reveal its methods. From a part of the church scene of Figure 34, LCIS creates a boundary pixel map (top left: boundary pixels are black) and a simplified image (top right), shown after contrast compression to make it displayable. A detail image (lower left) holds the input minus the simplified image. A detailed displayable image(lower right) is the sum of images at upper right and lower left.

to 0.32 spanned the entire range of interest for all my test scenes. Though w_{base} is dictated by desired display contrast, w_0 is usually best around 1.0 with intervening w values between the two, none are critical; a wide range of settings provides a pleasing visual appearance and small changes are not easy to notice. Usually $w_{color} = w_{base}$ looked pleasant, but some images looked better with exaggerated values: I set $w_{base} = 0.2$ in the hotel room scene, but used $w_{color} = 0.6$ to avoid a washed out appearance.

7.5 Results

As demonstrated in Figures 23 and 24 above and Figures 32, 33 and 34 below, the LCIS method both avoids halo artifacts (see Figure 29) and reveals an astonishing amount of subtle detail and scene content. For example, the Stanford Memorial Church ceiling dome has intricate gold filigree near the skylight, and these details are easily made visible by scaling original scene intensities to

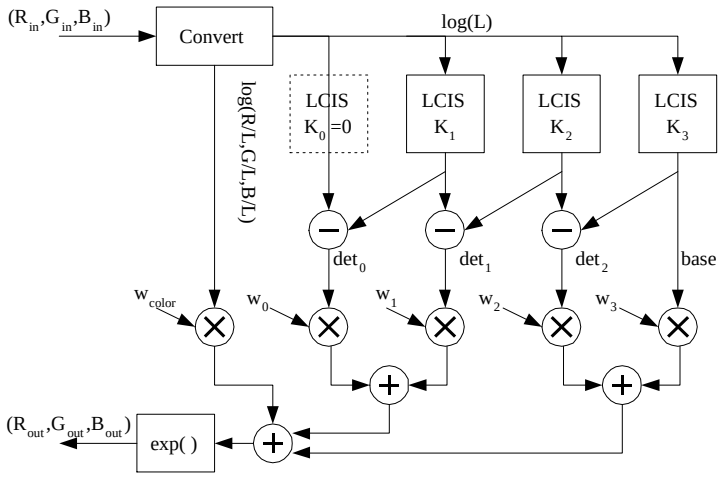

Figure 31: **LCIS Hierarchy Diagram**
Detail-preserving contrast reduction method using an LCIS hierarchy, as used for Figures 29, 34 and 35.

the display range. However, these details are difficult to preserve in the drastic contrast reductions necessary to display the entire high contrast scene on a low contrast display. With linear filter-based methods, these details are easily obscured or destroyed by strong halo artifacts caused by the nearby skylight. Compressing all contrasts uniformly with gamma compression as in Figure 33 reduces all low contrast details to invisibility. However, the LCIS method plainly shows this filigree in Figure 34. Also notice the details LCIS preserves in the alcove paintings of angels and the seams and marks in the stone columns. In the specular highlight on the floor at the lower right the LCIS method even reveals seams and ripples in the floor tiles.

Though the histogram-based result of Ward-Larson *et al.* [WLRP97] (also seen in [DM97]) is a beautiful and natural-looking depiction of Debevec's Stanford Church radiance map, comparisons of Figures 34 and 32 reveal it does not include many features clearly visible in the LCIS result, including stonework and filigree.

The LCIS method also reveals rich details in the hotel room scene of Figures 29 and 24. Previous depictions of this scene, perhaps made with gamma compression, made the carpet and bed covering look muted and subtle, but Figure 24 shows strong wood, bedspread and carpet textures. The elaborate shadow details behind the plant and on the ceiling also suggests the sun and sky lighting were approximated with several point sources.

Figure 32: **Histogram Adjustment Results**
Histogram adjustment [WLRP97] results are very appealing, but can lose details kept by LCIS in Figure 34 (radiance map courtesy Paul Debevec, University of California at Berkeley [DM97]).

Figure 33: **Gamma Compression Results**
Gamma compresion reduces all scene contrasts uniformly, but subtle details and textures may vanish (radiance map courtesy Paul Debevec, University of California at Berkeley [DM97]).

Figure 34: **LCIS Method Results**
The LCIS method preserves fine details and textures from the scene despite drastic contrast reduction (radiance map courtesy Paul Debevec, University of California at Berkeley [DM97]).

I confirmed that these details exist at these contrasts in both the church and hotel scene by using simple scaling and truncation to display scene radiances. In Figures 23 and 24, I include tiny versions of scaled, truncated scene radiances with $\gamma = 1.0$ and scale factor m (Equation 1) increasing by a factor of 10 for each successive image to help viewers understand the huge contrasts present in both scenes.

Computing costs for LCIS results are moderate in the current implementation. Written in Visual C++ without regard for speed or efficiency, using 32-bit floating point values for all images and computations, my code required 14 minutes 47 seconds to compute a 187x282-pixel image of Figure 35 on a 200 MHz Pentium Pro with 128MB RAM running Windows NT4.0. This timing result includes five LCIS simplifiers computing 500 timesteps each, the auxiliary calculations diagrammed in Figure 31 and windowing and display overhead of an interactive data-flow application. Most of my tests of LCIS behavior and choices of parameters such as curvature thresholds K, weighting factors w and the number of timesteps were made interactively on much smaller images (often 128 x 192) where computing delays rarely exceeded 30 seconds. I found that parameter choices at low resolutions invariably worked well for much larger images, and no parameter settings needed extensive or critical tuning. Better integration methods and software tuning may also greatly improve these initial LCIS speed ratings.

7.6 Discussion

LCIS smoothing and LCIS hierarchies offer a new way to decompose an image reversibly into a multiscale set of large features, boundaries and fine details. It permits a novel form of detail-preserving contrast reduction that avoids halo artifacts common to previous methods based on linear filtering. Detail extraction problems arise in many domains, and LCIS may also be useful for viewing other high contrast signals such as data from astronomy, radiology or seismology. The initial results presented in these images are promising, but my LCIS work has many open questions and opportunities for further research. Though I found the K and w controls of the LCIS hierarchy easy to use, they affect the image globally; would localized, paint-box-like controls to vary them within the scene be a useful artistic tool? Conversely, could psychophysical data control an LCIS

hierarchy automatically and perform as a true tone reproduction operator? How could a more thoughtful treatment of color better exploit the boundary and detail information it contains?

I have also discovered that even though LCIS forms sharp boundaries without smoothing across them, sometimes weak, "residual halo" artifacts can appear with strong contrast reductions. However, even the worst residual halos are far weaker than those from linear filters, and are caused by a different mechanism. Residual halos seem to form only in blurred image regions with high contrast, sparse shocks and low curvature such as a badly out-of-focus step image. If shocks do not form at the blurred boundary, LCIS further smoothes the region and reduces its curvature, causing a broad low-curvature component to appear in one or more detail images in Figure 31. Weak residual halos are visible in Figure 35 under the left lowermost tree branches.

I do not yet fully understand the relationship between curvature, spatial scale, contrast and shock formation for a given K, but suspect that the extensive published studies of anisotropic diffusion may offer help. Large scale, high contrast scene features such as soft shadows may have very low curvatures obscured by small scale high curvature textures; can LCIS-like methods find shocks for both? Is reasonable behavior for LCIS possible for all images?

Finally, extending LCIS to higher dimensions appears straightforward, and may be useful for revealing fine details in high contrast 3D scalar and vector fields or for motion estimation in high contrast scenes.

Chapter 8

Conclusions and Future Work

This dissertation presented arguments, evidence and examples supporting a thesis of two claims: first, for realistic depictions of high contrast scenes I claim that detail-preserving contrast reduction is *necessary*, and second, it is *acheivable* by separating the scene into components of compressible large features and preserved fine details. The necessity claim is supported by the first four chapters, and the next three support the achievability claim.

8.1 Thesis Support: Necessity

The necessity claim for detail-preserving contrast reduction is best supported by a careful restatement of the problem and a review of related background in earlier picture-making processes. Chapter 1 introduced the scene-to-display mapping problem, and showed by graphs and examples that the domain of safely viewable scene intensities dwarfs the range of available display intensities and contrasts; accordingly, some form of contrast reduction is unavoidable. With the notion of a tone reproduction operator, this introductory chapter also showed how any realistic displayed image, made by a machine, an artist or both, must include some measurement or intuitive understanding or reasonable assumptions about the nature of visual appearance.

In Chapter 2, I explain how film and television history has imposed superfluous constraints on most of our current digital picture-making processes, enforced by three seemingly obvious assumptions that make accurate high contrast scene depictions difficult with electronic imaging and displays, as discussed in Section 2.1. Until recently, new realistic digital imaging systems such as computer graphics rendering, radiance maps and electronic cameras have relied on digital imitations of photographic film, even though these new systems are not subject to the chemical or broadcast

constraints that dominate the scene-to-display mappings for film and television. As briefly reviewed in Sections 2.3 and 5.3, these film mappings must routinely discard details in highlights and shadows of high contrast scenes, details that are plainly visible to human observers and are important enough to be included in realistic renditions by skilled artists. Chapters 2 and 4 Section 4.1 notes that good photographers routinely distort lighting in the scene or distort scene-to-display mappings in the the darkroom to rescue these missing but important details in highlights and shadows, just as artists working in pen-and-ink, pastel and charcoal such as Chris Van Allsburg use their intuitive understanding of visual appearance to add details everywhere in their drawings of high contrast scenes. Apparently even the experts in film artistry such as Ansel Adams do not accept the film characteristic curve as the best scene-to-display mapping to capture visual appearance.

This intuitive understanding of visual appearance is largely lost to current practice in digital imaging. Section 2.4 re-examines the three digital imaging assumptions to find they closely fit the even more misleading notion that human eyes work as cameras. Clearly human vision is vastly superior, especially as detectors of fine details in even the most extreme lighting conditions where cameras fail badly, such as Figure 35. Chapter 3 provides additional concrete evidence from the psychophysics literature that easily distinguishes human vision from cameras and offers hints on how to better perform detail preserving contrast reduction in digital imaging systems.

Finally, Chapter 4 reviews more recent attempts in the computer graphics literature to improve the accuracy of digital imaging by incorporating explicit models of visual appearance. However, few of these models have directly addressed contrast reduction. Several use globally applied mapping functions similar to film, and like film they force a tradeoff between contrast compression and detail preservation. Others use locally varying attenuation functions, usually based on some form of linear bandpass filtering to separate fine details from large features, and each of these can suffer from strong and persistent halo artifacts that are not usually present in high contrast visual appearance. As such, new detail-preserving contrast reduction methods based on new ideas and background are badly needed.

8.2 Thesis Support: Achievability

The second claim of the thesis asserts that fine detail/large feature separations can achieve the desired detail-preserving contrast reduction properties, and is proven by example. As shown in the Results sections of Chapters 5– 7, all three new methods presented in this thesis reveal details in shadows and highlights that are often lost in the default scene-to-display mappings of film, television and computer graphics, but similar details are a strong distinguishing feature of the realistic works of skilled artists, whether they work at an easel, in a darkroom, or at a computer workstation placing synthetic lights.

I have long suspected that large feature/fine detail separations are not just conveniences, but are actually required to adequately model human visual appearance, as well as other scene decompositions such as illumination/reflectance or surface/boundary, but such claims are too speculative to act as a good thesis. Examined critically, the achievability claim may even seem tautological: once the details are separated from large features, preservation of either is trivial, but in this research the task of determining the *definitions* of large features and fine details was at least as important as devising the methods used separate them in the scene.

Local adaptation mechanisms in the retina enable human eyes to capture low contrast variations embedded within high contrast scenes, and these low-level mechanisms are fairly well measured and understood. They are so well studied, in fact, that Mead and Mahowald [Sch90] constructed an electronic "silicon retina" in 1987, a microelectronic array of photodetectors and analog circuitry that simulates local adaptation and motion sensing. Several other electronic retinal analogues have been constructed since then and more recently several research teams [RW98] are even now actively investigating electronic retinal implants to replace neural tissues destroyed or obscured by macular degeneration.

However, there is precious little direct evidence to explain how retinal ganglion signals leaving the eye are interpreted or assembled to form the visual appearance of high contrast scenes, and low-level measurements of the early stages of vision are easily misleading. For example, Hubel and Weisel's meticulous Nobel Prize-winning measurements and analysis of the striate cortex in macaque monkeys revealed an exquisitely orderly arrangement of size-, direction-sensitive cells [HW77], and

the receptive fields of the first stage, the so-called 'simple' cells bears a strong resemblance to the impulse response of linear bandpass filters and agrees with extensive psychophysical measurements of visual thresholds.

Accordingly, the literature of psychophysics, computational vision and image compression is filled with various bandpass filter decompositions of images. I suspect most of these models will be forever unverifiable and may not offer trustworthy predictions of appearance, except as predictors of their source data; thresholds of extremely weak stimuli on extremely simple backgrounds. Many of these models are demonstrably wrong as predictors of the appearance of complex, high contrast scenes, as illustrated by halo artifacts discussed in Chapter 3, and are intricate and computationally expensive.

All are carefully and rigorously built on the spectacular gains made in the past three decades in the understanding the early visual mechanisms, but I feel these are unwarranted extrapolations to higher visual processes that may be better understood by artists. By analogy, we seem to be cavemen with a computer, predicting the mysterious contents of the CPU based on a careful probing and analysis of only the keyboard. Despite the fundamental importance of their work, Hubel and Weisel themselves spoke of the vast distance remaining to a full understanding of visual appearance:

> "Our conclusions by no means solve the main question in vision–that is, how the brain makes sense of the image that falls on the retina, the image of a scene rich in form, color, depth and movement. We are only beginning to see a few steps along the way."
>
> *–David Hubel and Torsten Wiesel,*
> in their acceptance speech for the
> 1981 Nobel Prize in Medicine. [Wer84]

Instead, the layering, foveal and LCIS methods each relied on collections of more tentative, anecdotal evidence available for higher-level visual processes to define large features and fine details, as summarized in Chapter 3 Sections 3.1– 3.3. Any contrast reduction scheme must decide what portion of scene contents are discardable; for example, film discards small contrasts in highlights and shadows. However, as argued in the previous claim, these small contrasts are a vital part of visual appearance. Accordingly, the layering method discards high illumination contrasts by discarding its darkest and brightest extremes, the foveal method discards highlight and shadow details in peripheral vision and the LCIS method discards high contrasts of the simplest features. In each of

these cases, the discarded scene content is not missed and permits greater visual content than that available in displayed images made by more conventional means such as gamma compression.

The diversity of these three solutions has several plausible interpretations. Wildly different images of the same scene are commonplace in artistic depictions and our acceptance of such images seems to rely on the forgiving nature of human vision (or art audiences!). Given enough hints about the artist's intent in an image, viewers can often decode the scene contents. The human visual system yearns for order and will make plausible inferences from almost any visual signal, even an empty one; recall the infamous canals on Mars that Sir Percival Lowell insisted he saw through his pioneering telescopes. Are the layering, foveal and LCIS methods a visually appealing way to preserve scene detail only because they provide enough hints to let us figure out the rest of the scene?

I disagree, because our tolerance for distortions depends strongly on expectations; if we expect to see a line drawing then we forgive its lack of shading, but if we expect to see seamless integration of computer graphics and photography then even tiny mismatches in shading, lighting and movement are glaringly obvious. I argue that viewers of layering, foveal and LCIS results reasonably expect an accurate rendition of scene contents, and artifacts such as halos that are not a part of the visual appearance of the original scene are not easily tolerated in such images. Indeed, the exceedingly weak residual halo artifacts discussed at the end of the LCIS chapter are numerically small but draw the viewer's attention precisely because they are a slight divergence from the expected appearance of the displayed image.

Instead, I argue that the diversity of successful scene-to-display mapping methods shows that no one-to-one mapping exists from high contrast scenes to good display depictions of that scene, even so-called realistic depictions. As the displayed image can only indirectly suggest the contrasts it cannot reproduce, I would expect the visual system to collect these hints from a wide variety of scene cues; any combination of them might successfully convey scene contrasts. Again, artists offer instructive examples, as a night scene can be rendered in Rembrandt's deep dark oils, or in a simple water color, but both can convey darkness effectively. Accordingly, the layering, foveal

and LCIS methods are almost certainly just a few of the possible separation-based methods for detail-preserving contrast reduction.

8.3 Applications and Future Work

Each of the layering, foveal and LCIS methods has its own strengths, and each offers much room for improvements, extensions and new investigations.

The foveal method is well suited for interactive use and accepts high contrast scenes from any source, either measured real-world radiance-map-like data or synthetic scenes rendered by computer graphics, but it requires a computerized, interactive display, and its personalized response to a viewer's gaze makes layering unsuitable for static displays such as a printed page, and possibly annoying to viewers whose gaze direction does not match the cursor. I am particularly interested in how well the foveal method would work with an immersive display, especially if eye-tracking is available. Would display latency modify or corrupt the mentally assembled visual appearance of the scene? The foveal method might also benefit by using either of the other two methods to reduce contrasts and preserve more peripheral detail.

Unlike the foveal method, the layering method is well suited to static displays and multiple users, but requires more scene information than the foveal method. Until recently, only computer graphics renderings could supply the separate illumination, reflectance and transparency information about a scene needed to construct the tree of intrinsic image layers used by layering. However, a recent paper by Yu *et al.* [YDMH99] offers a method using controllable lights and radiance maps [DM97] to determine real-world scene lighting as well as its radiance; combining this work with layering offers interesting possibilities. Layering may also offer a useful artistic tool to aid in lighting synthetic scenes for display, as discussed at the end of Chapter 5, and when combined with a paint-box-like interactive tool layering might permit graphic artists to gracefully emphasize or de-emphasize selected scene components. I am also curious about how layering could be applied to animation; how should we adjust m and γ for each layer over time?

The LCIS method accepts scenes from any input and provides an output suitable for any display, but only in exchange for substantially increased computing costs. LCIS offers a new way to decompose an image reversibly into a multi-resolution set of boundaries, large features and fine details. Though the LCIS hierarchy of Figure 31 ignores the boundary pixel markings (See Chapter 7), these form a boundary hierarchy that may also prove useful. For example, current progressive image transmission methods on the World-Wide-Web send bandpass-like refinement signals, and the screen appearance resembles an aliased version of the focusing of a camera lens. Progressive transmission of LCIS boundary information would reveal the image as an artist might draw it; arriving first, the simple and important scene boundaries would be sharply defined, not blurred or aliased, and over time the smoothly shaded regions between them would fill with progressively finer boundaries and details. Scene components of boundaries, shadings and details may also prove useful for digital artists; a paint-box-like interface would allow cursor strokes to locally adjust weighting factors w in an LCIS hierarchy, 'painting' more or less detail wherever desired.

LCIS and its discrete implementation are easily extended to higher dimensions, and suggest some intriguing follow-on research. I addressed only static scenes, but good contrast reduction for animated scenes may very well require attention to both spatial and temporal boundaries and gradients in an image. Higher-dimensional LCIS might also prove useful for visualizing fine details in other high contrast data sets, including both scalar and vector fields of any dimension for use in astronomy, seismology, or visualizing finite element modeling results.

The layering, foveal and LCIS methods preserve details in displayed images of high contrast scenes, but are only preliminary steps towards the larger goal of perceptually valid displays. The three new methods each address only local luminance adaptation processes, but neglect many other equally important aspects of the visual appearance of high contrast scenes, some of them already addressed by other authors. Each of these methods would benefit from a more thoughtful treatment of color and color adaptation (see [PFFG98]), and color information must almost certainly contain additional boundary information that should be exploited in the LCIS method. Each of the three methods could be improved by modeling temporal response (see [FPSG96]) including time-varying local adaptation, especially the interactive foveal method. Neurophysiologists and psychophysicists

have identified and measured at least two multiplicative and one subtractive adaptation mechanism in the retina [WECH+90, HG92], each with its own time-constant; what are their effects on visual appearance? Previous computer graphics papers that address visual glare, blooming and diffraction [SSZG95, NKON90], after-images, masking effects [FPSG97] might also be exploited to improve any of the three methods in this dissertation. I have always been intrigued by the slowly varying forms of visual noise and uncertainty that cause the drifting indistinct patterns I see with my eyes closed. Similar patterns are present in night vision and are depicted by some artists in ways I have never seen in any computer graphics rendering. Although some authors have addressed loss of visual acuity in failing light [FPSG96, WLRP97], it is often modeled as a loss of high frequency response in the visual system, though its visual appearance is quite different and may also require models of visual uncertainty. Though I might not be able to read the largest letter of an eye chart lit only by starlight, the letter does not appear blurred; I can see that the chart and letter contains sharp boundaries, but I am unsure of their shape or position. How might we visually represent this uncertainty in a displayed image?

8.4 Closing Comment

The scene-to-display mapping problem is dauntingly broad, as it encompasses not just visual psychophysics and appearance, but higher forms of visual expression entirely within the artist's domain, probably out of the reach of science in my lifetime. Obviously no single scene-to-display mapping method is appropriate for all uses, and there are far more possibilities than just those offered by film or the three new methods presented here in Chapters 5– 7. My wandering explorations of cameras, displays and human vision have convinced me that scene-to-display mappings should be untangled and removed from cameras as much as possible, especially in the new digital electronic varieties that are free of the global, chemically imposed restrictions of film. Instead, cameras should only perform scene light measurements, capturing and recording scene radiances (and eventually, other scene properties such as depth, shape and BRDFs [YDMH99]) without regard to display device abilities. These scene-capturing tasks are a matter of objective physical measurements, well suited to electronic instruments, but the task of converting those scene measurements to a compelling

displayed image is a matter of artistry, taste and a deep understanding of visual appearance. Imagine how many opportunities for greater artistic expression are discarded by fixed, film-like transfer functions in new electronic cameras! Referring back to the introductory quote on page 1, how much more could an artist such as Degas achieve in a computerized darkroom with an accurate electronic record of scene properties and a large set of easily controlled scene-to-display mapping tools?

Figure 35: **Tree and Streetlight on a Foggy Night**
LCIS users can control the prominence of selected scene details. This scene-to-display mapping tool might help artists better capture "the atmosphere of lamps and moonlight."

Appendix A

Revised Tumblin-Rushmeier Tone Reproduction Operator

The foveal method presented in Chapter 6 depends on an improved tone reproduction operator made by revising and updating the method of Tumblin and Rushmeier [TR93]. I 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 I include a (γ) term to adjust contrasts just as m adjusts intensities:

$$L_d = m \cdot \left(\frac{L_w}{L_{wa}} \right)^{\gamma} \tag{51}$$

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.

Next, determine γ from Tumblin and Rushmeier's original tone reproduction operator [TR93], 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 [SS60, SS63], 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:

$$B = C_0 \left(\frac{L}{L_a} \right)^{\gamma} \tag{52}$$

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,

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. I have modified their γ term in three ways. As shown in equation 56 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 [SEC83]. I 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 52 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:

$$B_w = C_0 \left(\frac{L_w}{L_{wa}} \right)^{\gamma_w} = B_d = C_0 \left(\frac{L_d}{L_{da}} \right)^{\gamma_d}. \tag{53}$$

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

$$L_d = L_{da} \left(\frac{L_w}{L_{wa}} \right)^{\left(\frac{\gamma_w}{\gamma_d} \right)}. \tag{54}$$

The result has the same form as Equation 51, 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. I remove this anomaly by appending a new scale factor term $m(L_{wa})$.

My $m(L_{wa})$ function is built from a simple conjecture about visual appearance; I 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. I choose $m(L_{wa})$ to vary according the same log-linear expression Stevens used to find contrast sensitivity γ in Equation 56, forming an almost straight line (or a straight series of dots in Figure 13D) 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:

$$L_d = m(L_{wa}) \cdot L_{da} \cdot \left(\frac{L_w}{L_{wa}}\right)^{\left(\frac{\gamma_w}{\gamma_d}\right)} \tag{55}$$

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

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:

$$\gamma(L_a) = \begin{cases} 2.655 & \text{for } L_a > 100\ cd/m^2 \\ 1.855 + 0.4\log_{10}(L_a + 2.3 \cdot 10^{-5}) & \text{otherwise,} \end{cases} \tag{56}$$

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

$$m(L_{wa}) = \left(\sqrt{C_{max}}\right)^{(\gamma_{wd}-1)} \tag{57}$$

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 13. I apply the operator defined by Equations 55, 56 and 57 in Chapter 6.

Appendix B

LCIS Source Code

This appendix presents a fragment of C++ source code for the LCIS function used to make all results images shown in Chapter 7 and in the SIGGRAPH'99 paper "LCIS: A Boundary Hierarchy For Detail-Preserving Contrast Reduction" by Jack Tumblin and Greg Turk. Machine-readable versions of this code are also available on the ACM SIGGRAPH'99 Proceedings CD-ROM and is downloadable from my web-site at `http://www.gvu.gatech.edu/people/jack.tumblin/`.

The code is offered only as an aid to people who may want to implement LCIS themselves, and was written to test and explore LCIS easily, not for speed or efficiency. You may notice that the 'leakfix' and 'diffusion' passes through the image could be combined to cut code size almost in half, and that the code is full of expensive sqrt() calls and seemingly pointless casts between `PIXTYPE` and `double` etc.; these are vestiges of earlier experiments. The code is neither fast nor elegant, but instead was written to be readable and easy to debug. Please use, share and modify it in any way you wish.

A few items may need some explanation:

- `Raw2D` and `Raw3D` are my own floating-point image classes whose pixel datatype is set by `#DEFINE PIXTYPE float`. Both hold resizeable arrays of pixels; `Raw2D` holds a 2D array indexed by `(x,y)` and `Raw3D` holds a 1D array of `Raw2D` objects indexed by `z`.
- In the code below, Raw3D objects all hold three Raw2D objects.
- Raw2D and Raw3D member functions used here include:
  - `get(x,y)` returns the PIXTYPE pixel value at (x,y)
  - `put(x,y,val)` write PIXTYPE value

- `sizer(arg)` set array size to match 'arg'
- `wipecopy(src)` copy contents of 'src,' resize as needed.
- `ones()` set every pixel value to 1.0
- `zeros()` set every pixel value to 0.0

- The LCIS input image is read from `m_pA`, a pointer to a `Raw3D` image object. It holds three `Raw2D` objects, and the LCIS routine uses the middle one of three. For example, to read input pixel value at `(x,y)` use: `val = m_pA->get(x,y,1);`.

- An unseen Win32 dialog box class holds several user-settable parameters, including:

  number of timesteps to run: found at `m_pParamDlg->m_diffusCount`

  Conductance Threshold 'K': found at `m_pParamDlg->m_diffusKval`

  Timestep length 'T': found at `m_pParamDlg->m_diffusTstep`

  Leakfix enable/disable: found at `m_pParamDlg->m_diffusLeakfix`

- The LCIS output image is written to `m_pOut`, a pointer to a `Raw3D` object, which holds three `Raw2D` objects; two hold the current values for the leakfix multipliers (D_E, D_N in Chapter 7) and the other holds the result of the LCIS operation:

  `m_pOut->z[0]` holds the EW leakfix multiplier,

  `m_pOut->z[1]` holds the LCIS output image,

  `m_pOut->z[2]` holds the NS leakfix multipler.

- **Quick Overview:** The code begins by copying the input image to the output image object; it never touches the input object again. After initializing a few pesky details, the LCIS code below uses just a few image objects: `src`, `bound` and `m_pOut` (the output image object).

  The `src` object holds the result of the most recently completed timestep; use `src` as the source of all pixel information needed to compute results of the current timestep and write them to `m_pOut`. (`m_pOut` and `src` also hold the values of the leakfix multipliers for each pixel). Once finished finding all new pixel values for `m_pOut`, the code copies `m_pOut` object to `src` and start objects on the next timestep.

  For each timestep there are two passes through the entire image; The first pass computes the "leakfix multiplier" value for each link, and determines whether or not the link has become a

"boundary link." If it has, the code marks the pixels on either end of the link as a "boundary pixel" by setting its value in the Raw2D image object `bound` to the value `IS_BOUND_PIX`.

The second pass computes the flux through all links and applies the flux to the output image `m_pOut`. However, it does not compute or apply any flux at all through any link with a boundary pixel on either end. This accomplishes two goals: first, no flux will ever transform a ridge-like shock into a step-like shock (prohibited by the LCIS PDEs) and second, no curvature estimate will use pixels that straddle a shock, where curvature is (theoretically) infinite.

```
BOOL CFcnGroup::fcnLCIS4(CString & label)
//-------------------------------------------------------------------------
// Low Curvature Image Simplifier(LCIS)
//
//      K, timestep, leakfix enable, # iterations,
//      from SetFcnParam dialog class.
//
//-------------------------------------------------------------------------
//      --prevents any curvature or motive force estimates from using a set of
//         pixels that straddles a 'shock', and
//      --ensures that a 'shock' always forms a ridge discontinuity that never
//         (due to smoothing within a region) evolves into a step discontinuity,
//         mimicking the behavior of the PDE form of LCIS.
// Here's how:
//      --DEFINE: any pixel on either end of a link that contains a shock is a
//              'Boundary Pixel', (shock==leakfix multiplier is >10.0)
//      --APPLY: any link CONNECTED to a boundary pixel must have ZERO FLUX to
//              ensure the boundary pixel will not change.
//      Note that this scheme ADIABATIC and fully consistent with the
//      continuous PDE form of LCIS.
//-------------------------------------------------------------------------
// (Return 1 if successful; retn 0 if text 'label' does not match our internal
// name for this function; this is a double-check that index# used to select
// functions matches the desired function)
{
Raw2D src;                          // source img for each diffusion timestep.
Raw2D bound;                        // boundary-pixel map.

int i,iE1,iE2,iW,j,jN1,jN2,jS,k,imax,jmax,kmax;     // image indices
PIXTYPE N2val,N1val,Pval,S1val,W1val,E1val,E2val,NEval,NWval,SEval;
                                    // neighborhood pixels: north,south,..
                                    // (See LCIS paper, Figure 6)
PIXTYPE Pxx2,Pyy2,Exx2,Eyy2,Nxx2,Nyy2,Nxy2,Sxy2,Wxy2;
                                    // 2nd partial derivative ests. (squared)
double Nmag2,Emag2;                 // squared 'edginess' estimates m_EW, m_NS
double cXXX,cXXY,cXYY,cYYY;         // 3rd partial derivative ests.
double fN,fE;                       // Motive force for NS, EW links.
double condN,condE;                 // conductances for N,E links
double gainN,gainE,threshN,threshE; // leakfix temp. vars
double gcN,gcE;                     // temp: gain*conductance.
double fluxN,fluxE;                 // flow for this link, this timestep;
double lambtime;                    // actual timestep length
double k2inv;                       // 1.0/K^2 (K==diffusion constant)
#define BOUND_THRESH 10.0           // Boundary threshold for leakfix mpy.
#define IS_BOUND_PIX 0.0f           // flag value use to mark boundary pixels
                                    // in Raw2D 'bound'
//******MAGIC CONSTANTS******
double gainLimit = 1E+15;           // upper bound on leakfix gain boost
double tstep = 1.0/32.0;            // default timestep length

    if(0!=label.CompareNoCase(JT_LCIS4))    // if label mismatch,
    {
        BEEP;                               // COMPLAIN
        return(FALSE);
    }
    //-----------------------------------------------------initialize;
    m_pOut->wipecopy(*m_pA);        //  copy input image to output image.
                                    // use z=0,z=1 planes of output image
                                    // to store leakfix multiplier values:
    m_pOut->z[0].ones();            // set leakfix EWgain to 1.0.
    m_pOut->z[2].ones();            // set leakfix NSgain to 1.0.
    imax = m_pOut->getXsize();      // get image size,
    jmax = m_pOut->getYsize();      // calc a few quick constants;
```

```
k2inv = 1.0/(m_pParamDlg->m_diffusKval *  // 1/(Conductance Threshold K)^2
             m_pParamDlg->m_diffusKval);
lambtime = tstep * m_pParamDlg->m_diffusTstep;  // timestep length
kmax = m_pParamDlg->m_diffusCount;  // number of timesteps to run.
bound.sizer(&(m_pA->z[1]));         // boundary-pixel image (0,1 @all pix)
bound.ones();                       // set to 'no boundary pixels'.
//-------------------------------------------------Timestep loop;
// use 'src' for input data, write output data to'm_pOut'.
for(k=0; k<kmax; k++)           // before each iteration,
{
    src.wipecopy(&(m_pOut->z[1]));// make local image copy.
    if(m_pParamDlg->m_diffusLeakfix==TRUE)
                                //-------------------------leakfix;
    {                           // gain-setting;
        for(j=0;j<jmax; j++)        // For each scanline,
        {
            jN1 = j+1;              //(addr of North,South neighbor)
            jN2 = j+2;
            jS  = j-1;
            if(jN1>=jmax) jN1=jmax-1;       // (stay in-bounds)
            if(jN2>=jmax) jN2=jmax-1;
            if(jS<0) jS=0;
            for(i=0; i<imax; i++)   // and each pixel on scanline,
            {
                iE1 = i+1;          // (address of East,West neighbor)
                iE2 = i+2;
                iW  = i-1;
                if(iE1>=imax) iE1=imax-1;   // (stay in-bounds)
                if(iE2>=imax) iE2=imax-1;
                if(iW<0) iW=0;
                N2val = src.get(i,  jN2);  // find neighborhood pixels
                NWval = src.get(iW, jN1);
                N1val = src.get(i,  jN1);
                NEval = src.get(iE1,jN1);
                W1val = src.get(iW, j  );
                 Pval = src.get(i,  j  );
                E1val = src.get(iE1,j  );
                E2val = src.get(iE2,j  );
                S1val = src.get(i,  jS );
                SEval = src.get(iE1,jS );
                // find squared xx,yy curvatures through pixels P,E and N:
                Pxx2 = E1val+W1val-(PIXTYPE)2.0* Pval;
                Pxx2 = Pxx2*Pxx2;
                Pyy2 = N1val+S1val-(PIXTYPE)2.0* Pval;
                Pyy2 = Pyy2*Pyy2;
                Exx2 = E2val+ Pval-(PIXTYPE)2.0*E1val;
                Exx2 = Exx2*Exx2;
                Eyy2 = NEval+SEval-(PIXTYPE)2.0*E1val;
                Eyy2 = Eyy2*Eyy2;
                Nxx2 = NEval+NWval-(PIXTYPE)2.0*N1val;
                Nxx2 = Nxx2*Nxx2;
                Nyy2 = N2val+ Pval-(PIXTYPE)2.0*N1val;
                Nyy2 = Nyy2*Nyy2;
                // find squared xy curvatures through pixel midpoints N,S,W:
                Nxy2 = (NEval-E1val)-(N1val-Pval);
                Nxy2 = Nxy2*Nxy2;
                Sxy2 = (E1val-SEval)-(Pval-S1val);
                Sxy2 = Sxy2*Sxy2;
                Wxy2 = (N1val-Pval)-(NWval-W1val);
                Wxy2 = Wxy2*Wxy2;
            // Use wt'd. sums of these to est. squared link curv. mag;
                Emag2 = 0.25*(double)(Pxx2+Pyy2+Exx2+Eyy2)
```

```
                    + 0.5*(double)(Nxy2+Sxy2);
                Nmag2 = 0.25*(double)(Pxx2+Pyy2+Nxx2+Nyy2)
                    + 0.5*(double)(Wxy2+Nxy2);
                gainE = (double)m_pOut->z[0].get(i,j);
                                    // current E leakfix gain;
                gainN = (double)m_pOut->z[2].get(i,j);
                                    // current N leakfix gain;
                threshE = Emag2*k2inv;  // test;
                threshN = Nmag2*k2inv;  // test:
                                //  is edginess m > cond. thresh K?
                                // if not, shrink leakfix multiplier
                                // 1/5th of the way towards 1.0;
                if(threshE<1.0)     // gainE = gainE - 0.2*(gainE - 1.0);
                {                   // or put more succinctly,
                    gainE = 0.8*gainE + 0.2;
                }
                else                // but if so, BOOST leakfix multiplier
                {                   // (see LCIS paper, equations 21,22)
                    gainE = gainE*(1.0+ sqrt(Emag2));
                }
                if(threshN<1.0)
                {
                    gainN = 0.8*gainN + 0.2;
                }
                else
                {
                    gainN = gainN*(1.0+ sqrt(Nmag2));
                }
                                //  write min(gain,gainLimit)
                if(gainE>gainLimit) gainE = gainLimit;
                if(gainN>gainLimit) gainN = gainLimit;
                m_pOut->z[0].put(i,j,(PIXTYPE)gainE);
                m_pOut->z[2].put(i,j,(PIXTYPE)gainN);
                if(gainE > BOUND_THRESH)    // mark all boundary pixels
                {                           // (pixels on either end of
                                            // any link holding a shock)
                    bound.put(i , j,IS_BOUND_PIX);
                    bound.put(iE1,j,IS_BOUND_PIX);
                }
                if(gainN > BOUND_THRESH)
                {
                    bound.put(i,j  ,IS_BOUND_PIX);
                    bound.put(i,jN1,IS_BOUND_PIX);
                }
            }//for(i...)
        }//for(j...)
    }//if(leakfix...
                    //---------------------------diffuse
    for(j=0; j<jmax; j++)               // for each scanline,
    {
        jN1 = j+1;                      // addr. of North,South neighbors
        jN2 = j+2;
        jS  = j-1;
        if(jN1>=jmax) jN1=jmax-1;       // (stay in-bounds)
        if(jN2>=jmax) jN2=jmax-1;
        if(jS<0) jS=0;
        for(i=0; i<imax; i++)           // for each pixel on scanline,
        {
            iE1 = i+1;                  // addr. of East,West neighbors
            iE2 = i+2;
            iW  = i-1;
            if(iE1>=imax) iE1=imax-1;
            if(iE2>=imax) iE2=imax-1;
```

```
if(iW<0) iW=0;
fluxE = 0.0;                    // (clear flux values)
fluxN = 0.0;
            // If 'P' is a boundary pixel then NEITHER
            // the EW or NS link have any flux;
if(IS_BOUND_PIX != bound.get(i,j))
{                               // get common neighbor values;
    N1val = src.get(i,  jN1);
    NEval = src.get(iE1,jN1);
    W1val = src.get(iW, j  );
     Pval = src.get(i,  j  );
    E1val = src.get(iE1,j  );
    S1val = src.get(i,  jS );   // find curv. at pixel P
                                // (for link conductance calcs)
    Pxx2 = E1val+W1val-(PIXTYPE)2.0*Pval;
    Pxx2 = Pxx2*Pxx2;
    Pyy2 = N1val+S1val-(PIXTYPE)2.0*Pval;
    Pyy2 = Pxx2*Pxx2;
                                // find cross curv. between
                                // pixels P,N1,NE,E1;
    Nxy2 = (NEval-E1val)-(N1val-Pval);
    Nxy2 = Nxy2*Nxy2;
    if(IS_BOUND_PIX !=bound.get(iE1,j))
    {                           // if EW link doesn't connect
                                // to a boundary pixel, then
                                // find its flux;
        E2val = src.get(iE2,j  );   // get its pixel values
        SEval = src.get(iE1,jS );
                                // compute change-of-curvature
                                // for motive force fE;
        cXXX =  (E2val - W1val) + 3.0*(Pval - E1val);
        cXYY =  (NEval - N1val) + (SEval - S1val)
                                + 2.0*(Pval - E1val);
        fE = cXXX + cXYY;
        // For EW link conductance calcs, find curv. mag.^2
        // estimate Emag2:
        Exx2 = E2val+Pval-(PIXTYPE)2.0*E1val;
        Exx2 = Exx2*Exx2;
        Eyy2 = NEval+SEval-(PIXTYPE)2.0*E1val;
        Eyy2 = Eyy2*Eyy2;
                                // cross-curvature below link;
        Sxy2 = (E1val-SEval)-(Pval-S1val);
        Sxy2 = Sxy2*Sxy2;
        // Make weighted sum of squares to find a rotationally
        // invariant curvatures mag^2 estimate for EW link;
        Emag2 = 0.25*(double)(Pxx2+Pyy2+Exx2+Eyy2)
            + 0.5*(double)(Nxy2+Sxy2);
        if(m_pParamDlg->m_diffusLeakfix==TRUE)
        {           // leakfix: apply gain to curvature^2 ests.
            gainE = (double)m_pOut->z[0].get(i,j);
            gcE = Emag2 * gainE;
            condE = 1.0  / (1.0 + (k2inv*gcE));
        }
        else
        {
            condE = 1.0 / (1.0 + (k2inv*Emag2));
        }
        // flux=lambtime*conductance*motive force;
        fluxE = lambtime * condE * fE;
    }
    if(IS_BOUND_PIX !=bound.get(i,jN1))
    {                           // if NS link doesn't connect
                                // to a boundary pixel, then
```

```
                                    // find its flux;
                    N2val = src.get(i,  jN2);   // get its pixel values
                    NWval = src.get(iW, jN1);
                                    // compute change-of-curvature
                                    // for motive force fN;
                    cYYY =  (N2val - S1val) + 3.0*(Pval - N1val);
                    cXXY =  (NEval - E1val) +  (NWval - W1val)
                                    + 2.0*(Pval - N1val);
                    fN = cXXY + cYYY;
                    // For NS link conductance calcs, find curv. mag.^2
                    Nxx2 = NEval+NWval-(PIXTYPE)2.0*N1val;
                    Nxx2 = Nxx2*Nxx2;
                    Nyy2 = N2val+Pval-(PIXTYPE)2.0*N1val;
                    Nyy2 = Nyy2*Nyy2;
                                    // cross-curvature left of link
                    Wxy2 = (N1val-Pval)-(NWval-W1val);
                    Wxy2 = Wxy2*Wxy2;
                    // Make weighted sum of squares to find a rotationally
                    // invariant curvatures mag^2 estimate for NS link;
                    Nmag2 = 0.25*(double)(Pxx2+Pyy2+Nxx2+Nyy2)
                            +0.5*(double)(Wxy2+Nxy2);
                    if(m_pParamDlg->m_diffusLeakfix==TRUE)
                    {           // leakfix: apply gain to curvature^2 ests.
                        gainN = (double)m_pOut->z[2].get(i,j);
                        gcN = Nmag2 * gainN;
                        condN = 1.0 /  (1.0 + (k2inv*gcN));
                    }
                    else
                    {
                        condN = 1.0 / (1.0 + (k2inv * Nmag2));
                    }
                    // flux=lambtime*conductance*motive force;
                    fluxN = lambtime * condN * fN;
                }
                // Now apply flux to output image;
                m_pOut->addTo((PIXTYPE)(-fluxN-fluxE),i,j,1);// center pix
                m_pOut->addTo((PIXTYPE)fluxN,i,  jN1,1);    // North
                m_pOut->addTo((PIXTYPE)fluxE,iE1,j  ,1);    // East
            }//if(IS_BOUND_PIX...
        }//for(i...
    }// for(j...
}//for(k...)

#undef BOUND_THRESH                             // locals only!
#undef IS_BOUND_PIX
    return(TRUE);
}
```

Bibliography

[Act98] S. Acton. Multigrid anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):280–291, March 1998.

[Ade93] E. H. Adelson. Perceptual organization and the judgment of brightness. *Science*, 262:2042–2044, 24 December 1993.

[AH92] A. N. Akansu and R. A. Haddad. *Multiresolution Signal Decomposition*. Telecommunications, a book series. Academic Press, Boston, 1992.

[Are94] 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.

[Ash94] I. Ashdown. *Radiosity: A Programmer's Perspective*. John Wiley and Sons, New York, NY, 1994.

[BA83] 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.

[BB67a] C. J. Bartleson and E. J. Breneman. Brightness perception in complex fields. *Journal of the Optical Society of America*, 57(7):953–957, July 1967.

[BB67b] C. J. Bartleson and E. J. Breneman. Brightness reproduction in the photographic process. *Photographic Science and Engineering*, 11(4):254–262, July-August 1967.

[BD85] A. R. Biondini and M. L. F. DeMattiello. Suprathreshold contrast perception at different luminance levels. *Vision Research*, 25(1):1–9, January 1985.

[Bla46] H. R. Blackwell. Contrast thresholds of the human eye. *Journal of the Optical Society of America*, 36(11):624–643, November 1946.

[Bla81] H. R. Blackwell. *An Analytic Model for Describing the Influence of Lighting Parameters upon Visual Performance Volume I: Technical Foundations.* Commission Internationale De L'Eclairage, 1981.

[BM90] F. J. J. Blommaert and J. B. Martens. Mapping luminance into brightness: an object-oriented approach. In Bernice E. Rogowitz and Jan P. Allebach, editors, *Human Vision and Electronic Imaging: Models, Methods, and Applications*, number 1249 in Proceedings, pages 376–386, P.O.Box 10, Bellingham WA 98227, February 1990. Society of Photo-Optical Instrumentation Engineers, SPIE-The International Society for Opticdal Engineering. 12-14 February 1990, Santa Clara, CA.

[BT78] 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.

[Can86] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, November 1986.

[Car95] J. Carucci. *Capturing The Night With Your Camera: How To Take Great Photographs After Dark.* Amphoto Books, Watson-Guptill Publications, 1515 Broadway, New York, NY 10036, 1995. ISBN 0-8174-3661-8.

[CAS+97] C. J. Curtis, S. E. Anderson, J. E. Seims, K. W. Fleischer, and D. H. Salesin. Computer-generated watercolor. In Turner Whitted, editor, *SIGGRAPH 97*, Annual Conference Series, pages 421–430. ACM SIGGRAPH, Addison-Wesley, August 1997.

[Cat79] E. Catmull. A tutorial on compensation tables. In *Computer Graphics (Siggraph 79 Conference Proceedings)*, volume 13 of *Annual Conference Series*, pages 1–13. Addison-Wesley, July 1979.

[CF91] M. W. Cannon and S. C. Fullenkamp. A transducer model for contrast perception. *Vision Research*, 31(6):983–998, June 1991.

[CG84] M. A. Cohen and S. Grossberg. Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, 36(5):428–456, 1984.

[CHJ80] B. H. Carroll, G. C. Higgins, and T. H. James. *Introduction to Photographic Theory: The Silver Halide Process*. A Wiley-Interscience Publication. John Wiley and Sons, New York, NY, 1980.

[CHS+93] 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.

[Cor70] T. N. Cornsweet. *Visual Perception*. Academic Press, New York, NY, 1970.

[CW93] M. F. Cohen and J. R. Wallace. *Radiosity and Realistic Image Synthesis*, chapter 9.5.1: Gamma Correction, page 268. Academic Press Professional, 955 Massachusetts Avenue, Cambridge, MA 02139, 1993.

[Dau90] J. G. Daugman. An information-theoretic view of analog representation in striate cortex. In Eric L. Schwartz, editor, *Computational Neuroscience*, chapter 31, pages 403–423. "A Bradford Book", 1990. A collection of papers presented at a 1987 Symposium.

[Deg98] E. Degas. about 1895, discussing photography in a letter to a friend, quoted in a caption to one of his works in the show *Capturing Time: A Celebration of Photography* at The Getty Museum, Los Angeles CA, December 16, 1997–March 22, 1998.

[DeM72] L. E. DeMarsh. Optimal telecine transfer characteristics. *Journal of the SMPTE*, 81(10):784–787, October 1972. Society of Motion Picture and Television Engineers.

[DeM77] L. E. DeMarsh. Evaluation of color rendering in film and television. *SMPTE Journal*, 86(9):624–625, September 1977. Society of Motion Picture and Television Engineers.

[DeM93] L. E. DeMarsh. TV display phosphors/primaries-some history. *SMPTE Journal*, 102(12):1095–1098, December 1993. Society of Motion Picture and Television Engineers.

[DG81] E. T. Davis and N. Graham. Spatial frequency uncertainty effects in the detection of visual sinusoidal gratings. *Vision Research*, 21:705–712, 1981.

[DM97] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In Turner Whitted, editor, *SIGGRAPH 97*, Annual Conference Series, pages 369–378, New York, NY, August 1997. ACM SIGGRAPH, Addison-Wesley.

[Dow87] J. Dowling. *The Retina: An Approachable Part of the Brain.* Harvard University Press, Cambridge, MA, 1987.

[DuB92] J. M. H. DuBuf. Modelling spatial vision at the threshold level. *Spatial Vision*, 6(1):25–60, January 1992.

[FDB92] B. V. Funt, M. S. Drew, and M. Brockington. Recovering shading from color images. In G. Sandini, editor, *Proceedings of the European Conference on Computer Vision (ECCV '92)*, pages 124–132. Springer-Verlag, May 1992.

[FPSG96] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual adaptation for realistic image synthesis. In Holly Rushmeier, editor, *SIGGRAPH 96*, Annual Conference Series, pages 249–258, New York, NY, August 1996. ACM SIGGRAPH, Addison-Wesley. Held in New Orleans, LA, 4–9 August 1996.

[FPSG97] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual masking for computer graphics. In Turner Whitted, editor, *SIGGRAPH 97*, Annual Conference Series, pages 143–152. ACM SIGGRAPH, Addison-Wesley, August 1997.

[FvDFH96] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics Principles and Practice.* Systems Programming Series. Addison-Wesley Publishing Company, Reading, MA, 2nd edition, 1990,1996.

[Gei84] W. S. Geisler. Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, 1(7):775–782, July 1984.

[Gei89] W. S. Geisler. Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, 96(2):267–314, February 1989.

[Geo79] M. Georgeson. Spatial fourier analysis and human vision. In *Tutorial Essays in Psychology: Guide to Recent Advances*, chapter 2, pages 39–88. Lawrence Erlbaum Assoc., 1979. Sutherland, S. editor.

[Gil90] 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.

[Gil94] A. L. Gilchrist. *Lightness, Brightness, and Transparency.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.

[GJ84] A. L. Gilchrist and A. Jacobsen. Perception of lightness and illumination in a world of one reflectance. *Perception*, 13:5–19, 1984.

[Gla95] A. Glassner. *Principles of Digital Image Synthesis*, volume 1 of *Computer Graphics and Geometric Modeling*, chapter 3.4:Monitors, page 100. Morgan-Kaufman Publishers, Inc., San Francisco, CA, 1995.

[GM98] E. J. Giorgianni and T. E. Madden. *Digital Color Management: Encoding Solutions.* Addison-Wesley Longman, Inc., Reading, MA, 1998.

[Gol96] E. B. Goldstein. *Sensation and Perception.* Brooks/Cole Publishing, Pacific Grove, CA, 4th edition, 1996.

[Gra89] N. V. S. Graham. *Visual Pattern Analyzers.* Oxford Psychology Series. Oxford University Press, New York, NY, 1989.

[Gro90] S. Grossberg. A model cortical architecture for the preattentive perception of 3-d form. In Eric L. Schwartz, editor, *Computational Neuroscience*, chapter 31, pages 117–138. "A Bradford Book", 1990. A collection of papers presented at a 1987 Symposium.

[GT88] S. Grossberg and D. Todorovic: Neural dynamics of 1-d and 2-d brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics*, 43:241–277, 1988.

[GTS+97] D. P. Greenberg, K. E. Torrance, P. Shirley, J. Arvo, J. A. Ferwerda, S. Pattanaik, E. P. F. Lafortune, B. Walter, S. C. Foo, and B. Trumbore. A framework for realistic image synthesis. In Turner Whitted, editor, *SIGGRAPH 97*, Annual Conference Series, pages 477–494. ACM SIGGRAPH, Addison-Wesley, August 1997.

[GW87] R. C. Gonzalez and P. Wintz. *Digital Image Processing.* Addison-Wesley, Reading, MA, 2nd edition, 1987.

[Hai93] N. D. Haig. Local gain control and focal accommodation in the self-similar stack vision model. *Spatial Vision*, 7(1):15–34, January 1993.

[Hal89a] R. Hall. *Illumination and Color in Computer Generated Imagery*, chapter 5.4:Monitor Alignment and Calibration, pages 140–153. Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010, 1989.

[Hal89b] R. Hall. *Illumination and Color in Computer Generated Imagery*, chapter 5.1.2:Gamma Correction for Display, pages 120–125. Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010, 1989.

[Hee87] D. J. Heeger. Model for the extraction of image flow. *Journal of the Optical Society of America A*, 4(8):1455–1471, August 1987.

[Hen35] R. H. Henneman. A photometric study of the perception of object color. *Archives of Psychology*, 179:5–89, 1935.

[HF86] 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.

[HG92] D. C. Hood and N. Graham. Modeling the dynamics of light adaptation: The merging of two traditions. *Vision Research*, 32(7):1373–1393, July 1992.

[HS93] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume I, II. Addison-Wesley, 1993.

[Hun75] R. W. G. Hunt. *The Reproduction of Colour*. Kings Langley:Fountain Press, 3rd edition, 1975. (5th edition is now available).

[HW77] D. H. Hubel and T. N. Wiesel. Ferrier Lecture: Functional architecture of macaque monkey visual vortex. *Proceedings of the Royal Society of London B.*, 198:1–59, July 1977.

[HWS71] 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.

[JC48] L. A. Jones and H. R. Condit. Sunlight and skylight as determinants of photographic exposure: I. luminous density as determined by solar altitude and atmospheric conditions. *Journal of the Optical Society of America*, 38(2):123–178, February 1948.

[JC49] L. A. Jones and H. R. Condit. Sunlight and skylight as determinants of photographic exposure: II. scene structure, directional index, photographic efficiency of daylight and safety factors. *Journal of the Optical Society of America*, 39(2):94–135, February 1949. see page 123.

[JG87] L. E. Arend Jr. and R. Goldstein. Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A*, 4(12):2281–2285, December 1987.

[JG90] L. E. Arend Jr. and R. Goldstein. Lightness and brightness over spatial illumination gradients. *Journal of the Optical Society of America A*, 7(10):1929–1936, October 1990.

[JN42] L. A. Jones and C. N. Nelson. Control of photographic printing by measured characteristics of the negative. *Journal of the Optical Society of America*, 32(10):558–619, October 1942.

[JRW97a] 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.

[JRW97b] 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.

[Kan90] G. Kanisza. Subjective contours. In I. Rock, editor, *The Perceptual World*, pages 155–163. W. H. Freeman, New York, NY, 1990.

[Kau86] 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.

[KM88] F. Kingdom and B. Moulden. Border effects on brightness: A review of findings. *Spatial Vision*, 3(4):227–263, April 1988.

[KPC93] J. K. Kawai, J. S. Painter, and M. F. Cohen. Radioptimization - goal based rendering. In *SIGGRAPH 93*, Annual Conference Series, pages 147–154, New York, NY, August 1993. ACM SIGGRAPH, Addison-Wesley. Held in Anaheim, CA, 1–6 August 1993.

[LC94] X. Li and T. Chen. Nonlinear diffusion with multiple edginess thresholds. *Pattern Recognition*, 27(8):1029–1037, August 1994. Pergamon Press.

[LKG94] D. Luo, M. King, and S. Glick. Local geometry variable conductance diffusion for post-reconstruction filtering. *IEEE Transactions on Nuclear Science*, 41(6):2800–2806, December 1994.

[LM71] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, January 1971.

[LM88] E. H. Land and J. J. McCann. Lightness and retinex theory. In Whitman Richards, editor, *Natural Computation*, chapter III.13, pages 177–193. A Bradford Book, The MIT Press, 1988. Reprint of 1971 JOSA paper.

[Mar82] D. Marr. *Vision*. W. H. Freeman and Co., 1982.

[MC91] N. A. Macmillan and C. D. Creelman. *Detection Theory: A User's Guide*. Cambridge University Press, New York, NY, 1991.

[Mey91] G. W. Meyer. *An Inexpensive Method of Setting the Monitor White Point*, volume II of *Graphics Gems*, chapter III.7, pages 159–162. Academic Press, Harcourt Brace Jovanovic, Publishers, Boston MA, 1991. Edited by James Arvo.

[Mil72] G. Millerson. *The Technique of Lighting for Television and Motion Pictures*. Communication Arts books: The Library of Communication Techniques. Hastings House, New York, NY, 1st edition, 1972.

[Mor50] J. Moranz. *The Professional Guide to Drawing and Illustration: Mastery of Drawing*. Grosset, New York, NY, 1950.

[Nal93] V. S. Nalwa. *A Guided Tour of Computer Vision*. Addison-Wesley Publishing, Reading, MA, 1993.

[Nel66a] C. N. Nelson. The theory of tone reproduction. In T. H. James, editor, *The Theory of the Photographic Process*, chapter 22, page 498. The Macmillan Company, New York, London, third edition edition, 1966. See reference on page 498 to "F. Hurter and V. C. Driffield, *J. Soc. Chem. Ind.(London),* **10,** 100 (1891)." I could not obtain a copy of this work.

[Nel66b] C. N. Nelson. The theory of tone reproduction. In T. H. James, editor, *The Theory of the Photographic Process*, chapter 22, pages 464–498. The Macmillan Company, New York, London, 3rd edition, 1966.

[NKON90] E. Nakamae, K. Kaneda, T. Okamoto, and T. Nishita. A lighting model aiming at drive simulators. In *Computer Graphics (SIGGRAPH 90 Conference Proceedings)*, Annual Conference Series, pages 395–404, New York, NY, August 1990. ACM SIGGRAPH, Addison-Wesley. Held in Dallas, TX 6–10 August 1990.

[NS92] M. Nitzberg and T. Shiota. Nonlinear image filtering with edge and corner enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):826–833, August 1992.

[OSS68] 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.

[Pel90] E. Peli. Contrast in complex images. *Journal of the Optical Society of America A*, 7(10):2032–2040, October 1990.

[Pel91] E. Peli. Effect of luminance on suprathreshold contrast perception. *Journal of the Optical Society of America A*, 8(8):1352–1359, August 1991.

[PFFG98] S. N. Pattanaik, J. A. Ferwerda, M. D. Fairchild, and D. P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In Michael Cohen, editor, *SIGGRAPH 98*, Annual Conference Series, pages 287–298. ACM SIGGRAPH, Addison-Wesley, July 1998.

[PG80] D. H. Pritchard and J. J. Gibson. Worldwide color television standards– similarities and differences. *SMPTE Journal*, 89(2):111–120, February 1980.

[PM87] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. In *Proceedings of the IEEE Computer Society Workshop on Computer Vision*, pages 16–22. IEEE Computer Society, November 1987.

[PM90] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[Poy93] C. A. Poynton. "Gamma" and its disguises: The nonlinear mappings of intensity in perception, crts, film and video. *SMPTE Journal*, 102(12):1099–1108, December 1993. Copyright 1993, Society of Motion Picture and Television Engineers, Inc.

[Poy98a] C. A. Poynton. Frequently asked questions about Gamma. Web page document, January 1998. see `http://www.inforamp.net/~poynton/GammaFAQ.html`.

[Poy98b] C. A. Poynton. The rehabilitation of Gamma. In B. E. Rogowitz and T. N. Pappas, editors, *Proceedings of SPIE–The International Society for Optical Engineering*, volume 3299 of *Human Vision and Electronic Imaging III*, pages 232–249, held in San Jose, CA, January 1998. Bellingham, Wash.:SPIE, 1998.

[RJW96] 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.

[RW69a] 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.

[RW69b] M. L. Rubin and G. L. Walls. *Fundamentals Of Visual Science*, chapter The Measuring of Light, pages 42–47. Charles C. Thomas, Springfield, IL, 1969.

[RW98] J. F. Rizzo and J. L. Wyatt. Retinal prosthesis. *Age-Related Macular Degeneration*, pages 413–432, 1998.

[Sch86] W. F. Schreiber. *Fundamentals of Electronic Imaging Systems*, volume 15 of *Springer Series in Information Sciences*. Springer-Verlag, Berlin, Germany, 1986.

[Sch90] Eric L. Schwartz, editor. *Computational Neuroscience*, chapter 25: A Silicon Model of Early Visual Processing, pages 331–339. MIT Press, 1990. Collection of papers presented at a 1987 Symposium.

[Sch92] J. P. Schaefer. *Basic Techniques of Photography: an Ansel Adams Guide*. Little Brown, Boston, MA, 1992.

[Sch95] 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.

[SDS96] E. J. Stollnitz, A. D. DeRose, and D. H. Salesin. *Wavelets for Computer Graphics Applications*. Morgan Kaufmann Series in Computer Graphics and Geometric Modeling. Morgan Kaufmann, San Francisco CA, 1996.

[SEC83] R. Shapley and C. Enroth-Cugell. Visual adaptation and retinal gain controls. In Neville N. Osborne and Gerald J. Chader, editors, *Progress in Retinal Research, Volume 3*, chapter 9, pages 263–346, see p. 283. Pergamon Press, Oxford, 1983.

[SFAH92] E. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, March 1992.

[SG92] J. Schubert and A. L. Gilchrist. Relative luminance is not derived from absolute luminance. *Investigative Opthalmology and Visual Science*, 33(4):1258, April 1992.

[SP94] F. X. Sillion and C. Puech. *Radiosity and Global Illumination*, chapter A.1.16:Gamma Correction, pages 218–219. Morgan Kaufman Publishers, Inc., 340 Pine Street, Sixth Floor, San Francisco CA 94104, 1994.

[SS60] 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.

[SS63] 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.

[SSZG95] G. Spencer, P. Shirley, K. Zimmerman, and D. P. Greenberg. Physically based glare effects for digital images. In *SIGGRAPH 95*, Annual Conference Series, pages 325–334, New York, NY, August 1995. ACM SIGGRAPH, Addison-Wesley. Held in Los Angeles CA. 6–11 August 1995.

[Ste61] S. S. Stevens. To honor Fechner and repeal his law. *Science*, 133(12):80–86, January 13 1961.

[Sto72] T. G. Stockham. Image processing in the context of a visual model. *Proceedings of the IEEE*, 60(7):828–842, July 1972.

[TH94] P. Teo and D. Heeger. Perceptual image distortion. In *First IEEE International Conference on Image Processing*, pages 982–986, November 1994. Held in Los Alamitos, CA, 11–16 November 1994.

[THG99] J. Tumblin, J. Hodgins, and B. Guenter. Two methods for display of high contrast images. *ACM Transactions On Graphics*, 18(1):56–94, January 1999.

[TO97] 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.

[TR93] J. Tumblin and H. Rushmeier. Tone reproduction for computer generated images. *IEEE Computer Graphics and Applications*, 13(6):42–48, November 1993.

[TT99] J. Tumblin and G. Turk. LCIS: a boundary hierarchy for detail-preserving contrast reduction. In Alyn Rockwood, editor, *SIGGRAPH 99*, Annual Conference Series, pages 83–90. ACM SIGGRAPH, Addison-Wesley, August 1999.

[VA84] C. Van Allsburg. *The Mysteries of Harris Burdick*. Houghton Mifflin, New York, NY, 1984.

[WAJA89] A. B. Watson and Jr. A. J. Ahumada. A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex. *IEEE Transactions on Biomedical Engineering*, 36(1):97–106, December 1989.

[Wan95] B. A. Wandell. *Foundations of Vision*, chapter 5:The Retinal Representation, page 148. Sinauer Associates, Inc., Sunderland, MA, 1995.

[War91] G. Ward. Real pixels. In James Arvo, editor, *Graphics Gems II*, chapter Image Processing: II.5, pages 80–85. AP Professional, Cambridge MA 02139, 1991.

[War94a] G. Ward. A contrast-based scalefactor for luminance display. In Paul S. Heckbert, editor, *Graphics Gems IV*, chapter Frame Buffer Techniques: VII.2, pages 415–421. AP Professional, Cambridge MA 02139, 1994.

[War94b] G. Ward. The RADIANCE Lighting simulation and rendering system. In *SIGGRAPH 94*, Annual Conference Series, pages 459–472, July 1994.

[Wat87a] A. B. Watson. The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics And Image Processing*, 39:311–327, 1987.

[Wat87b] A. B. Watson. Efficiency of a model human image code. *Journal of the Optical Society of America A*, 4(12):2401–2417, December 1987.

[WC69] P. Whittle and P. D. C. Challands. The effect of background luminance on the brightness of flashes. *Vision Research*, 9:1095–1110, 1969.

[WC92] 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.

[WECH+90] 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.

[Wer84] L. T. Wertenbaker. *The Eye: Window to the World.* The Human Body. Torstar Books, Inc., New York, NY, 1984.

[WLRP97] 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.

[WP93] R. Whitaker and S Pizer. A multi-scale approach to nonuniform diffusion. *CVGIP: Image Understanding*, 57(1):99–110, January 1993.

[WW92] A. Watt and M. Watt. *Advanced Animation and Rendering Techniques: Theory and Practice*. ACM Press. Addison-Wesley Publishing Company, New York, NY, 1992. ISBN 0-201-54412-1.

[Yar67] A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, New York NY, 1967. Translated from Russian by Basil Haigh. Original Russian Text published for the Institute of Information Transmission of the Academy of Sciences of the USSR by Nauka Press, Moscow, in 1965.

[YDMH99] Y. Yu, P. Debevec, J. Malik, and T. Hawkins. Inverse global illumination: Recovering reflectance models of real scenes from photographs. In Alyn Rockwood, editor, *SIGGRAPH 99*, Annual Conference Series, pages 215–224. ACM SIGGRAPH, Addison-Wesley, August 1999.

[YXKT95] Y. You, W. Xu, M. Kaveh, and A. Tannenbaum. On ill-posed anisotropic diffusion models. In *Proceedings of the International Conference On Image Processing*, volume 2 of 3 of *8186-7310*, pages 268–271. IEEE Computer Society, October 1995. Washington D.C.

[ZC89] C. Zetzsche and T. Caelli. Invariant pattern recognition using multiple filter image representations. *Computer Vision, Graphics, and Image Processing*, 45:251–262, 1989.

[ZS87] C. Zetzsche and W. Schönecker. Orientation selective filters lead to entropy reductions in the processing of natural images. *Perception*, 16:225, 1987.

Vita

Jack Tumblin, born in Durham, North Carolina in 1955, lived in northeast Brazil from ages 1-5 and 12-13 but otherwise grew up in Atlanta. At Towers High School he won scholastic honors and several state-wide science fair awards, and in his senior year he attended classes at Georgia Tech.

As a Cooperative Plan undergraduate, Jack worked at the Engineering Experiment Station (GT-EES, later GTRI), where he helped design an image generator for missile testing. He was the WREK-FM (student radio station) Music Director in 1977, took music classes at Georgia State University, patented a music theory training device (#4,321,853) and earned a BSEE in 1978.

After a semester of film school at the University of Southern California in 1980, he worked at KTLA-TV/Los Angeles, then did contract work on custom TV equipment. In early 1983 Jack's work helped raise venture-capital funding for IVEX Corporation and he was the first employee hired.

In 1984, his work on a videodisc-based 3D flying video game yielded two patents (#4,752,836 and #4,873,585). In 1985-7 his work on the VDS-1000, an electronic image generator for airline pilot training, was filed in patent #4,807,158. The VDS-1000-series products enabled IVEX to grow to over 45 employees. Scientific-Atlanta Inc. subcontracted Jack in 1988-9 and filed patents (#5,142,575 and #4,924,498) on the split-sync video scrambling method he invented for them.

In 1987 Jack began part-time course work and earned a Masters degree (MS) in Electrical Engineering from Georgia Tech in 1990. At the end of 1991 he left IVEX for full-time doctoral studies in Computer Science. In 1994, Jack accepted a one-year internship at Microsoft Research. After returning to Atlanta, Jack helped IVEX design a new flight simulator during 1996-7 and then completed his Ph.D in 1999.