Manipulating Scale-Dependent Perception of Images

by

Matthew Trentacoste

B.Sc. Computer Science, Carnegie Mellon University, 2003

M.Sc. Computer Science, University of British Columbia, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Computer Science)

The University Of British Columbia

(Vancouver)

November 2011

Abstract

The purpose of most images is to effectively convey information. Implicit in this assumption is the fact that the recipient of that information is a human observer, with a visual system responsible for converting raw sensory inputs into the perceived appearance. The appearance of an image not only depends on the image itself, but the conditions under which it is viewed as well as the response of human visual system to those inputs. This thesis examines the scale-dependent nature of image appearance, where the same stimulus can appear different when viewed at varying scales, that arises from the mechanisms responsible for processing spatial vision in the brain. In particular, this work investigates changes in the perception of blur and contrast resulting from the image being represented by different portions of the viewer's visual system due to changes in image scale. These methods take inspiration from the fundamental organization of spatial image perception into multiple parallel channels for processing visual information and employ models of human spatial vision to more accurately control the appearance of images under changing viewing conditions. The result is a series of methods for understanding the blur and contrast present in images and manipulating the appearance of those qualities in a perceptually-meaningful way.

Preface

All publications, along with the relative contributions of the collaborating authors, that have resulted from the research presented in this thesis are listed in the following.

Scale-Dependent Perception of Countershading *M. Trentacoste, R. Mantiuk and W. Heidrich.*

This work is currently under submission [Trentacoste et al., 2011b] and presented in Chapter 6. The author had the initial idea, implemented the software for both the perceptual experiment and the algorithm, wrote the majority of the paper and compiled the submission video. Dr. Mantiuk analyzed the results of perceptual experiment, contributed to the writing of the paper and provided insights in discussions. Dr. Heidrich supervised the project, aided in writing the paper and contributed to discussions.

Synthetic Depth-of-Field for Mobile Devices *M. Trentacoste, R. Mantiuk, W. Heidrich.*

This work is currently awaiting resubmission [Trentacoste et al., 2011c] and presented in Chapter 4. The author had the initial idea, implemented the software, and wrote the paper. Dr. Mantiuk provided input in discussions. Dr. Heidrich supervised the project, contributed ideas and aided in drafting the paper.

Blur-Aware Image Downsampling *M. Trentacoste, R. Mantiuk, W. Heidrich.*

The contents of this paper [Trentacoste et al., 2011a] are split between chapters, where the details of the blur estimation algorithm are included in Chapter 4 while Chapter 5 presents the blur-aware downsizing operator. The author had the

initial idea, implemented the software for both the perceptual experiment and the algorithm, wrote the majority of the paper, compiled the submission video and presented the work at Eurographics 2011. Dr. Mantiuk analyzed the results of perceptual experiment, contributed to the writing of the paper and provided insights in discussions. Dr. Heidrich supervised the project, contributed ideas and aided in writing the paper.

Quality-Preserving Image Downsizing *M. Trentacoste, R. Mantiuk, W. Heidrich.*

This poster [Trentacoste et al., 2010b] described an earlier version of the work contained in Trentacoste et al. [2011a,c], and contained additional details on the appearance of noise in downsampled images. The pertinent contributions are also included in Chapter 4. The author had the initial idea, implemented the software, created the poster and presented at the SIGGRAPH Student Research Competition. Dr. Mantiuk contributed to discussion. Dr. Heidrich supervised and contributed ideas.

Defocus Techniques for Camera Dynamic Range Expansion *M. Trentacoste, C. Lau, M. Rouf, R. Mantiuk, W. Heidrich.*

The contents of this paper [Trentacoste et al., 2010a] are included Chapter 7. Dr. Heidrich had the initial concept for the project, aided in the writing of the paper, supervised the project and contributed ideas. Cheryl Lau implemented portions of the software, provided input into the evaluation and during discussions. The author implemented the majority of the software, conducted the evaluation, drafted most the paper, and presented the at Electronic Imaging 2010. Dr. Mantiuk and Mushfiqur Rouf provided input in discussions.

Portions of this thesis were conducted under *Perception of image characteristics and differences*, Ethics Certificate Number **H09-02278** of the UBC Research Ethics Board.

Table of Contents

List of Tables

List of Figures

Glossary

CSF contrast sensitivity function

DOF depth-of-field

HDR high dynamic range

HDTV high-definition television

HVS human visual system

JND just-noticable-difference

LDR low dynamic range

LGN lateral geniculate nucleus

MRScale minimum reliable scale

MURA Modified Uniformly Redundant Array

PSF pointspread function

PSNR peak signal-to-noise ratio

SLR single-lens reflex

SNR signal-to-noise ratio

UMO unsharp masking operator

WLS weighted least-squares

Acknowledgments

Completing my degree has been quite a journey. As such, there are quite a few people along the way to which I owe a debt of gratitude. Fives years is a long time. Times change, people move apart. I've almost certainly forgotten someone that helped me out at one point or other. This list is by no means complete and I am sure I owe thanks to many more people than I list here.

To all the people of the Imager and PSM labs who have given me guidance and camaraderie: Cheryl Lau, Brad Atcheson, Nasa Rouf, Tyson Brochu, Allan Rempel, Anika Mahmud, Lukas Ahrenberg. In particular, Gordon Wetzstein, who beared more than his share of my existential angst and provided much guidance in how to work on the right things.

To my friends who have supported me over the years, through making me dinner, buying me a drink, letting me rant or whatever else was necessary at the moment including (in no particular order) Max Ulis, Hannah Holmes, Keenan O'Conner, Jamie Abugov, Drew Smith, Tru Moreland, Justine Townsend, Tom Schulz, Ross Kakushke, Erin Caton, Jody Marshall, Jen Mackie, Annie Chiavaroli, Kat Southam, Nicole Sanches, Jules Emmerson, Matthew Harris, Marin Patenaude, Nathan Fenwick, Hendrik Kueck, Kim Kowaliuk and Brian Leroux. To my housemates, Adam Barlev and Marlo Carpenter, who have been infinitely patient with my late nights, frantic running to and fro and generally leaving a small disaster in my wake. Finally, to the members of my adoptive West Coast family: Josh Williams, Kelsey Nash, Monica Pearson and Shelley Oliver, who I can't even begin to summarize all the gratitude I owe.

To my supervisory committee, Michiel van de Penne and Bob Woodham, for all their suggestions and input and encouragement. To my collaborator, Rafał Mantiuk,

for all the time he spent helping my half-baked ideas come to fruition and all the guidance he gave me on how to be a better researcher. Finally, to my supervisor, Wolfgang Heidrich, for supervising the entirety of my graduate career. For all of his input and insights, patience through my missteps and mess-ups, encouraging me and guiding me and continuing to believe in me even I wasn't sure if I did myself.

Lastly, I would like to thank my family. To my sisters, for the encouragement and belief that I could make it through, for always having the time to listen to whatever I needed to complain about, and taking time to celebrate my minor victories along the way. And last but certainly not least, my parents, Kathy and Michael. For, of all the things they've given me, the realization that success doesn't depend on intelligence or luck, but hard work and sticking with things you believe in no matter how uncomfortable it may feel at the moment.

Chapter 1

Introduction

1.1 Perceived Appearance

Consider the scenario of a professional photographer documenting a scene of interest: the subject could be candid portraits in low light, fast-moving sports or a macro shot of an insect. In any case, there is the chance of some undesired blur degrading the image, so the photographer diligently reviews the images on the viewfinder to ensure the necessary parts are in focus. After inspecting the captured images and finding them to be satisfactory, the photographer wraps up and heads back to their studio.

However, when the photographer views the images at full-size on their computer, they appear blurry, despite the smaller versions having appeared sharp. Most digital camera users have encountered this phenomenon on one or more occasions, where the viewfinder on the camera does not accurately represent the full-size image captured. When shown at a small size, the image often appears sharper than its larger-scale counterpart. The pixel content of the image did not change, but the image somehow looks different. While the image itself has remained unaltered, some change in how the image was viewed caused it to be understood differently.

Any discussion about image understanding includes assumptions about the perceptual makeup of the viewer, in this case the visual system of a human observer. One of the most powerful features of the human visual system (HVS) is its ability to operate effectively across a vast array of conditions, adapting to provide not only

meaningful but consistent results. This property can easily be seen by our ability to read this page in both direct sunlight and a dimly lit room while maintaining a nearly constant appearance of the page across such a significant change.

The goal of visual perception is to provide a meaningful representation of the scene being observed. Due to various biological limitations, the neural channels in our visual system are not capable of conveying the vast range spanned by the physical quantities of that scene. To overcome these limitations and provide a representation of the scene that we recognize, the HVS must continually adapt to changing conditions. This adaptation relies upon representing how the image appears in a manner not based on the underlying physical amount of light received by the eye, but a more general set of perceived quantities.

The combination of these perceptual quantities describe the *visual appearance* of the scene, as understood by the viewer. The adaptation process inherent in perception preserves the visual appearance of the text on this page, allowing it to be equally readable in direct sun or dim lighting. We define the visual appearance of a stimulus as:

Visual appearance The supra-threshold perception of a stimulus produced by a viewer's visual system defined in terms of perceptual units like brightness and colorfulness.

Specifically, we distinguish the visual appearance of a stimulus from the measurements of the underlying physical object being observed. Some aspects of visual appearance, such as perceived contrast, directly relate an underlying property (physical contrast), while other aspects, such as colorfulness, are more abstract. We refer to the collection of physical properties of an object as the *physical appearance*, to distinguish between the physical stimuli and the perceived phenomenon.

The concept of visual appearance, and by extension image appearance, is a complex issue. Terms like color and brightness are simultaneously universal and elusive, intuitively understood by everyone, yet hard to define rigorously. The HVS performs exceeding well at the task of providing a consistent visual appearance of our surroundings and we generally go about our lives with the intuitive notion that objects appear the same regardless of how we happen to be viewing them. However, there are numerous examples in everyday life where this assumption does not hold:

- Car headlights, barely noticeable during the day, are nearly blinding at night.
- Scenes appear more colorful and of higher contrast on a sunny day.
- Different color matte boards cause artwork to take on different appearances.

None of these examples can be explained by physical measurements of the subject matter alone. In all of these cases, we consider the visual appearance of the object of interest to have changed, while the underlying physical representation of that object has generally remained constant. Sometimes that change is consistent with our notions of the world, such as the examples of the car headlights. Sometimes we tend not to even consider the change, such as with scenes appearing more colorful on a sunny day. In both cases, the visual appearance of the scene changed, and we understood the change and accounted for it.

In other cases, the change in visual appearance can be unexpected and even confusing, as is the case with the appearance of the artwork and the apparent change in sharpness of the photographs described at the beginning. In all cases, the visual appearance changed, but in these two examples the appearance changed in a way that we did not expect and our understanding of the phenomena changed with it. Even after a lifetime of experience, the mind has not completely learned to account for all the nuances of the perceptual mechanisms it relies on to experience the world.

Consider the example of the matted artwork. Part of the reason the change in the appearance is so unexpected is that none of the physical properties of the artwork itself changed. If there is a change in the physical properties of an object, we intuitively understand that we will perceive it differently. In this case, a change in one part of the scene (the matte) caused another part of the scene (the artwork) to change in appearance. This particular scenario brings us to an important realization:

> *The perception of a stimulus does not occur in isolation and the visual appearance of any given subject depends on more than just the inherent properties of the object being observed.*

The relationship implies that extrinsic factors, beyond the pixel content of the image, affect its perception. The fields of perception and color science have have conducted considerable amounts of research into quantifying and correcting these

changes in appearance. The increases of colorfulness and contrast of a scene on a sunny day are respectively known as the *Stevens effect* and the *Hunt effect*, while the artwork changing in appearance as a result of the matte is know as the *Bartleson-Breneman effect* [see Fairchild, 2005].

1.2 Scale-Dependence

In terms of extrinsic factors, the field of color science is mostly concerned with the perception of global properties, such as image-wide contrast and luminance. More complete models [Fairchild and Johnson, 2004], that address the complexities in image appearance can account for some additional factors but still fall short of capturing the full complexity of visual perception. None of the current models of image appearance account for what the photographer observed with regards to the sharpness of the images.

This thesis investigates how the scale at which an image is viewed can alter the perception of sharpness and contrast features in that image and how understanding image appearance can improve image processing algorithms. Any change to the geometry of a given viewing condition, such as the size of the displayed image or the distance from the viewer, will cause a proportional change in the size of the image projected on the retina of the observer. As a result, the physical relation between the viewer and the image can determine the scale at which image features are perceived.

Evidence from both psychophysical studies [Blakemore et al., 1970] and neurological recordings [Hubel and Wiesel, 1959] suggests that the visual system contains mechanisms that only respond to a narrow range of spatial frequencies. That the threshold of human contrast sensitivity changes as a function of spatial frequency [Blakemore et al., 1970] further supports this hypothesis. Changes in the physical relation of image and viewer can change the visual channel to which a given feature is mapped. As a result, differences in how individual channels represent visual information can cause perceived properties like sharpness and contrast to change in appearance when the image projected onto the retina is altered.

The size of the image formed on the retina is affected by aspects of the display-viewer relationship that include: the physical size of the display, the distance between the viewer and the display, and the pixel dimensions of the displayed image. Often

changes in image appearance are a result of a change in more than one of these aspects. Walking three times closer to a monitor and switching from a screen in a movie theater to an iPhone result in considerable changes in the size of the image projected on the retina. Knowing how large the display will be and how far away the viewer will be is crucial to conveying the right appearance of that image.

The dimensions (in pixels) at which an image is displayed are even more important to consider, since they change even more frequently. Very few cameras remaining on the market, including mobile devices, capture an image at low enough resolution to show on a 1080p HDTV display without resizing. As a result, image downsampling has become synonymous with image display.

1.3 Manipulating Image Appearance

The purpose of most images is to effectively convey information to a human observer. However, most computer graphics research focuses on producing the images that are either the most physically correct or mathematically optimal. Physical correctness and mathematical optimality do not necessarily capture the complex relations required for an image to best communicate information.

Involving elements of human visual perception in the development of image processing algorithms can ensure that the result can appear the same as the original and be understood correctly. The vast majority of images are produced to be viewed, interpreted and understood by human observers. Given that humans do not directly perceive the underlying physical representation of an image, but some analogous representation of appearance, it follows that most image processing algorithms could produce better results if they focused on producing the image with the desired appearance instead of optimizing another metric.

In most cases, the appearance of an image matches the physical representation closely enough that conventional appearance-naïve image processing methods can produce acceptable results. However, a perceptually-focused approach is crucial in the examples above where the perceived appearance of an image deviates significantly from its physical appearance. In these cases, appearance-aware algorithms can better account for the complex nonlinearities of visual perception and more accurately produce the desired interpretation.

In our work, we begin by quantifying the change in appearance that a viewer observes when the image is altered, building models of how that image is perceived under different conditions. We then use those models to develop image processing algorithms that incorporate perceptual considerations and produce results that may not be the most optimal with respect to simple mathematical metrics, but instead are the most visually appropriate. The result is a series of methods that first produce a perceived image according to some desired criteria and then solve for the physical image that best yields that result.

This approach parallels how color perception has informed color management practices. The transformations between device color profiles not only account for differences between display response or gamut, but for changes in image perception. Color management algorithms must account for changes in display intensity or viewing environment altering the appearance of the image. Likewise, the methods we present address issues related to image perception across the vast size range of current displays, covering everything from cinema screens to mobile devices.

1.4 Novel Contributions

The algorithms all share the conceptual framework of *computational image analysis/synthesis*. At the core of each approach is a method that analyzes an image and produces an estimate of how some specific physical attribute changes across the image. The estimate is recast in another (usually perceptually-motivated) representation. We can preserve the perceived appearance when altering an image by solving for the appearance-preserving modifications to restore the perception of the transformed image. Alternatively we can alter the perceived appearance of an image by transforming the image appearance then solving for the modifications to the underlying image that will cause the desired change in appearance. Finally, given the original image and these modifications, we synthesize the final result. Figure 1.1 depicts this framework.

The main contribution of these methods are as follows:

Spatially-Variant Blur Estimation An efficient technique for robustly estimating the amount of blur present at each pixel in an image.

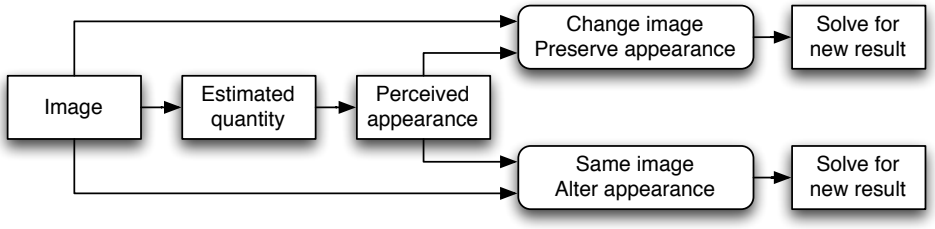


Figure 1.1: *Depiction of image analysis/synthesis framework shared by our contributions. In each case, we analyze some quantity in the image, then synthesize a new result that either preserves the appearance of that quantity or has a novel appearance.*

Synthetic Depth-of-Field for Mobile Devices An image enhancement technique to overcome the limitations of the optics in mobile devices and create a narrower depth of field than can be physically captured.

Blur-Aware Image Downsampling A perceptually-based approach for preserving the appearance of blur when downsizing images.

Scale-Dependent Perception of Countershading A model of visual appearance of countershading in complex images and the conditions under which countershading is considered objectionable.

Defocus Dynamic Range Expansion An investigation into increasing the effective dynamic range of image sensors by optically blurring the captured image and restoring it with deconvolution.

In the case of the *Synthetic Depth-of-Field* and *Blur-Aware Downsampling* algorithms, we analyze the image to produce a spatially-variant estimate of the blur present at each pixel. This estimate is transformed according to a model representing the desired change in image appearance to produce the blur present in the desired image. In the case of *Synthetic Depth-of-Field* the model represents a change in the aperture of the camera, while in the case of *Blur-Aware Downsampling* the model represents the change in perceived blur. Finally, the resulting image is synthesized from the original image and the result of the model.

With the *Scale-Dependent Perception of Countershading* paper, we conduct a perceptual experiment to determine what countershading parameter combinations in-

troduce halos and are thus considered objectionable. We demonstrate how this perceptual model applies to a number of image processing algorithms that either introduce or modify the appearance of countershading, including unsharp masking, tone mapping and image resizing. Additionally, we present some novel applications of our model including artifact-free countershading and viewer-adaptive displays.

In the case of *Defocus Dynamic Range Expansion* we investigated several techniques for deconvolving an optically blurred image to synthesize a final sharp image with a dynamic range that exceeded the physical limitations of the imaging sensor. The deconvolution algorithms we employ implicitly generates an estimate of the blur present at each pixel and synthesize the un-blurred result, concurrently adding any expansion in dynamic range. The overall approach is similar to analysis/synthesis but this method does not explicitly formulate the quantity being estimated.

1.5 Thesis Organization

This thesis begins with an overview of visual perception, image appearance, blur and contrast perception, and how these topics relate to the field of computer graphics.

Chapter 2 provides background on the makeup of the human visual system, the mechanisms of spatial vision, blur and contrast perception, image appearance. We begin with an overview of the vision concepts that form the foundation of our work, and build upon that with a discussion of how blur and contrast are represented in spatial vision. We continue with how these concepts related to image appearance, discuss specific examples such as the Cornsweet illusion, and conclude with a survey of existing attempts to model image appearance.

Chapter 3 covers existing research related to the design of optical systems and image processing algorithms that relate to our work. This survey includes model of blur formation in images, techniques for blur estimation and methods of restoring and altering image appearance.

In the following four chapters, the focus shifts to the topics in computer graphics related to our contributions. Chapter 4 presents the work of *Synthetic Depth-of-Field for Mobile Devices*, a method of spatially-variant blur estimation that is efficient enough for mobile devices yet robust enough to handle large amounts of noise. We use this blur estimation method to create images with a shallower depth-of-field than can be

captured by the camera optics.

Chapter 5 discusses how *Blur-Aware Image Downsampling* uses the same blur estimation method to preserve the appearance of blur in images when downsizing. Replacing the model of lens optics with a perceptual model of blur appearance, this method creates downsampled images that better resemble the appearance of the original.

Chapter 6 explains how *Scale-Dependent Perception of Countershading* extends this concept of preserving appearance to contrast perception. We present a model of how the width and contrast of countershading profiles affect how they are perceived and use this model to influence novel image processing methods.

Chapter 7 examines the feasibility of *Defocus Dynamic Range Expansion*, using modified optics and deconvolution to improve the effective dynamic range of image sensors. We take a blurry image captured by a lens with an aperture filter and attempt to recover a sharp image with a larger dynamic range than the original.

Finally, Chapter 8 concludes the thesis with a summary of all the material presented and discusses future directions of investigation.

Chapter 2

Visual Perception

This chapter presents elements of visual perception related to our work and discusses the physiological basis and algorithmic work concerning how images are perceived. These topics motivate our contributions detailed in Chapters 4 through 7.

To begin, Section 2.1 discusses the foundations of visual perception that relate to image appearance. In addition to image appearance, we present details on spatial vision and the band-limited nature of the channels involved in visual perception.

Sections 2.2 and 2.3 discuss the mechanisms in the HVS which mediate the perception of blur and contrast, respectively. The sections discuss how the underlying visual channels affect our perception of the phenomena, including sensitivity, and discuss the subtle differences between sensitivity to a perceptual quality and the perceived appearance of that quality.

Section 2.4 examines several higher-order visual phenomena arising from more complex edge profiles. These profiles, along with spatial contrast perception explain the appearance of contrast and blur in complex images.

Finally, Section 2.5 surveys image appearance models, computational attempts at modeling elements of visual perception. These algorithms incorporate models of how the HVS processes imagery, and attempt to produce a representation matching our perception of image appearance.

2.1 Foundations of Vision

2.1.1 Image appearance

As stated in the Introduction, we differentiate between the *visual appearance* of an object, as perceived by a human observer, and the underlying physical properties of that object. Visual perception includes complex properties like depth, occlusion and disparity that must be integrated to make sense of the wide array of visual stimuli present in the real world. In this thesis, we are primarily concerned with *image appearance*, the perception of two dimensional renditions on displays, printed pages, or other physical media.

Before attempting to inform the algorithms we construct to consider image appearance, we must first consider how the portions of an image can be perceived. In this section, we provide several examples of how the perceived appearance can differ from the underlying physical image. These examples provide context to the details of the HVS we present in the remainder of this chapter.

Within a single image, regions with identical physical properties can be interpreted completely differently, depending on the surrounding context. Adeleson's now-famous checkerboard [Adelson, 1995], reproduced in Figure 2.1, is one case where similar portions of an image can be perceived in different ways. In this image, the patches labeled **A** & **B** are the same intensity and will produce the same luminance on the retina. However, the other squares around **A** & **B** define separate contexts that alter the understanding, resulting in the luminance shared by **A** & **B** is interpreted as two different brightnesses. Adelson [1999] demonstrates that surrounding regions can have a larger effect on the perception of brightness than the region itself and the conditions under which an image is viewed strongly determine its appearance. However, once the context is changed by adding bars of the same color to connect the two regions, it becomes obvious that they are in fact the same intensity.

Likewise, the perception of the image also depends on the portion of the visual system responsible for conveying its appearance. The visual channels along which the HVS conveys information are not identical and do not reproduce the exact same response. As a result the appearance of an image is in part determined by which

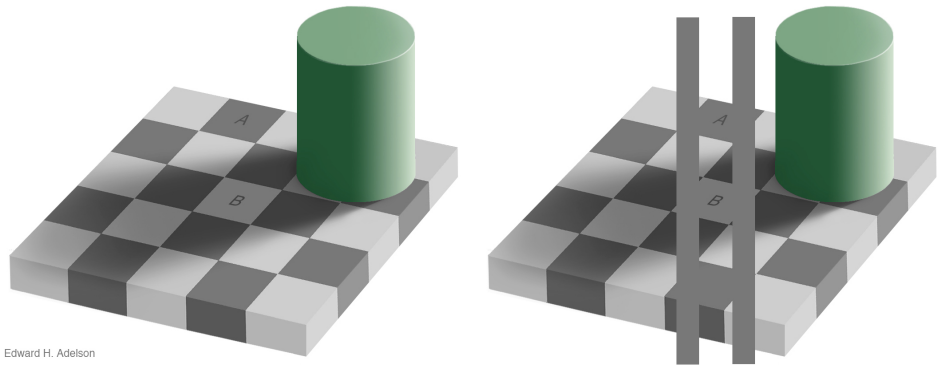


Figure 2.1: *Adelson checkersquare illusion. Patches **A** & **B** have the same luminance, but appear different brightnesses. Image copyright Edward Adelson.*

pathways were responsible for transmitting it to the conscious mind.

The *sensitivity* of the HVS to a quality of the image appearance differs depending on the conditions under which that image is perceived. We define sensitivity as the ability to observe changes in a particular visual stimulus, defined in terms of the threshold that must be exceeded in order for the stimulus to be detectable. The more sensitive the viewer, the lower the threshold and the more they are able to discern changes in an image.

The hybrid images by Oliva et al. [2006], shown in Figure 2.2 are one particularly surprising example of how changes in sensitivity can affect image appearance. Both the large left and small right image are the same, however the content of the image has specifically been designed to cause some details be visible at one scale while other details be visible at another. The change in contrast sensitivity between different visual channels causes the image to appear like Albert Einstein when viewed at a large size but Marilyn Monroe when viewed at a small size or from far away.

The sensitivity of the HVS to an image attribute and the appearance of that attribute are clearly related. However, the HVS is a complex nonlinear system with various elements of the perceptual process continually adapting and compensating for the performance of other elements. A difference in the sensitivity of the HVS under two sets of conditions does not necessarily imply that the resulting images will appear different. Likewise, a change in image appearance does not necessarily imply that the sensitivity of the visual system has changed. We explore these relations in

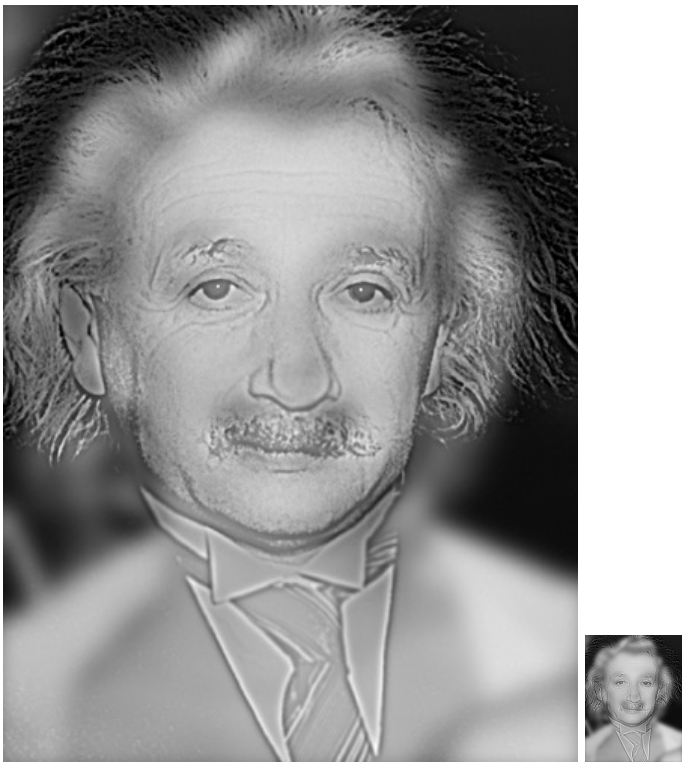


Figure 2.2: *Hybrid images of Oliva et al.. The image contains the high frequency content of Albert Einstein and the low frequency content of Marylyn Monroe. One image or the other is more apparent depending on which frequency content is closer to the sensitivity range of the* HVS. *Image copyright Aude Oliva.*

depth after a brief overview of the related low-level aspects of visual perception.

2.1.2 Spatial Vision

Hybrid images provide another concrete example of how the difference between visual channels representing an image in the HVS can cause changes in appearance. This section discusses the structures of the visual system that give rise to these visual channels. *Spatial Vision* by De Valois and De Valois [1990] provides detailed coverage of all levels of spatial vision. For a general discussion of the psychophysics and neurology related to visual perception, refer to *Vision Science* by Palmer [1999] and *Principles of Neural Science* by Kandel et al. [2000].

The act of visual perception begins at the retina, where the photons corresponding to the image are converted into neural impulses by photoreceptive rods and cones. The signals produced by these photoreceptive cells are transmitted to retinal ganglion cells, the next level of visual processing in the retina. The connections between photoreceptors and retinal ganglion cells form a complex network, where each photoreceptor sends, and each ganglion cell receives, information from multiple of their respective counterparts. Each portion of the retina has several functionally distinct subsets of ganglion cells conveying signals from the same photoreceptors in parallel [Kandel et al., 2000]. Even at this very early stage in the visual pathway, information from regions that extend beyond a single cell is being collected and transmitted in neural impulses from individual units.

The ganglion cells combine signals from several photoreceptors in ways that depend on precise spatial (and temporal) patterns [Kandel et al., 2000]. The signals from the ganglion cells travel along the visual pathway to the lateral geniculate nucleus (LGN), where the neurons connect to one of two kinds of cells, either *magnocellular* (M-cells) or *parvocellular* (P-cells) [Palmer, 1999]. The receptive fields of M-cells correspond to larger regions of the retina and respond stronger to stimuli of lower spatial frequencies. On the other hand, the receptive fields of P-cells correspond to smaller regions of the retina and respond stronger to stimuli of higher spatial frequencies. While the exact spatial sensitivities of both M- and P-cells vary and the two types overlap, it is apparent the HVS continues to partition ranges of possible spatial frequencies in more discrete parallel pathways.

The neural signals produced by the M- and P-cells leave the LGN and continue on to the visual cortex. Within the visual cortex, these signals are received by neurons known as *simple cells*. Similar to the pattern of connections linking photoreceptors, retinal ganglion cells and the cells of the LGN, there are multiple sets of connections between the M- and P-cells and the simple cells of the visual cortex operating in parallel. Within the visual cortex, this pattern of interconnectedness continues as larger and larger structures are linked together: Groups of simple cells are connected to complex cells. Groups of interconnected complex cells form columns in the visual cortex. Groups of interconnected columns form hypercolumns. Groups of adjacent hypercolumns are connected to each other.

The increasing interconnectedness at each stage of the visual system allows a

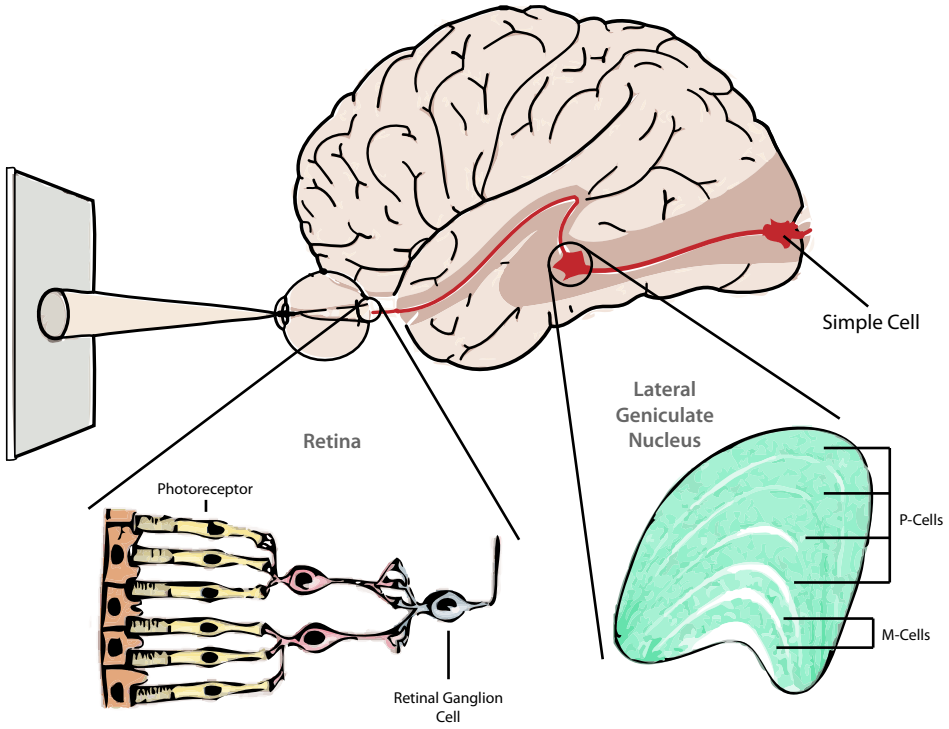


Figure 2.3: *Neurological bases for visual channels. The connection patterns of photoreceptors and ganglion cells, as well as the organization of the* LGN *are examples of how visual channels are formed in the* HVS.

greater range of possible scales to be represented. A photoreceptor on the retina will only respond to the light from a single point in the scene. On the other hand, a complex simple cell could respond to a range of different region sizes, depending on how large a group of photoreceptors are connected to it. Some cells will collect responses from only a small region of retinal cells while some cells will collect responses from a larger region. As a result, these cells respond to different-sized features [De Valois and De Valois, 1990]. Features too small for a given region are averaged out as the signals are summed together and features too large for that region cannot be represented represented by the the underlying mechanisms.

This configuration of multiple parallel connections gives rise to visual channels present in the HVS. Each channel acts as a band-pass filter, carrying information about a particular spatial frequency band (and orientation) [De Valois and De Valois, 1990].

This interpretation is support by neurological measurements by Hubel and Wiesel [1959] of cells in the V1 area of the visual cortex that selectively respond to different spatial frequencies and orientations, as well as the existence of horizontal connections between hypercolumns in the visual cortex [Hubel and Wiesel, 1959]. Further evidence is provided by psychophysical studies regarding selective adaptation of subjects to spatial frequency. In the seminal experiments by Blakemore et al. [1970], subjects shown gratings of a specific spatial frequency adapted to that frequency and demonstrated a temporary reduction in sensitivity to that spatial frequency. All of this evidence makes a compelling case for the existence of parallel channels of information processing in the HVS. Figure 2.4 shows an image decomposed into a Laplacian pyramid, a series of band-pass images. This representation is one possible interpretation of what information is carried in separate visual channels and discussed in Section 2.3.

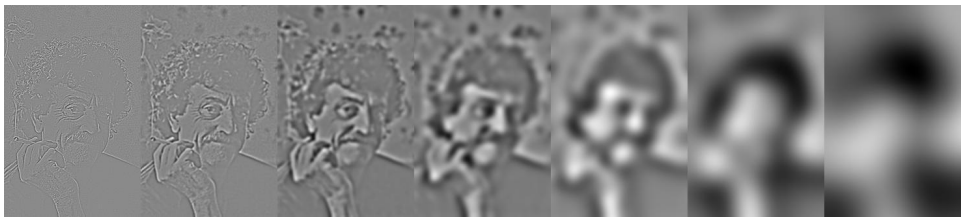


Figure 2.4: *Example of band-pass filters. Each image of Kurt Vonnegut has been processed by a different band-pass filter and contains separate spatial frequency content.*

These mechanisms have served as inspiration for a variety of image processing techniques. The concept of local or band-limited contrast inherent in the parallel visual channels representation has proven to be a powerful image processing technique. Gabor filters [Feichtinger and Strohmer, 1997] best model the responses of the simple cells in the visual cortex. While accurate representations of local spatial contrast and orientation sensitivity, Gabor filters are difficult to use computationally and other representations are often employed. There are a number of sub-band transforms that mimic properties of visual channels. The cortex transform by Watson [1987] provides a simpler, yet reasonably accurate model of the ability of the HVS to decompose visual stimuli into separate spatial frequencies and orientations. The Gaussian and Laplacian pyramids [Gonzalez and Woods, 2001] are more common

methods that ignore orientation yet retain segmentation of spatial frequencies.

2.1.3 Edges

Having presented evidence that the HVS transmits visual information via multiple channels, we now address the exact nature of the information these channels convey. As we have stated, these parallel channels begin with the connections between the photoreceptors and the retinal ganglion cells. Retinal ganglion cells connect to two separate sets of photoreceptors, an inner region surrounded by an outer region. The pattern in which these cells are connected displays *lateral inhibition* [De Valois and De Valois, 1990], where light falling on a given photoreceptor causes that cell to excite a ganglion cell and simultaneously inhibit the signals sent to the same ganglion cell by adjacent photoreceptors. The strength of impulses the cell sends increases with the difference between light falling on the inner and outer regions. The ganglion cells do not respond to changes in the absolute intensity of the light, but to the contrast between adjacent regions of the retina.

This relationship is maintained at subsequent stages of the visual system. The simple cells present in the visual cortex provide further evidence. While simple cells combine the information received from the retinal ganglion cells in a variety of configurations, all share the same three properties: they respond to a specific retinal position, they contain discrete excitatory and inhibitory zones and have a specific axis of orientation. Figure 2.5 contains several examples of the receptive fields associated with simple cells.

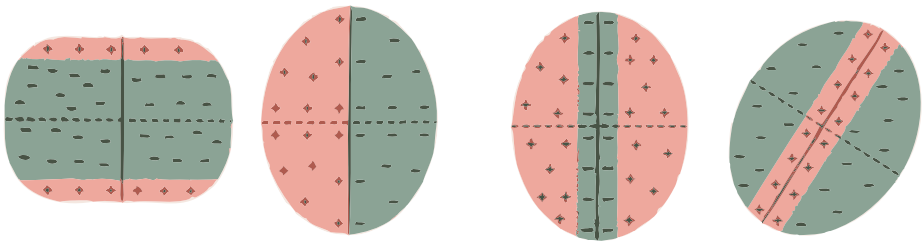


Figure 2.5: *Sample receptive fields of simple cells. Each contains discrete excitatory and inhibitory zones and have a specific axis of orientation.*

In particular, all simple cells are tuned to respond to specific combinations of

retinal ganglion cells that represent linear features. Since retinal cells respond primarily to contrast, the simple cells respond almost exclusively to the boundaries of an object. In other words, the simple cells in the visual cortex directly represent image edges. The boundaries between image regions are explicitly coded in the visual system and are a fundamental primitive of the HVS. The importance of which was best stated by David Hubel, neural physiologist [Palmer, 1999]:

> *Many people, including myself, still have trouble accepting the idea that the interior of a form … does not itself excite cells in our brain … that our awareness of the interior as black or white … depends only on cells' sensitivity to the borders. The intellectual argument is that perception of an evenly lit interior depends on the activation of the cells having fields at the borders and on the absence of activation of cells whose fields are within the borders, since such activation would indicate the interior is not even lit. So our perception of the interior as black, white, gray or green has nothing to do with cells whose fields are in the interior – hard as that may be to swallow … What happens at the borders is the only information you need to know: the interior is boring.*

Edges define regions of images and are the basis of our understanding of visual stimuli. The full characterization of an edge includes orientation, curvature and chromatic and other attributes. However, we focus on two specific properties: the blur and the contrast of the edge, depicted in Figure 2.6. We define the edge blur as the extent over which a luminance change occurs and the edge contrast as the magnitude of luminance change across the image region subtended by the edge.

The combination of magnitude (contrast) and extent (blur) of an edge is the primary factor determining the appearance of that edge. The blur of that feature determines which of the visual channels carry the information of a particular image feature. Likewise, the strength at which any visual channel is activated is representative of the contrast of the associated portion of the image. In the next three sections we will discuss the underlying properties of blur and contrast perception and how they relate to image appearance, first individually then in combination.

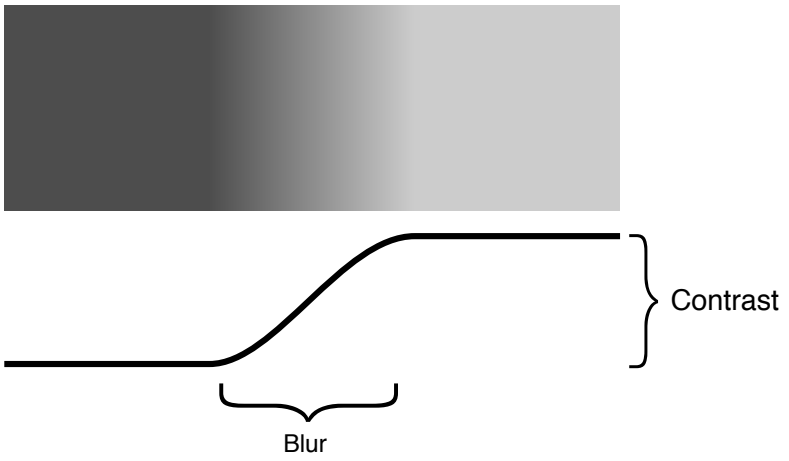


Figure 2.6: *Definition of edge contrast. Contrast is the magnitude of a transition. Blur is the spatial extent over which that transition occurs.*

2.2 Blur Perception

We examine the response of the visual system to changes in physical stimuli, starting with the perception of blur. There are two aspects of the response of the HVS to a stimulus: detection and discrimination. *Detection* (threshold effects) is the change required for a given stimuli to be visible as distinct from the surround while the *discrimination* (supra-threshold effects) is the difference at which two already-visible stimuli will be considered distinct. Both threshold and supra-threshold responses constitute an inverse relationship, where lower thresholds imply higher sensitivity to visual phenomena.

Blur is the spatial extent over which a change in contrast occurs. For a given magnitude change in contrast, the change of an edge with little blur will occur over a small region while the change of an edge with a larger blur will change over a larger region. In both cases, the total change in luminance is equal but the distance over which the change occurs varies.

In terms of complex images, blur is the absence or suppression of fine details in the image. Convolution with certain pointspread functions (PSFS) results in the attenuation of high frequency components of the image and the edges cannot contain abrupt changes in contrast. This absence of high frequency information is apparent

when an image is analyzed in the Fourier domain. Blur can also be interpreted as the spatial frequency above which there is not significant information:

$$\mathcal{F}(\omega) \approx 0, \text{ for } \omega > c \tag{2.1}$$

establishing a rough cut-off frequency c for the content of the image. This representation generalizes the effect of convolving an image by one of many individual PSF kernels, such as a box filter or Gaussian function and Figure 2.7 demonstrates the change in cutoff frequency for several Gaussian blurs. The exact appearance of the result can vary subtly, depending on exactly which frequencies are attenuated and by how much, but the end result is the high frequency content of the image will be attenuated.

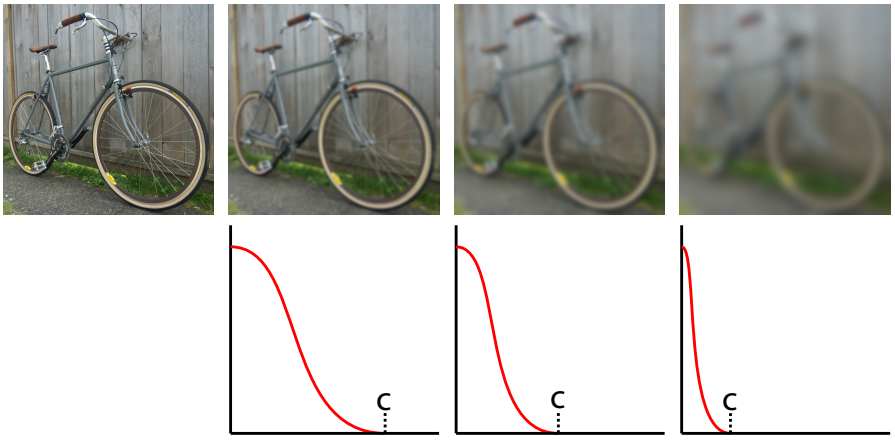


Figure 2.7: *Examples of different blur filter frequency cutoffs. The upper row contains representative images while the lower row contains the respective plots of spatial frequency content. The lower the cutoff c, the less high frequency information present in the image.*

2.2.1 Blur Sensitivity

Much of the work quantifying human blur sensitivity has come from the optometry community, related to deficiencies in the optics of the visual system, and myopia (nearsightedness) in particular. Hamerly and Dvorak [1981], and more recently Wang and Ciuffreda [2005b], report the HVS is capable of detecting a difference between a sharp edge and a blurred edge width of 25 seconds of arc. Interestingly, both report

discrimination between two blurred edges is more accurate than discriminating a blurred edge from a sharp one, and the threshold for blur discrimination is roughly half that of detection.

Both Hamerly and Dvorak and Wang and Ciuffreda tested blur discrimination on a relatively narrow ranges of edge blurs, corresponding to the first few just-noticable-differences (JNDS) of blur. More recently, Chen et al. [2009] conducted a similar study covering a significantly wider range of blur radii covering Gaussian edges with values of σ ranging from 0-.525°. Chen et al. report a blur detection threshold of σ =.03° for a sharp edge, increasing to roughly twice the sensitivity at σ =.04°, then sharply decreasing in sensitivity past that. The shape of the blur sensitivity plot is shown in Figure 2.8.

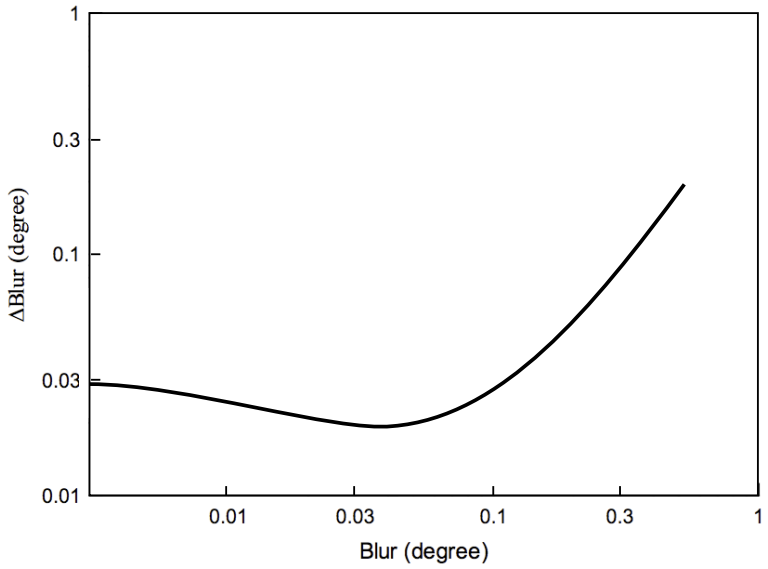


Figure 2.8: *Blur discrimination threshold from Chen et al.. The sensitivity increases then decreases sharply past a point. Image reproduced from Chen et al..*

The sensitivity to blur depends on a number of different factors. Hamerly and Dvorak found that blur sensitivity increased with contrast and subjects were able to detect smaller changes when the stimulus was higher contrast. Both Wang and Ciuffreda [2005a] and Wang et al. [2006a] studied how blur sensitivity changed with regards to position on the retina, and found that the threshold of detection increases

slightly for stimuli further from the fovea, corresponding to the decrease in photoreceptor density. At each retinal position, the blur discrimination thresholds were similar to each other, and they were approximately 60% of the blur detection threshold magnitude. Wang et al. also found that blur detection sensitivity decreased roughly twice as fast as blur discrimination. Wuerger et al. [2001] studied the difference in blur sensitivity for chromatic edges, including isoluminant red/green and yellow/blue stimuli in addition to luminance. The thresholds for red/green stimuli are similar to those of luminance stimuli, while thresholds for blue/yellow stimuli were significantly higher.

Ciuffreda et al. conducted research linking low-level aspects of blur sensitivity to higher-order perceptual constructs. This work includes task-specific blur sensitivity [Ciuffreda et al., 2006], derived from experiments where subjects rank the point at which blur was detectable, objectionable and significantly degraded performance on visual recognition tasks. The study served as the inspiration for the perceptual experiment we conducted in Chapter 6. The authors also presented [Ciuffreda et al., 2007] a conceptual model of blur perception unifying a number of these individual findings. Their work combines the measurements of detection and discrimination thresholds for different retinal positions and constructs volumes in the visual field, shown in Figure 2.9, where two stimuli would be within a JND of blur from each other.

Much of the sensitivity to blur of the HVS is determined by neurological factors, not the optics of the eye, suggesting that long-term differences in the visual stimuli a subject is exposed to can affect their ability to resolve blur differences. In particular, myopic (near-sighted) viewers and emmetropic viewers (not requiring correction) should display different sensitivity and adaptation to the presence of blur, given that one regularly experiences blurred stimuli while the other does not.

Cufflin et al. [2007] and Wang et al. [2006b] studied how the blur discrimination thresholds change when subjects have adapted to out-of-focus stimuli. In their experiments, subjects' vision was blurred by 1-2.5 diopters, and their visual acuity was tested immediately after the blur was introduced, then after an adaptation period. Initially, both myopes and emmetropes displayed a loss of visual acuity, but recovered some of the acuity after the adaptation period. George and Rosenfield [2004] noted that myopes displayed a larger increase in sensitivity after adaptation than em-

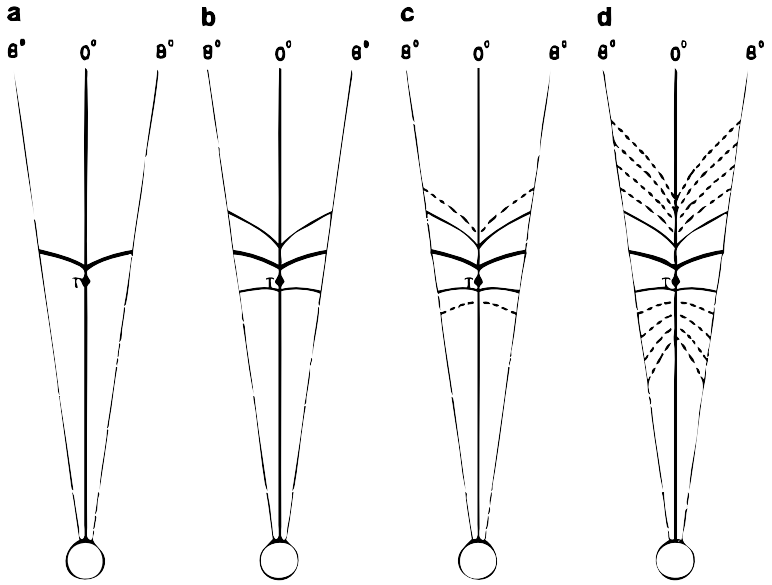


Figure 2.9: *Conceptual model of blur perception of Ciuffreda et al. The thick solid line represents the plane focus. The thin solid line represents the distance corresponding to blur detection while the dashed lines correspond to* JNDS *of blur discrimination. Image reproduced from Ciuffreda et al..*

metropes, possibly due to more exposure to blurred stimuli. However, while myopes were more able to adapt to the presence of blur, even with corrected vision they could not match the visual acuity of emmetropes [Rosenfield et al., 2004]. Rosenfield and Abraham-Cohen [1999] noted that this improvement in visual acuity came without a change in the optical refraction associated with accommodation of the eye, implying changes in the neurological gains of different visual channels.

2.2.2 Blur Appearance

In addition to altering the sensitivity of viewers, adaptation to differently blurred stimuli can also change the appearance of images. Webster et al., in 2001a and 2002, found that viewers adapted to the amplitude spectra of blurred and sharp images, causing subsequently viewed natural images with a conventional $1/f^2$ frequency distribution to appear altered. Webster et al. first showed subjects images that had either been blurred or sharpened, giving them time to adapt. Normal images shown

afterwards appeared too sharp or too blurry, respectively. Webster et al. also showed that the same phenomenon existed for simultaneous contrast [see Fairchild, 2005, chap. 6], where a normal image surrounded by blurred images would seem too sharp and vice versa, shown in Figure 2.10.

Figure 2.10: *Simultaneous blur contrast. The center face is the same in both images, but the surrounding faces affect its appearance. Images copyright Michael Webster.*

In natural settings, defocus blur most often results from the current focal distance of the observer, and there is strong relation between the amount of blur perceived and the distance between the viewer and the object. Mather and Smith [2002] investigated how blur discrimination affected the ability of the HVS to determine depth relationships in a scene and demonstrated blur variation alone is sufficient to determine the apparent depth ordering. More recent work by Vishwanath and Blaser [2010] agrees with Mather and Smith, additionally suggesting the blur is likely used as a quantitative cue to distance consistent with the observer relationship between retinal blur and distance, and not the result of a precise computation. Both authors found that the ability of observers to discriminate different levels of blur was rather poor, consistent with our own results. This lack of sensitivity limits the use of retinal blur to a relatively coarse, qualitative depth cue.

Held et al. [2010] presented a probabilistic model of how viewers may use defocus blur in conjunction with other pictorial cues to estimate the absolute distances to objects in a scene. Their model suggests how the visual system might use a com-

bination of blur pattern and relative depth cues to determine the apparent scale of the content of images. Held et al. validated their model with a series of perceptual experiments and found viewer's estimates of model distance consistent with their predictions. The authors also proposed an algorithm for introducing blur into an image to achieve a novel perceived object scale, as seen in Figure 2.11.

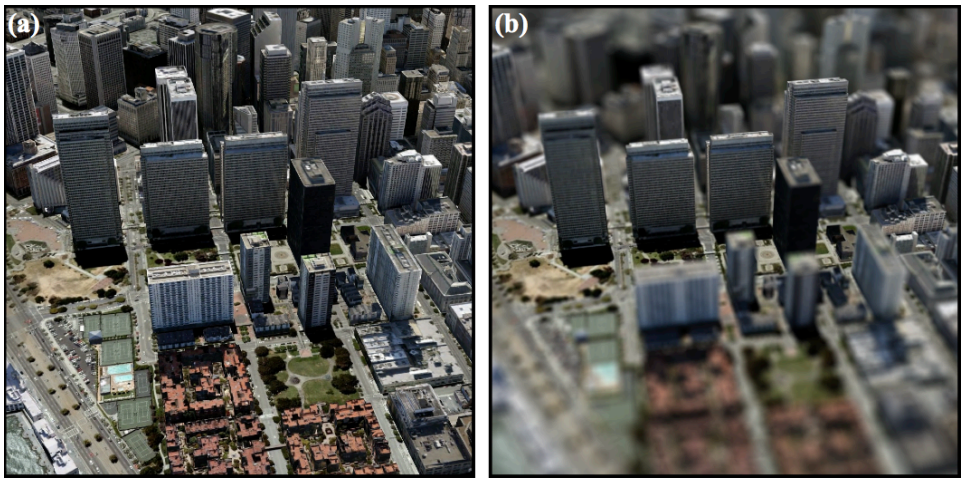


Figure 2.11: *Perceived object scale of Held et al.. Simulating different defocus patterns can alter the perception of scale and distance of objects. Image copyright Robert Held.*

The degree to which a given stimulus is blurred determines the visual channel responsible for processing that stimulus. Differences between the responses of individual visual channels affect the perception of that blur, resulting in changes in blur sensitivity and appearance across the range of possible scales. Many other aspects of image perception are tied to the underlying visual channel, including contrast. The next section details how the contrast sensitivity and the appearance of contrast in image features is determined by spatial scale.

2.3 Spatial Contrast Perception

Contrast, in the most general sense, is the spatial variance in visual properties that causes one element of an image to be distinguishable from the surrounding elements. This thesis is primarily concerned with *luminance contrast*, how the intensity of light changes between one region and another. Contrast is specified in terms of ratios

between light and dark sides of an edge, removing any dependence on absolute magnitude of the stimuli. Features with vastly different magnitudes can have the same contrast as long as the relative difference remains the same:

$$\mathscr{C}_{Ratio} = \frac{Y_{max}}{Y_{min}}. \tag{2.2}$$

Other metrics of contrast have arisen from specific stimuli that have appeared repeatedly in literature. For simple sinusoid gratings with maximum and minimum luminances, Y_{max} and Y_{min}, contrast is measured by *Michelson contrast*:

$$\mathscr{C}_{Michelson} = \frac{Y_{max} - Y_{min}}{Y_{max} + Y_{min}}. \tag{2.3}$$

For tests concerning the contrast between a simple target and surrounding area contrast is measured by *Weber contrast*, and describes the increment or decrement ΔY from some base level of illumination Y. Weber contrast is defined as the ratio of change in luminance to base luminance:

$$\mathscr{C}_{Weber} = \frac{\Delta Y}{Y}. \tag{2.4}$$

Both of these formulations are closely related and each can easily be expressed with the terms of the other:

$$\mathscr{C}'_{Michelson} = \frac{\Delta Y}{2Y + \Delta Y} \tag{2.5}$$

$$\mathscr{C}'_{Weber} = \frac{Y_{max} - Y_{min}}{Y_{min}}. \tag{2.6}$$

However, these contrast measures are lacking. Neither measure can address the fact that natural scenes contain a diverse range of spatial scales that vary across the image. Contrast is a local phenomenon that changes across the image, where small-scale changes in local contrast correspond to changes in surface texture and larger-scale changes in local contrast correspond to geometry or illumination. A better definition of contrast would adapt to the changes in intensity present at different scales in different regions of an image, similar to the visual channels present in the HVS.

Pyramid approaches [Adelson et al., 1984] are one of the most common means of processing images that accurately capture differences in scale. The simplest measure of spatial contrast is the Laplacian pyramid [Burt and Adelson, 1983]. This representation decomposes an image into discrete frequency bands by computing differences between Gaussian-filtered images. Each level of a Laplacian pyramid is defined as the difference between subsequent octaves of a Gaussian pyramid, with the first Laplacian level comprised of the difference between the original image and the topmost Gaussian level. There is an additional base level containing the remaining frequency information. The Laplacian pyramid decomposes an image into n band-pass images b_i and a single low-pass image l, resembling the local spatial processing and band-limited nature of human vision:

$$I = \sum_{i=0}^{n-1} b_i + l. \tag{2.7}$$

One shortcoming of the Laplacian pyramid is that it is based on differences in luminance, as opposed to ratios of luminance. Peli [1990] proposed a similar approach, basing the measure of contrast on ratios of luminance. Peli constructs a series of band-pass images of the ratio of luminance to a local average of the following structure

$$I = l_0 + \sum_{i=1}^{n-1} a_i \qquad\qquad l_i = l_0 + \sum_{j=i}^{i-1} a_j, \tag{2.8}$$

where a_i is the band-pass filtered image, l_i is the local mean consisting of all frequencies below that band, with l_0 being the lowest band. The definition of contrast is then the ratio of the band-pass image to the local average:

$$\mathscr{C}_{Peli} = \frac{a_i}{l_i}. \tag{2.9}$$

Pyramid representations provide efficient approximations of the local band-limited contrast processing of the HVS. If these metrics are insufficient, more accurate measures of local contrast based on models of human contrast sensitivity should be employed.

2.3.1 Contrast Sensitivity

One of the most fundamental issues related to contrast perception concerns *contrast sensitivity*, or how large a luminance difference must be in order to be visible. Perceptual psychologists have dedicated considerable effort to measuring the dependence of contrast threshold on spatial frequency and luminance. The contrast sensitivity function (CSF), shown in Figure 2.12, quantifies the ability of the HVS to detect changes in relative intensity under a specific set of conditions. The full model of contrast sensitivity depends on many factors, but the spatial frequency of a stimulus plays one of the most significant roles in determining the sensitivity of the HVS to the details present.

The CSF displays a peak sensitivity between 2-6 c/deg with a sharp drop in sensitivity for higher frequencies and gentler but still significant decrease for lower frequencies [De Valois and De Valois, 1990]. Reasons for reduced contrast of high spatial frequencies due to the optical quality of the lens in the eye and the limited density of photoreceptors in the retina, as well as reduced contrast for low spatial frequencies due to the limited spatial extent over which low-level cortex cells can merge visual input.

The shape of the CSF is directly visible from looking at frequency gratings like Figure 2.12. Combinations of spatial frequency and contrast that appear gray are below the visibility threshold of the HVS while combinations that show the grating are in the range of sensitivity. The similarity with blur perception is apparent, and different visual channels do not exhibit exactly the same responses to stimuli of equal contrast. The overall characteristic of the CSF is a composite of sensitivities of the underlying visual channels and its shape describes the relative performance of those channels. Individual channels are still responsible for the processing of visual stimuli, but the CSF embodies summation of channel contrast sensitivity.

2.3.2 Contrast Appearance

The CSF, however, only describes the behavior of the visual system to *threshold contrasts*. Most regularly encountered visual stimuli are supra-threshold and easily detectable. There is no guarantee that the sensitivity of the HVS to contrast of a certain spatial frequency is representative of how that contrast actually appears.

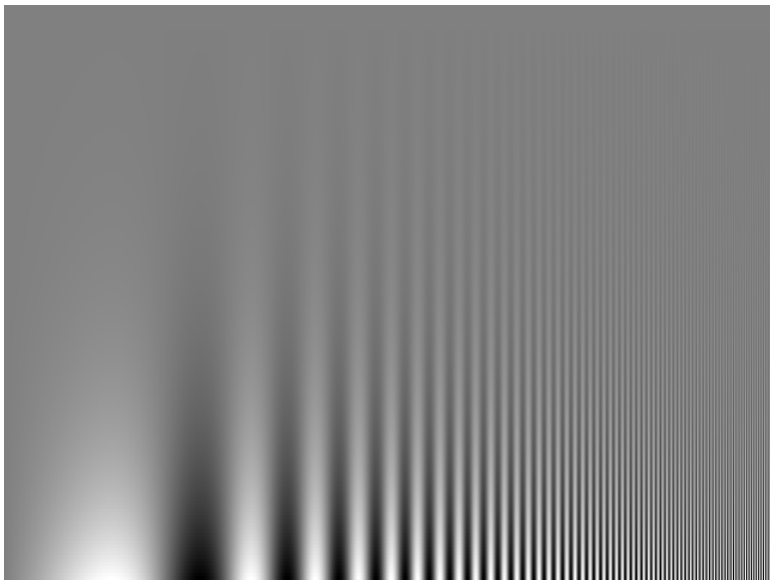


Figure 2.12: *Grating demonstrating the contrast sensitivity of the human visual system. The apparent reduction of contrast at high and low frequencies is a result of our visual system, not the image.*

Blakemore et al. [1973] studied the appearance of supra-threshold contrast and reported the relative apparent contrast of sine waves at different contrasts and spatial frequencies. Subjects adjusted the contrast of sinusoidal gratings of various spatial frequencies to match the contrast of a 5 cpd grating. At threshold contrast, the curves match the CSF. As contrast increases, low frequencies appear less attenuated up to a Michelson contrast of 0.7, at which point they appear equal contrast to the reference. Higher frequencies remain attenuated similar to the CSF.

Georgeson and Sullivan [1975] were the first to note that high (and to a lesser extent low) spatial frequencies never appear as faint or as fuzzy as mid-spatial frequency gratings. The phenomenon, known as *contrast constancy*, refers to the ability of the visual system to account for differences in contrast sensitivity between different spatial frequencies. Instead of a stimulus at threshold contast appearing barely detectable, it appears near full contrast. Cannon [1985] corroborated that two equal physical contrast stimuli have an equal perceived contrast, even if the respective thresholds are quite different, implying the mechanisms mediating threshold detection and supra-

threshold perception may be different. It is as if the visual system has an internal model of the CSF it uses to partially invert the effects of contrast sensitivity on the perceived result, as seen in Figure 2.13.

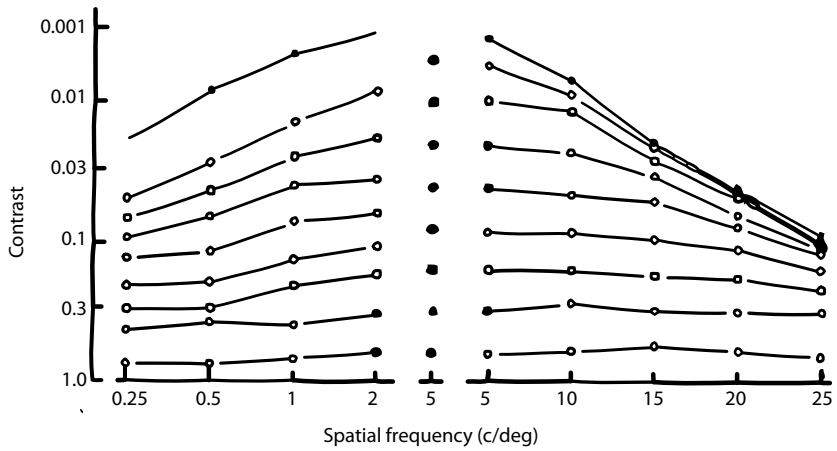


Figure 2.13: *Plot of contrast constancy. Even though the threshold (topmost line) increases for high and low frequency, the appearance of contrast appears the same as that of middle frequencies with lower threshold. Image reproduced from Georgeson and Sullivan.*

Psychophysical studies have shown contrast constancy to hold under a wide range of conditions. Subsequent studies by Peli et al. [1996; 1991] have shown that the HVS is capable of preserving the appearance of contrast across a vast range of visual conditions, reducing in effect only for dim or very high frequency targets. Tiippana and Näsänen [1999], likewise, found that for high contrasts, matches were nearly independent of stimulus spatial-frequency bandwidth up to about 6 octaves, even though detection thresholds increased with bandwidth. Brady and Field [1995] derived a model of contrast constancy that, in addition to explaining experimental results, also predicts perceived contrast will be approximately constant across scale for scenes whose spectra fall as 1/f, as is typical of natural scenes. In some cases, the compensatory mechanisms can be overdriven and frequencies with a lower sensitivity can appear to have more contrast than higher-sensitivity stimuli, referred to as contrast overconstancy by Georgeson [1991].

The perception of contrast in complex images depends on factors beyond the amplitude of the spatial frequency components. While it is convenient to treat the

visual system as a set of quasi-linear operations, numerous exceptions exist. Examples like the contrast constancy phenomenon have demonstrated that the appearance of stimuli is effectively decoupled from the sensitivity of the underlying visual channels responding to a stimulus. The contrast between one region of an image and adjacent regions can also affect the *apparent contrast*.

Apparent contrast is the relative brightness of visual stimuli, where *brightness* is formally defined as "the visual sensation according to which an area appears to emit more or less light" [Poynton, 2003]. Brightness could be thought of as apparent luminance, and much like the ratio between luminances determines physical contrasts, the ratio between brightnesses determines apparent contrast.

The most commonly known example of these higher-level phenomena is the shift in color appearance of a stimulus when the color of its background is changed, known as *simultaneous contrast*. In the case of Figure 2.14, identical gray patches presented on different backgrounds appear distinct. Darker backgrounds cause the grey patch to appear brighter, while lighter backgrounds cause the grey patch to appear dimmer.

Figure 2.14: *Example of simultaneous contrast. The internal gray squares appear of different brightness depending on the surrounding area.*

2.4 Edge Profile Perception

The amplitudes of spatial frequencies that comprise a stimulus have an obvious effect on the perceived blur and contrast. This fact is especially relevant in the context of edges, which consist of all spatial frequencies. In particular, the edge forming the boundary between regions can have significant effects on the perception of those regions. The change in intensity across that boundary, known as the *luminance profile*,

can substantially alter the appearance of both the edge and adjacent regions.

Figure 2.15: *Reduction in contrast of regions bounded by blurred edges demonstrated, by Kim et al. [2011]. The inside of* ***(a)*** *and* ***(b)*** *are the same color, but the blurred edges in* ***(b)*** *reduce the perceived contrast. Kim et al. also present a method to correct for the loss of contrast, demonstrated in* ***(c)****. Images copyright Min Kim.*

The previous section addressed how individual spatial frequencies are perceived, and we now address how groups of spatial frequencies are perceived in the context of edges. By definition, blurred edges, with attenuated high spatial frequencies, appear less well defined than sharp edges. Likewise, contrast thresholds increase as the width of a blurred edge increases [Shapley and Tolhurst, 1973], in accordance with the CSF. What is surprising, however, is blurred edges also appear to have less contrast than step edges of equal Michelson contrast [O'Brien, 1958], the effects of which were recently studied for complex images by Kim et al. [2011] and depicted in Figure 2.15.

Edges profiles with amplified high frequencies, known as *countershaded edges*, are more interesting still. The relative intensity of amplified spatial frequencies can alter the perception of both sharpness and contrast of the edge. Countershading increases local contrast by adding a high-pass image $\mathscr{H}_\sigma(Y)$ to the original image

$$\mathscr{Y} = Y + \lambda \mathscr{H}_\sigma(Y) = (1+\lambda)Y - \lambda g_\sigma * Y, \tag{2.10}$$

where λ is the contrast of the countershading. The process is illustrated in Figure 2.16. In some cases, the underlying edge may be attenuated, or omitted entirely, and only the high-pass image used.

Depending on the parameters used, the result of countershading causes different changes in the appearance of the image. The choice of a small σ increases

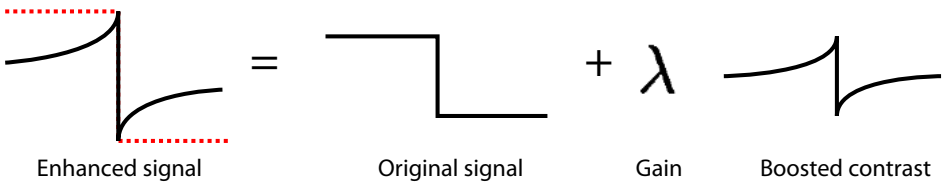


Figure 2.16: *Demonstration of countershading process. A high-pass image modulated by* λ *is added the underlying image. Image reproduced from Smith.*

the acutance [Neycenssac, 1993], or apparent sharpness, of the image (Figure 2.17), which is the effect usually referred to as "unsharp masking" [Badamchizadeh and Aghagolzadeh, 2004]. However, narrow profiles do not significantly alter the perception of contrast [Kingdom and Moulden, 1988]. The choice of a large σ increases the contrast of the image (Figure 2.17), which is usually referred to as countershading.

Increased Sharpness Normal Increased Contrast

Figure 2.17: *Comparison of perceived effect of countershading. Narrow profiles can increase perceived sharpness while wide profiles can increase perceived contrast. Original image copyright Greg Ward.*

For both changes in sharpness and contrast appearance, the perception of countershaded edges again falls into two categories: threshold sensitivity and supra-threshold appearance. The threshold, when equal-contrast countershaded edges and step edges can be differentiated, and the supra-threshold appearance, the amount of contrast or sharpness increase that a countershaded edge appears to possess, are determined by countershading parameters σ and λ.

The studies on the Cornsweet illusion by Sullivan and Georgeson [1977], Campbell et al. [1978; 1971] and Burr [1987] determined the contrast at which a coun-

tershaded edge becomes distinguishable from a step edge of equivalent contrast, the so called *scalloping threshold*. The thresholds apply only to larger σ-values because the countershading is always distinguishable from a square wave at spatial frequencies above 1 c/deg [Sullivan and Georgeson, 1977]. Thus narrow countershading that increases apparent sharpness will always appear distinct from their step edge counterparts.

Interestingly, research has almost exclusively focused on the perception of wider, contrast-enhancing countershading profiles and Lin et al. [2006] conducted one of the only studies regarding the perceived change in apparent sharpness of countershaded edges. Subjects ranked the perceived quality of images for contrasts of different magnitudes and Lin et al. computed the most desirable and highest tolerated contrast. Findings with Laplacian edge-enhancement filters show that the baseline preferred sharpness is about 2.6× that of the local just-noticeable difference, and the actual preferred sharpness is also dependent on the contrast increase in the surrounding areas.

On the other hand, a substantial body of perceptual literature exists on the effect of countershading on perceived contrast. The best known and most dramatic example is the *Craik-O'Brien-Cornsweet illusion* [Cornsweet, 1970; Craik, 1966; O'Brien, 1958], the name commonly given to a family of related illusions that induce an increase in perceived brightness by introducing countershading. The family, often referred to collectively as the Cornsweet illusion, results from a sufficiently-wide countershading profile. The illusion, shown in Figure 2.18, causes the region on the bright side of a countershaded edge to appears brighter and the region on the dark side of the the edge to appears darker.

Tolhurst [1972] and Shapley and Tolhurst [1973] found that the threshold contrasts of countershaded edges were the same as that of step edges, provided the width of the high-pass image was above a critical value of .5°. Below the critical width, sensitivity to Cornsweet edges became progressively lower than that of step edges. Campbell et al. 1978; 1971 and Sullivan and Georgeson [1977] found that the contrast thresholds for the missing fundamental illusion, shown in Figure 2.19, to be the same as those for square waves provided the slope width was less than roughly .25 c/deg. Differences between the findings can be attributed to differing luminance profiles used in their respective experiments.

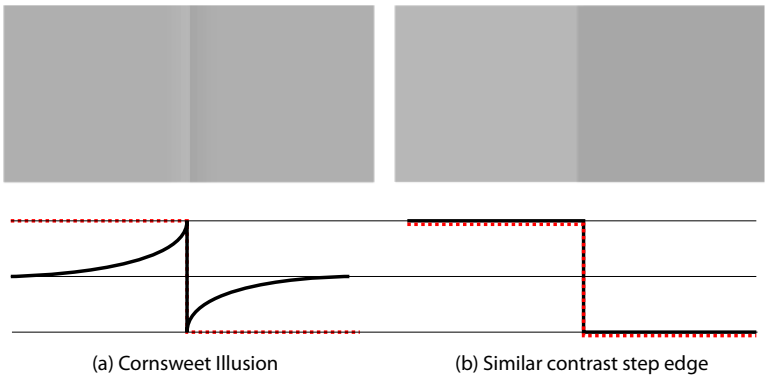


Figure 2.18: *Comparison of step and Cornsweet edges.* Left: *a Cornsweet edge.* Right: *a standard edge. The regions on either side of the edges appear to have roughly the same brightness, even though they are of the same brightness for the Cornsweet edge. Image reproduced from Smith.*

Seminal work by Dooley and Greenfield [1977] resulted in the graph in Figure 2.20 that relates the apparent contrast of a Cornsweet profile to the physical contrast of its luminance profile. The apparent contrast of the edge is shown to be a function of the width of the edge profile measured in angular degrees, and amplitude measured in Michelson contrast. The strength of the illusion diminishes as the width of the Cornsweet profile decreases, and as amplitude increases, the apparent contrast of the edge becomes nonlinear and eventually begins to decrease. The choice of a larger σ or a smaller λ results in a larger relative increase in contrast.

The Cornsweet illusion is additive, as noted by Burr [1987], and a countershading profile can be added to an existing step edge to further increase the apparent contrast of the edge. Dooley and Greenfield [1977] noted that the strength of the Cornsweet illusion decreases as the contrast of the existing step edge increased. Additionally, the Cornsweet illusion is cumulative, and multiple Cornsweet profiles will result in brightness that approximates a set of step edges, where each patch will appears brighter than the previous.

Countershading is not limited to changes in perceived brightness; it has been observed to be a property of several aspects of visual perception. Of particular interest to image processing algorithms is the fact that the illusion is present in similar edge profiles of color change [van den Brink and Kleemink, 1976; Ware and Cowan, 1983],

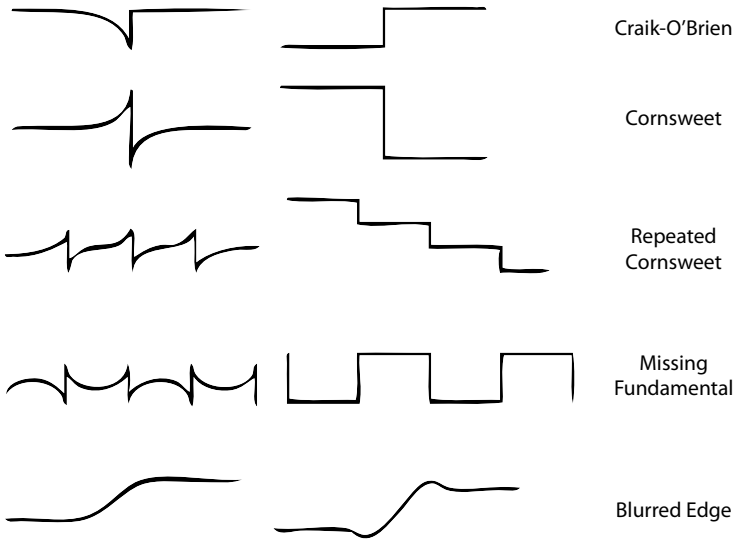


Figure 2.19: *Examples of different physical edge profiles (left) and their perceived result (right). Image reproduced from Kingdom and Moulden.*

and can be created from changes in hue, saturation and chromaticity [Wachtler and Wehrhahn, 1997]. Additionally, Anstis and Howard [1978] presented the existence of the illusion with regards to stereoscopic depth, Loomis and Nakayama [1973] and Walker and Powell [1974] described the relation of the effect to motion perception, while Purves et al. [1999; 2004] provided some compelling examples of how the effect relates to our perception of 3D scenes. Ihrke et al. [2009] performed a perceptual evaluation of the work on 3D unsharp masking by Ritschel et al. [2008]. The authors tested the preferred value of their magnitude parameter λ for profiles of several different widths, measured on object surfaces, opposed to image-space.

2.5 Image Appearance Models

The perceptual basis of the Cornsweet illusion is the subject of ongoing debate and there are competing theories on what aspect of visual processing mediates the illusion. Dooley and Greenfield [1977] suggest that the effect is due to the shape of the CSF and based the fact that given the lack of sensitivity of the CSF to low frequencies, a Cornsweet edge and a step edge have the same response. Another class of theories

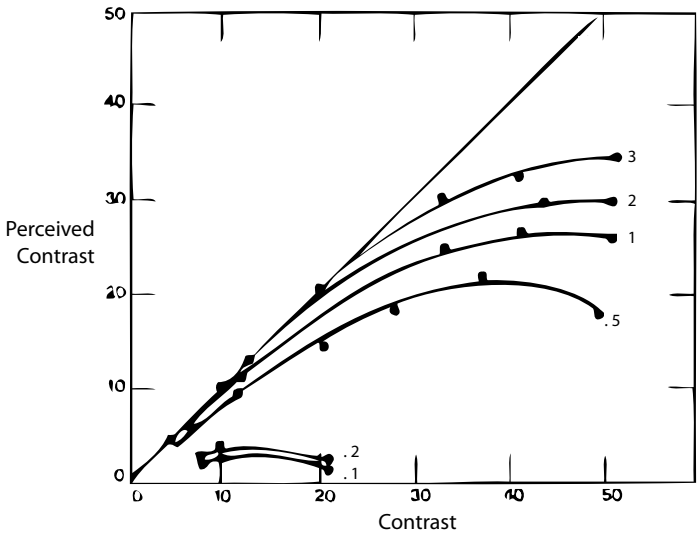


Figure 2.20: *Strength of the Cornsweet illusion as a function of width and contrast. As width increases, so does the strength of the illusion. Increasing the contrast increases the strength of the effect up to a point, then the illusion fails. Image reproduced from Georgeson and Sullivan.*

attempts to explain the effect in terms of lightness constancy like Retinex [Land and McCann, 1971], described below. A third interpretation suggests that the effect is the result of local analysis of edge features and uses sets of filters to detect and extract information about edges and uses that information to extrapolate global effects. Davey et al. [1998] has reported that the illusion is consistent with the effects observed from the "filling in" mechanism of our vision perception, but that has been disputed by Devinck [Devinck et al., 2007]. Kingdom and Moulden [1988] provide a comprehensive comparison of all the competing theories.

Attempts to model perceptual processes extend beyond countershading and numerous approaches have been proposed to model how the human visual system understands blur and contrast. These methods share the goal of modifying an image such that it approximates the HVS perception of brightness. One significant class of methods is known as lightness algorithms. The most commonly known version is Retinex [Land and McCann, 1971], but a number of similar approaches exist [Blake, 1985; Horn, 1974]. Hurlbert [1986] provides a summary of methods and presents several formal connections between their formulations.

The key observation these methods make is that human perception of brightness roughly corresponds to the reflectances of surfaces in the scene. However, the observed image is the product of the reflectance and the illumination. Given this observation, if one separates the contributions of reflectance and illumination from the observed image and only considers the reflectance, they could have an estimate of how the HVS perceives the brightness.

The original Retinex paper presented by Land and McCann was based on the observation that changes in illumination correspond to smooth gradients that vary slowly while changes in reflectance correspond to edges and are abrupt. This assumption leads to the approach shared across many of the lightness algorithms: first differentiate the image to recover change in luminance, then employ a threshold operation to remove small changes that correspond to smooth gradients, and finally integrate to recover a modified image that approximates the surface reflectance.

Land's original paper focused on 2D patches of uniform reflectance called *Mondrians*, shown in Figure 2.21. The approach performs random walks between different patches of the 2D Mondrian, summing the thresholded differences along the path to approximate the log of the ratio of surface reflectances at the beginning and end of the path. This operation is repeated along numerous random paths to try and find region of highest reflectance. The approach was effective at predicting the appearance of the Mondrian patches it was designed for, but did not accurately predict more complex scenes.

To handle more complex inputs, a second version [Jobson et al., 1997; Land, 1986] of Retinex was proposed based on a center-surround formulation. This version keeps the assumption that the illumination is smoothly varying, and concludes that some spatially weighted average of the observed scene will approximate that illumination. If the observed scene S is the product of the illumination I and reflection r, the Retinex output is defined as:

$$R = \log\left(\frac{Ir}{\bar{I}\bar{r}}\right) \tag{2.11}$$

$$R = \log\left(\frac{r}{\bar{r}}\right) \tag{2.12}$$

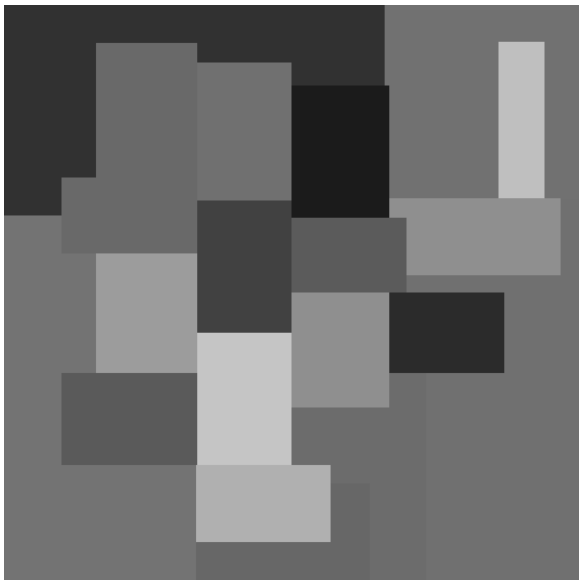


Figure 2.21: *Example Mondrians of Land and McCann [1971].*

as long as $I \approx \bar{I}$. Instead of randomly exploring patches of a Mondrian to try and determine region of maximum reflectance, this version computes the normalization constant from a Gaussian weighted average of a large surrounding neighborhood. The scale of the Gaussian allows a tradeoff between amount of reduction of contrast and global correspondence of lightness.

Despite additions by Jobson and Woodell [1995]; Rahman [1995], no single Retinex scale works for all images, or even different portions of a single image. Rahman et al. [1996a] proposed a multiple-scale version that alleviates some of the issues of the center-surround formulation. Multi-scale Retinex is equivalent to the center-surround form, but the output is a combination of a weighted average of several different scales of the single-scale version of Retinex. Numerous researchers [Barnard and Funt, 1997; Funt et al., 1997; Rahman et al., 1996b] have studied and improved on the multi-scale version, but there are still issues. Many variants of the algorithm attempt to solve for changes in luminance and color simultaneously and have trouble separating side effects between the two. The photographic tonemapping operator by Reinhard et al. [2002] is very similar to luminance multi-scale Retinex, but removes several unnecessary steps and includes automatic selection of the proper scale

for each pixel.

The field of color science has made significant progress on mimicking how the HVS converts physical values in perceptually-salient units. Perceptually uniform color spaces like L^*ab space and luminance quantizations [Mantiuk et al., 2004] make it possible to compare perceived differences in physical quantities of light. Color appearance models like the CIECAM02 [CIE, 2004] are capable of exhibiting the complex behavior of our visual system like simultaneous contrast. Image appearance models extend this further to model the spatial and multi-band components of image perception. The s-CIELAB [Zhang and Wandell, 1996], multi-scale adaptation [Pattanaik et al., 1998] and iCAM [Fairchild and Johnson, 2004] models extend color appearance models to incorporate some aspects of the HVS when viewing complex images. Reinhard et al. [2008] provides a comprehensive overview of these models.

While covering a multitude of dimensions of visual perception, these frameworks are still incomplete. None of these models accurately capture the complex scale-dependent relationships of blur and contrast perception we have observed. It is this deficiency in the ability of existing models to represent the nuances of the HVS that has motivated our contributions in Chapters 5 and 6.

Chapter 3

Related Work

The previous chapter detailed the significance blur/contrast perception plays in image understanding. Unsurprisingly, given this importance, the fields of image processing and computer graphics have investigated the manipulation of blur and contrast in images. This chapter surveys methods and computational approaches to understanding and altering blur and contrast in complex images.

Section 3.1 covers computational photography, methods for extending camera capabilities of image capture. In particular, we examine the ability of aperture filtering and deconvolution techniques for manipulating blur and contrast when capturing images. Then Section 3.2 provides a summary of techniques for estimating the blur and contrast present in images, with an emphasis on local multi-scale operators. This coverage includes the estimation of blur and contrast, as well as how those operators relate to edges and edge detection. Next, Section 3.3 describes methods for manipulating blur and contrast in images in spatially-variant manners, including blur magnification and applications of countershading. Finally, Section 3.4 details other applications that relate to the spatial perception of blur and contrast in natural images, including resizing operators and the restoration of detail.

3.1 Computational Photography

3.1.1 Computational Image Capture

As digital sensors have become the de facto means of image capture, methods of acquisition have evolved to leverage the processing capabilities inherent in this augmented imaging pipeline. This new envisioning of photography extends beyond the conventional constructs of a single lens focusing the scene onto a static sensor plane. The paradigm of computational photography circumvents the limitations inherent to normal methods of image acquisition by including additional encoding and decoding steps into the conventional imaging process. In this framework, the acquisition setup, including optics, lighting and exposure, is altered to encode additional information in the captured image. The captured image is then decoded by software to recover the conventional pixel values and use the encoded information to synthesize a novel image.

Many of these computational techniques have addressed aspects of how the formation of contrast and blur is dictated by the physics inherent to lens optics. To overcome the limitations of focus in a conventional lens system Ng introduced the plenoptic camera [Ng et al., 2005], based on the principles of Fourier-slice photography [Ng, 2005]. This new camera design involved an array of microlenses over the image sensor that encoded a series of images that could later be used to refocus the image or change the aperture or viewpoint. Subsequent designs by Veeraraghavan et al. [2007] and Lumsdaine and Georgiev [2009] have recast the problem in the frequency domain to improve the utility of the original construction. All of these techniques address the removal of defocus blur, and Levin et al. [2009] analyzed the use of computational cameras for depth-of-field (DOF) extension.

The same principles apply equally to the temporal domain, and information can be encoded in the shutter sequence to address motion blur. Raskar et al. [2006] used a coded exposure pattern, opening and closing the shutter across a single exposure, to better decode motion blur. Similarly, Levin et al. [2008] devised a parabolic camera motion that represented motion blur in a velocity-independent manner, and Agrawal et al. [2009] proposed a similar coded approach for video sequences.

Figure 3.1: *Plenoptic camera of Ng et al. [2005] (left). The combination of microlenses and processing allow the image to be refocused after capture Images copyright Ren Ng.*

3.1.2 Coded Aperture Imaging

All of these techniques allow the photographer to exert considerable control over the blur present in the image after it has been captured. Similar to how coded exposure patterns address motion blur, coded aperture imaging allows for better removal of defocus blur. Coded aperture imaging first appeared in the context of astronomical x-ray and gamma ray imaging as a solution to constraints in feasible optical systems for those telescopes. Compared to visible light, the high energy photons simply pass through media without refracting, rendering conventional lenses useless. At first, the standard practice was to use a pinhole aperture to produce a sharp resulting image, but this approach blocks the majority of the energy from the source and has a very poor signal-to-noise ratio (SNR).

The first improvement over the pinhole aperture was the random aperture arrays proposed by Dicke [1968]. Instead of a single pinhole, they used a two-dimensional array of randomly positioned pinholes, resulting in numerous shifted copies of the image overlaid on top of each other, which they attempted to correct with a decoding step. This process significantly improved the SNR of the imaging system, increasing the amount of light captured by the number of pinholes in the array, but did so at the cost of the resolving power. It is impossible to completely undo the cumulative effects of the random array, and there was always some residual blur. This method was improved upon by structured patterns such as URA [Fenimore and Cannon, 1978] and MURA [Gottesman and Fenimore, 1989]. These patterns retain the multiple aperture holes of the random array but are constructed in such a way that the placement of the holes has a unique and complete means of decoding the

signal. This design permits the complete restoration of the original image, and the patterned array effectively acts as a single pinhole with the SNR of multiple apertures.

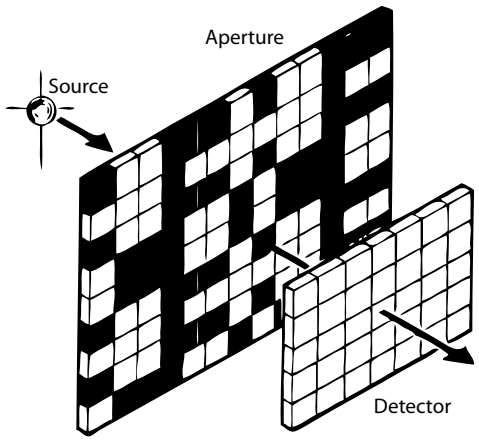


Figure 3.2: *Coded aperture pattern for x-ray imaging. Image reproduced from Gottesman and Fenimore.*

More recently, the principles of coded aperture imaging have been employed in visible light photography. Levin et al. [2007] made use of aperture filters to accurately determine the amount of blur in order to refocus an image. Zhou and Nayar [2009] analyzed which coded aperture patterns were best suited for deconvolution of defocus blur, providing a comparison of the suitability of each at recovering the original image. In related work, Cossairt et al. [2010] analyzed the depth invariance of different coded aperture designs and proposed an improved diffuser-based construction that still preserved image spatial frequencies.

3.1.3 Deconvolution

The coded aperture design in visible light optical systems differs from that of systems for x-ray astronomy. Several of the underlying assumptions regarding the optical setup no longer hold, and simple methods of restoring the original image cannot be used. Instead, deconvolution is used to reverse the effects of the blur introduced by the camera optics. In deconvolution, it is assumed that some desired image has been distorted (blurred) by some known function, and the goal is to recover that original image. Mathematically this relation is represented as:

$$f = f_0 \otimes k + \eta, \tag{3.1}$$

where the recorded image f is the result of convolving the original image f_0 by some pointspread function (PSF) k with additive noise η. Numerous solutions have been proposed over the years, from frequency-space methods such as Wiener filtering [Wiener, 1964] to iterative methods such as Richardson-Lucy deconvolution [Lucy, 1974; Richardson, 1972] and expectation-maximization.

It is exceedingly difficult to recover the original image at good quality. The system of equations does not have an exact solution due to the presence of noise and image information is suppressed by the frequency response of the filter. Additionally, the system is ill-posed, resulting in an infinite number of possible solutions. Results often include ringing around edges, amplified noise and other artifacts. In the case of coded aperture imaging, the PSF k is known and specifically chosen to aid in conditioning the deconvolution. While this design reduces the size of the solution space, it is still no easy task, and additional measures are required to guide the process towards the correct solution.

New methods have incorporated *natural image statistics* into deconvolution algorithms to better recover the original image. The power spectra of images of real world scenes all roughly follow the same distribution. Individual images exhibit variation, but the overall trend carries strongly across all natural images. A number of recent papers have utilized deconvolution algorithms incorporating these assumptions to better recover the original image. Levin et al. [2007] used a combination of coded aperture and enhanced deconvolution to recover depth and refocus images. More recently, Bando and Nishita [2007] presented a means of recovering the original image from defocus blur without a coded aperture, while Yuan et al. [2008] proposed more advanced techniques for removing artifacts.

Most deconvolution methods assume the PSF that degraded the image is already known and use it to invert the blur. Blind deconvolution methods, on the other hand, such as those by Fergus et al. [2006]; Lam and Goodman [2000]; Shan et al. [2008a] iteratively estimate the shape of the blur while restoring lost image details. To do so, blind methods require additional information about the blur, and most assume that the blur kernel is invariant across the entire image, as is the case with

Figure 3.3: Left*: image degraded by motion blur (*PSF *shown in insert.)* Right*: deconvolution results of Yuan et al.. Images copyright Lu Yuan.*

most camera shake.

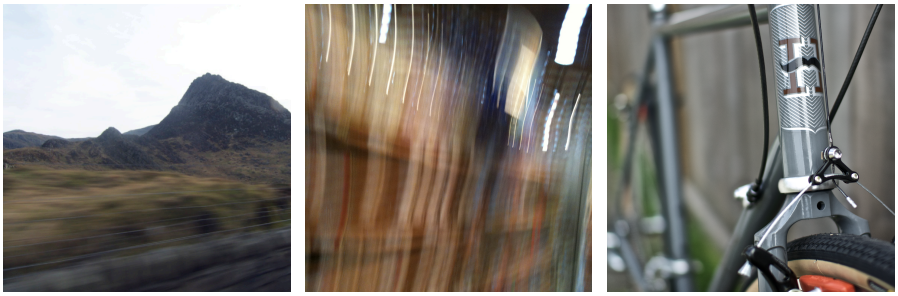
Any blurring operation spreads the luminance incident on a single pixel over a larger area, averaging the values of neighborhoods of pixels together and suppressing high frequency variations. As adjacent pixels share more of the same light falling upon then, the contrast between them is also reduced. Deconvolution is most commonly employed to restore image sharpness lost when degraded by blurring. Our inquiry in Chapter 7 investigates the feasibility of using deconvolution methods to restore the contrast lost by blurring.

3.2 Blur Estimation and Manipulation

All deconvolution methods restore image detail lost to blurring and thus require an accurate model of the PSF that degraded the image. Blind deconvolution methods perform the additional step of estimating the shape of that blur kernel. Blur estimation has proven to be of great utility for a number of applications besides the restoration of lost image detail, including computer vision and image quality assessment. This section surveys different techniques of estimating blur in images as well as other applications made possible by blur estimates.

Blur is introduced into images by a number of different processes. Object shape, illumination and capture method can all cause different kinds of blur to an image. We focus our discussion on three kinds blurs resulting from the image capture process, each with their own properties: motion blur, camera shake and defocus blur. *Motion blur* is a linear blur caused by the subject moving in the frame and while it only

applies to that object, tends not to vary spatially. *Camera shake* is caused by rotation of the camera during the exposure and, depending on the particular rotation, may closely resemble a motion blur or may be a much more complex shape. Camera shake is often uniform, but may vary spatially depending on which axis the rotation is around. *Defocus blur* is circular blur that depends on the distance of each point in the scene from the focal plane, and each portion of the image is blurred by the scaled shape of the aperture.



Motion blur Camera shake Defocus blur

Figure 3.4: *Examples of different kinds of motion blur.*

These three types of blur motivate several approaches to blur estimation. In some cases, like camera shake, the entire image shares the same blur PSF. In cases of motion blur, the blur shape varies across the image, but can be divided into large regions sharing the same PSF. When addressing defocus blur, estimation must consider each pixel individually.

Individual algorithms may address one or more of these approaches, depending on the circumstances responsible for introducing the blur into the image. However, estimation techniques can be roughly divided into blur estimation, techniques that estimate the uniform blur of a given region, and edge detection, techniques that estimate the location (and blur) of edges in the image.

3.2.1 Blur Estimation

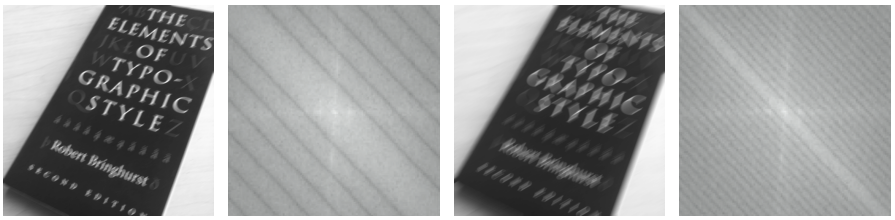
Blur estimation methods vary in terms of the amount of information they ascertain about the blur. Some routines merely estimate the fine detail absent from the image. Other routines attempt to characterize the parameters of blur, such as direction and

magnitude of motion blur. The simplest algorithms perform a global analysis of the blur characteristics, while more advanced methods analyze regions and attempt to segment the image based on the kind of blur present, such as identifying a moving car in front of a stationary background.

Any blur estimation procedure must contend with noise present in the image. Both noise and image detail consist of high-frequencies, but the presence of high-frequency information does not imply image detail and blur estimation routines must ensure that high frequency information is not due to noise. Methods such as those by Rooms et al. [2001] and Tong et al. [2004] decompose the image using a wavelet transform to separate noise from the rest of the image content and only consider high-frequency information that is consistent with lower-frequency information, as is the case in sharp edges. More recently, Ilic et al. [2009] compute statistics based on the average cone ratio of wavelet coefficients to improve noise performance further. While resilient to noise, none of these methods estimate the shape of the blur present, only the degree to which it has suppressed image detail.

If the shape of the blur is simple, in the case of motion blur, Fourier analysis methods can be used to ascertain blur parameters from how the spatial frequencies of the image deviate from that of natural images. Moghaddam and Jamzad [2007] observe that a linear motion blur is equivalent to a temporal box-filter, resulting in the spatial frequencies of the image being multiplied by a **sinc** filter orthogonal to the direction of the blur. Moghaddam and Jamzad use the Radon transform [Deans, 1993] to determine the orientation of the blur, then compute the distance between the zeros of the **sinc** function to determine the length. However, this approach is only valid for linear motion blurs with uniform velocity. Ji and Liu [2008] extended the approach to robustly identify the blur kernel for a number of motion types and accurately determine whether or not the motion is of uniform velocity.

Spatially-variant blur estimation schemes operate similar to conventional blur estimation, but have the additional challenge of trying to identify one or more regions of the image degraded by the same blur. Liu et al. [2008] propose a blur classification framework for automatically detecting blurred regions and recognizing the blur types for those regions without performing blur kernel estimation or image deblurring. Chakrabarti et al. [2010] derive a local measure of the likelihood of a small neighborhood having been blurred by a candidate blur kernel. The authors employ



Small blur Small blur Large blur Large blur

Figure 3.5: *Relationship between motion blur and spatial frequencies demonstrated by Moghaddam and Jamzad [2007]. The angle of the zeros in the Fourier transform corresponds to the direction of blur and the spacing of zeros is inversely proportional to the length of the blur.*

this measure to simultaneously select a motion blur kernel and segment the region that kernel affects for a given image. Dai and Wu [2008] devise a matte-based blur estimation approach where the object and background are highlighted and the algorithm derives the blur parameters based on the assumption that the blurred outline of the object is a blending of it and the background.

Finally, researchers have estimated the motion blur between the sequential frames of a video. In this case, the motion of the blur and the motion of the object between frames will be coherent, and the additional frames provide more information for robustly computing the parameters of the blur. Trussell and Fogel [1992] compare the motion between portions of sequential images to segment the images into regions of consistent blur and estimate the parameters of that blur, subsequently attempting to restore the images of the video sequence. Ben-Ezra and Nayar [2004] combine a high-speed but low-resolution video camera with a conventional video camera to capture the PSF of the motion blur in higher precision. Like Trussell and Fogel, Ben-Ezra and Nayar segment the image into different blur regions and attempt to restore the original image.

3.2.2 Edge Detection

Blur estimation routines consider all the information present in a given region to characterize an estimate of the blur for that region. Edge detection methods, on the other hand, focus on using purely local information to determine the sparse set positions representing edges present in the image.

The most straightforward definition of an edge is any location with a rapid intensity variation. Mathematically, this construction lends to defining edges in terms of the gradient of pixel intensities:

$$\mathbf{J}(x) = \nabla I(x) = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)(x), \tag{3.2}$$

where $\mathbf{J}$ indicates the direction of steepest ascent, perpendicular to the edge direction, and the magnitude of $\mathbf{J}$ denotes the contrast of the edge.

While it is convenient to define edges as locations of perfectly sharp intensity change, a cursory inspection of natural images reveals that the majority of edges present are not perfectly sharp transitions in luminance. Various phenomena, including focal blur inherent in camera optics, penumbral blur of shadows, or shading variations of smooth objects, all lead to the luminance variation of an edge extending beyond a single pixel. As a result, all edge detectors must consider a range of potential widths of edge transitions. Additionally, operators must isolate the edge at a given location within the transition that corresponds to the *maxima* of edge strength (gradient magnitude).

The Canny edge detector [Canny, 1986] is one of the earliest and still best-performing edge detection algorithms. At its heart the Canny edge detector consists of the following stages, also depicted in Figure 3.6:

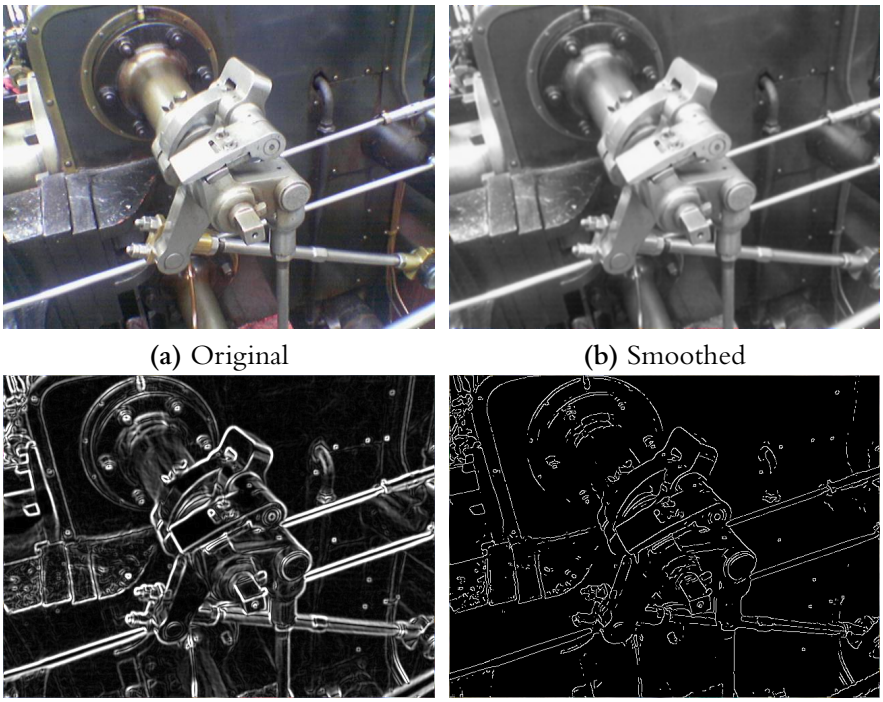
1. A smoothing filter to reduce the degree to which noise and high frequencies affect the subsequent estimation of image gradients.

2. A gradient magnitude operator to estimate the strength of edge contrast at each pixel in the image.

3. A suppression of non-maximum gradient values to determine the precise location of edges.

4. A linking operation to chain individual pixels into edges.

This high-level structure of the algorithm has served as a template for a host of later approaches. Subsequent algorithms have refined the method, including differential geometry and variational geometry formulations [Lindeberg, 1998], but the



(a) Original **(b)** Smoothed

(c) Gradient magnitude **(d)** Non-maxima suppression

Figure 3.6: *Steps of Canny edge detector. The original image* (a) *is first low-pass filtered to remove noise* (b)*. The gradient magnitude* (c) *is then computed, followed by the suppression of non-maxima values* (d)*. Images copyright Wikipedia user Simpsons contributor.*

original Canny algorithm remains one of the most-utilized methods of edge detection.

Any choice of smoothing operator used in edge detection requires the selection of a spatial scale σ. If only the detection of sharp edges is desired, the choice of σ can be determined from the noise characteristics of the image [Canny, 1986; Elder and Zucker, 1998]. Alternatively, a more principled approach employs a *scale-space* [Koenderink, 1984; Witkin, 1987] to reliably detect edges at a variety of scales. Objects in real-world scenes can appear in images at a large range of scales, owing both to the multitude of object sizes and the array of viewing conditions under which the object is observed. The HVS has developed much of the spatial vision adaptations detailed in Chapter 2 to account for this change in object size and scale-space

frameworks provide an analogous ability to computer vision algorithms.

Scale-space methods represent images as a family of smoothed images depending on some spatial parameter, and have the advantage that for a chosen spatial scale σ, features smaller than $\sqrt{\sigma}$ will be suppressed. The scale-space function is defined as:

$$N_n(x;\sigma) = \sigma^{\alpha} I(x) \otimes \frac{\partial^n G(x;\sigma)}{\partial^n x}, \tag{3.3}$$

where the image I is convolved by a nth-derivative Gaussian of width σ, and normalization term α such that the function peaks at scale of the desired feature [Lindeberg, 1996]. Furthermore, in comparison to partial representations like Gaussian or Laplacian pyramids, complex image features can be recognized by either local maxima, minima, or zero crossings at respective locations across scales. Scale-space methods also have the additional benefit of recovering a spatially-variant estimate of the blur present in an image in the process of adapting the scale according to image content.

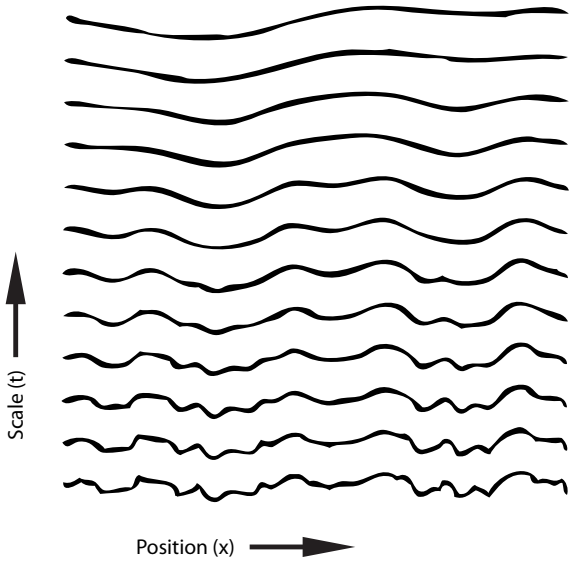


Figure 3.7: *Conceptualization of a Gaussian scale-space. In this figure, a* 1D *signal is depicted at different levels of the scale-space, t. As the scale increases, more of the high-frequency content is filtered out. Image reproduced from Perona and Malik.*

The scale-space framework has proven invaluable for adapting detection and recognition tasks to the different scales at which objects appear in natural images.

Detection algorithms for edges, ridges, corners and blobs have been extended by use of scale-spaces to automatically adapt to changing sizes of objects in images. Additionally, this conceptualization provides a hierarchical organization of image features, attaching smaller-scale features to the larger-scale features in which they reside. We have found two particular methods to be of utility in our work: the minimum reliable scale computation of Elder and Zucker [1998] and the blur estimation method of Samadani et al. [2010], which we describe below in more detail. This survey of edge detection methods is far from complete and we direct the reader to comprehensive overview of Basu [2002].

3.2.3 Elder and Zucker Minimum Reliable Scale

The minimum reliable scale (MRSCALE) of Elder and Zucker [1998] acts as a precursor to their scale-space edge detection operator. MRSCALE compares the gradient magnitude to the noise level and determines the scale at which edge detection should occur to ensure accuracy. The operator takes advantage of the fact that for any given pixel the portion of the gradient due to noise is independent of adjacent pixels, while the portion of the gradient due to a blurred edge is related to surrounding pixels. This difference in structure implies the portion of the gradient due to noise will decrease faster than the portion due to the edge when the image is blurred, and the signal detection properties will improve when blurring more.

Elder and Zucker use a partial scale-space representation, operating at scales corresponding to powers of two, and estimate the minimum scale they can accurately recover the gradient. Their method proceeds as follows: first use local scale control to estimate the gradient and its orientation at each pixel, then use local scale control to estimate the second derivative along the gradient direction at each pixel, and localize edges using the second derivative. This information allows them to reliably estimate the location, intensity, and extent of edges in real images with noise and other artifacts.

Elder and Zucker demonstrate that for a Gaussian edge G of amplitude A, offset B and width σ_b captured by a sensor with white noise $n(x,y)$ of variance v_n^2:

$$G(\sigma,x) = \left(\frac{A}{2}\right)\left(\operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma_b}\right)+1\right)+B+n(x,y) \tag{3.4}$$

they can determine a value for a given blur scale s above which they can be confident of the estimated gradients, and define that critical value as:

$$c(s) = \frac{1.1 v_n}{s^2}, \tag{3.5}$$

where the factor 1.1 is determined by their assumptions of the estimation reliability; please see the original paper for more details. Combining this formulation with their edge definition, one can solve for the blur scale $\hat{s}$ at which the gradient can be reliably detected:

$$\hat{s}^2 = \frac{v_n}{A}\left(5.4 + \sqrt{28.9 + 15.2\left(A\sigma_b / v_n\right)^2}\right) \text{pixels}^2. \tag{3.6}$$

In practice, when estimating the blur of an edge, the amplitude A is also unknown. Elder and Zucker approximate the MRSCALE by computing gradient magnitude estimates at a series of scales and selecting the smallest scale at which the gradient magnitude exceeds the magnitude of the noise present:

$$\hat{s}(x,y) = \inf\{s : r_s(x,y) > c(s)\}. \tag{3.7}$$

The result is a map of how much each region of a noisy image must be blurred by to allow for accurate estimation of the blur in that region.

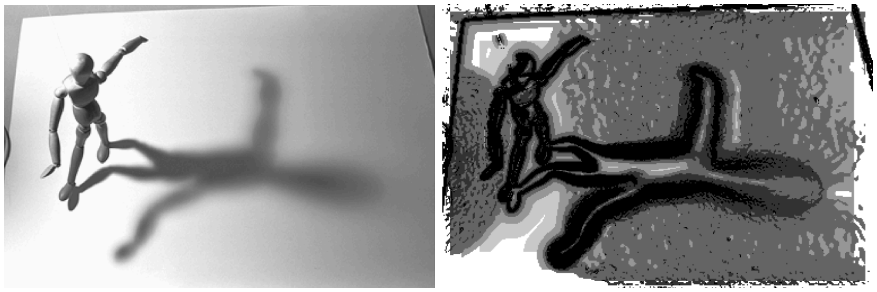


Figure 3.8: *Example of Elder and Zucker minimum reliable scale.* Left: *original image.* Right: *map of minimum reliable scale required to accurately estimate gradient. Dark regions correspond to smaller scales while light regions correspond to larger scales. Images copyright James Elder.*

3.2.4 Samadani et al. Blur Estimation

The blur estimation method of Samadani et al. [2010] produces a spatially-varying map of the blur present at each pixel in the image at the resolution of the thumbnail. The algorithm first generates a standard thumbnail, t_s, and produces a scale-space of thumbnails blurred by different amounts. Image features are computed for the high resolution original as well as for each of the thumbnail images in the scale-space. The amount of blur is determined by the level of the scale-space with feature values most similar to those of the original image features.

These features are computed as the maximum absolute difference between a pixel and its eight neighbors. In the case of the original image, the feature values are downsampled using a maximum filter to produce a thumbnail resolution *range map*, r_o. The levels of the thumbnail scale-space l_{σ_j} are created by convolving the standard thumbnail with a set of Gaussian kernels of standard deviations σ_j,

$$l_{\sigma_j} = g(\sigma_j) \otimes t_s \tag{3.8}$$

where l_{σ_0} is the unblurred, original thumbnail. For each of these images l_{σ_j}, a low resolution range map r_{σ_j} is generated.

The estimate of the blur present in an image is represented by the blur map m. Each pixel i of the blur map is determined as

$$m(i) = \min_j \left\{ j \mid r_{\sigma_j}(i) \leq \gamma r_o(i) \right\}, \tag{3.9}$$

where γ is a user-specified parameter that controls which r_{σ_j} most closely matches r_o and in turn adjusts the amount of blur added. An example of the blur map is shown in Figure 3.9. The final image is synthesized by selecting the pixel from l_{σ_j} that corresponds to $m(i)$. However, the result does not correspond to the σ of the blur in pixels. Our method in Section 4.2 calibrates the method to provide an absolute measure of the estimated blur.

3.2.5 Blur Magnification

In the context of image processing, blur is mostly considered undesirable and most uses of estimated blur focus on the deconvolution methods described previously to

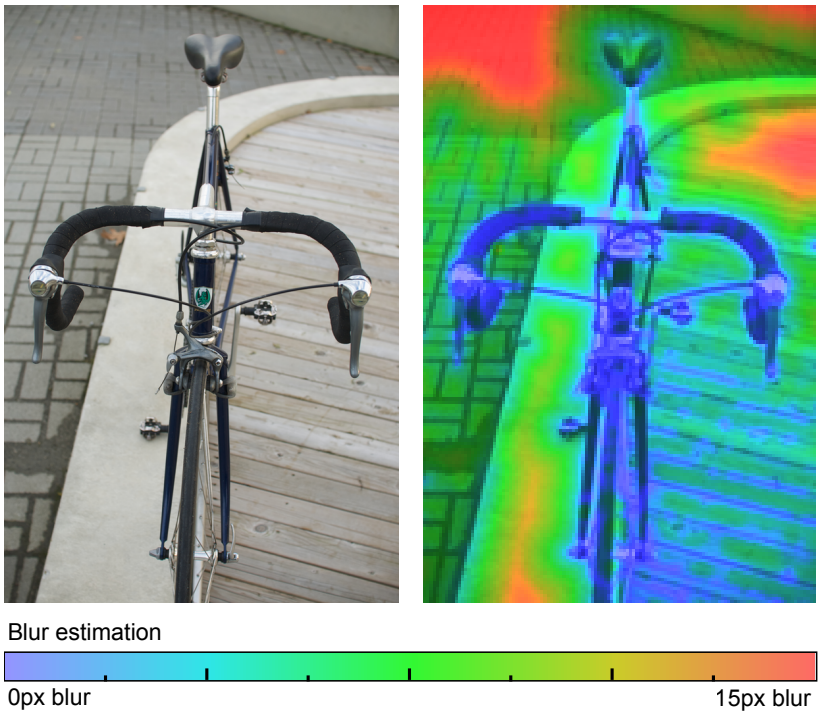


Figure 3.9: *Input image (left) and the associated blur map (right). Note that while the curb is the same distance as the wooden boards, it is estimated as blurrier due to the lack of detail in that area. The blur map is visible only in color.*

restore lost image details. However, appropriately-applied blur can be used to great artistic effect, including separating foreground from background and conveying sense of motion. In these circumstances it is advantageous to magnify the blur present in an image. Examples include preserving the appearance of blur that would otherwise be lost to some other image processing operation and introducing blur that could not be captured, in the case of small apertures. We detail two uses related to our work: the preservation of the appearance of blur to generate representative thumbnails and magnification of blur to simulate a wider aperture than used to capture the image.

Due to the limited display resolution and computational expense of displaying many images, lower-resolution thumbnails are regularly viewed in place of full-sized originals on the viewfinder of a digital cameras or in photo management software like iPhoto. However, thumbnail images are not always representative of their high-

resolution counterparts. In particular, the sparser pixel sampling of thumbnails makes it is impossible to differentiate between sharp and slightly blurred edges. Downsampled images often appear sharper than their respective originals and it is common to encounter sharp and blurred versions of an image that result in identical thumbnails.

Samadani et al. [2010] utilize their spatially-variant estimate of image blur to reintroduce blur into thumbnails lost during the downsampling process. Their approach assumes the thumbnail will be sufficiently small that all blur present in the original image will be removed. Samadani et al. then estimate the blur present in the original image, and reintroduce that blur into the thumbnail. The downsampled image is used to generate a scale-space of increasing amounts of blur. The blur estimated from the full-size image is then used to select the level of the scale-space that best represents the original image blur and produce a thumbnail better representing the blur present in the original.

Depending on the desired application, the defocus blur resulting from the choice of aperture in the lens either enhances or detracts from the image. A wide aperture with large defocus blur is considered desirable for portraits, where it can separate the subject from the background. Conversely, the same would be considered undesirable for photographing a landscape and a narrow aperture with minimal blur would be more appropriate. In both cases, the photographer wants as much control of the DOF as possible. However, a large aperture is not feasible in all cases, such as having to choose a small aperture for excessively bright conditions or using a device with a small sensor like a mobile device. In both examples, the limited aperture size prevents the capture of images with narrow DOF.

While Samadani et al. [2010] focus on magnifying blur for low-resolution thumbnails, Bae and Durand [2007] address the high-quality magnification of defocus blur in full-resolution images. Building upon the edge detection of Elder and Zucker [1998], Bae and Durand obtain an accurate estimate of the scale of blur present at the location of each edge in the image. This estimate is propagated to non-edge pixels using an image colorization scheme [Levin et al., 2004]. Once they have obtained a full-image map of the amount of blur present, Bae and Durand then use a model of defocus blur resulting from camera optics to solve for the amount of blur to introduce to the image associated with the desired wider aperture. Finally, they synthesize the resulting image with the narrower DOF.

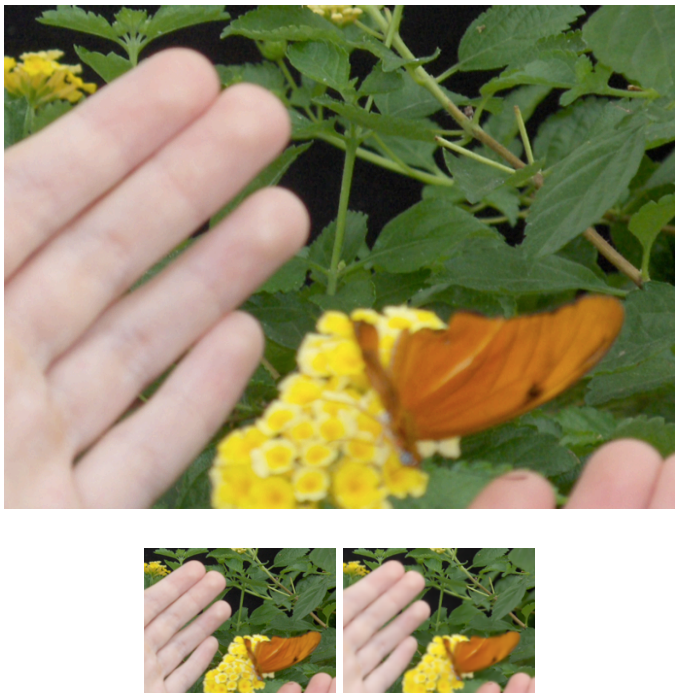


Figure 3.10: *Example of Samadani et al. [2010] blur magnification. The blur in the input image* (top) *is not visible in the conventional thumbnail* (bottom left). *The thumbnail created by Samadani et al.* (bottom right) *retains the appearance of blur. Images copyright Ramin Samadani.*

The methods of Samadani et al. [2010] and Bae and Durand [2007] are not without limitations. The method of Bae and Durand [2007] is computationally very expensive and thus not suitable for many applications, such as a digital viewfinders. In both cases, the amount of blur is increased by a single scale factor, specified by the user. As our experiment in Section 5.2 shows, the perception of blur is more complex than this relationship. Neither method can ensure that the appearance of blur will remain constant if the image is resized, motivating our approach.

3.3 Countershading Operations

In terms of both the estimation and manipulation of image blur, the most versatile methods are those that are able to operate on individual pixels using purely local

Original Estimated blur Narrowed DOF

Figure 3.11: *Example of input image* (left)*, estimated blur map* (center) *and Bae and Durand [2007] blur magnification* (right)*. Images copyright Soomin Bae.*

neighborhood information. This feature is a necessity for complex spatially-variant operations like DOF magnification. The most powerful contrast manipulation techniques operate in a similar fashion, modifying each pixel based on local information of its respective neighborhood.

In terms of the estimation of contrast, two of the methods already presented are capable of determining spatially-variant estimates. Local, band-limited contrast measures (Section 2.3) inherently represent image contrasts in a local sense. Likewise, edge detection methods that rely on scale-spaces (Section 3.2) recover both the width and contrast at pixels corresponding to edge locations. The latter is a sparse set of pixels but since edges, by definition, are changes in contrast and the representation provides complete coverage of local changes in image contrast.

Local contrast enhancement is a powerful image processing technique, fundamental to many aspects of computer graphics such as image editing and tonemapping of high dynamic range (HDR) images. The same basic techniques can be applied to improve the recognition of objects in a scene, to aid in identifying brightness of regions or to accentuate specific image details. In many cases, images with enhanced contrast appear more aesthetically pleasing [Calabria and Fairchild, 2003a,b].

As presented in Section 2.4, one of the most common approaches to enhancing local contrast in images is countershading, where the local edge contrast is increased by adding gradients to either side of the edges. This approach is common across numerous classes of algorithms. Many of these algorithms, whether explicitly or implicitly, resemble the effect of the unsharp masking operator (UMO). Unsharp masking increases local contrast by adding a high-pass image $\mathscr{H}_\sigma(Y)$ to the original

image

$$\mathscr{Y} = Y + \lambda \mathscr{H}_\sigma(Y) \tag{3.10}$$

where λ is the contrast of the countershading and the high-pass image $\mathscr{H}_\sigma(Y)$ is produced by subtracting a Gaussian blurred image $g_\sigma * Y$ from the original image Y. Most commonly, σ and λ are determined by the user and do not vary across the image. This simple technique has proven to be incredibly versatile and, depending on the choice of blur employed in the high-pass filter, can produce a variety of effects.

The choice of a small σ increases the accutance [Neycenssac, 1993], or apparent sharpness, of the image, which is the effect usually referred to as "unsharp masking" [Badamchizadeh and Aghagolzadeh, 2004]. The choice of a large σ increases the contrast of the image, which is usually referred to as countershading. In this text, we refer to the technique as unsharp masking and the effect of local contrast enhancement it produces as countershading, regardless of the choice of σ. Figure 2.17 presents a comparison of different countershading profiles.

3.3.1 Manipulating Sharpness

The addition of narrow countershading profiles resulting from a small σ will increase the apparent sharpness of the edge. Increasing the apparent sharpness, known as accutance, by the introduction of countershading differs from the deconvolution techniques of Section 3.1.3. Deconvolution techniques, in particular those that employ natural image statistics, restore frequency content absent from the image. Narrow countershading profiles, on the other hand, only increase the contrast of the edge, but do not restore lost image details. However, due to the relation between blur and sharpness inherent in the HVS, higher contrast edges *appear* sharper than low contrast edges, and the observer's impression is that the detail of the image has increased. The strength of the effect on perceived sharpness can appear to hold even more than the increase in perceived contrast.

However, this increase in sharpness can also introduce artifacts and unsharp masking is known to excessively amplify the contrast of small features, especially noise. Neycenssac [1993] noted that unsharp masking introduces contrast through two sep-

arate mechanisms: the addition of countershading profiles around edges and the amplification of existing image features. Lindeberg [1994a] notes that convolution with a Gaussian of width σ will remove all the features smaller than $\sqrt{\sigma}$. Thus any feature smaller than $\sqrt{\sigma}$, including noise, will be present in the high-pass image $\mathscr{H}_\sigma(Y)$ at their original contrast. When the high-pass image is added to the original image, only features larger than $\sqrt{\sigma}$ receive countershading profiles, while features smaller than $\sqrt{\sigma}$ are amplified. In the process of increasing edge contrasts the UMO preferentially amplifies high frequency details in the image. The noise predominantly consists of frequencies smaller than $\sqrt{\sigma}$, for even very small σ, and as such will be magnified by unsharp masking more than other features.

Original image Blurred Resulting high-pass

Figure 3.12: *Amplification of details in high-pass image. Noise present in the original image is removed when the image is blurred in the process of creating the high-pass image. The absence of high-frequency detail in the blurred image causes the noise to be present in the high-pass image and reintroduced during countershading.*

Given the effectiveness of unsharp masking for increasing the apparent detail in images, image processing researchers have dedicated substantial effort to determining optimal parameters to enhance contrast while avoiding unwanted amplification of noise. The most successful approaches have adapted the parameters to local image content. The visual channels in the HVS are not completely independent of each other and frequency content at one scale can affect the perception of details at another scale. As a result of this attribute, known as *contrast masking*, the visual system is less sensitive to noise in regions around edges [De Valois and De Valois, 1990].

Adaptive unsharp masking operators adjust their parameters to local gradient magnitude, increasing the strength of countershading near edges where it will increase accutance, and decreasing it further away where it will increase the appearance of noise. Polesel et al. [2000], Ramponi et al. [1996] and Kim and Allebach [2005]

have all proposed methods of adaptively determining the optimal countershading contrast, λ. Likewise Wang et al. [2001] and Nowak and Baraniuk [1998] developed techniques adaptively determining the scale σ. Badamchizadeh and Aghagolzadeh [2004] evaluated a selection of unsharp masking techniques comparing their performance at contrast improvement and noise amplification. However, most of this work attempts to address the issue of artifacts resulting from noise amplification, and not the artifacts caused by objectionable magnitude of countershading (haloes), which is the focus of Chapter 6.

3.3.2 Manipulating Contrast

While the choice of a small unsharp masking σ parameter will increase the perceived sharpness of an image, the choice of a large σ parameter will increase the local contrast of the image. For a given choice of λ, both parameter choices result in an equal increase in contrast at the location of the edge, but the narrower profile is perceived as a change in detail and the wider profile is perceived as a change in contrast. Sufficiently wide unsharp masking profiles can even induce the Cornsweet illusion (Section 2.4), where the entirety of adjacent regions change in appearance.

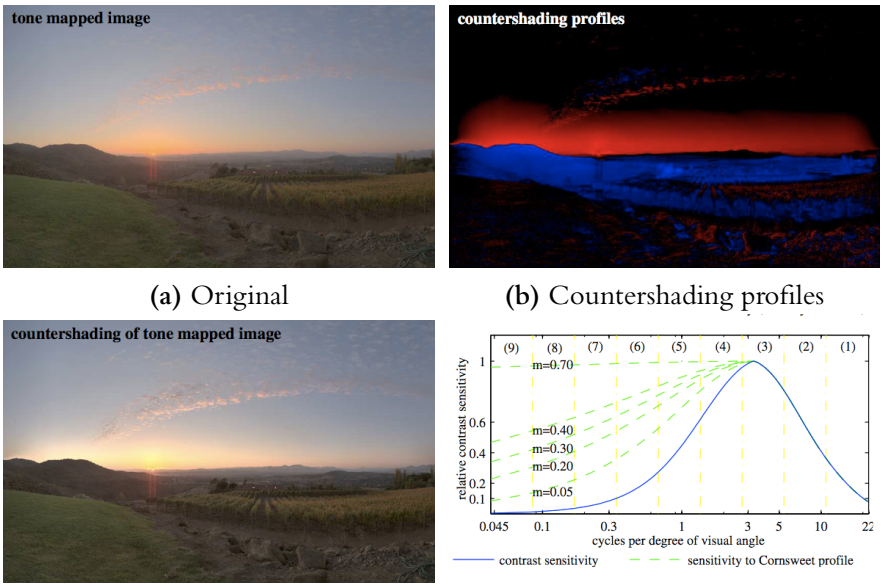
In many cases, increasing the perceived contrast of an image increases viewers' preference [Calabria and Fairchild, 2003a,b] and image processing algorithms have used countershading to increase the aesthetic quality of the image. Additionally, local contrast enhancement can be applied to improve the recognition of objects in a scene, increasing the contrast between them and the surrounding areas [Luft et al., 2006].

Countershading is frequently employed in local tonemapping operators, although not always intentionally. The operator by Chiu et al. [1993] explicitly computed a high-pass image with a large σ to reduce global contrast while retaining edge contrast. However, while reducing the dynamic range this naïve approach introduced significant haloes. More recent operators like Durand and Dorsey [2002] and Fattal et al. [2002] can also introduce countershading profiles, though they do not view them as a benefit. In general, countershading in local tonemapping is considered synonymous with halo artifacts or "gradient reversals". On the other hand, we view the particular combination of contrast and scale of the countershading, rather than the presence of any profiles, to be responsible for the loss of image quality.

Figure 3.13: *Many tonemapping operators employ countershading to reduce the contrast of a* HDR *image* (left) *to make details more visible* (right). *Images copyright Kevin McCoy Wikipedia contributor Darxus.*

Most related to our work is that of Krawczyk et al. [2007] on restoring luminance contrast lost during tonemapping. They propose an automated method that introduces countershading to a low dynamic range (LDR) tonemapped image to match the contrast of a reference HDR image. In an attempt to avoid introducing objectionable artifacts, they propose a perceptual model of just-detectable countershading based on the model of Dooley and Greenfield [1977]. However, such a model relates to the detection of countershading profiles, not whether the countershading is considered objectionable, and is overly conservative. Additionally, the method of Krawczyk et al. is meant to reproduce contrast of an HDR image rather than enhance contrast without an HDR reference. We discuss their model in more detail in Section 6.4.

Countershading has been used in several other capacities to restore lost contrast and enhance scene understanding. Smith et al. have used countershading to restore color saturation lost during tonemapping [2006] and while converting from color to greyscale [2008]. Similarly, Luft et al. [2006] and Ritschel et al. [2008] added countershading based on depth values to improve the recognition of objects in scenes. Finally, work by Akyüz and Reinhard [2008] used the Cornsweet illusion as a means of evaluating the perception of contrast alterations introduced by tone-mapping operators.



(a) Original **(b)** Countershading profiles

(c) Restored contrast **(d)** Countershading detection

Figure 3.14: *Restoration of lost contrast by Krawczyk et al. [2007]. A tone mapped image* (a) *doesn't always reproduce the appearance of contrast of its* HDR *counterpart. Krawczyk et al. introduce countershading profiles* ***(b)*** *to produce a higher contrast result* ***(c)*** *that more closely matches the original, using a model of human perception* (d) *to avoid the countershading being deemed objectionable. Images copyright Gregorz Krawczyk.*

3.4 Related Applications

Finally, our investigations on image resizing and the creation of representative images relate to a selection of other topics in image processing and computer graphics. These subjects include image upsampling, seam carving, visualization of large images and noise estimation.

In the context of image resizing, Fattal [2007], Kopf et al. [2007] and Shan et al. [2008b] developed techniques for intelligently *upsampling* images. These methods use assumptions about the image statistics to invent additional information based on existing details to provide a more natural rendition than resampling with a reconstruction-filter. Fattal [2007] define a edge-frame continuity modulus to impose a certain set of statistics on the edges in the upsampled image, retaining sharpness where appropriate. Kopf et al. [2007] use existing high-resolution images to aid in the upsampling

of the result of a computation performed at lower resolution. Lastly, Shan et al. [2008b] extended the concept of upsampling to extend across multiple images and handle video sequences.

Upsampling input Upsampling result

Figure 3.15: *Intelligent upsampling tries to hallucinate more realistic image details than a reconstruction filter. Images copyright Raanan Fattal.*

Seam carving methods, such as the one by Avidan and Shamir [2007], resize images by inserting or removing pixels in the least important regions of the image, preserving the overall structure. A number of additional methods have been presented, including extensions to video [Krähenbühl et al., 2009; Pritch et al., 2009]. Seam-carving mostly focuses on adjusting aspect ratios and is combined [Rubinstein et al., 2009] with regular downsampling operations for extreme changes in resolution. Rubinstein et al. [2010] provides a comprehensive overview of existing methods, and conducts both perceptual and objective analyses of each, comparing their performance.

Image resampling is not the only means of generating smaller versions of the image and depending on the intended task, a different representation of the content present in the image may be preferable. These algorithms produce smaller versions of an image that are not derived from a straight resizing of the original. Suh et al. [2003] and Santella et al. [2006] have proposed automatic cropping methods for selecting the most salient region of image to highlight in a thumbnail. Berkner and Erol [2008] present a method of creating thumbnails of pages of documents that are more recognizable than simple downsampled versions of the page. They focus on

Seam carving input | Scaled image | Retargeted result

Figure 3.16: *Image retargeting tries to change the aspect ratio of the image without scaling by removing the least salient portions of the image. Images copyright Wikipedia contributor Newton2.*

preserving the shape and position of blocks of text, as well as giving added emphasis to the position and content of images.

Lastly, resizing operations can remove additional features besides the blur present in the full-resolution original. When resizing, the full-size image is low-pass filtered to avoid introducing aliasing into the thumbnail. However, this operation will remove any noise present in the large image and the thumbnails for a pair of low-noise and high-noise images of the same scene will be identical. If the resizing operation wants to preserve the appearance of noise in the original, it will need to accurately estimate the quantity present. Any adaptive denoising technique includes some means of estimating the noise level of an image region, and wavelet denoising techniques by Donoho et al. [1993] and Simoncelli and Adelson [1996] are a common example. Other approaches attempt to characterize the noise as a function of the camera's response, such as methods by Liu et al. [2006] and Shin et al. [2005]. These techniques have applications beyond denoising and resizing. Adjusting the smoothing step of edge detection to the level of noise present is one such example.

Chapter 4

Synthetic Depth-of-Field for Mobile Devices

The depth-of-field (DOF) present in a photograph can significantly alter how that image is perceived. Skillful photographers can use a shallow depth-of-field to great effect and dramatically emphasize the subject matter of the photograph. Cell phones and similar mobile devices are rapidly becoming the primary camera for many users, and the optical packages of these devices cannot produce a narrow depth-of-field. The size of the optical package that can be constructed in such a confined area prevents the lens from producing a large defocus.

4.1 Introduction

The limited space available for cameras in mobile devices places constraints on the design of the lens and the optics of which it is comprised. One effect of this confined space is the the lens is unable to achieve a significant defocus blur, preventing the camera from capturing images with a narrow DOF. It is very difficult to overcome this constraint with the physical construction of the lens, and instead we propose a method of producing a narrower DOF synthetically. Our method builds on existing approaches to robustly determine a spatially-variant estimate of the defocus blur present in a captured image when significant noise is present, shown in Figure 4.1. We then modify that estimate to produce the blur associated with an image of a

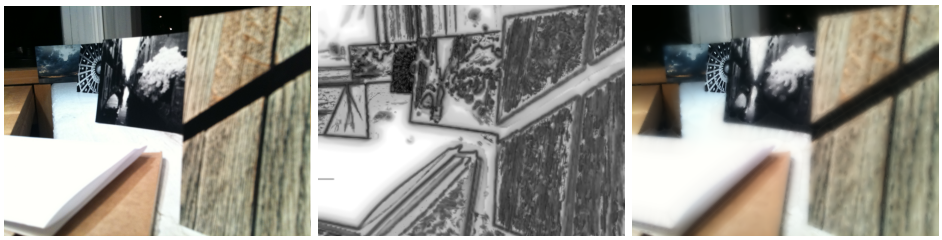


Figure 4.1: *A comparison of the original image (left), estimated blur map (center), and our synthetic depth-of-field algorithm (right). The depth-of-field is sufficiently wide in the original image to make determining the card most in focus difficult, while the depth-of-field in our result is narrow enough to make it obvious that the third photograph is in focus.*

narrower DOF and use that new blur to synthesize an image with the desired amount of defocus blur. The result, also shown in Figure 4.1, is a means of producing a narrower DOF than could physically be captured suitable for mobile devices, in terms of both computational efficiency and robustness to the noise present in captured images.

Since the constraints on the DOF cannot be easily solved with a replacement optical design, we propose to narrow the depth-of-field by synthetically magnifying the defocus blur. While methods for magnifying the blur present in an image already exist, they are either too computationally expensive to be realized in a mobile device [Bae and Durand, 2007], or are sensitive to the noise levels common in cellphone cameras and produce low-quality results [Samadani et al., 2010].

Drawing on existing methods, we propose an efficient and robust algorithm for estimating the blur present in an image. Given a noisy image, we first compute how much we must reduce the variance of the noise at each pixel location to accurately estimate the blur present. We estimate the blur at a number of different scales and choose the most reliable estimate given the noise present. We then use that blur estimate to synthesize a new image with magnified defocus blur, simulating a reduced DOF for devices with aperture-limited optical packages.

4.2 Blur Estimation

To create an image with the desired DOF, our algorithm first determines a spatially-variant estimate of the blur present in the original image, then modifies that blur

estimate to correspond to a narrower depth-of-field, and finally synthesizes the resulting image. As a starting point for our estimation method we use the algorithm by Samadani et al. [2010], detailed in Section 3.2.4, because of its simplicity and computational efficiency. Figure 4.2 provides a flowchart of the operation of the algorithm.

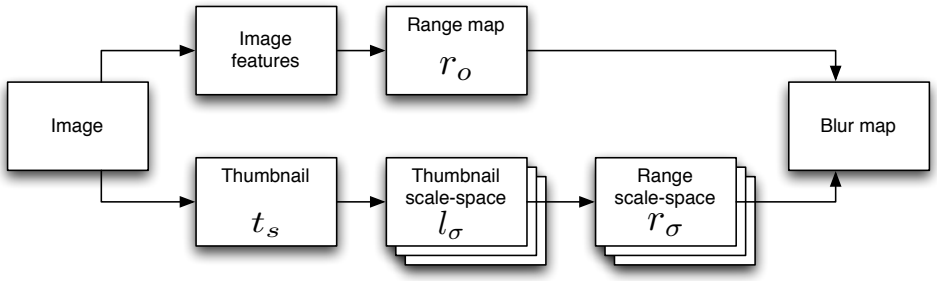


Figure 4.2: *Flowchart of Samadani et al. blur estimation. Features from the full-size image are compared to corresponding features from the thumbnail scale-space to determine the appropriate blur for each pixel.*

While Samadani et al.'s blur estimation provides a means of controlling the relative increase in the amount of blur in the resulting thumbnail, it does not provide an absolute measure of the blur present in the large image. In order to amplify the defocus blur, we need to know the scale of the blur present in the original image. To do so, we extend their local image features to a general relationship between the width of a blurred edge and the corresponding derivatives at different resolutions, which we can use to recover the scale of the original image blur.

In the case of a 1D Gaussian blurred edge of normalized contrast, the edge profile is the integral of the Gaussian function. The derivatives of this profile follow the Gaussian function, with the peak lying at the center point of the edge. For a Gaussian profile with standard deviation σ to have a contrast of 1, the derivative of the edge cross-section will be:

$$g(\sigma, x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}. \tag{4.1}$$

This scaling factor establishes a relationship between the width of the Gaussian profile and the scale of its derivatives. If the width of the edge profile changes by a factor

k, the derivatives must change by a factor of $1/k$ to retain the same contrast.

The range map operator in Samadani et al. approximates the gradient magnitude. For an edge with blur σ_o, the corresponding range map will equal a Gaussian distribution at the edge location with an amplitude of $1/\sqrt{2\pi\sigma_o^2}$. After downsampling that image to obtain r_o, the value at the edge location is still equal to $1/\sqrt{2\pi\sigma_o^2}$.

Due to downsampling, the effective width of that edge in the thumbnail l_{σ_0} will differ by the downsample factor d. The pixel corresponding to the edge location in the thumbnail range map r_{σ_0} will be

$$\frac{1}{\sqrt{2\pi\left(\frac{\sigma_o}{d}\right)^2}} \tag{4.2}$$

due to the relationship between the width and scale of a Gaussian mentioned above. Additionally, that σ_o/d will be further altered by the Gaussian filtering that generates the scale-space images l_{σ_j}. Using the convolution formula for Gaussian functions:

$$g(n_1,\sigma_1)\otimes g(n_2,\sigma_2) = g\left(n_1+n_2,\sqrt{\sigma_1^2+\sigma_2^2}\right), \tag{4.3}$$

the width of the edge in thumbnail scale-space image l_{σ_j} will thus be

$$\sqrt{\left(\frac{\sigma_o}{d}\right)^2+\sigma_j^2}. \tag{4.4}$$

We construct the scale-space from a series of blurs with a uniform spacing of β, which implies $\sigma_j = \beta j$. This way, the choice of β along with the maximum j determines the range and quantization of the scale-space. The end result is two different values for corresponding pixels of the two range maps:

$$r_o = \frac{1}{\sqrt{2\pi\sigma_o^2}} \text{ vs. } r_{\sigma_j} = \frac{1}{\sqrt{2\pi\left[\left(\frac{\sigma_o}{d}\right)^2+(\beta j)^2\right]}}. \tag{4.5}$$

Figure 4.3 depicts the relationship between r_o and the scale-space r_{σ_j} for an edge of increasing blur.

For the algorithm to select the correct value for the blur map $m(i)$, the two values of Equation 4.5 must be equal. In complex images, adjacent image features

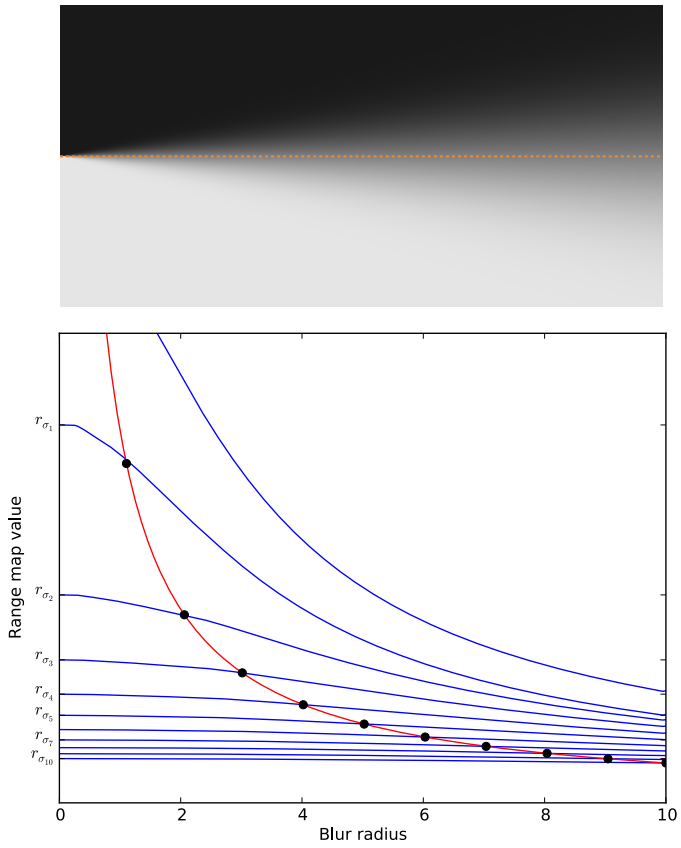


Figure 4.3: *Demonstration of calibrating blur estimation. Top: image of a step edge with blur increasing from* $0 \ldots 10$ *along the x-axis. Bottom: range map values along dotted line for the original image* r_o*(red), and the downsampled scale-space* r_{σ_j} *(blue). Note that the intersection between* r_o *and* r_{σ_j} *(black dots) happens at* $x = j$*.*

alter the derivative values at this edge location and a direct solution would misestimate σ_o. Our approach determines σ_o based on the correspondence between the range map and the levels of the scale-space. Adjacent features alter the gradient magnitude in both range maps in the same fashion, and the correspondence between them is preserved. Additionally, image features smaller than $\sqrt{\sigma}$ are suppressed on the scale-space level with a Gaussian blur of σ, eliminating some of overlapping features [Lindeberg, 1994b].

To determine the value of the blur map $m(i)$, Samadani et al. employ the user-

specified parameter γ to bias the selection of values for $m(i)$ towards more or less blurred levels of r_{σ_j}. Noting that the relation between the two range maps depends on the downsample factor, we instead determine the value of γ that correctly scales the range map of the original image and the range maps of the downsampled scale-space. We solve the following equation between r_o and r_{σ_j}

$$\gamma \frac{1}{\sqrt{2\pi\sigma_o^2}} = \frac{1}{\sqrt{2\pi\left[\left(\frac{\sigma_o}{d}\right)^2 + (\beta j)^2\right]}} \tag{4.6}$$

for the value of γ that ensures the index of the blur map selected with Equation 3.9 is equal to the width of the blur in the original large image, $m(i) = j$ if $\sigma_o = j$. Canceling terms and solving for γ yields:

$$\gamma = \frac{1}{\sqrt{\left(\frac{1}{d}\right)^2 + \beta^2}}. \tag{4.7}$$

The result is a value of γ automatically chosen for a given downsample factor and scale-space resolution. In our method, we use 25 levels of blur ($j = 1, \ldots, 25$) and a value of $\beta = 0.4$. The resulting values for downsamples of 2, 4, and 10 are $\gamma_{d=2} = 1.55$, $\gamma_{d=4} = 2.11$ and $\gamma_{d=10} = 2.48$.

4.3 Noise-Robust Estimation

However, the resulting estimate is highly susceptible to any noise present in the original image. Consider an image that contains a blurred edge and some amount of noise. The larger the blur of the edge, the smaller the corresponding gradient magnitudes will be, until the gradient magnitude, and resulting blur estimate, is determined more by noise present than the edge. This situation can occur for even small amounts of noise, and when it does the estimated blur is biased towards low blur scales or is mis-estimated as perfectly sharp, as seen in Figure 4.4.

As a result, Samadani et al. alone cannot accurately estimate the blur in images captured with mobile devices due to the noise levels present. To address this shortcoming, we can determine the MRSCALE of Elder and Zucker [1998], described in Section 3.2.3, at which we can accurately estimate the blur. The MRSCALE of an image region corresponds to the largest amount of blur that can be reliably detected

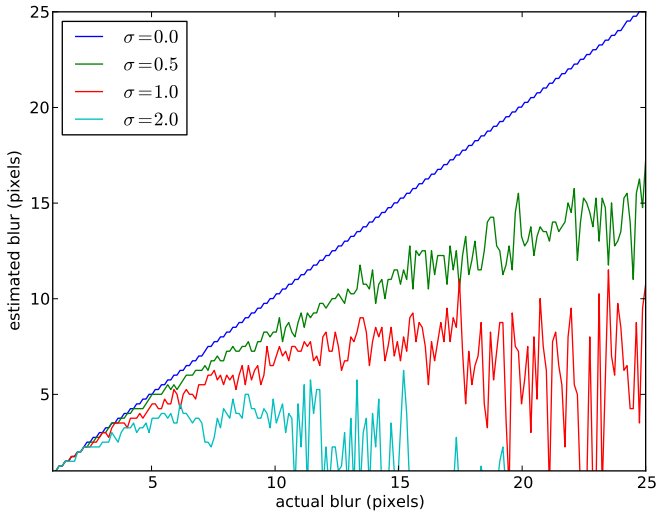


Figure 4.4: *Comparison between actual edge blur ($\sigma = 0$) and the blur estimated by Samadani et al. for various noise levels σ. As the noise level increases, the estimated blur is biased towards smaller scales, eventually estimating the entire image is sharp.*

in that image region. This scale is dependent on the SNR, and hence the image intensities in that region. Elder and Zucker prove that for a given edge profile and noise intensity, the image can be blurred by some amount to sufficiently reduce the noise to allow accurate estimation of the blur.

Combining the methods by Samadani et al. and Elder and Zucker, we can produce a robust estimate of the blur in noisy images as follows, and shown in Figure 4.5:

1. For each scale s of the set of minimum reliable scales, convolve the original image by a Gaussian of scale s to produce a blurred copy $I_s(x,y)$.

2. For each $I_s(x,y)$, obtain the estimated blur map $m_s(x,y)$ using Samadani et al., as described in Section 4.2.

3. For each scale s, compute the gradient magnitude $r_s(x,y)$ of the corresponding blurred image $I_s(x,y)$.

4. Compare all the gradient magnitudes $r_s(x,y)$ to determine the appropriate scale at which to estimate the blur $\hat{s}(x,y)$ for each pixel.

5. Select the appropriate $m_s(x,y)$ for each pixel according to the scale map $\hat{s}(x,y)$ to produce the final blur estimate $m(x,y)$.

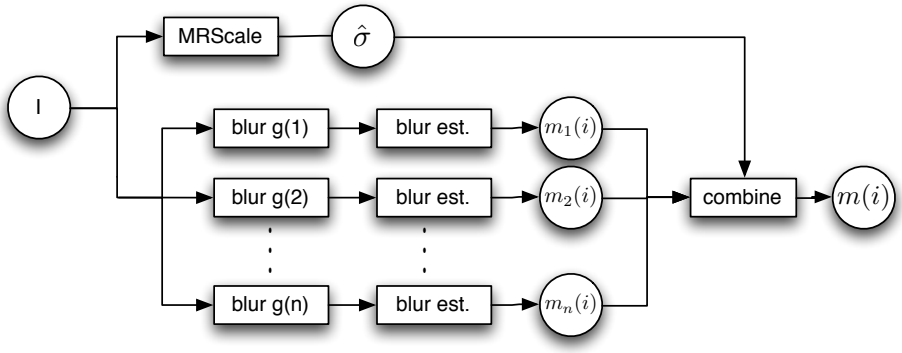


Figure 4.5: *Flowchart overview of our combined minimum reliable scale and blur estimation. The gradient magnitude $r_s(x,y)$ and blur estimate $m_s(x,y)$ are computed at scale s. Then the gradient magnitudes $r_s(x,y)$ are compared to produce the minimum reliable scale map $\hat{s}(x,y)$, which is used to select the appropriate blur estimate $m_s(x,y)$ for each pixel to produce the final blur estimate $m(x,y)$. This process is repeated for each minimum reliable scale level being considered.*

The blur estimate corresponding to each blurred copy $I_s(x,y)$ includes the amount of blur that was added to the image to ensure a reliable estimation. We use the convolution formula of Gaussians from Equation 4.3 to determine the blur present in the original image before creating the final blur estimate $m(x,y)$. Once the reliable blur estimate has been obtained, we modify the final blur estimate to represent the narrower DOF. Finally, we use the modified blur map to synthesize the desired image.

4.4 Efficient Estimation

While conceptually simple, the approach outlined in the previous section is inefficient. Compared to the method of the previous section, our optimized method exploits redundancies in the intermediate results of computing each stage of the method. Both the MRSCALE computation and the blur estimation corresponding to each scale constructs a scale-space [Lindeberg, 1994c] of Gaussian-blurred images used in the computation of their respective final results. We optimize two portions

of the method: results shared between a given MRSCALE level and the blur estimation associated with that level — $r_s(x,y)$ vs. $m_s(x,y)$ — and results shared between the blur estimations associated with different scales — $m_{s_1}(x,y)$ vs. $m_{s_2}(x,y)$.

First, we consider the results shared between the computation of a given minimum reliable scale level and the blur estimation associated with that level. To determine the final map of the minimum reliable scale $\hat{s}(x,y)$ and blur estimate $m(x,y)$ for a given image I, two outputs from each scale s of the MRSCALE computation are required: the gradient magnitude map $r_s(x,y)$ and estimated blur map $m_s(x,y)$ corresponding to that scale.

Both the MRSCALE computation and the blur estimation require the gradient magnitude $r_s(x,y)$ corresponding to that scale. Additionally, the blur estimation requires the original image I convolved by a Gaussian of radius s, $g(s)$. We first blur I by $g(s)$ to obtain the blurred image $I_s(x,y)$. We then compute $r_s(x,y)$ from the partial differences of $I_s(x,y)$. Once the input to the blur estimation $I_s(x,y)$ and gradient magnitude $r_s(x,y)$ have been computed, the blur estimation proceeds as normal. Our method is shown in the top half of Figure 4.6.

On the other hand, the naïve formulation computes each step in different ways, as shown in the bottom half of Figure 4.6. The MRSCALE computes the gradient magnitude from the result of convolving the image with steerable Gaussian basis filters $g_x(s)$, $g_y(s)$. To compute the same quantity for the blur estimation, the image has to first be blurred by a Gaussian $g(s)$, then partial differences are computed to determine the gradient magnitude. Our method computes 2 less convolutions of the full-size I per scale s.

Second, we consider the results shared between the blur estimations associated with different scales. For each scale s, the image blurred by that scale $I_s(x,y)$ is passed to the corresponding instance of blur estimation. In the construction of each of the blur estimation scale-spaces, that input $I_s(x,y)$ is further blurred by range of values $\sigma_j = 0 \ldots j$, resulting in a final blur of

$$I_{s,\sigma_j} = I \otimes g\left(\sqrt{s^2+\sigma_j^2}\right), \tag{4.8}$$

by Equation 4.3. A resulting blur of $\sqrt{s^2+\sigma_j^2}$ will occur in the blur estimations of multiple scales. We compute all the possible combinations of $\sqrt{s^2+\sigma_j^2}$, round

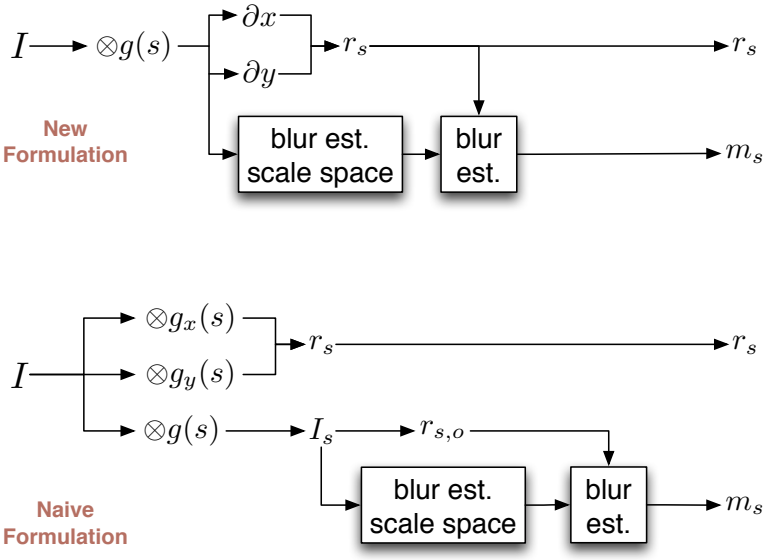


Figure 4.6: *Comparison of our formulation and the original formulation for computing $r_s(x,y)$ and $m_s(x,y)$. Both* MRSCALE *and the blur estimation share $r_s(x,y)$, but compute it in different ways. We rearrange the terms to share the computation between both portions with minimal extra computation.*

each to the nearest integer, and only compute the final blur once and use it in all blur estimations that require that value. At worst, this is $O(\sqrt{s^2+\sigma_j^2})$ convolutions, opposed to $(s \times j)$ convolutions.

Finally, there are several minor additional optimizations. All of the Gaussian blurs are computed using the recursive implementation by Young and van Vliet [1995]. We compute all the gradient magnitude maps $r_s(x,y)$ using the partial differences, compared with the maximum difference between a pixel and the 8 neighbors of Samadani *et al.*, reducing the number of memory reads by a factor of $4\times$.

While neither of these optimizations seems like much on its own, they add up to a notable difference in required computation. Consider the specifics of an Apple iPhone 4, with a resolution of 5 megapixels. We use 7 scales in the minimum reliable scale computation: $s \in .5, 1, 2, 4, 8, 16, 32$, while we use a blur estimation scale-space with a maximum of 30: $\sigma_j = 0 \ldots 30$.

As with the naïve implementation, all of the $r_s(x,y)$ are combined to form the minimum reliable scale map $\hat{s}(x,y)$. In turn, that map is used to select the appropriate

operation	our method	naïve
full-size image blurs	7	21
blur-estimation scale-space blurs	33	182
full-size gradient magnitude diffs	14	56
downsampled gradient magnitude diffs	66	1456

Table 4.1: *Comparison of the number of operations for different steps in the full blur estimation operation, specifically the number of Gaussian blurs and pixel differences that happen at the original resolution and the downsampled resolution used in Samadani* et al.

blur estimate $m_s(x,y)$ for each pixel of the final blur estimate $m(x,y)$. Lastly, as noted in Section 4.2, the estimated blur is computed on a downsampled image, and we resize the final blur estimate $m(x,y)$ to the size of the original image.

4.5 Blur Synthesis

Given an accurate estimate of the blur present in an image, we need to synthesize the final image with the desired narrower DOF. To do so, we scale the estimated blur map to represent the target DOF $m_d(x,y)$, and compute the amount of additional blur $m_a(x,y)$ we must introduce at each pixel to achieve that amount of blur. Finally, we produce the image with that additional blur to obtain the desired DOF.

In optical terms, the estimated blur map $m(x,y)$ approximates the radius of the circle of confusion due to defocus blur at each point. Decreasing the DOF of the image corresponds to linearly scaling all values of the blur map by a value greater than 1 to increase the circle of confusion. To adjust the DOF to be equal to an f-number different from that of the physical aperture, we use the relationship between the diameter of the circle of confusion c and f-number N:

$$c = \frac{|S-D|}{S} \cdot \frac{f^2}{N(D-f)} \tag{4.9}$$

where D is the focus distance, S is the distance of a point in the scene, and f is the focal length of the lens. We are interested in the proportional increase between the circles of confusion of the desired f-number N_d and original f-number N_o. The increase b applied is equal to the ratio of c_d and c_o, canceling terms, we can discard all distances, and this quantity is equal to the ratio of the f-numbers:

$$b = \frac{c_d}{c_o} = \frac{N_o}{N_d}. \tag{4.10}$$

So the desired blur is given as $m_d(x,y) = b\,m(x,y)$.

Even though the $m(x,y)$ approximates the blur at each pixel in the image, that blur does not necessarily represent the defocus or depth. If an image region is a flat color, we cannot determine whether that is a detailed region that is out of focus or it is in focus but lacks any detail. If a point in the scene is out-of-focus we know that the corresponding pixel will be estimated as blurred, but if a pixel is estimated as blurred we cannot assume the corresponding point in the scene is out of focus.

One implication of this ambiguity is that the estimated blur of adjacent pixels can differ significantly. While synthesizing the final image, the difference in amount of additional blur those pixels receive will be even larger, due to the scaling b applied to the blur map. This inconsistency between adjacent blurs implies the differences the respective neighborhoods can introduce contrast inversions in the final image, as seen in Figure 4.7.

Uncorrected blur map Original image Corrected blur map

Figure 4.7: *Example of the contrast inversion resulting from synthesizing the final image with an uncorrected blur map. The uncorrected blur map can have vastly different blurs adjacent to each other, causing contrast inversions where the blur amount changes, like the eyes, while the corrected blur map does not. Original image copyright Flickr user hpj7173.*

One means of addressing this issue is to solve a sparse linear system [Levin et al., 2004] like Bae and Durand [2007] to propagate the estimated blurs until they are consistent. While this approach successfully avoids any artifacts for a large range of desired DOFs, it is too computationally expensive for our targeted application. In-

stead, we enforce the simple condition that no pixel in the desired blur $m_d(x,y)$ may differ from its neighbors by more than some quantity. For each $\sigma = 0 \ldots \max(m_d)$ in the range of desired blurs, we compute the distance transform for the set of pixels with a desired blur of σ. Each pixel in the modified map m'_d is the minimum of all the distance values the corresponding to that pixel. The result of this operation is shown in Figure 4.8.

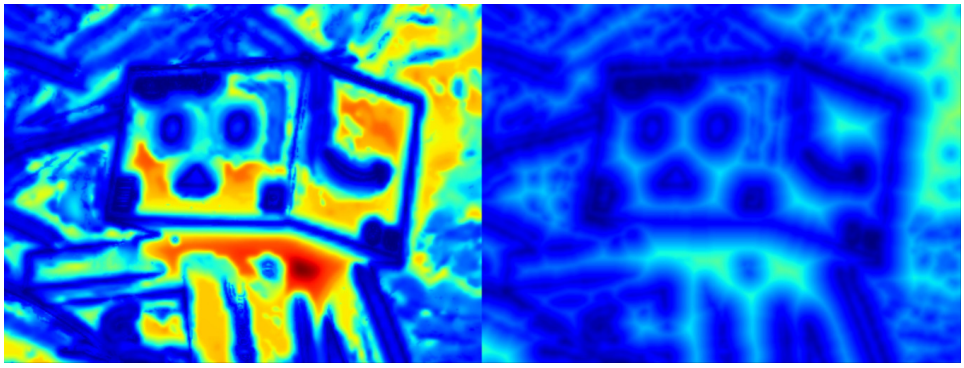


Figure 4.8: *False-color image comparing the desired blur map m_d (left) and the modified version m'_d (right). The large estimated blurs on the character's face and chest are reduced to be consistent with nearby smaller values. Original image copyright Flickr user hpj7173.*

We use that new blur $m'_d(x,y)$ and the blur present $m(x,y)$ to compute the additional blur required $m_a(x,y)$ at each pixel with Equation 4.3. To produce the final image, we construct a Gaussian scale-space corresponding to the range of additional blurs $\sigma_a = 0 \ldots \max(m_a)$ For each level, we blur the original image I by the corresponding amount of additional σ_a, then linearly blend those blurred images together to approximate non-integer values of m_a.

The Gaussian model differs from the geometric model for defocus blur. However, it has been argued that Gaussian blur better accounts for artifacts in actual cameras, and it has been used widely in computer vision [Pentland, 1987; Subbarao, 1992]. While more accurate spatially-variant blur synthesis is possible, such as Popkin et al. [2010], we haven't noticed any artifacts requiring such methods.

4.6 Evaluation

Most of the parameters of our algorithm can be determined from the properties of the imaging system of the mobile device. We use the Apple iPhone 3GS as the example for our evaluation, the specifications of which are in Table 4.2.

Resolution	2048 x 1536
Aperture	$f/2.8$
Focal length	3.85 mm
Sensor size	4.37 mm (diag)

Table 4.2: *Specifications of the Apple iPhone* 3GS *camera.*

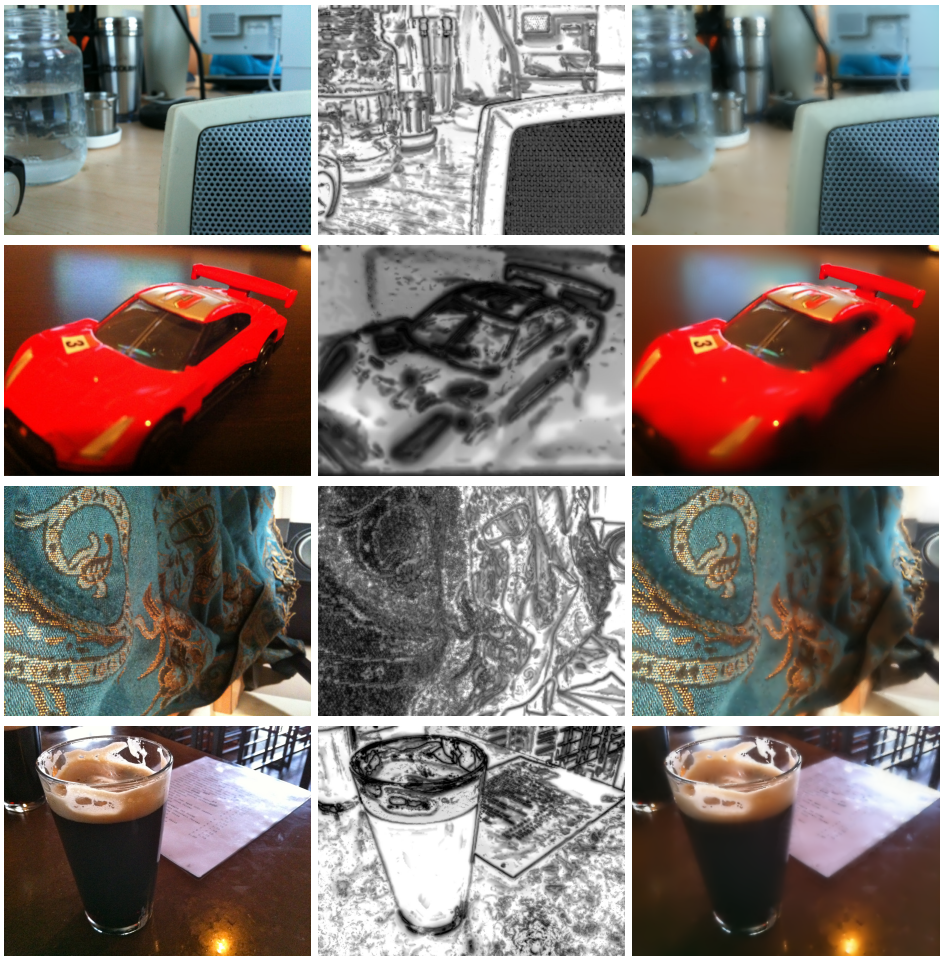
Even though the iPhone 3GS lens has an aperture of $f/2.8$, the defocus blur produced by that aperture is the same as an $f/28$ aperture for the equivalent focal length on a 35mm camera. Given that circle of confusion, the maximum defocus blur that can be achieved is a disc filter with radius of 15 px, by Equation 4.9.

For the blur estimation, we use a downsample factor of $4\times$ and a scale-space with levels corresponding to Gaussian scales ($j = 0 \ldots 15$) and a value of $\beta = 0.4$. For the minimum reliable scale computation, we estimate the noise present in the image using the method by Ibenthal [2007].

We synthesize the final image with $2\times$ the circle of confusion found in the original image. This increase in circle of confusion equates to an aperture of $f/1.4$ ($f/14$ in 35 mm terms), and results in a maximum of additional blur of $\sigma = 25$. Artifacts begin to become apparent beyond this scale, and narrower DOFs would require better blur synthesis methods such as those used by Bae and Durand [2007]. Figure 4.9 contains results of our method.

While our method can produce reasonable results in many cases, it has several limitations. In order to synthesize a narrower depth-of-field, there has to be a range of defocus blurs detectable in the image. Unless the foreground is very close to the camera, the amount of defocus blur will be not be sufficiently large for our algorithm to distinguish between blurred and sharp edges. In this case, magnifying the estimated blur will have little effect on the DOF.

Additionally, our method can accurately estimate the blur of edges when noise is present, but it has difficulty with textured regions. Depending on the estimated noise



Original image | Estimated blur map | Narrower-DOF result

Figure 4.9: *A selection of images produced with our method, comparing the conventional thumbnail (left), estimated blur map (center), and our synthetic depth-of-field algorithm (right). In all images, the desired blur $m_d(x,y)$ was determined by doubling the estimated blur.*

level, our method cannot easily distinguish between fine texture detail and noise, as seen in Figure 4.10. Significant noise present in the image causes many textured regions to be estimated as blurry and all detail is blurred out when synthesizing the final image. This confusion between noise and texture would be avoided in the future by using an analytic noise model for the given camera at the specific exposure and gain levels used to capture the image.

Figure 4.10: *Our blur estimation method cannot distinguish between noise and texture detail, and regions with fine texture detail will result in a blurry estimate. The texture in the foam (left) is indistinguishable from the noise in the beer, and both are blurred in the synthesized image (right).*

Neither our method nor Samadani et al. can accurately detect motion blur or camera shake. Our method estimates the blur in proportion to the maximum gradient magnitude. Large gradient magnitudes correspond to small blurs, while small gradient magnitudes correspond to large blurs. Both motion blur and camera shake result in a linear blur, which still has large gradient magnitudes orthogonal to the blur direction, and such image regions are estimated as sharp.

4.7 Conclusion

In this chapter we have combined the blur estimation of Samadani et al. with the robustness of Elder and Zucker's minimum reliable scale to produce an efficient means of estimating spatially-variant blur in noisy images. We have shown that refactoring the implementation of the method can reduce the amount of computations necessary, especially in terms of the number of Gaussian blurs required.

The optical design of cameras included in mobile devices makes it impossible to physically capture a narrow DOF. Using our estimate of the blur present in the

image, we synthesize a new image with narrower DOF than the camera could actually capture. While our method has limitations in terms of how accurately it can estimate blur, and how much it can enhance the blur it does find, it is capable of producing compelling results.

Chapter 5

Blur-Aware Image Downsizing

Synthesizing a narrower DOF is not the only application of blur magnification. The spatially-variant blur estimation of the previous chapter can be used with other models than camera optics. Resizing to a lower resolution can alter the appearance of an image. In particular, downsampling an image causes blurred regions to appear sharper. It is useful at times to create a downsampled version of the image that gives the same impression as the original, such as for digital camera viewfinders. To understand the effect of blur on image appearance at different image sizes, we conduct a perceptual study examining how much blur must be present in a downsampled image to be perceived the same as the original. We find a complex, but mostly image-independent relationship between matching blur levels in images at different resolutions. We incorporate this model in a new appearance-preserving downsampling algorithm, which alters blur magnitude locally to create a smaller image that gives the best reproduction of the original image appearance.

5.1 Introduction

One pervasive trend in imaging hardware is the ever increasing pixel count of image sensors. Today even inexpensive cameras far outperform common display technologies in terms of image resolution. For example, very few cameras remaining on the market, including mobile devices, capture an image at low enough resolution to show on a 1080p high-definition television (HDTV) display without resizing. In an extreme case, the preview screen on one Nikon professional digital single-lens

reflex (SLR) can only display 1.5% the pixels captured by the sensor. The cellphone camera owner and the 4K cinematographer face the same problem of getting an accurate depiction of the image when they can't see all the pixels.

While high resolution images are needed for a number of applications such as on-camera previewing, print output or cropping, the image is often previewed on a display of lower resolution. As a result, image downsampling has become a regular operation when viewing images. Conventional image downsampling methods do not accurately represent the appearance of the original image, and lowering the resolution of an image alters the perceived appearance. In particular, downsampling can cause blurred regions to look sharp and the resulting image often appears higher quality than its full-size counterpart. While the higher quality images can be desirable for purposes such as web publishing, the change is problematic in cases where the downsampled version is to be used to make decisions about the quality of the full-scale image, for example in digital view finders.

In this chapter, we aim to develop an image downsampling operator that preserves the appearance of blurriness in the lower resolution image. This is a potentially complex task — the human visual system's ability to differentiate blurs is dependent on spatial frequency, and edges blurred by different amounts may be perceived as different at one scale but equal at another. Additionally, there is potential for content-dependent blur perception where the same amount of blur is perceived differently, depending on the type(s) of object(s) shown.

We approach this problem by conducting a perceptual study to understand the relationship between the amount of blur present in an image, and the perception of blur at different image sizes. Our study determines how much blur must be present in a downsampled image to have the same appearance as the original. We find a complex and mostly image-independent relationship between matching blur levels in images at different resolutions. The relationship can be explained by a linear model when the blur magnitude is analyzed in terms of spatial frequency.

Using the results of this study, we develop a new image resizing operator that amplifies the blur present in the image while downsampling to ensure it is perceived the same as the original. While our algorithm is compatible with any combination of methods for producing a spatially-variant estimate of image blur and spatially-variant image filtering, we employ the blur estimation scheme presented in Chapter 4. The

result is a fully-automatic method for downsampling images while preserving the appearance of blur, the performance of which we verify with another user study.

5.2 Experiment Design

The basic premise of our work is that the blur in an image is perceived differently when that image is downsampled. In order to create a downsampled image that preserves the appearance of the original image, we must quantify that change in perception. The experiment was intended to measure the amount of blur that needs to be present in a thumbnail image in order to match the appearance of blur in a full-size version of the same image. This relationship was measured in a blur-matching experiment.

Observers were presented a full-size image, as well as two thumbnail images. They were asked to adjust the amount of blur in both thumbnail images such that the first thumbnail was just noticeably blurrier than the full-size image, and the second thumbnail was just noticeably sharper. An example stimulus is shown in Figure 5.1. We found that this 'bracketing' procedure resulted in more accurate measurements than direct matching and was necessary due to the relatively wide range of blur parameters that result in approximately equal appearance. Such variation of the method of adjustment was used before to measure a just noticeable blur in the context of the depth of focus of the eye [Yi et al., 2010] and the brightness of the glare illusion [Yoshida et al., 2008]. The matching blur amount was computed as the mean of the 'less-' and 'more-blurry' measurements.

An alternative experiment design that would produce more accurate results, is the 2-alternative-forced-choice procedure, in which the observers are asked to select a blurrier/sharper image when presented the original and downsampled version and the amount of blur is randomly added or removed from the smaller image. Such a procedure, although more accurate, consumes much more time (is on average 5–20 longer) and thus is not effective with a larger group of observers. The objective of our study was to gather data for an 'average' observer, thus it was more important to collect a larger number of measurements for a larger population, rather than fewer but more accurate measurements.

Figure 5.1: *Screen capture of the stimuli used in the experiment. Subjects adjust the blur in the small images on the right to match the blur in the large image on the left.*

Viewing conditions. The images were presented on a 27" Dell 2707WFPc display with 1920×1200 resolution. The experiment was run in a dimmed room with no visible display glare. The viewing distance was 1 m, resulting in a pixel Nyquist frequency of 30 cycles per visual degree.

Image selection. A pilot study was run to observe how the blur estimates differ between images, and in order to identify a possibly small set of images that would still reflect image-dependent effects. For the pilot experiment we selected 20 images containing people, faces, animals, man-made objects, indoor and outdoor scenes. The pilot experiment was run with 10 blur-levels and only 7 observers. The results were averaged for each test condition (blur level × downsampling level) to form a vector value. Then, the Euclidean distance was computed between vectors for each pair of images, to build a difference matrix. The difference data was then projected onto a 2D space using multi-dimensional scaling [Kruskal and Wish, 1978] in order to produce the plot in Figure 5.2. The plot reflects image-dependent differences in blur perception for the same blur parameters. To maximize diversity in image content, we selected five images which were located far apart on the plot and thus were likely to be the most different in terms of produced results.

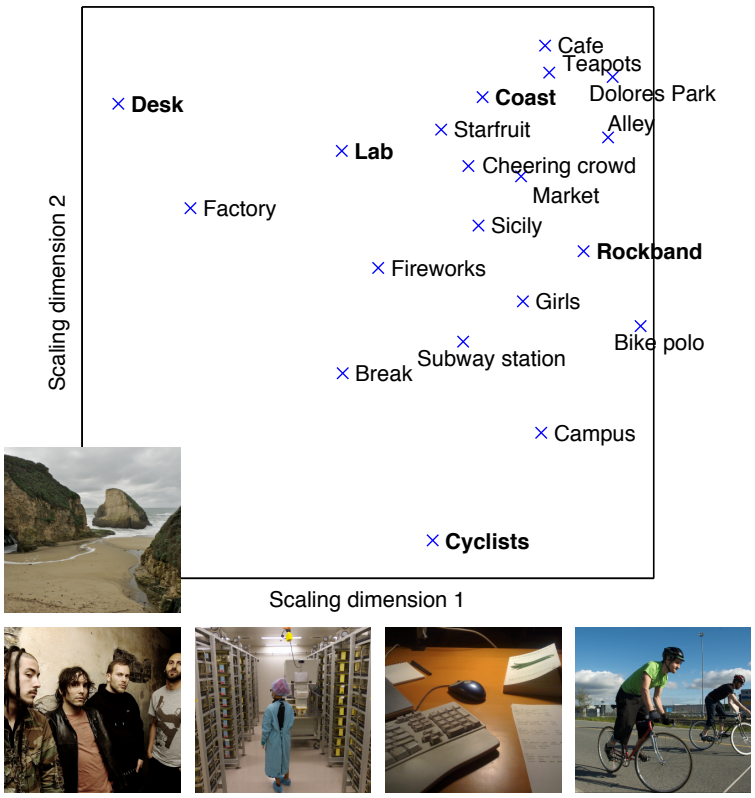


Figure 5.2: *The result of multi-dimensional scaling on the differences between per-image results collected in a pilot study. The pilot study was intended to identify the representative images that had the highest potential to reveal any image-dependencies in the study. The images selected for use in the main study are shown at the bottom.*

Stimuli. For both the pilot study and the full experiment, differently blurred versions of images were generated by introducing synthetic blur to full-size images with no noticeable blur in them. Since we could not control where in the image users were looking to make their judgements, we introduced uniform blur to completely in-focus images to avoid any ambiguities in response. For this purpose we used a Gaussian kernel of a specified standard deviation ς (in this chapter, we use ς to denote standard deviations of blur kernels expressed in visual degrees, and σ for blur kernels expressed in pixels). Thumbnail images were produced by the same process, except that the convolution of a full-size image was followed by nearest-neighbor

resampling. We chose a nearest neighbor filter for this step in order to not distort the experiments by introducing additional low-pass filtering. However, as a result, some amount of aliasing was present for small blur kernels under large downsampling factors (also see Section 5.3). The reported ς-values are given in visual degrees to make them display-independent. The five selected images were shown at 10 blur levels, ranging from 0 do 0.26 visual degrees, and at three downsampling factors: 2, 4, and 8.

Observers. 24 observers (14 male and 10 female) participated in the study. They were paid and unaware of the purpose of the experiment. The observer age varied from 21 to 38 with the average 28. All observers had normal or corrected-to-normal vision.

Experimental procedure. Given a reference image with blur ς_r, the observers were asked to adjust the matching blur to be just noticeably stronger in one and just noticeably weaker in the other thumbnail image. Each observer repeated the measurement for each condition three times, but each observer was assigned a random subset of 30 out of 150 conditions to reduce workload (150 = 3 downsampling factors $\times$ 10 blur levels $\times$ 5 images). In total over 2,100 measurements were collected. The experiment was preceded with a training session during which no data were recorded, followed by three main sessions with voluntary breaks between them. The breaks were scheduled so that each session lasted less than 30 minutes.

5.3 Experiment Results

The results of the experiment, averaged over the five selected images and for each image separately, are shown in Figure 5.3. The results are very consistent regardless of the image content, but averaging across all images is necessary to reduce variability in the data. Because both ς values are reported with respect to the blur in the full-size image, the $y = x$ line (dashed black line in the plot) is equivalent to retaining the blur of the original image. For all data points the matching blur is larger than the blur in the original images (points above the dashed black line). This is because images look sharper after downsampling and they need to be additionally blurred to

match the appearance of full-size images.

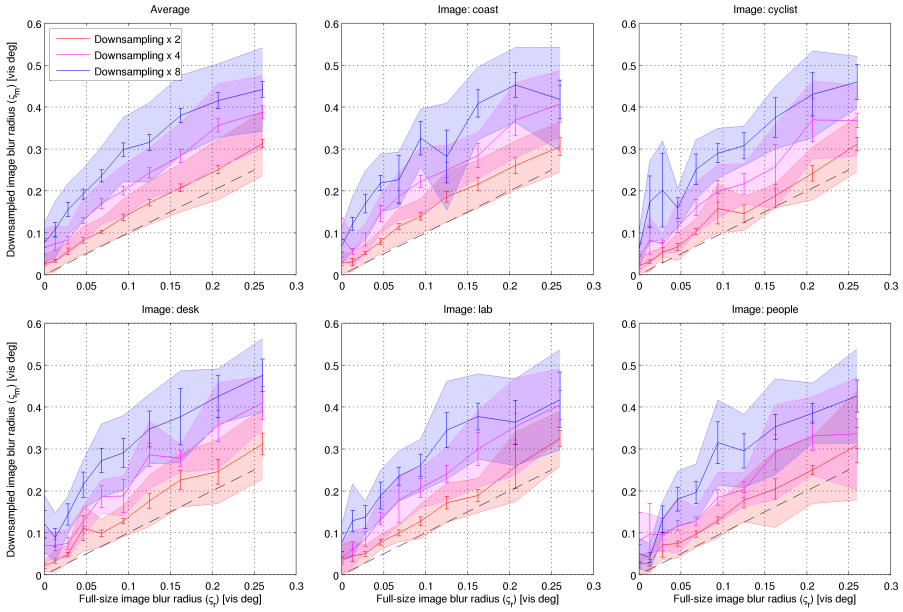


Figure 5.3: *The results of the blur matching experiments, plotted separately for averaged data ς_m (top-left) and for each individual image. The continuous lines are the expected magnitudes of matching blurs found by computing the average between two measurements for 'more blurry' and 'less blurry'. The error bars represent a 95% confidence interval. The edges of the shaded region correspond to the mean measurement for 'more blurry' and 'less blurry'.*

The experimental curves also level off at higher downsampling levels and for larger blur amounts. This effect is easy to explain after inspecting actual images, in which the amount of blur is so large that it sufficiently conveys the appearance of the full-size image and no additional blurring is needed.

It is important to note that the reported values also include the blurring necessary to remove aliasing artifacts. As mentioned in the previous section, we used a simple nearest-neighbor filter to resample the blurred high-resolution images so that the results are not confounded with an anti-aliasing filter. If the blur was not sufficient to prevent aliasing in the downsampled image, the result appeared sharper than the original. We observed that when no blur was present in the large image, subjects adjusted the amount of blur in the thumbnail to a value close to the optimal low pass filter for the given downsampling factor.

5.4 Model for Matching Blur Appearance

In this section we introduce a model that can predict our experimental results. The plot curves in Figure 5.3 suggest a non-linear relation for matching blur in original and downsampled images. However, we show that the averaged measured data ς_m is well explained by the combination of an anti-aliasing filter ς_d and a model $\mathscr{S}$, which is linear in *spatial* frequencies (measured in cycles per visual degree):

$$\widehat{\varsigma}_m = \sqrt{\varsigma_d^2 + \mathscr{S}^2}. \tag{5.1}$$

The $\widehat{\varsigma}_m$ is the model prediction of the experimental blur-matching data for an average observer. The term ς_d approximates the effect of an ideal anti-aliasing filter. The standard deviation ς_d of the Gaussian filter that provides a least-squares fit of the sinc filter is

$$\varsigma_d = d\frac{\sqrt{3\log 2}}{\pi \cdot p}, \tag{5.2}$$

where d is the downsampling factor, while p is a conversion factor that maps from image units (pixels) to visual degrees, which is equal to the number of pixels per visual degree. In our experiments, we had subjects sit further away from the screen than usual, to prevent limitations in screen resolution from affecting the results. As a result, $p = 60$ for our experimental data.

To motivate our choice of model, we remove the anti-aliasing component ς_d from the experimental data and plot it in terms of spatial frequency $1/\varsigma$ in Figure 5.4. The plot shows the experimental data expressed as the $\mathscr{S}$ component of Equation 5.1. All data points are now well aligned and mostly in linear relation, except several measurements at high frequencies and for the $2\times$ downsampling factor. We attribute these inaccuracies to the measurement error, which is magnified in this plot because of the $f = 1/\varsigma$ transform. The plot demonstrates that the remaining term $\mathscr{S}$ can be modeled as a set of straight lines when expressed in terms of spatial frequencies. Moreover, the lines cross at approximately the same point. The model that provides the best least-squares fit of the experimental data in terms of ς-values is

$$\mathcal{S}(\varsigma_r, d) = \frac{1}{2^{-0.893\log_2(d)+0.197}(\frac{1}{\varsigma_r} - 1.64) + 1.89}, \tag{5.3}$$

where d is the downsample amount and ς_r is the amount of the reference blur in the original image.

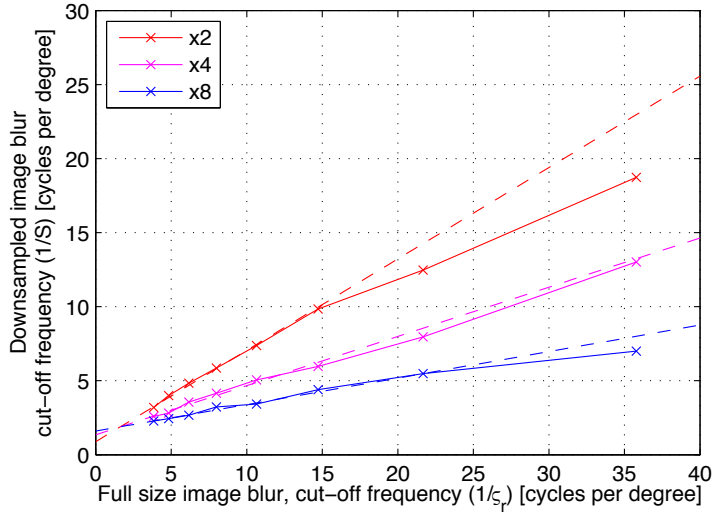


Figure 5.4: *Average matching blur data (Figure 5.3 first panel), with the anti-aliasing component ς_d removed, replotted as the roll-off frequency. The matching blur follows straight lines, except for the small blur amounts (high frequency roll-off), where aliasing dominates. The two lowest value ς_r points were omitted from the plot as the values were excessively large due to the $1/\varsigma$ transform.*

Figure 5.5 plots the combined blur model $\widehat{\varsigma}_m$ as compared to the results from our experiments. The figure shows that the fitting error is quite acceptable, even for low-ς (high-frequency) points, which did not follow the linear relation in Figure 5.4. When comparing plots, note that the large frequency values correspond to small ς-values. While a higher-order function could provide a better fit, our experimental data do not provide enough evidence to justify such a step. Moreover, we believe that a linear model in terms of spatial frequency is more plausible than a higher-order function.

Note that the combined model of matching blur $\widehat{\varsigma}_m$ is the absolute amount of blur that needs to be present in the full-size image before downsampling and is

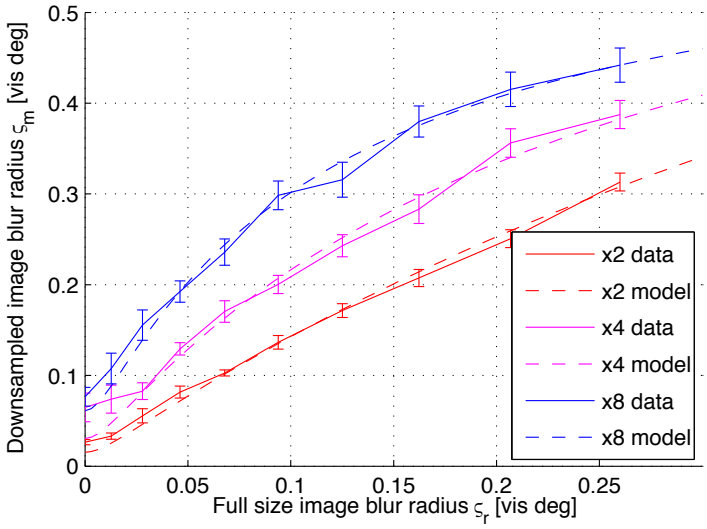


Figure 5.5: *Blur model $\widehat{\varsigma}_m$ (dashed lines) compared with the experiment results ς_m (continuous lines and error bars).*

expressed in units of visual degrees. In Section 5.5 we explain how to compute the amount of blur that needs to be added to a downsampled image.

5.5 Perceptually Accurate Blur Synthesis

The goal of our algorithm is to use the results of our experiment to automatically produce a downsampled image that preserves the appearance of the original blur. We first compute a spatially-varying estimate of the amount of blur present in the full-size image. Given that estimate, we use the results of our study to determine how much additional blur is needed for the specified downsample. Finally, we synthesize a new downsampled image with the amount of blur required to preserve the appearance of the image.

Our overall approach can work with any method that provides a spatially variant estimate of image blur. We considered the method by Elder and Zucker [1998], but it only produces estimates at edge locations and requires the work of Bae and Durand [2007] to provide a robust estimate of the blur at all pixels. While the approach produces high-quality results, it operates at the resolution of the original image and is computationally intensive. Instead, we rely on the algorithm presented

in Section 4.3 because of its simplicity and computational efficiency, which is a result of performing most work at the resolution of the downsampled image. The method provides blur estimates at each pixel location in terms of the Gaussian kernel σ before synthesizing the final image using our model of perceived blur. To synthesize the final image, we employ the same process as Section 4.5, replacing the model of camera optics with the perceptual model derived in Section 5.4.

While our method recovers a blur map for the image, it is worth noting that this blur map does not necessarily represent the defocus or depth of pixels. It can better be understood as a map of relative gradient magnitude per image region. While rapidly changing derivatives correspond to sharp regions, there is an ambiguity with slowly changing derivatives. If an image region is a flat color, we cannot determine whether that is a detailed region that is out of focus or it is in focus but lacks any detail. Both situations are equivalent from the viewpoint of this algorithm.

With an accurate estimate of the blur present at each pixel of the large image, we use our model from Section 5.4 to compute the amount of blur desired in the downsampled image. To produce the appearance-matching image, we reduce its resolution by downsampling it by a factor of d using the standard technique with an antialiasing filter. Because the anti-aliasing is now accounted for, we use the aliasing-free component of the model $\mathscr{S}(\varsigma_r, d)$ from Equation 5.3, rather than the complete model $\widehat{\varsigma}_m$. Given the existing blur in the full-size image σ_o, the amount of blur that needs to be added to a *downsampled* image is expressed as

$$\sigma_a = \sqrt{\left(\frac{\mathscr{S}(\sigma_o \cdot p^{-1}, d) \cdot p}{d}\right)^2 - \sigma_o^2}. \tag{5.4}$$

The downsampling factor d reduces the blur amount as we work on a lower-resolution downsampled image. The conversion factor p, which is equal to the number of pixels per one visual degree, converts visual degrees used in the model to pixels used in the blur estimation. For a computer monitor seen from a typical distance, p is approximately $30\,pix/deg$.

To produce the final image, for each level of the scale-space σ_j we blur the downsampled image by the corresponding amount of additional σ_a, then linearly blend sequential pairs of those blurred images together to approximate non-integer values of σ_a. While more accurate spatially-variant blur synthesis is possible, such as

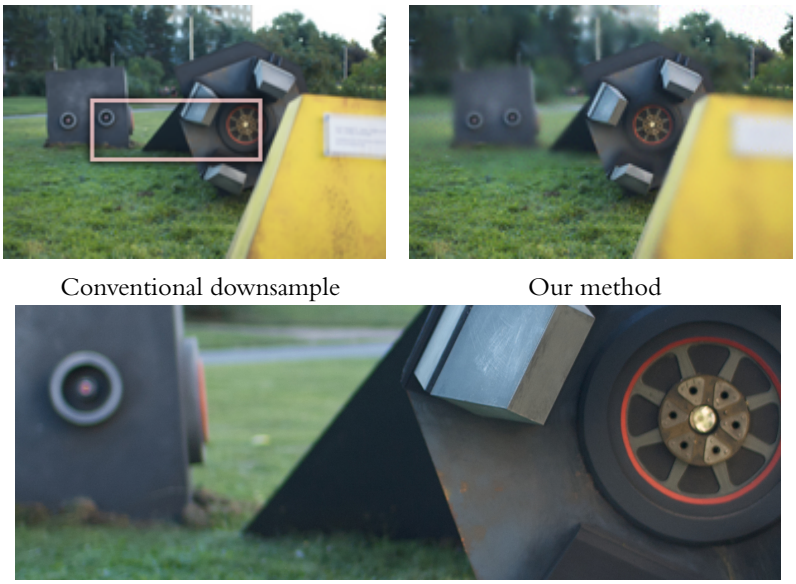


Figure 5.6: *Comparison of conventional downsample and our method. The bottom row contains cropped portions of the image at the original resolution (see pink boxes in conventional thumbnail). Note the blur present in the eye of the robot sculpture and cardboard box is visible in our result, but appears sharp in the conventional thumbnail.*

Popkin et al. [2010], we haven't noticed any artifacts requiring such methods.

5.6 Evaluation

In this section, we provide results of our method and compare our approach to that of Samadani et al. [2010]. We encourage the reader to look at the electronic versions of the images, which represent the fine details better than prints.

Figures 5.6 and 5.7 compares the results of our algorithm to those of a conventional downsampling method of low-pass filtering the image followed by nearest-neighbor sampling. In both the example of the robot sculpture and the art supplies, objects that appear in focus (such as the head of the robot or cardboard box) in the conventionally-downsampled image are in fact blurry, as can be seen in the zoomed portions. Our method accurately detects this blur and preserves the appearance in the downsampled image.

Figure 5.7: *Comparison of conventional downsample and our method. The bottom row contains cropped portions of the image at the original resolution (see pink boxes in conventional thumbnail).*

Figure 5.8 demonstrates the effectiveness of our algorithm at preserving the appearance of blur in images across multiple downsample factors. In this image, the original images are downsampled by a factor of 2 and 4, and the smaller versions retain the same impression of the depth of field.

Figure 5.9 compares the results of our method to those of the original method of Samadani et al. [2010]. If the value of γ is manually chosen for the image, their method can approximate our own. However, if the value of γ is incorrectly chosen, their method will either introduce too much blur and remove detail from the branches in the upper left or not introduce enough blur and retain all the details in the flowers. Even with a correctly chosen value of γ their method can only linearly scale the amount of blur, and cannot model the more complex relationship between existing blur and desired blur observed in the user study.

Additionally, we conducted a second user study to verify the effectiveness of our method. Previously, Samadani et al. [2010] performed a preference study to

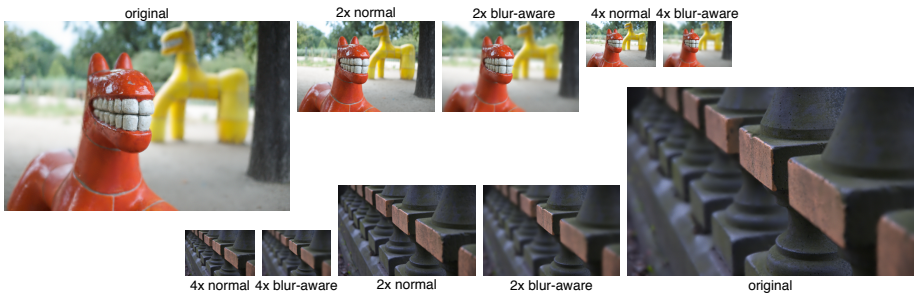


Figure 5.8: *Comparison of appearance of blur at multiple downsample levels. All of our results retain roughly the same amount of blur as the original while the conventionally downsampled appear to get progressively sharper.*

determine whether subjects felt their method was more representative of the original image than a conventional thumbnail. This study showed that users did prefer the method of Samadani *et al.* over standard thumbnails. We instead chose to conduct a task-based survey to determine the extent to which our method improves users' ability to make accurate comparisons of how objects in a scene are blurred.

In this 2-alternative-forced-choice study, we photographed a series of objects with increasing amounts of defocus blur. Thumbnail versions of these images were created using both our algorithm and the conventional downsample process to downsample by a factor of 8. Subjects were shown pairs of images with different amounts of defocus blur and asked to specify in which of the thumbnails the object appeared sharper. Figure 5.10 contains an example stimuli. A total of 5 observers participated in the study, performing a total of 240 trials for each of the downsampling algorithms.

Overall, subjects correctly identified the sharper object 67% of the time when viewing conventional thumbnails, while they correctly identified the sharper object 83% of the time when viewing the results of our method.

Our method outperforms conventional downsampling when the blur is small enough that the object will appear sharp in a standard thumbnail and blurred in our result. However, both methods exhibit similar performance if the blur is small enough for the object to appear sharp at the original resolution, and thus in both thumbnail versions. Likewise, the performance of the two methods will be the same if the blur is large enough that the object will appear blurred in both thumbnails.

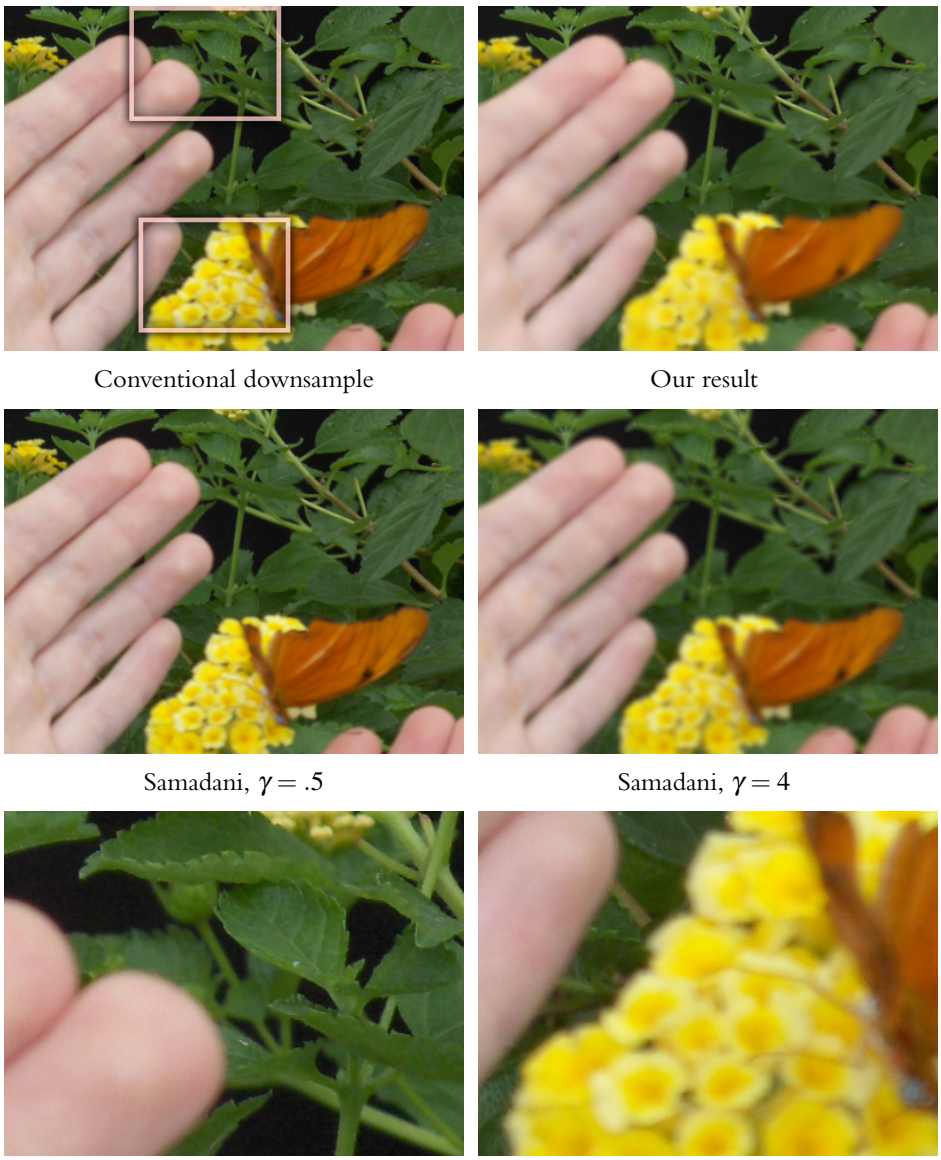


Figure 5.9: *Comparison of naive downsampling and our method (top row) to Samadani et al. with too little blur ($\gamma = .5$) and too much blur ($\gamma = 4$) from incorrect choices of γ. The bottom row contains a cropped region of the original image (see pink boxes) for comparison. Original image copyright Ramin Samadani.*

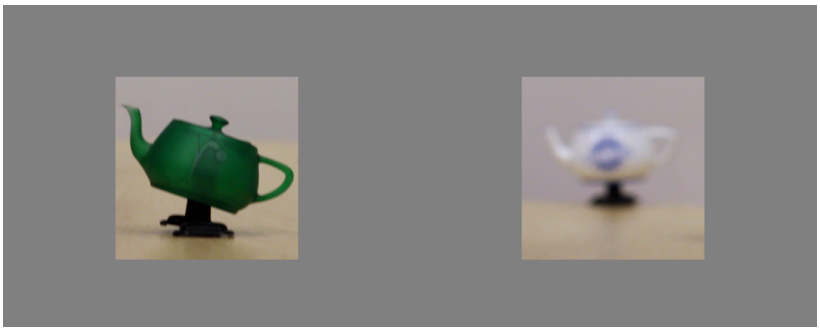


Figure 5.10: *A screenshot of the verification study containing a pair of thumbnails with different amounts of defocus blur. Subjects had to choose which image contained the object more in focus.*

We used a uniform distribution of blur amounts, so our experiment covers all three of these cases.

5.7 Conclusion

In this chapter, we have presented a perceptually-based model of how the perception of blur in an image changes as the size of that image is reduced. This model is based on a linear relationship between the perceived blur magnitude and the blur present in the image, when analyzed in terms of spatial frequency.

We have used that model to create a new image-resizing operator that preserves the perception of blur in images as they are downsampled, ensuring that the new image appears the same as the original. To do so, we modified an existing blur estimation algorithm by Samadani et al. [2010] to provide estimates of the original image in absolute units.

More generally, any form of downsampling involves discarding information present in the image. The convention in graphics and image processing is to attempt to produce the highest quality result, which usually involves throwing away higher frequency detail to avoid any aliasing artifacts. Due to the disparity between sensor resolution and display resolution, users often view images and make image assessments based on lower-resolution versions that might not represent their full-size counterparts. We have proposed an approach that considers how the image is perceived, and preserves that appearance rather than producing a higher-quality but less

representative result.

Chapter 6

Scale-Dependent Perception of Countershading

The previous chapter discussed how the perception of image blur changes with the size at which the image is presented. This chapter investigates similar changes in the appearance of edge profiles of different scales. Countershading is a double-edged sword: while correctly chosen parameters for a given viewing condition can significantly improve the image sharpness or trick the human visual system into perceiving a higher contrast than physically present in an image, wrong parameters, or different viewing conditions can result in objectionable halo artifacts.

In this chapter we analyze the circumstances under which countershading turns from image enhancement to artifact for a range of parameter settings and viewing conditions. Our experimental results can be modeled as a function of the width of the countershading profile. We employ this empirical function in a range of applications such as image resizing, view dependent tone mapping, and countershading analysis in photographs and works of fine art.

6.1 Introduction

Local contrast enhancement is a powerful image processing technique, fundamental to many aspects of computer graphics such as image editing and tonemapping of HDR images. The same basic techniques can be applied to improve the recognition

of objects in a scene, to aid in identifying brightness of regions or to accentuate specific image details. In many cases, images with enhanced contrast appear more aesthetically pleasing.

One of the most common approaches to enhancing local contrast in images is countershading, where the local edge contrast is increased by adding gradients to either side of the edges. This approach is common across numerous classes of algorithms. Many of these algorithms, either explicitly or implicitly, resemble the effect of the unsharp masking (UM) operator.

This simple technique has proven to be incredibly versatile and, depending on the choice of blur employed in the high-pass filter, can produce a variety of effects. Unsharp masking with a narrow high-pass filter can increase the acutance, also known as apparent sharpness, of the image, making fine details easier to identify. On the other hand, unsharp masking with a wide high-pass filter can increase the contrast of the regions adjacent to the edge, altering the overall impression of contrast in the image. Sufficiently wide unsharp masking profiles can even induce the Cornsweet illusion, where the entirety of adjacent regions change in appearance. However, unsharp masking can also introduce objectionable countershading around an edge, frequently referred to as haloes. In these cases, the contrast enhancement detracts from the image, providing neither improved understanding nor aesthetic quality.

We desire a better understanding of what causes local contrast enhancement to be considered objectionable. Our inquiry begins from the simple observation that the same basic operation of unsharp masking, and countershading in general, can lead to multiple, disagreeing interpretations of its effect on the image. In some cases countershading is interpreted as an enhancement, while in others it is interpreted as an artifact. We investigate the acceptable contrast for different width countershading profiles and conduct a perceptual study to determine the amount of countershading that can be introduced at different scales without it becoming objectionable.

We discover an "uncanny valley" of countershading profiles, where certain width profiles are considered unacceptable even if only slightly visible, separating adjacent regions of both wider and narrower countershading profiles with considerably higher levels of tolerated contrast. Figure 6.1 illustrates how the regions of indistinguishable and objectionable countershading vary with the width of the countershading profile and its amplitude. The important observation is that countershading that is indistin-

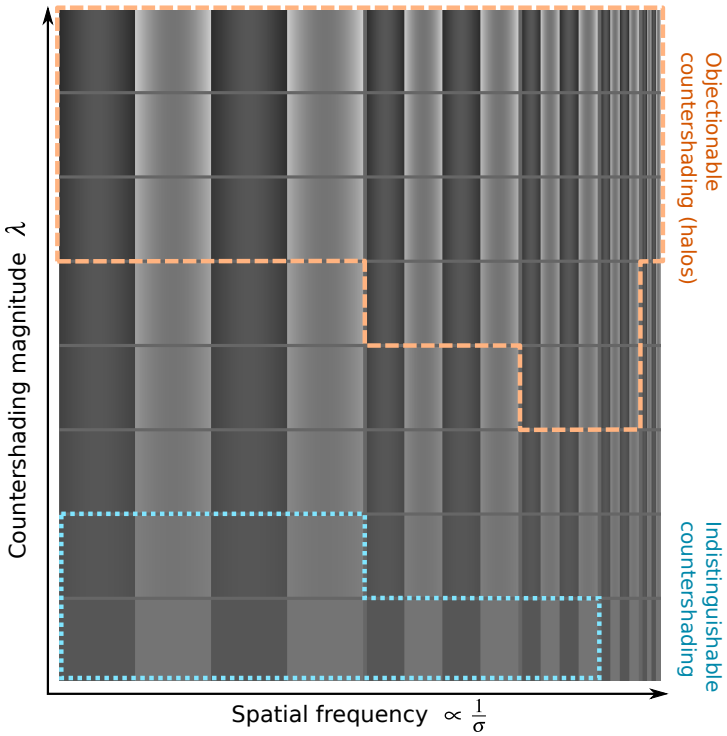


Figure 6.1: *The square-wave enhanced by a countershading profiles. The regions of indistinguishable (from a step edge) and objectionable countershading are marked with dotted and dashed lines of different color. The higher magnitude of countershading produces higher contrast edges. But if it is too high, the result appears objectionable. The marked regions are approximate and for illustration and actual regions will depend on the angular resolution of the figure.*

guishable from a plain edge provides only very limited contrast enhancement; thus the region below objectionable countershading needs to be used to achieve good quality results. We find that these regions correspond to the various semantic descriptions of the images resulting from different unsharp masking parameters. The "valley" consists of profiles considered haloes, while narrower profiles were perceived to sharpen image features and wider profiles were perceived to enhance contrast.

The remainder of the chapter is organized as follows. In the next section, we discuss unsharp masking and various applications of countershading in graphics and image processing. Section 6.2 describes the design of the perceptual experiment

we conducted, while Section 6.3 discusses its results. In Section 6.4, we relate our measurements to other studies on the perception of countershading profiles. Finally, we demonstrate the applicability of our results and the understanding they afford with a number of applications, including size-aware image resizing, control of tone-mapping parameters, artifact-free unsharp masking, detection of countershading in images, and a viewer-adaptive display.

6.2 Experiment Design

In our perceptual experiment, we target the most general case of countershading operations: determining the magnitude at which the countershading profile of a given width becomes objectionable. In designing our study, we take inspiration from work by Ciuffreda et al. [2006] on determining the level of "bothersome blur." While parameters of σ and λ responsible for the shape of the countershading profile smoothly vary over the space of possible values, a semantic shift from "enhancement" to "artifact" occurs along a boundary within that space. Our goal is to determine the boundary between the region of enhancements and the region of artifacts within the parameter space of the unsharp masking operator. Our study does not attempt to determine any of the other aspects of local contrast perception such as local contrast appearance, detection thresholds of countershading, or user preference.

View setup. The images were presented on a 26" NEC LCD2690WUXi display with 1920×1200 resolution with a black level of .45 cd/m^2 and a peak intensity of 213 cd/m^2. The experiment was run in a darkened room with no visible display glare. The viewing distance was 1 m, resulting in a pixel Nyquist frequency of 30 cycles per visual degree.

Stimuli. The study consisted of six images, three test patterns of a single step edge of various contrasts, and three complex scenes, as shown in Figure 6.2. The contrast of images of complex scenes was linearly scaled down to retain headroom for the countershading without saturating pixels. The process of adding countershading enhancement to an image is illustrated in Figure 6.3. Given a linear luminance image, the countershading is applied in the logarithmic domain to produce the profiles with

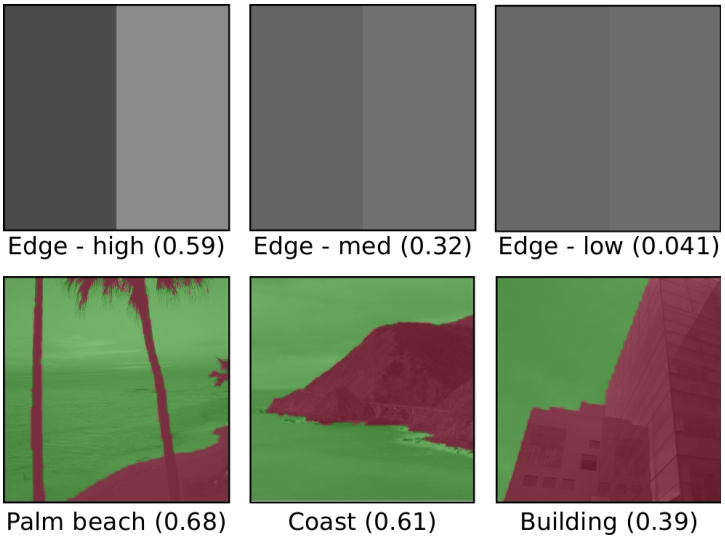


Figure 6.2: *Images used in the perceptual experiment: 3 edges of different contrast and 3 images of complex scenes. The semi-transparent red-green color mask is the edge template. The numbers in parenthesis denote the Michelson contrast of the edge.*

the most symmetric appearance of lightness. To produce a high pass image for the enhancement, a Gaussian-filtered image is subtracted from an original. However, the enhancement is computed based on an edge-template, as opposed to using the original image. The template image contains the edges to be enhanced with smooth regions in between, and can be produced by an edge-preserving filter such as that of Farbman et al. [2008]. Using an edge-template ensures a constant increase in contrast along the edge and, more importantly, avoids the amplification of high frequency detail noted by Neycenssac [1993] which could distract subjects from evaluating the appearance of the countershading profiles. Ten countershading profile widths were used in the trials, ranging from .009 to 4.6 visual degrees, equivalent to $.5 - 256$ px at 1 m viewing distance, increasing by factors of two.

Observers. 15 observers (9 male and 6 female) participated in the study. They were paid and unaware of the purpose of the experiment. The observer age varied from 23 to 37 with the average 29. All observers had normal or corrected-to-normal vision.

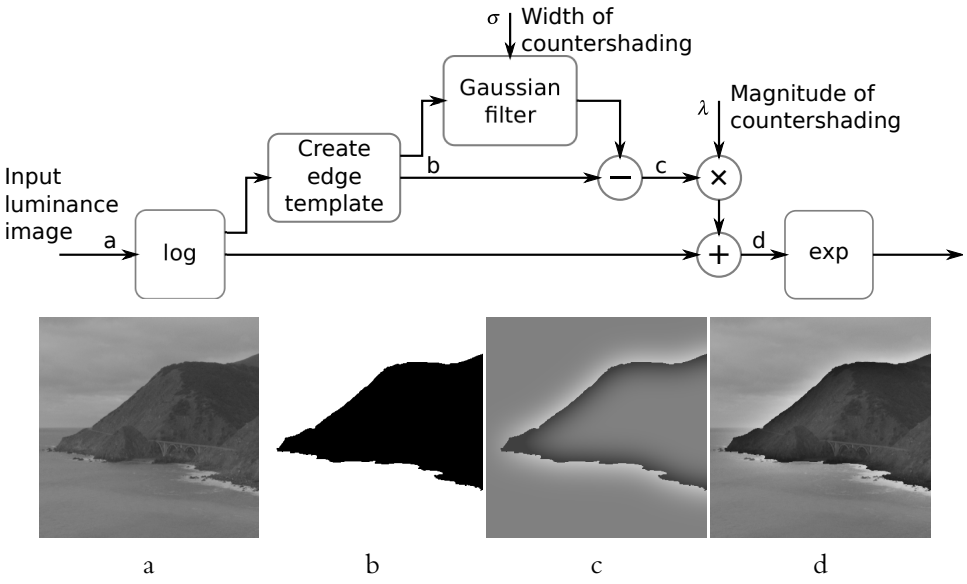


Figure 6.3: *Generation of the countershading enhancement in an image. An edge template is used to precisely mark selected edges in the experiment, but is normally generated with an edge-preserving filter.*

Procedure. After being presented an image with a countershading profile of width σ, the observers were asked to adjust the magnitude λ of the countershading to the maximum level not considered an artifact. Both "artifact" and "objectionable" are subjective terms, and we relied on a no-reference measure of the artifacts, where if the subject saw the image without knowing the original, they would say that it contained undesirable countershading. Each observer repeated the measurement for each of the 60 conditions (6 images $\times$ 10 profile widths) twice, for a total of 120 trials each. In total over 1700 measurements were collected. The experiment was preceded with a training session to familiarize participants with the task. No data were recorded during the training phase, which was followed by three main sessions with voluntary breaks between them. The breaks were scheduled so that each session lasted less than 20 minutes.

Screening and outlier removal. Because of the subjective nature of the experiment, erroneous and inconsistent measurements are likely to be found in the data. To remove

erroneous measurements, we firstly screened the participants using the procedure recommended for magnitude estimation experiments ITU-R-BT.500-11 [2002, Sec. 2.3.1], then we eliminated those measurements for which the intra-observer standard deviation exceeded two times the mean standard deviation of all the measurements. The screening eliminated the data of two participants and the outlier removal removed 9% of the measurements.

6.3 Experimental Results

To illustrate individual variations, the data for individual observers is shown in Figure 6.4. The main inter-observer variation is visible as the vertical shift of curves on the plot, suggesting that different individuals have different notions of 'just-objectionable' countershading. The data becomes more consistent once the differences in the mean values are compensated, as shown on the right of Figure 6.4. For the remaining considerations we use the data averaged over all observers. But the strong inter-observer variation indicates that the 'objectionable' level is subjective and the algorithms may choose to include a user-defined scaling parameter for our mean-observer data.

Figure 6.5 shows the results averaged over all observers.The most salient feature of all plots is the U-shaped characteristic, indicating a reduced tolerance to the halo effect for the medium profile widths, with the trough around 0.2 visual degrees.

One of the most interesting observations is the difference in the characteristic between an isolated-edge and complex images, which is shown as the difference between the plots in Figures 6.5a and 6.5b. At large countershading widths a larger magnitude of enhancement was selected for a complex image than for an isolated edge. This could be due to visual masking, which was present in complex images, but not in the case of isolated edges. The difference between the two plots suggests that the measurements for simplified stimuli do not generalize to complex images.

The results for individual images are better aligned if the countershading profile is generated irrespectively of the contrast of the underlying edge. The λ-values on the plots represent the magnitude of the countershading generated from an edge template (refer to Figure 6.3) with a fixed $\log_{10}$-contrast of 1, rather than the contrast of an edge. This observation suggests that profiles of the same magnitude should be

used regardless of the contrast of the underlying edge.

The remaining variation between images can be explained by the small effect of the underlying edge contrast. The effect is mostly visible for complex images and narrow countershading profiles. The data, however, is not sufficiently accurate to model this effect. Moreover, the effect disappears for large σ-values, which are the most relevant for an effective contrast enhancement.

For convenience of use, we fit a polynomial function with a linear segment to the values averaged across all complex images. The λ-values for the just-objectionable countershading can be found from:

$$\lambda = \begin{cases} -.249\varsigma^3 - .233\varsigma^2 + .377\varsigma + .674 & \text{if } \varsigma \leq .418 \\ .048\varsigma + .752 & \text{if } \varsigma > .418 \end{cases} \tag{6.1}$$

where $\varsigma = \log_{10}(\sigma)$. The model-fit is shown as a black continuous line in Figure 6.5b.

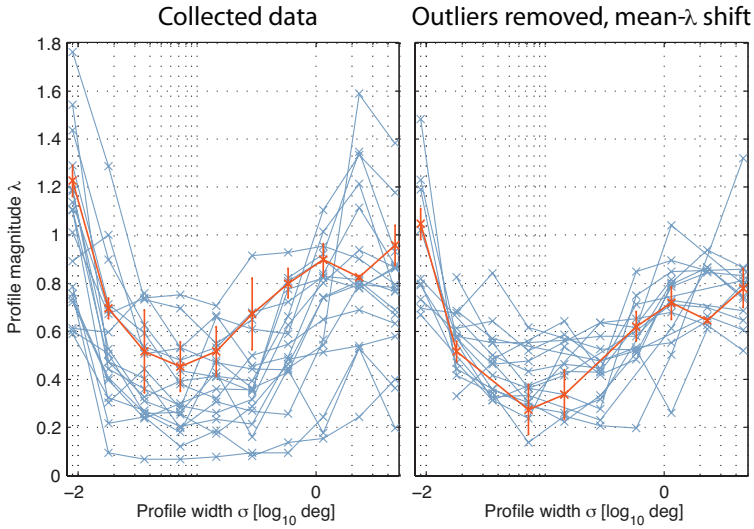
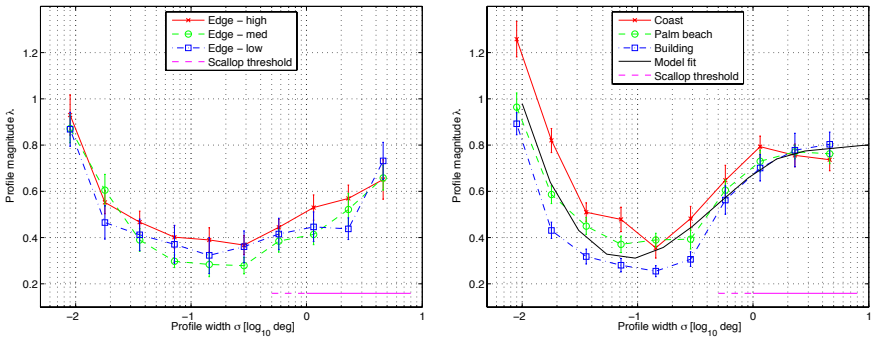
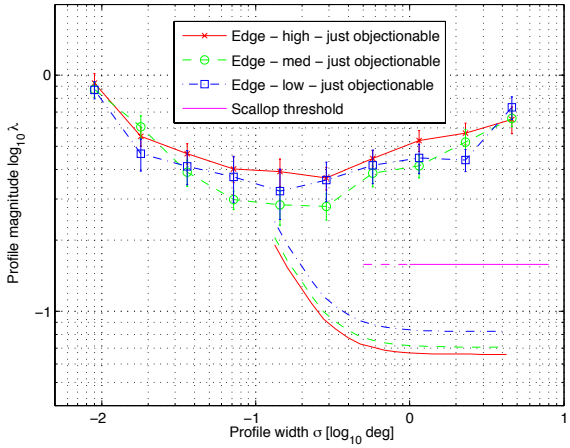


Figure 6.4: Left: *Results for individual participants for the* Palm beach *image.* **Right:** *The same results but after outlier removal and compensating for the mean λ-value between the participants. Standard error of the mean is plotted for a single participant for better clarity (in orange).*



(a) *Edges* (b) *Complex images*



(c) *Comparison with Krawczyk et al. [2007]*

Figure 6.5: *The results averaged over all participants,* ***(a)*** *for isolated edges, and* ***(b)*** *for for complex images. The black line in* ***(b)*** *represents our model fit.* ***(c)*** *Our just-objectionable measurements (edges) compared to the just-distinguishable thresholds from Krawczyk et al. [2007] (lines with no markers).* $\log_{10}$ *scale was used for* λ *in this plot for better visualization.*

6.4 Discussion and Relation to Other Studies

The perceived effects of countershading and local contrast enhancement are well-studied. Work includes studies related to the sensitivity of the human visual system to local contrast enhancement, as well as the amount preferred by subjects under various conditions. We review a selection of key work and discuss its relation to our own findings.

The studies on the Cornsweet illusion by Sullivan and Georgeson [1977], Campbell et al. [1978, 1971] and Burr [1987] determined the contrast at which a countershaded edge becomes distinguishable from a step edge of equivalent contrast, the so called scalloping threshold. The horizontal magenta lines in Figures 6.5a-6.5c show the approximated values of the scalloping threshold, after Kingdom and Moulden [1988]. The thresholds apply only to larger σ-values because the countershading is always distinguishable from a square wave at spatial frequencies above 1 c/deg [Sullivan and Georgeson, 1977], which approximately corresponds to $\sigma < 1$ for our Gaussian-based profiles. Very low scalloping thresholds clearly show that the countershading is likely to be noticeable for most practical cases; and thus the just-objectionable threshold is more relevant for contrast enhancement applications. For a more complete discussion of scalloping, see Kingdom and Moulden [1988] for a comprehensive review.

Lin et al. [2006] studied the perceptual impact of edge sharpness. In their study, they processed images using an unsharp masking filter with a fixed $\sigma = 1.4$ px at edge locations only to avoid noise amplification. This width equates to .04 spatial degrees in their viewing setup, roughly corresponding to the third of our tested widths. The filter was applied to all edges in the image, and the amount of countershading introduced was proportional to the underlying edge contrast. Subjects ranked the perceived quality of images for contrasts of different magnitudes and Lin et al. computed the most desirable and highest tolerated contrast. Without knowing the contrast of the original image edges, we cannot compute the equivalent λ to plot a direct comparison.

Based on the theoretical model of countershading perception by Dooley and Greenfield [1977], Krawczyk et al. [2007] proposed a visual model of just-detectable countershading, which was used to adaptively introduce countershading in images. The main assumption behind their algorithm is that the countershading becomes objectionable as soon as the profiles become visible. The comparison of our data with the scalloping thresholds in Figures 6.5a and 6.5b demonstrates that it is not the case. We reproduced their model and computed its predictions for the isolated edges from our experiment, assuming the t.v.i. value equal to 5% and no masking. The model predictions, plotted in Figure 6.5c, show little correlation with our experiment results and seem to be too conservative even for the scalloping threshold

reported in the literature.

Ihrke et al. [2009] performed a perceptual evaluation of the work on 3D unsharp masking by Ritschel et al. [2008]. The authors tested the preferred value of their magnitude parameter λ for profiles of several different widths, measured on object surfaces, opposed to image-space. Interestingly, they did not find their results varied with the width of the countershading profiles. However, they only tested images of complex scenes with a narrow set of profiles (corresponding to the trough of the valley we find). Additionally, the depth-dependency of their algorithm means that profiles of varying widths were added to their scenes, making it hard to ascertain the magnitude level associated with a single countershading width. Their largest σ was the only width large enough to cause the Cornsweet illusion. Similar to Lin et al., it is difficult for us to plot a direct comparison to our findings because λ and σ vary across the image.

6.5 Applications

We have implemented a number of simple applications of our model of objectionable countershading for use in adjusting image scale and contrast, HDR tone mapping, artifact-free unsharp masking, and countershading profile analysis. In the rest of this section, we briefly describe these methods and their results. The purpose of these examples is to demonstrate the breadth of topics for which our measurements are relevant. Any of these tools could be made more sophisticated, but are sufficient to demonstrate the associated approach.

As noted in Section 6.4, the perception of countershading profiles strongly depends on the spatial frequency at which they appear. All the images assume viewing of this thesis as printed on letter-size paper (8.5"×11" / 11.6×27.9 cm) page viewed from a distance of 18" (45.7 cm). When evaluating the images, please ensure you are viewing under similar conditions. Additionally, some of the effects are subtle and we suggest looking at the electronic copy of the document.

6.5.1 Avoiding Haloes When Resizing

As shown in the Section 6.3, increasing the magnitude of a countershading profile can not only boost the perceived edge contrast, but also move the appearance into

the objectionable region. Similarly, operations that change the width of the profile can cause the same issue.

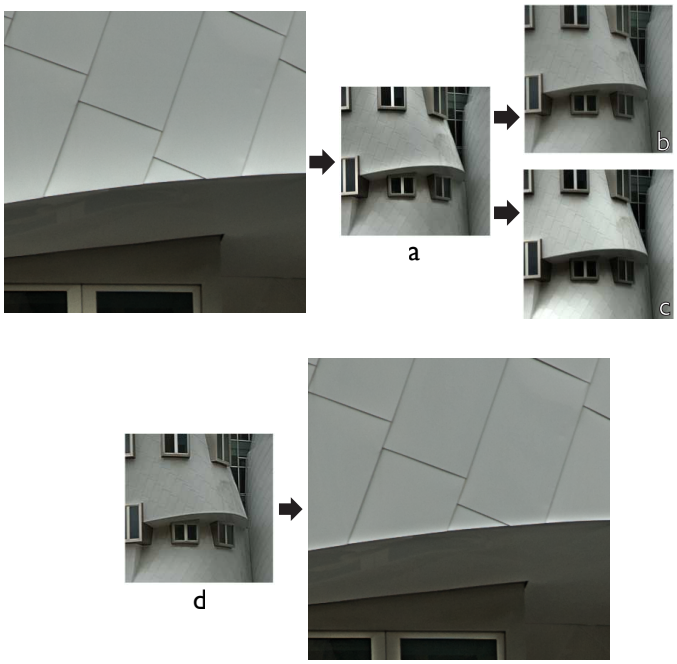


Figure 6.6: *Downsizing an image (a) transforms a 1.84° countershading profile to a smaller one, where the threshold of objectionable magnitude is significantly reduced, causing objectionable haloes. These artifacts can be remedied by adjusting the magnitude of the countershading (b) to correspond to the new angular size of the profiles or adjusting the width of the countershading (c) to compensate for the downsample factor, or some combination thereof. Neither image in this simple case appears exactly the same as the full-size, but they do not include objectionable artifacts. The same issue occurs in the opposite case, when a sharpened image is enlarged (d).*

Resizing an image with acceptable countershaded edges can cause the profiles around those edges to move from the acceptable to objectionable region, as shown in Figure 6.6. Figure 6.6a shows an edge with 1.84° countershading profiles near the just-acceptable contrast. Shrinking that image by a factor of 10× causes those profiles to move into the objectionable region and appear as haloes. In order to regain acceptable profiles, the current combination of profile contrast and width must be projected back outside the objectionable region. Fig 6.6b and 6.6c show the result

of projecting along only the contrast axis or only the width axis, respectively. The resulting contrast appearance depends on the projection, where primarily adjusting contrast reduces the overall image contrast enhancement, while primarily adjusting width approximates a global contrast adjustment. Algorithms can choose how to adjust those parameters based on desired image appearance.

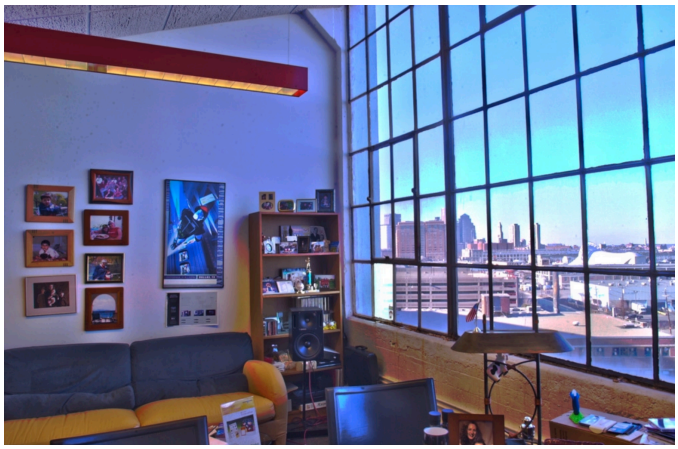
The same issue also must be considered when enlarging images. Contrasts acceptable for very narrow profiles will become objectionable as they are widened, as shown in Figure 6.6d. While less frequent, this scenario can occur if a user is adjusting the contrast of a sharpening filter in a preview of less than 100% of the full image size.

6.5.2 Local Tonemapping Operators

The same consideration to the scale of countershading from the observer's perspective can be applied to other operations. Local tonemapping is frequently associated with introducing objectionable haloes. In some cases, poor algorithm performance, regardless of parameters, is responsible for the artifacts. In others, it may be a mismatch between the scale for which the image was produced and the scale at which it is being displayed.

We use the *Office* image from Durand and Dorsey [2002], which is 2000×1312 px in size. Their algorithm calls for the σ_s of the bilateral filter to be 2% of the image size, in this case 26 px. Viewing the image 1-to-1 pixels on a 26" 1920×1200 display from a distance of 2", σ_s is equivalent to 0.70°. We observe the artifacts to be acceptably low under these conditions. However, resizing that image to fit within the margins of this document at our 18" viewing distance, the same σ_s is equivalent to 0.28°, a spatial frequency at which we are considerably more sensitive to haloes. The contrast of the countershading profiles depends on the interaction of several algorithm parameters and is difficult to specify. On the other hand, the width of the profiles only depends on the choice σ_s in the bilateral filter, so we project back out of the objectionable region along the profile width axis.

After fixing the σ_s to be equivalent to 0.70° (66 px in this case) the haloes are below the objectionable threshold for the image in this thesis, as can be seen in Figure 6.7. We have not entirely removed the haloes from the image, as can be seen when viewing the image from a large distance. We have simply ensured the profile



(a)

(b)

Figure 6.7: *A comparison between* ***(a)*** *the* σ_s *specified in Durand and Dorsey [2002] and* ***(b)*** *our* σ_s *chosen for the size of the image in this document. The contrast of the countershading has not changed, the profiles have been sufficiently widened so the contrast is no longer objectionable. Office image copyright Fredo Durand.*

contrasts are below the objectionable threshold for the given viewing conditions. This approach provides an alternative to that of Farbman et al. [2008] when considering situations involving small displays or requiring computational efficiency, such as mobile devices.

6.5.3 Artifact-Free Unsharp Masking

We can also use the insight from our experimental procedure to develop improved unsharp masking that does not introduce artifacts. Neycenssac [1993] noted that unsharp masking introduces contrast through 2 separate mechanisms: the addition of countershading profiles around edges and the amplification of existing image features. An ideal countershading operator should introduce acceptable magnitudes of countershading at edges without introducing other artifacts. While the results derived from our experiment provide a model of acceptable countershading magnitudes, we still must address the amplification of image features.

A naïve unsharp masking operation amplifies small features, especially noise, as visible in Figure 6.8c. Lindeberg [1994a] notes that convolution with a Gaussian of width σ will remove all the features smaller than $\sqrt{\sigma}$, so those features, including noise, will be present in the high-pass image $\mathscr{H}_\sigma(Y)$ at their original contrast. When the high-pass image is added to the original image, only features larger than $\sqrt{\sigma}$ receive countershading profiles, while features smaller than $\sqrt{\sigma}$ are amplified. To achieve effective countershading, rather than detail amplification, it is necessary to suppress unwanted details in the high-pass image.

Many attempts have been made to adaptively scale the unsharp masking parameters λ and σ based on local image content to avoid the unwanted features present in $\mathscr{H}_\sigma(Y)$ from being introduced into the final image. Taking inspiration from the edge templates employed our experiment, we consider a different approach: if small features are amplified rather than countershaded, they should not be present in the high-pass image. That way the unsharp masking operator does not need to remove them after the fact.

We replace the conventional high-pass image $\mathscr{H}_\sigma(Y)$, based on the difference between the image and a Gaussian-blurred copy g_σ

$$\mathscr{H}_\sigma(Y) = Y - g_\sigma * Y \tag{6.2}$$

with a modified version $\mathscr{H}'_\sigma(Y)$ based on a template function T_σ that removes features smaller than the $\sqrt{\sigma}$ from the image while retaining high-frequency edges:

$$\mathscr{H}'_\sigma(Y) = T_\sigma(Y) - g_\sigma * T_\sigma(Y). \tag{6.3}$$

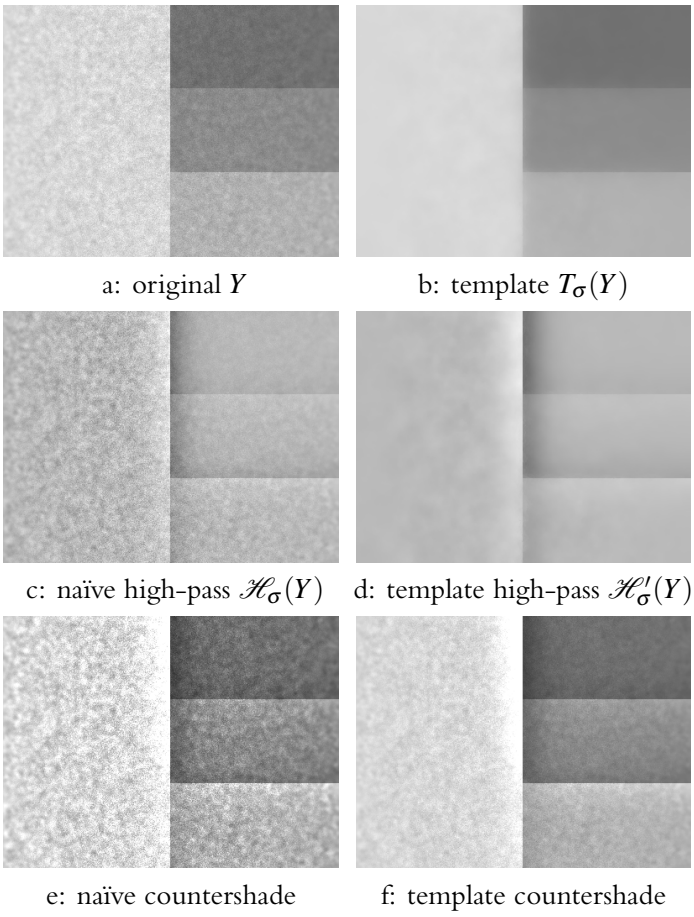


Figure 6.8: *Comparison of the original image Y with the result of the* WLS *filter $T_\sigma(Y)$ for the generated high-pass image and countershaded result. The naïve high-pass image retains the noise of the original and amplifies it in the result, while the template version does not. In this case the countershading magnitude has been chosen to make the difference between operators easily visible.*

Edge-preserving smoothing filters, especially the WLS framework of Farbman et al. [2008], provide a very good approximation of T_σ. In the case of Farbman et al., the parameters must be calibrated such that the frequency response of the filter corresponds to that of g_σ used by $\mathscr{H}$. The frequency response of a Gaussian filter of width σ is another Gaussian of width $1/\sigma$:

$$\mathscr{G}_\sigma(\omega) = e^{-\frac{\omega^2}{2/\sigma^2}} \tag{6.4}$$

From Farbman et al., the frequency response of WLS for a region without significant edges is

$$\mathscr{F}_{\gamma,\alpha}(\omega) = \frac{1}{1+\gamma\omega^2\left(|\partial\ell|^\alpha+\varepsilon\right)^{-1}}, \tag{6.5}$$

where γ controls the spatial extent of the function, α is the edge-stopping parameter, $\partial\ell$ is the average partial difference of the log-luminance of the input image and ε is a small term to avoid division by zero. We choose Farbman et al.'s value of $\alpha = 1.2$ and an average pixel difference of .04 to approximate the gradient magnitude of flat regions of images in the range $[0,1]$ and solve the γ that minimizes the least-squares difference between $\mathscr{G}_\sigma$ and $\mathscr{F}_\gamma$. For $\sigma = 1$, the equivalent value of $\gamma_1 = 0.027$, given our choice of parameters. From Farbman et al., values equivalent to other σ are determined by $\gamma_\sigma = \sigma^2\gamma_1$.

The corresponding WLS filter F_λ will remove features smaller than $\sqrt{\sigma}$ while preserving edges, the desired behavior of our template function T_σ. Figure 6.8 compares conventional unsharp masking to the template image approach, which successfully removes small details from the high-pass image. Finally, we use our model from Section 6.3 to ensure the added countershading is acceptable.

This operator can be viewed as the complement to using the WLS filter for multi-scale tone manipulation. Because WLS does not smooth across edges, the differences in contrast between scales only occur in smoothly-changing regions and WLS tone manipulation predominantly adjusts the contrast of regions without sharp edges. Conversely, our operator predominantly adjusts the contrast of regions around sharp edges.

6.5.4 Countershading Analysis

Many images already contain countershading, including some works of fine art such as Seurat's *Le Bec du Hoc*, shown in Figure 6.9. In fact, countershading originated in the fine arts, where artists overcame the limitations of the medium and gave images the appearance of higher contrast than they could otherwise convey. When this painting is seen at a different visual resolution, for example when reproduced on a web-page, the appearance of countershading may be very different from the

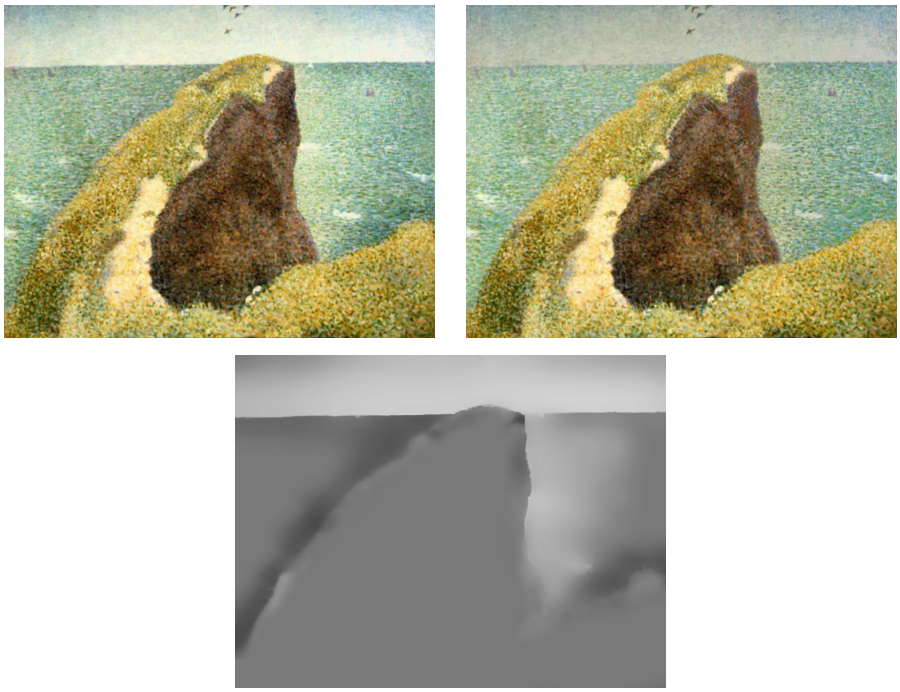


Figure 6.9: *Comparison of Seurat's original* Le Bec du Hoc *(left) with our adjusted version (right). We performed a multi-scale decomposition of the painting (left) and estimated the countershading profiles on each side of the edges separated the land, sea, and sky portions (bottom), then demonstrate the difference by synthesizing the new image (right) with those profiles removed.*

artist's intent. However, it is possible to compensate for the difference in viewing conditions by analyzing the countershading present and then reintroducing adjusted countershading profiles.

To accomplish this task we use an extreme case of image abstraction followed by an *analysis-by-synthesis* estimation of the countershading profiles. We observe that if a single iteration of an edge-preserving filter removes texture details, multiple iterations remove any low-amplitude intensity changes, including the countershading profiles. We employ the WLS framework of Farbman et al. [2008], using the iterative version of their algorithm to obtain a texture-free layer D_s, and then repeat several more iterations to obtain a countershading-free layer D_t. The layer D_t consists of nearly uniform regions of color separated by sharp edges like the template

images described earlier. We then segment the template image and solve for the Gaussian centered at each edge position that best approximates the countershading present: $D_t - D_s$. Estimation is performed independently for each side of the edge to account for asymmetries in the profiles. The result is an approximation of profile width and magnitude at each edge location, shown in Figure 6.9 (right). Given this representation, we are able to remove existing countershading and synthesize a new set of profiles for the desired conditions. This approach not only provides a better means to reproduce artwork containing countershading, but also may afford a deeper insight into how it is used artistically.

6.5.5 Viewer-Adaptive Display

The underlying theme of these applications is that the perception of countershading strongly depends on the width of the profile from the point of the viewer. Algorithms can easily account for the dimensions and resolution of the display, but fail to account for the fact that the perceived profile width also depends on the distance between observer and display. We created a setup with a viewer-adaptive display, which determines the distance of a viewer from the screen using head-tracking, and then adjusts countershading profiles accordingly (details in video at http://matttrent.com/research/thesis). The goal is to maximize contrast enhancement without causing haloes to appear, especially at larger viewing distances (refer to Figure 6.5). We can achieve the adjustment in two different ways: either by keeping the width of the profiles constant on the screen and adjusting the magnitude of the distortion, or by changing the width of the profiles on the screen so that their angular size from the viewer point stays the same. We found the first approach to be less disruptive to image content while the second to provide stronger contrast enhancement.

6.6 Conclusion

In this study we measured conditions under which countershading profiles are perceived as objectionable. We found a strong effect of the width and magnitude of the profile and a much weaker effect of the underlying image content. Unlike previous studies on the detection of countershading profiles (Sullivan and Georgeson [Sul-

livan and Georgeson, 1977], and Krawczyk *et al.* [Krawczyk et al., 2007]) or their matching contrast (Dooley and Greenfield [Dooley and Greenfield, 1977]), the focus of our work is the aesthetics of countershading. In particular, the perceived quality is strongly affected by changes in profile width, including those resulting from changes in viewing conditions such as distance. Image downsampling also affects profile width and may easily convert acceptable profiles into objectionable ones, implying this observation is very relevant to a large group of image enhancement algorithms. It is our suspicion that a number of algorithms said to introduce haloes may in fact produce acceptable results, but suffer from the disparity in size between authors' monitors and their printed reproductions.

We have shown several applications where our model, combined with edge-preserving smoothing, can be used to improve upon existing countershading approaches, as well as enable some new possibilities. Improvements to existing methods include more accurate resizing of countershading profiles, halo-free local tonemapping, and a new unsharp masking operator that avoids issues of noise amplification. We also present a means of estimating and modifying countershading in existing images, including fine art.

Chapter 7

Defocus Techniques for Camera Dynamic Range Expansion

Chapters 5 and 6 presented perceptual models relating the appearance of blur and contrast. In this chapter, we investigate the use of optical blur to capture a greater contrast range with conventional sensors.

Defocus imaging techniques, involving the capture and reconstruction of purposely out-of-focus images, have recently become feasible due to advances in deconvolution methods. This chapter evaluates the feasibility of defocus imaging as a means of increasing the effective dynamic range of conventional image sensors. Blurring operations spread the energy of each pixel over the surrounding neighborhood; bright regions transfer energy to nearby dark regions, reducing dynamic range. However, there is a trade-off between image quality and dynamic range inherent in all conventional sensors.

The approach involves optically blurring the captured image by turning the lens out of focus, modifying that blurred image with a filter inserted into the optical path, then recovering the desired image by deconvolution. We analyze the properties of the setup to determine whether any combination can produce a dynamic range reduction with acceptable image quality. Our analysis considers both properties of the filter to measure local contrast reduction, as well as the distribution of image intensity at different scales as a measure of global contrast reduction. Our results show that while combining state-of-the-art aperture filters and deconvolution

methods can reduce the dynamic range of the defocused image, providing higher image quality than previous methods, rarely does the loss in image fidelity justify the improvements in dynamic range.

7.1 Introduction

The fact that the range of luminances found in the real world greatly exceeds the capabilities of imaging sensors is a fundamental problem encountered during the acquisition of digital images. Real world scenes contain values brighter and darker than the range that can be captured at any one time by conventional image sensors, and as a result, over-exposed and under-exposed pixels commonly occur in photographs. Conventional image sensors cannot match the dynamic range of the scene, and can only capture a subset of the luminances present. Although specialized high dynamic range image sensors can capture the range of luminances found in most real world scenes, they suffer from lower signal-to-noise ratios (SNR) or slower read-out speeds [Yang et al., 1999].

Photographers have contended with this problem since the advent of photography, and the most common solution is the concept of exposure to control the amount of light that falls on the sensor. While controlling the amount of light reaching the sensor by adjusting the aperture and the exposure time, photographers are able to select which subset of the scene luminances they wish to capture without undesirably over- or under-exposing the image. However, adjusting the exposure does not improve the limited dynamic range that can be acquired. The subset of luminances that can be accurately captured can be thought of as a slice through the entire range of luminances found in the scene. Adjusting the exposure can move the slice up and down the range of scene luminances, and pixels with luminances above the top of the slice are recorded as white, while pixels with luminances below the bottom of the slice are recorded as black. A correctly chosen exposure can minimize the number of over- and under-exposed pixels, resulting in a properly exposed image, but it is the dynamic range of the sensor that controls the width of the slice. The fact remains that if the dynamic range of the scene exceeds that of the sensor, some pixels will not be recorded accurately.

Most existing techniques capture multiple slices of the luminance range and com-

bine them into a single image representing a wider slice, and all of these methods require some tradeoff to extend the width of the slice. Multi-exposure high dynamic range reconstruction [Debevec and Malik, 1997; Robertson et al., 1999] takes a sequence of slices distributed in time, trading off temporal resolution for a larger dynamic range. Similarly, placing an array of different neutral density filters onto the sensor [Nayar and Mitsunaga, 2000] can trade spatial resolution for a wider slice of the dynamic range. The best option is to develop new sensor technology [Acosta-Serafini et al., 2005; IMS Chips, 2004; Yang et al., 1999] that is directly capable of capturing a wider slice of the dynamic range, but these sensors are still some way off from commercial availability and currently suffer from resolution and quality issues.

7.2 Overview

The majority of the existing methods attempt to expand the dynamic range of the sensor to match a fixed range of real scene luminances incident upon the sensor. We investigate the opposite, reducing the dynamic range of the scene to better fit in the limited range of the sensor.

The method we investigate attempts this reduction in a two-part, combined optical and software approach. First, we optically blur the image to reduce the dynamic range of the scene incident on sensor. Then we restore the original image detail and dynamic range in software. Blurring is a convolution operation, where the energy that would fall on a single photosite of the sensor is spread over a local neighborhood, and reciprocally, that photosite receives some amount of energy from its neighborhood. This exchange reduces the difference between the pixel and its neighbors, thus producing an image with less local contrast.

The underlying assumption of this approach is that the most extreme luminance values are sufficiently spatially distributed such that the local contrast reduction from convolution will reduce the number of pixels outside the sensor's dynamic range. A good candidate for this method would be small point sources, such as streetlights at night, while a bad candidate would be large bright areas, such as daytime sky viewed through a window.

We analyze, both in terms of possible dynamic range reduction and resulting image quality, the properties of the optical-digital system composed of a blurred im-

age obtained by an aperture filter inserted into the optical path then restored by a deconvolution method. The results are quantified to determine whether any combination can produce a meaningful reduction of dynamic range required to capture by a sensor while maintaining acceptable image quality.

7.3 Physical Setup

While the optical system in a camera lens is complex, the entire collection of lenses can be approximated as a pair of thick lenses with the aperture in between them. This pair of lenses focuses a bundle of rays coming from points in the scene to points on the sensor, while the size and shape of the aperture controls which rays in the bundle reach the sensor.

The optical system focuses the image by directing all the rays in a bundle that originate from a point on the focal plane in the scene to converge to the same point on the imaging sensor. If a point lies in front of or behind the focal plane, the bundle of rays do not converge to a point on the sensor. Instead, the sensor plane will intersect the cone of light exiting the rear lens element, resulting in a circular pattern projected on the sensor. The amount of defocus determines the radius of the blur circle, and points further from the focal plane are proportionally further from focusing on the sensor and are more blurred since they intersect a bigger slice of the cone.

As shown in Figure 7.1, the aperture sits in the middle of the imaginary pair of optical elements we treat as the lens. Light rays pass directly through it. The circular blur pattern normally observed in out-of-focus images results from the circle shape of the aperture in a normal lens. With a different aperture shape, if a point is out of focus, the sensor still intersects with the cone of out of focus rays, but the aperture shape has blocked some rays traveling through it and the shape of the blur matches that pattern of the aperture.

7.4 Coded Aperture

As mentioned in the previous section, defocus blur convolves the image of the scene by aperture shape. Thus, the Fourier transform of the captured image has the frequency characteristics multiplied with the Fourier transform of the aperture shape.

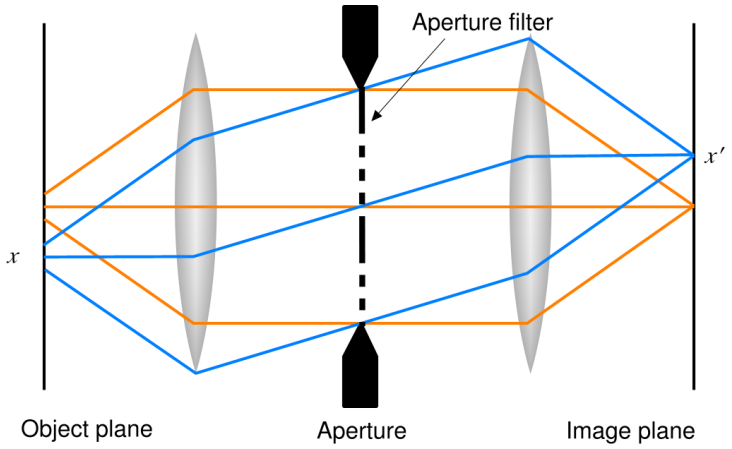


Figure 7.1: *Placement of aperture filter in optical setup.*

The goal of inserting an aperture filter into the optical system is to produce a blur pattern that has good frequency preservation properties.

Convolution is equivalent to multiplying the Fourier transform of the scene image by the aperture optical transfer function (OTF), the frequency space representation of the aperture shape. Therefore, in a noise-free case, deconvolution can be viewed as dividing the Fourier transform of the scene image by the aperture OTF. Zeros and very small values in Fourier transform of the filter result in division by zero and excessive error in the deconvolved image; these values are responsible for many of the artifacts observed.

The question of "what constitutes a good coded aperture pattern?" involves a number of criteria. We take this inquiry to specifically consider the aperture pattern that yields the highest quality reconstruction of the original image with the least number of artifacts. As mentioned above, the zeros in the frequency response of the filter prevent information from the scene being captured by the camera.

The ideal coded aperture pattern has a frequency spectrum with as few near-zero values as possible, with the locations of the existing near-zero values located as far away from the low frequency bins and each other as possible. Additionally, that property must hold for a number of different scales of the filter, as points at different depths in the scene will be blurred by different amounts.

There are several practical issues beyond these theoretical considerations. Suffi-

ciently small holes in the aperture pattern will cause the light passing through them to diffract, causing ringing artifacts in the captured image. Another desired property is that the aperture pattern transmits as much light as possible. Given these considerations, we will now analyze several coded aperture patterns (shown in Figure 7.2) and discuss their relative merits:

Standard Gaussian Veeraraghavan Zhou Levin

Figure 7.2: *Aperture filters evaluated.*

Standard (circular) aperture: The circular aperture found in conventional lenses corresponds to a first-order Bessel function in frequency space. The Bessel function resembles a dampened **cos** function, with near-zero values around the numerous zero-crossings of the function. This function causes very poor results with conventional deconvolution algorithms but can be made to yield acceptable results with a better deconvolution algorithm utilizing natural image statistics. However, as far as coded aperture patterns go, the standard aperture is very poor, and the only advantage is that it has the highest light transmission of any pattern.

Circular Gaussian: The Gaussian function has two compelling reasons for its use: the value of the function never reaches zero and the Fourier transform of a Gaussian function is another Gaussian. However, there are two equally significant caveats. First, while the Gaussian doesn't ever reach zero, it quickly reaches the noise floor of the camera. For the Fourier transform of a Gaussian function to be sufficiently broadband, the aperture pattern is nearly a pinhole. Second, while the Gaussian function has infinite extent, the actual filter shape would be a truncated Gaussian, the Gaussian multiplied by a box filter. The result would have the same zero-crossing issues as the standard aperture.

Veeraraghavan et al. [2007] started with a Modified Uniformly Redundant Array (MURA) pattern, then improved it with an optimization function. They employed gradient decent to select the binary pattern that best fits the difference in configuration between conventional cameras and x-ray telescopes. Modeling such differences as the linear convolution that occurs in the optical systems of conventional cameras, they iteratively searched for the binary pattern with the greatest minimum value of the filter frequency response.

Zhou and Nayar [2009] improve upon the work of Veeraraghavan et al. and formulate a definition of the quality of the aperture filter based directly on the quality of the deconvolved image as opposed to the properties of the filter power spectrum alone. They assume that the camera will be used to capture natural images, and they make use of the $1/f$ relation of frequencies in natural images to further enhance their filter choice. Additionally, they note that the amount of noise present in the captured image significantly affects the quality of the deconvolution, and they use a genetic algorithm to search for the optimal coded aperture for a series of different noise levels. Both this work and Veeraraghavan et al. assume that the radius of the blur is determined by another method.

Levin et al. [2007] conducted similar work but optimized a filter for a different set of criteria. In their work, they wanted to accurately determine the amount of defocus present at every pixel and use that to recover the depth. Opposed to the work of Veeraraghavan et al. and Zhou and Nayar, who either directly or indirectly optimized filter patterns with a minimal number of near-zero values, Levin et al. constructed a filter such that there was the maximum possible difference in the position of the zero-crossing in the frequency spectrum for different filter radii. This allowed them accurately recover the amount of blur at every pixel by looking at the missing information, which they were able to fill in with their deconvolution algorithm.

7.5 Deconvolution

Deconvolution is used to restore the blurred image recorded by the sensor to its original sharpness and dynamic range. As discussed in Section 7.5, more recent deconvolution routines based on natural image statistics have made significant progress towards restoring the sharpness of blurred images.

While the use of a coded aperture filter improves the spatial frequency properties of the captured image, the fact remains that the deconvolution problem is ill-posed. The blurred image could be the result of any one of an infinite number of images that produce the same result when convolved by the chosen filter. However, none of those images will be exactly the same as the perfect (not blurred) photograph of a real world scene.

Real world scenes all share some very specific properties that can be used to guide the result of the deconvolution towards a physically-plausible result. Specifically, natural images tend to consist of large areas of nearly-constant values with sharp divisions between them. Described in a more formal way, the derivatives of natural images have a heavy-tailed distribution with a narrower peak and longer tail than a Gaussian function. This distribution implies that most pixels have values close to zero, but some small number of pixels, especially those lying next to edges, have significantly larger values.

Both Bando and Nishita [2007] and Levin et al. [2007] have presented deconvolution algorithms that make use of these natural image statistics. They solve a system of equations that includes a weighting term that corresponds to the prior assumption of a natural-image distribution. Bando and Nishita modify the WaveGSM [Bioucas-Dias, 2006] algorithm to operate in gradient space and perform expectation maximization on the resulting non-linear system of equations using the second-order stationary iterative method. Levin et al. [2007] approach the problem as finding the maximum likelihood explanation, employing iteratively reweighted least-squares to solve for the non-linear sparse prior term.

7.6 Evaluation

Our goal is to determine whether any combination of aperture filter and deconvolution algorithm can produce a meaningful increase in effective dynamic range

while maintaining acceptable image quality. Specifically, we compare filters by Veeraraghavan et al. [2007], Zhou and Nayar [2009], and Levin et al. [2007] in addition to the Gaussian and standard aperture. We evaluate deconvolution methods by Bando and Nishita [2007] and Levin et al. [2007] in addition to classical Wiener filtering [Wiener, 1964] and Richardson-Lucy [Lucy, 1974; Richardson, 1972] methods.

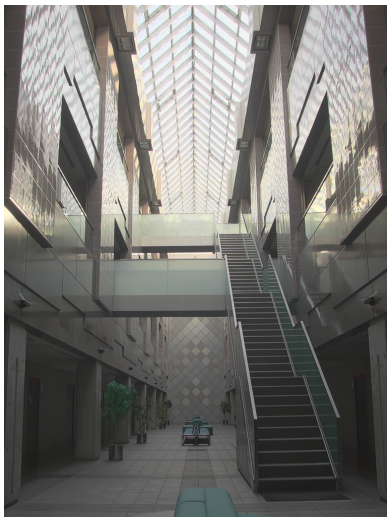
In our evaluation, we measure the *reduction* in contrast between the blurred image and the original image as opposed to the expansion in contrast between the result of the deconvolution and the blurred image. Errors in images produced by deconvolution often appear as ringing artifacts around contrast edges, artificially increasing the contrast in those regions. This approach avoids mistaking those artifacts with meaningful increases in dynamic range. The quality of the deconvolution algorithm is still determined as the difference between the original image and the reconstructed image.

The structure of the image has an impact on the effectiveness of the algorithm. If the size of a bright or dark image feature exceeds the given filter diameter, there will be no reduction in dynamic range since energy is only exchanged between pixels within the local neighborhood. Concerning dynamic range reduction, our analysis considers both properties of the filter used as a measure of local contrast reduction as well as the distribution of intensity values in the image at different scales as a measure of global contrast reduction.

Our primary interest was validating the proof of concept, and we conduct our evaluation using synthetic results instead of real optical systems. This approach introduces less complexity in quantifying the performance of the method while still determining if the upper bound of the performance is of sufficient quality.

We present two of the images we used to evaluate the performance in this proposal, *Atrium Morning* and *Atrium Night*, shown in Figure 7.3. This pair was chosen to demonstrate how the performance of the system depends on the spatial distribution of luminance values in the image. While *Atrium Night* has a larger dynamic range than that of *Atrium Morning*, its extreme values are located in small bright light sources in contrast to the broad skylights in *Atrium Morning*.

Table 7.1 shows the radii of filters used in the evaluation and the change in dynamic range of the image as a result of being blurred by a standard aperture disc filter of different radii. The night image results in a larger reduction in dynamic range



Atrium Morning

Atrium Night

Figure 7.3: *Sample images used in evaluation. Images copyright Frederic Drago.*

since the point light sources distribute energy over neighborhoods of significantly dimmer values. The tables and plots included in this section all reference the differences in dynamic range in terms of the photographic concept of exposure value (EV) stops. In photography, a change of 1 stop or 1 EV represents a unit change on the $\log_2$ scale, where 1 EV $= \log_2(L_1) - \log_2(L_2)$.

We evaluated the quality of the images reconstructed by the deconvolution algorithms in terms of peak signal-to-noise ratio (PSNR) [Thomos et al., 2006]. Typical values for the PSNR in lossy image and video compression are between 30 and 50 dB, where higher is better [Thomos et al., 2006; Xiangjun and Jianfei, 2007]. However, the images used in this evaluation have a larger dynamic range than that of conventional 8-bit images. From evaluating the quality of the results we have chosen 35 dB to be the lower bound on acceptable image quality for deconvolved images. These images have visible artifacts, but all of the features are still clearly visible. Similarly, we have chosen a dynamic range reduction of 2 stops to be the minimum acceptable reduction in dynamic range.

Our evaluation proceeded as follows. First, each image was convolved by each of the aperture filters at a number of different radii. The minimum and maximum

Radius	Atrium Morning			Atrium Night		
	min	max	**reduction**	min	max	**reduction**
Original	0.00	11.0		0.00	12.0	
1	0.00	10.8	**0.200**	0.452	12.0	**0.452**
2	0.00	10.6	**0.424**	0.622	12.0	**0.622**
3	0.00	10.3	**0.716**	1.163	11.8	**1.34**
4	0.02	10.0	**1.00**	1.436	11.4	**1.99**
5	0.08	9.94	**1.14**	1.589	11.4	**2.23**
6	0.15	9.92	**1.24**	1.731	11.2	**2.51**
8	0.31	9.83	**1.48**	1.890	10.8	**3.13**
9	0.40	9.79	**1.61**	1.950	10.5	**3.41**
11	0.66	9.71	**1.94**	2.08	10.3	**3.74**
13	0.86	9.67	**2.19**	2.18	10.1	**4.13**
16	1.04	9.59	**2.45**	2.26	9.61	**4.65**

Table 7.1: *Amount of reduction in dynamic range as a function of radius of a standard aperture (disc) filter in pixels. The minimum and maximum values are the result of convolving the original image with a disc filter of the specified size, showing how much of the dynamic range reduction was from reducing the intensity of highlight regions versus increasing the intensity of shadow regions. All units are in terms of powers of two, referred to as exposure value (EV) stops.*

values of the original image and the blurred image were compared to compute the amount of dynamic range reduction for that size filter. Next, different amounts of Gaussian noise were added to simulate the random nature of the image acquisition process for different sensor sizes. Then, all of the combinations of filter, radius and noise were deconvolved by each of the deconvolution algorithms. Finally, all results were compared to the original image to compute the PSNR. Figures 7.4 and 7.5 summarize our results across the combinations of aperture filter and deconvolution method for images without any noise added, while Figures 7.6 and 7.7 summarize our results for images with Gaussian noise $\sigma = 1$ added.

The deconvolution algorithm by Levin et al. [2007] was able to produce acceptable results for images with small bright areas, such as *Atrium Night*, when paired with one of the filters by Levin et al. [2007], Veeraraghavan et al. [2007], or Zhou and Nayar [2009]. However, it was only able to do so at noise levels below those

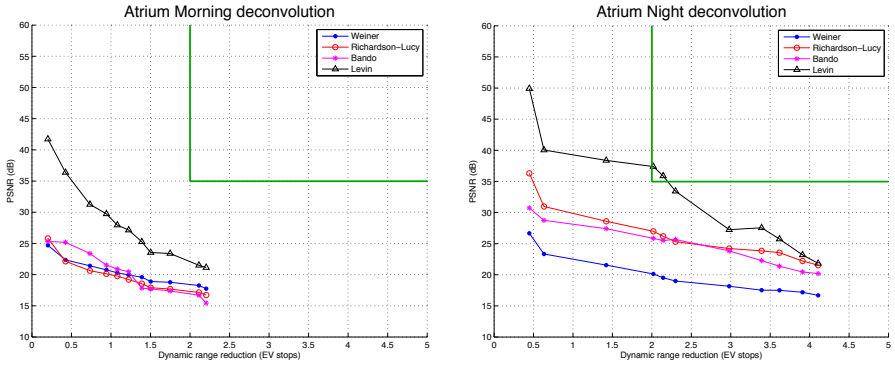


Figure 7.4: *Comparison of all four deconvolution algorithms on the Atrium test scenes without any noise added. All results were computed using the aperture filter proposed by Zhou and Nayar [2009]. Values above and to the right of the green bars pass our acceptance criteria.*

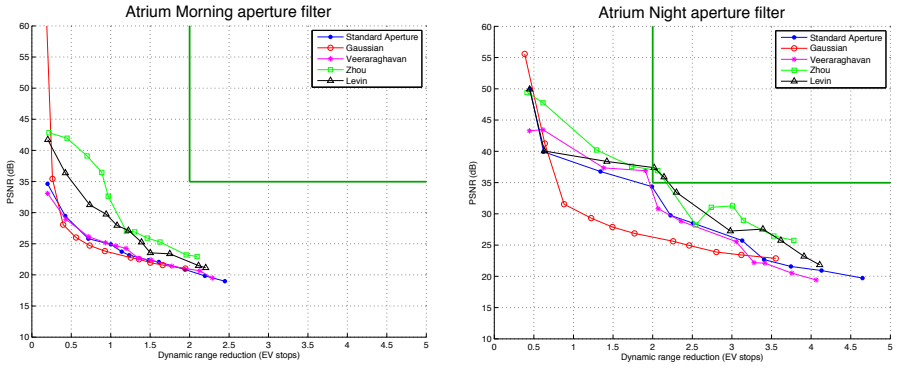


Figure 7.5: *Comparison of all aperture filters on the Atrium test scenes without any noise added. All results were computed using the deconvolution algorithm proposed by Levin et al. [2007] Values above and to the right of the green bars pass our acceptance criteria.*

found in existing cameras.

Additionally, the Levin et al. [2007] method performed worse than Richardson-Lucy for images with a few bright points like *Atrium Night* at realistic noise levels. While it is able to reconstruct very fine details, the method of Levin et al. [2007] tends to introduce high frequency ringing. The amount of ringing associated with a given feature is insignificant relative to its overall magnitude, but if that feature is bright enough, the ringing will destroy detail in the dark regions of the image.

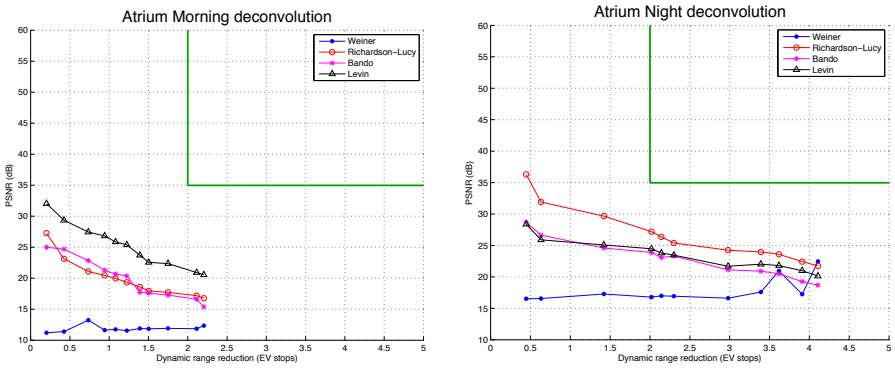


Figure 7.6: *Comparison of all four deconvolution algorithms on the Atrium test scenes with additive Gaussian noise of* $\sigma = 1$*. All results were computed using the aperture filter proposed by Zhou and Nayar [2009]. Values above and to the right of the green bars pass our acceptance criteria.*

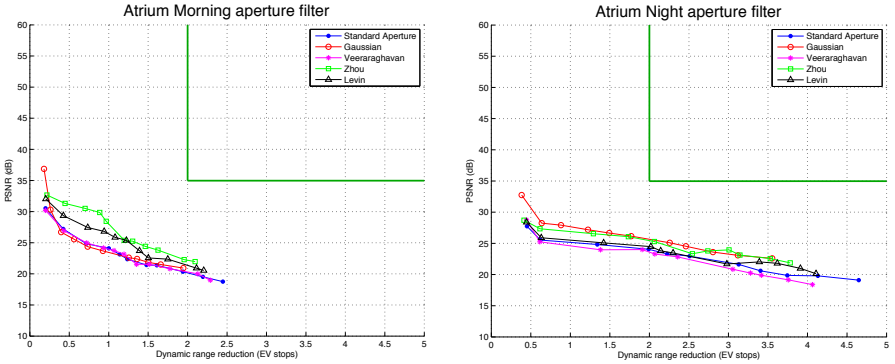


Figure 7.7: *Comparison of all aperture filters on the Atrium test scenes with additive Gaussian noise of* $\sigma = 1$*. All results were computed using the deconvolution algorithm proposed by Levin et al. [2007] Values above and to the right of the green bars pass our acceptance criteria.*

Increasing the smoothing parameter of the algorithm might have produced less noise in the dark regions but would not have recovered any additional detail.

The deconvolution algorithm by Bando and Nishita [2007] performed worse than expected, given its success on conventional 8-bit images. This could be caused by the selection of its normalization parameters, which we have not been able to optimize for the given images. However, in all tests on conventional (8-bit) images,

it performed at best equal with that of Levin et al. [2007], and we are confident that it could not do significantly better here.

The filters by Levin et al. [2007], Veeraraghavan et al. [2007] and Zhou and Nayar [2009] outperformed the other filters for images with low noise levels. As the amount of noise increased, the difference between the capability of all the different filters decreased, and eventually all the filters had the same performance as the conventional aperture. This confirms the observation of Zhou and Nayar [2009] that the optimized aperture filters they designed became simpler and more like the conventional aperture as the noise level increased.

Overall, the results show that no current combination of aperture filter and deconvolution algorithm can deliver an acceptable combination of dynamic range reduction and image quality for images with large bright areas like *Atrium Morning* at any noise level, while the desired performance was only possible for images with small bright areas with unrealistically low amounts of noise. Additionally, the quality of the final result depended more on the deconvolution method than on the choice of aperture filter for realistic noise levels.

7.7 Conclusions

None of the possible combinations of aperture filter and deconvolution algorithm were able to consistently reduce the dynamic range of the captured image without excessively degrading image quality. The combination of algorithm and filter that did work did so under very controlled conditions. Without advances to either aperture filtering or image reconstruction, the approach is not applicable to general circumstances.

The efficiency of this defocus imaging approach is scene-dependent. The method is good for small over-exposed regions that are just above the maximum photosite capacity of a sensor but performs worse on large overexposed areas or in recovering exceedingly bright regions. The more complex deconvolution algorithms performed better than traditional methods but at significant computational cost. These algorithms took on the order of several minutes to produce results for megapixel images. The marginal improvement in dynamic range at acceptable image quality does not justify the amount of computation required by the method.

We could also consider different means of evaluating image quality. High PSNR is rarely the goal of deconvolution. Images may look acceptable, or even better, if they are sharper despite the fact that the PSNR score would be lower than a different version. We could obtain a more accurate estimate of the image quality using different metrics or subjective user preference means of evaluation.

Subsequent collaborative work [Rouf et al., 2011] took a novel view of HDR capture inspired by this combination blurring-deconvolution approach. In this work, we first optically encode both the LDR reflectance portion of the scene and highlight information into the image captured with a conventional image sensor. This step is achieved using a cross-screen or "star" filter. Second, we decode, in software, both the low dynamic range image and the highlight information. Lastly, these two portions can be combined to form an image of a higher dynamic range than the regular sensor dynamic range.

Chapter 8

Discussion and Conclusions

This dissertation has presented several methods for manipulating the perception of blur and contrast in images. These methods take inspiration from the fundamental organization of spatial image perception into multiple parallel channels for processing visual information and conveying image appearance. In addition to taking inspiration from, some of the methods employ models of human spatial vision to more accurately control the appearance of images under changing viewing conditions.

8.1 Benefits and Limitations

Beyond the individual contributions of each chapter, there are several other benefits. The first benefit of the work presented in this dissertation is the conceptual framework of understanding scale-dependent effects of image perception. We have demonstrated that the parallel spatial frequency channels present in every viewer's visual system affect the perceived appearance of displayed images. The resolution of the image on the retina, not in terms of pixels but of angular resolution, determines the mapping of spatial frequency information in the image to specific visual channels.

With this understanding, we are able to identify which elements of image display affect the mapping from spatial frequencies to visual channels, the physical arrangement of viewer and display in particular. It is this relationship, along with models of image appearance based on those spatial frequency channels that allow us to predict how an image is perceived and how that perception will change depending on changes in viewing conditions.

We have used this model to create new methods of manipulating blur and contrast in images. Accurate estimates of spatially-variant image blur allow us to synthesize images with defocus patterns prohibited by the constraints of lens optics design. When coupled with a model of perceived image blur, the same estimation routine allows us to preserve the appearance of images when downsampling.

A similar perceptual model of whether countershading is perceived as acceptable or objectionable has provided us new insights into the use of unsharp masking and similar image processing operations. Our work has also suggested novel ways of introducing countershading that avoid the amplification of high frequency details that plagued previous approaches. Finally, we have provided some initial attempts at the scale-aware display of images with monitors that present images not only calibrated to the size and resolution of the display, but to changes in the physical distance to the viewer.

The contributions presented in this work are not without limitation. Several of the methods presented rely on accurate estimation of image features when altering image appearance. Regardless of the quality of the image appearance models, if the estimation falters, the result will contain artifacts.

The method of blur estimation presented is still not entirely robust to noise. It can confuse noise with fine texture detail. Our means of estimating the noise level in the image could be improved and more complex approaches would distinguish between noise and texture detail based on statistics of image regions. Currently, we just compare gradient magnitude of regions. Patched-based methods that estimate local histograms could better distinguish between random noise and fine image structures. Additionally, our method can only estimate the amount of defocus blur. Motion blur and camera shake do not alter the gradient magnitude in an analogous way and still appear sharp to the estimation algorithm. Extending our model of blur would allow us to enhance the appearance of more than defocus blur effects.

There is even more work regarding the estimation of countershading profiles. Our example of Seurat's *Le Bec du Hoc* was only a proof of concept and can not easily be extended to automatically handle arbitrary images. While the sharp transition of a countershading profile is easy to measure in images, the slow-changing gradients are much harder to accurately estimate. The close proximity of step edge and low-magnitude gradients defy existing multi-scale estimation techniques such as Elder and

Zucker [1998] and more robust means of estimating very small gradients in images need to be developed.

Providing correct-appearing images to the viewer is another challenge. The scale at which the image appears on the viewer's retina is in part determined by the physical relationship between the viewer and the display. Display-specific attributes such as dimension and resolution can be known in advance and precomputed. The distance of the viewer continually changes and the display must adapt to the change in distance. Our scale-aware display provides a proof of concept capable of adapting the image content to the viewer's distance, but only for a single viewer. However, basing the approach on head-tracking technology is not suitable for everyday use. Better methods of tracking multiple viewers and providing them with specifically-tailored images are necessary for truly scale-aware image display. Some new technology, whether lightfield display or something else is necessary to enable this possibility.

8.2 Future Work

There are a number of potential avenues of further investigation. Each individual contribution suggests next possible steps, as well as some other directions that draw on lessons from the body of work as a whole.

On the topic of blur estimation, future work includes improving the algorithm's ability to differentiate between textured and noisy regions and preserve fine details. In the context of mobile devices, information is available from additional types of sensors and it may be possible to use the accelerometer found in many devices to more accurately estimate blur resulting from camera shake.

Future work in image resizing includes extending the concept of preserving the perceived appearance to other image attributes. From an image-quality perspective, accurately preserving the appearance of noise when downsampling can be as important as blur.

We would like to conduct a more comprehensive study of the perception of countershading, including more images and larger variation of the underlying edge contrast. The existing models of perceived contrast resulting from the Cornsweet illusion would enable us to accurately determine the new combinations of λ and σ to preserve the perceived edge contrast while still avoiding haloes.

The investigation of Camera Dynamic Range Expansion did not yield positive results, in part due to the quality of deconvolution algorithms used, which are not well-tuned for high dynamic range images. The human visual system is sensitive to relative change and can detect small changes in dark regions. However, deconvolution algorithms are linear and minimize absolute error without accounting for how salient that error is to the final observer. An interesting line of research would be attempting to devise a deconvolution algorithm that gives preference to relative error and correctly weights the importance of dark regions.

Additionally, while we treat the visual channels in the brain as independent, we know this in not the case. Lateral inhibition, found between adjacent photoreceptors in the retina, applies equally to adjacent spatial frequency channels in the HVS at large. The presence of strong contrast at one frequency inhibits the response of nearby visual channels and can even amplify the response of more distant ones. We want to extend our models of the perception of blur, contrast and other image attributes to incorporate models of spatial perception accurate enough to capture these subtle effects.

Finally, the main theme of this dissertation is centered on the observation that the same image can appear different depending on the scale at which it is viewed. The field of color science has developed models capable of accounting for changes in color perception under different conditions. Similarly, we envision the development of an overarching pipeline that employs accurate models of human spatial vision to account for all changes in image appearance resulting from the diverse scales at which that image can be displayed.

8.3 Conclusions

Over the last decade, the paradigm of computational photography has made great strides increasing the capabilities of image capture. All of these approaches have involved some change in the physical setup to visually encode more information with a change in the processing of that image to recover that information and produce the desired result. The next decade will see a similar increase in the computational abilities available for image display and it is worth considering what principles would best guide the development of future display algorithms.

At the heart of the issue, we are fundamentally asking ourselves "What are the reasons to display, or even produce images?" Imaging is synonymous with communication. In the end, images are produced to communicate some notion or concept, no matter how abstract, with the viewer. The implicit assumption is we are communicating with a human observer possessing a human visual system that will process visual stimuli and the resulting perceived qualities will form their interpretation.

Computational photography constructs novel physical setups to encode information and designs matching algorithms to decode that information. Likewise, the emerging field of computational display must recognize since the HVS is responsible for decoding all information received by the viewer, the means of image display capable of conveying the most information will be matched to its respective decoder, the viewer's visual system. Accurate and communicative display of images must rely on knowledge of human perception, for the act of perception is the only means of receiving information available to viewers.

The conventional imaging pipeline includes the capture, manipulation and display of image content. In the future, we foresee the development of a new imaging pipeline that extends beyond the production of photons by the monitor. This new pipeline will recognize that any displayed image is viewed under some specific conditions by the visual system of an observer which responds to both the image and the conditions. Any communicated information will result from that specific response of the visual system to those stimuli. By considering the perceived appearance of images we can develop an *image understanding pipeline* as the basis of more effective visual communication.

Bibliography

Acosta-Serafini, P., Masaki, I., and Sodini, C. (2005). Single-chip imager system with programmable dynamic range. U.S. Patent 6,977,685. → pages 123

Adelson, E. H. (1995). Checkershadow illusion. → pages 11

Adelson, E. H. (1999). Lightness perception and lightness illusions. In Gazzaniga, M., editor, *The New Cognitive Neurosciences*, chapter 24, pages 339–351. MIT Press, Cambridge, MA, 2nd edition. → pages 11

Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., and Ogden, J. M. (1984). Pyramid methods in image processing. *Engineer*, 29(6):33–41. → pages 27

Agrawal, A., Xu, Y., and Raskar, R. (2009). Invertible motion blur in video. *ACM Trans. Graph.*, 28:95:1–95:8. → pages 42

Akyüz, A. and Reinhard, E. (2008). Perceptual evaluation of tone-reproduction operators using the Cornsweet-Craik-O'Brien illusion. *ACM Transactions on Applied Perception (TAP)*, 4(4). → pages 63

Anstis, S. M. and Howard, I. P. (1978). A Craik-O'Brien-Comsweet illusion for visual depth. *Vision Research*, 18:213–217. → pages 36

Avidan, S. and Shamir, A. (2007). Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 26(3):10. → pages 65

Badamchizadeh, M. A. and Aghagolzadeh, A. (2004). Comparative study of unsharp masking methods for image enhancement. In *Proceedings of International Conference on Image and Graphics*, pages 27–30, Washington, DC, USA. IEEE Computer Society. → pages 33, 60, 62

Bae, S. and Durand, F. (2007). Defocus magnification. *Computer Graphics Forum*, 26(3):571–579. → pages xi, 57, 58, 59, 68, 78, 80, 93

Bando, Y. and Nishita, T. (2007). Towards digital refocusing from a single photograph. In *Proceedings of Pacific Conference on Computer Graphics and Applications*, pages 363–372. → pages 45, 128, 129, 133

Barnard, K. and Funt, B. (1997). Analysis and improvement of multi-scale retinex. *Proceedings of the IS&T/SID Fifth Color Imaging Conference: Color Science, Systems and Applications*, pages 221–226. → pages 39

Basu, M. (2002). Gaussian-based edge-detection methods - a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 32(3):252–260. → pages 53

Ben-Ezra, M. and Nayar, S. K. (2004). Motion-based motion deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:689–698. → pages 49

Berkner, K. and Erol, B. (2008). Adaptation of document images to display constraints. In *Human Vision and Electronic Imaging XIII*, volume 6806, page 68061C. SPIE. → pages 65

Bioucas-Dias, J. (2006). Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Transactions on Image Processing*, 15(4):937–951. → pages 128

Blake, A. (1985). On lightness computation in Mondrian world. In Ottoson, T. and Zeki, S., editors, *Central and Peripheral Mechanisms of Color Vision*, pages 45–49, New York. Macmillan. → pages 37

Blakemore, C., Carpenter, R. H. S., and Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228:37–39. → pages 4, 16

Blakemore, C., Muncey, J. P., and Ridley, R. M. (1973). Stimulus specificity in the human visual system. *Vision Res.*, 13:1915–1931. → pages 28

Brady, N. and Field, D. J. (1995). What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns. *Vision Res.*, 35:739–756. → pages 30

Burr, D. (1987). Implications of the Craik-O'Brien illusion for brightness perception. *Vision Research*, 27(11):1903–1913. → pages 33, 35, 110

Burt, P. J. and Adelson, E. H. (1983). The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540. → pages 27

Calabria, A. J. and Fairchild, M. D. (2003a). Perceived image contrast and observer preference i: The effects of lightness, chroma, and sharpness manipulations on contrast perception. *Journal of Imaging Science & Technology*, 47:479–493. → pages 59, 62

Calabria, A. J. and Fairchild, M. D. (2003b). Perceived image contrast and observer preference ii: Empirical modeling of perceived image contrast and observer preference data. *Journal of Imaging Science & Technology*, 47:494–508. → pages 59, 62

Campbell, F. W., Howell, E. R., and Johnstone, J. R. (1978). A comparison of threshold and suprathreshold appearance of gratings with components in the low and high spatial frequency range. *J. Physiol. (Lond.)*, 284:193–201. → pages 33, 34, 110

Campbell, F. W., Howell, E. R., and Robson, J. G. (1971). The appearance of gratings with and without the fundamental Fourier component. *J. Physiol. (Lond.)*, 217:17P–18P. → pages 33, 34, 110

Cannon, M. W. (1985). Perceived contrast in the fovea and periphery. *J Opt Soc Am A*, 2:1760–1768. → pages 29

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 679–698. → pages 50, 51

Chakrabarti, A., Zickler, T., and Freeman, W. T. (2010). Analyzing spatially-varying blur. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2512–2519. → pages 48

Chen, C.-C., Chen, K.-P., Tseng, C.-H., Kuo, S.-T., and Wu, K.-N. (2009). Constructing a metrics for blur perception with blur discrimination experiments. In *Proc. SPIE, Image Quality and System Performance VI*, number 724219. → pages x, 21

Chiu, K., Herf, M., Shirley, P., Swamy, S., Wang, C., and Zimmerman, K. (1993). Spatially nonuniform scaling functions for high contrast images. In *Proc. Graphics Interface*, pages 245–253. → pages 62

CIE (2004). A colour appearance model for colour management systems: CIECAM02. Technical Report Technical Report CIE 159:2004, Commision Internationale De L'Eclairage, Vienna. → pages 40

Ciuffreda, K. J., Selenow, A., Wang, B., Vasudevan, B., Zikos, G., and Ali, S. R. (2006). "Bothersome blur": a functional unit of blur perception. *Vision Res.*, 46:895–901. → pages 22, 104

Ciuffreda, K. J., Wang, B., and Vasudevan, B. (2007). Conceptual model of human blur perception. *Vision Research*, 47(9):1245 – 1252. → pages x, 22, 23

Cornsweet, T. (1970). *Visual Perception*. Academic Press, New York. → pages 34

Cossairt, O., Zhou, C., and Nayar, S. (2010). Diffusion coded photography for extended depth of field. *ACM Trans. Graph.*, 29:31:1–31:10. → pages 44

Craik, K. J. W. (1966). *The nature of psychology; a selection of papers, essays, and other writings*. Cambridge University Press, Cambridge. → pages 34

Cufflin, M. P., Mankowska, A., and Mallen, E. A. H. (2007). Effect of blur adaptation on blur sensitivity and discriminations in emmetropes and myopes. *Investigative Ophthalmology & Visual Science*, 48(6):2932–2939. → pages 22

Dai, S. and Wu, Y. (2008). Estimating space-variant motion blur without deblurring. In *Proceedings of the International Conference on Image Processing*, pages 661–664. → pages 49

Davey, M., Maddess, T., and Srinivasan, M. (1998). The spatiotemporal properties of the Craik-O'Brien-Cornsweet effect are consistent with 'filling-in'. *Vision Res*, 38:2037–2046. → pages 37

De Valois, R. and De Valois, K. (1990). *Spatial Vision*. Oxford University Press. → pages 13, 15, 17, 28, 61

Deans, S. R. (1993). The radon transform and some of its applications. → pages 48

Debevec, P. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. *Proceedings of ACM SIGGRAPH*. → pages 123

Devinck, F., Hansen, T., and Gegenfurtner, K. (2007). Temporal properties of the chromatic and achromatic Craik-O'Brien-Cornsweet effect. *Vision Research*, 47(27):3385–3393. → pages 37

Dicke, R. (1968). Scatter-hole cameras for x-rays and gamma rays. *Astrophysical Journal*, 153. → pages 43

Donoho, D., Johnstone, I., and Johnstone, I. M. (1993). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455. → pages 66

Dooley, R. and Greenfield, M. (1977). Measurements of edge-induced visual contrast and a spatial-frequency interaction of the Cornsweet illusion. *Journal of the Optical Society of America*, 67(6):761–765. → pages 35, 36, 63, 110, 120

Durand, F. and Dorsey, J. (2002). Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.*, 21:257–266. → pages xii, 62, 113, 114

Elder, J. and Zucker, S. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7):699–716. → pages xi, 51, 53, 54, 57, 72, 73, 93, 137

Fairchild, M. and Johnson, G. (2004). iCAM framework for image appearance, differences, and quality. *Electronic Imaging*. → pages 4, 40

Fairchild, M. D. (2005). *Color Appearance Models*. Wiley-IS&T, Chichester, UK, 2nd edition. → pages 4, 24

Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.*, 27:67:1–67:10. → pages 105, 114, 116, 117, 118

Fattal, R. (2007). Image upsampling via imposed edge statistics. *ACM Trans. Graph.*, 26(3):95. → pages 64

Fattal, R., Lischinski, D., and Werman, M. (2002). Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256. → pages 62

Feichtinger, H. G. and Strohmer, T. (1997). *Gabor Analysis and Algorithms*. Birkhäuser. → pages 16

Fenimore, E. E. and Cannon, T. M. (1978). Coded aperture imaging with uniformly redundant arrays. *Applied Optics*, 17(3):337–347. → pages 43

Fergus, R., Singh, B., Hertzmann, A., Roweis, S., and Freeman, W. (2006). Removing camera shake from a single photograph. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 25(3):787–794. → pages 45

Funt, B., Barnard, K., Brockington, M., and Cardei, V. (1997). Luminance-based multi-scale retinex. *AIC International Color Association*. → pages 39

George, S. and Rosenfield, M. (2004). Blur adaptation and myopia. *Optom Vis Sci*, 81:543–547. → pages 22

Georgeson, M. and Sullivan, G. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology*, 252(3):627–656. → pages 29, 30, 37

Georgeson, M. A. (1991). Contrast overconstancy. *J Opt Soc Am A*, 8:579–586. → pages 30

Gonzalez, R. C. and Woods, R. E. (2001). *Digital Image Processing*. Addison-Wesley, 2nd edition. → pages 16

Gottesman, S. and Fenimore, E. (1989). New family of binary arrays for coded aperture imaging. *Applied Optics*, 28(20):4344–4352. → pages xi, 43, 44

Hamerly, J. R. and Dvorak, C. A. (1981). Detection and discrimination of blur in edges and lines. *J. Opt. Soc. Am.*, 71(4):448–452. → pages 20, 21

Held, R. T., Cooper, E. A., O'Brien, J. F., and Banks, M. S. (2010). Using blur to affect perceived distance and size. *ACM Trans. on Graph.*, 29(2):19:1–16. → pages x, 24, 25

Horn, B. (1974). On lightness. Technical report, MIT Artificial Intelligence Lab. Membo 295. Massachusetts Institute of Technology. → pages 37

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148:574–591. → pages 4, 16

Hurlbert, A. (1986). Formal connections between lightness algorithms. *Journal of the Optical Society of America*, 3(10):1684–1693. → pages 37

Ibenthal, A. (2007). Image sensor noise estimation and reduction. Technical Report ITG 3.2, University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen. → pages 80

Ihrke, M., Ritschel, T., Smith, K., Grosch, T., Myszkowski, K., and Seidel, H.-P. (2009). A perceptual evaluation of 3D unsharp masking. *Human Vision and Electronic Imaging*, page 12. → pages 36, 111

Ilic, L., Pizurica, A., Vansteenkiste, E., and Philips, W. (2009). Image blur estimation based on the average cone of ratio in the wavelet domain. volume 7248, page 72480F. SPIE. → pages 48

IMS Chips (2004). HDRC VGAx. http://www.hdrc.com. → pages 123

ITU-R-BT.500-11 (2002). Methodology for the subjective assessment of the quality of television pictures. → pages 107

Ji, H. and Liu, C. (2008). Motion blur identification from image gradients. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8. → pages 48

Jobson, D., Rahman, Z., Woodell, G., Center, N., and Hampton, V. (1997). Properties and performance of a center/surround retinex. *IEEE Transactions on image processing*, 6(3):451–462. → pages 38

Jobson, D. and Woodell, G. (1995). Properties of a center/surround retinex part two: Surround design. *Citeseer*. → pages 39

Kandel, E. R., Schwartz, J. H., and Jessell, T. M., editors (2000). *Principles of Neural Science*. McGraw-Hill, New York, 4th edition. → pages 13, 14

Kim, M. H., Ritschel, T., and Kautz, J. (2011). Edge-aware color appearance. *ACM Transactions on Graphics (presented at SIGGRAPH2011)*, 30(2):13:1–9. → pages 32

Kim, S. H. and Allebach, J. P. (2005). Optimal unsharp mask for image sharpening and noise removal. *J. of Electronic Imaging*, 14(2):023005. → pages 61

Kingdom, F. and Moulden, B. (1988). Border effects on brightness: A review of findings, models, issues. *Spatial Vision*, 3(4):225–262. → pages 33, 36, 37, 110

Koenderink, J. (1984). The structure of images. *Biological Cybernetics*, 50(5):363–370. → pages 51

Kopf, J., Cohen, M. F., Lischinski, D., and Uyttendaele, M. (2007). Joint bilateral upsampling. *ACM Trans. Graph.*, 26. → pages 64

Krähenbühl, P., Lang, M., Hornung, A., and Gross, M. (2009). A system for retargeting of streaming video. In *ACM SIGGRAPH Asia 2009 papers*, SIGGRAPH Asia '09, pages 126:1–126:10, New York, NY, USA. ACM. → pages 65

Krawczyk, G., Myszkowski, K., and Seidel, H. P. (2007). Contrast restoration by adaptive countershading. In *The European Association for Computer Graphics Annual Conference (EUROGRAPHICS)*. → pages xi, 63, 64, 109, 110, 120

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Sage Publications. → pages 87

Lam, E. Y. and Goodman, J. W. (2000). Iterative statistical approach to blind image deconvolution. *J. Opt. Soc. Am. A*, 17(7):1177–1184. → pages 45

Land, E. (1986). An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, 83(10):3078–3080. → pages 38

Land, E. and McCann, J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*. → pages x, 37, 38, 39

Levin, A., Fergus, R., Durand, F., and Freeman, W. (2007). Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 26(3). → pages 44, 45, 127, 128, 129, 131, 132, 133, 134

Levin, A., Hasinoff, S. W., Green, P., Durand, F., and Freeman, W. T. (2009). 4d frequency analysis of computational cameras for depth of field extension. *ACM Trans. Graph.*, 28:97:1–97:14. → pages 42

Levin, A., Lischinski, D., and Weiss, Y. (2004). Colorization using optimization. *ACM Trans. Graph.*, 23:689–694. → pages 57, 78

Levin, A., Sand, P., Cho, T. S., Durand, F., and Freeman, W. T. (2008). Motion-invariant photography. *ACM Trans. Graph.*, 27:71:1–71:9. → pages 42

Lin, W., Gai, Y., and Kassim, A. (2006). Perceptual impact of edge sharpness in images. *Vision, Image and Signal Processing, IEE Proceedings* -, 153(2):215 – 223. → pages 34, 110, 111

Lindeberg, T. (1994a). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1):225–270. → pages 61, 115

Lindeberg, T. (1994b). Scale-space theory in computer vision. → pages 71

Lindeberg, T. (1994c). *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers. → pages 74

Lindeberg, T. (1996). Edge detection and ridge detection with automatic scale selection. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 465–470. → pages 52

Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116. → pages 50

Liu, C., Freeman, W., Szeliski, R., and Kang, S. B. (2006). Noise estimation from a single image. In *Proc. CVPR*, volume 1, pages 901–908. → pages 66

Liu, R., Li, Z., and Jia, J. (2008). Image partial blur detection and classification. In *Proc. CVPR*, pages 1–8. → pages 48

Loomis, J. M. and Nakayama, K. (1973). A velocity analogue of brightness contrast. *Perception*, 2:425–428. → pages 36

Lucy, L. (1974). An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79(6):745–754. → pages 45, 129

Luft, T., Colditz, C., and Deussen, O. (2006). Image enhancement by unsharp masking the depth buffer. *ACM Trans. Graph.*, 25(3):1206–1213. → pages 62, 63

Lumsdaine, A. and Georgiev, T. (2009). The focused plenoptic camera. In *In Proc. IEEE ICCP*, pages 1–8. → pages 42

Mantiuk, R., Krawczyk, G., Myszkowski, K., and Seidel, H. (2004). Perception-motivated high dynamic range video encoding. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 23(3):733–741. → pages 40

Mather, G. and Smith, D. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception*, 31(10):1211–1219. → pages 24

Moghaddam, M. E. and Jamzad, M. (2007). Linear Motion Blur Parameter Estimation in Noisy Images Using Fuzzy Sets and Power Spectrum. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9. → pages 48, 49

Nayar, S. and Mitsunaga, T. (2000). High dynamic range imaging: Spatially varying pixel exposures. *International Conference on Computer Vision*, 1. → pages 123

Neycenssac, F. (1993). Contrast enhancement using the laplacian-of-a-gaussian filter. *CVGIP: Graphical Models and Image Processing*, 55(6):447 – 463. → pages 33, 60, 105, 115

Ng, R. (2005). Fourier slice photography. *ACM Trans. Graph.*, 24:735–744. → pages 42

Ng, R., Levoy, M., Bredif, M., Duval, G., Horowitz, M., and Hanrahan, P. (2005). Light field photography with a hand-held plenoptic camera. Technical Report CSTR 2005-02, Stanford University. → pages xi, 42, 43

Nowak, R. D. and Baraniuk, R. G. (1998). Adaptive weighted highpass filters using multiscale analysis. *IEEE Trans Image Process*, 7:1068–1074. → pages 62

O'Brien, V. (1958). Contour perception, illusion and reality. *Journal of the Optical Society of America*, 49:112–119. → pages 32, 34

Oliva, A., Torralba, A., and Schyns, P. G. (2006). Hybrid images. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 25(3):527–532. → pages x, 12, 13

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, Mass. → pages 13, 14, 18

Pattanaik, S., Ferwerda, J., Fairchild, M., and Greenberg, D. (1998). A multiscale model of adaptation and spatial vision for realistic image display. In *Proceedings of ACM SIGGRAPH*, pages 287–298. → pages 40

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America*, 7(10):2032–2040. → pages 27

Peli, E., Arend, L., and Labianca, A. T. (1996). Contrast perception across changes in luminance and spatial frequency. *J Opt Soc Am A Opt Image Sci Vis*, 13:1953–1959. → pages 30

Peli, E., Yang, J. A., Goldstein, R., and Reeves, A. (1991). Effect of luminance on suprathreshold contrast perception. *J Opt Soc Am A*, 8:1352–1359. → pages 30

Pentland, A. P. (1987). A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9:523–531. → pages 79

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639. → pages 52

Polesel, A., Ramponi, G., and Mathews, V. (2000). Image enhancement via adaptive unsharp masking. *IEEE Trans. Image Proc.*, 9(3):505 –510. → pages 61

Popkin, T., Cavallaro, A., and Hands, D. (2010). Accurate and efficient method for smoothly space-variant Gaussian blurring. *IEEE Trans. Img. Proc.*, 19(5):1362–1370. → pages 79, 95

Poynton, C. (2003). *Digital Video and HDTV: Algorithms and Interfaces*. Morgan Kaufmann. → pages 31

Pritch, Y., Kav-Venaki, E., and Peleg, S. (2009). Shift-map image editing. In *ICCV'09*, pages 151–158, Kyoto. → pages 65

Purves, D., Shimpi, A., and Lotto, R. (1999). An empirical explanation of the Cornsweet effect. *Journal of Neuroscience*. → pages 36

Purves, D., Williams, S. M., Nundy, S., and Lotto, R. B. (2004). Perceiving the intensity of light. *Psychol Rev*, 111:142–158. → pages 36

Rahman, Z. (1995). Properties of a center/surround retinex part one: Signal processing design. *NASA Contractor Report*, 198194. → pages 39

Rahman, Z., Jobson, D., and Woodell, G. (1996a). Multi-scale retinex for color image enhancement. *Image Processing, 1996. Proceedings., International Conference on*, 3. → pages 39

Rahman, Z., Jobson, D., and Woodell, G. (1996b). A multiscale retinex for color rendition and dynamic range compression. *SPIE Proceedings: Applications of Digital Image Processing XIX*, 2847. → pages 39

Ramponi, G., Strobel, N. K., Mitra, S. K., and Yu, T.-H. (1996). Nonlinear unsharp masking methods for image contrast enhancement. *Journal of Electronic Imaging*, 5(3):353–366. → pages 61

Raskar, R., Agrawal, A., and Tumblin, J. (2006). Coded exposure photography: motion deblurring using fluttered shutter. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 25(3):795–804. → pages 42

Reinhard, E., Khan, E. A., Akyüz, A. O., and Johnson, G. M. (2008). *Color Imaging: Fundamentals and Applications*. A K Peters, Ltd. → pages 40

Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic tone reproduction for digital images. *ACM Transactions on Graphics (special issue SIGGRAPH 2002)*, 21(3):267–276. → pages 39

Richardson, W. (1972). Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1). → pages 45, 129

Ritschel, T., Smith, K., Ihrke, M., Grosch, T., Myszkowski, K., and Seidel, H.-P. (2008). 3D unsharp masking for scene coherent enhancement. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, 27(3). → pages 36, 63, 111

Robertson, M., Borman, S., and Stevenson, R. (1999). Dynamic range improvements through multiple exposures. In *Proceedings of International Conference on Image Processing (ICIP) 1999*, pages 159–163. → pages 123

Rooms, F., Pizurica, A., and Philips, W. (2001). Estimating image blur in the wavelet domain. In *International Conference on Acoustics, Speech, and Signal Processing*. → pages 48

Rosenfield, M. and Abraham-Cohen, J. A. (1999). Blur sensitivity in myopes. *Optom Vis Sci*, 76:303–307. → pages 23

Rosenfield, M., Hong, S. E., and George, S. (2004). Blur adaptation in myopes. *Optom Vis Sci*, 81:657–662. → pages 23

Rouf, M., Mantiuk, R., Heidrich, W., Trentacoste, M., and Lau, C. (2011). Glare encoding of high dynamic range images. In *Computer Vision and Pattern Recognition (CVPR)*. → pages 135

Rubinstein, M., Gutierrez, D., Sorkine, O., and Shamir, A. (2010). A comparative study of image retargeting. *ACM Trans. Graph.*, 29:160:1–160:10. → pages 65

Rubinstein, M., Shamir, A., and Avidan, S. (2009). Multi-operator media retargeting. *ACM Trans. Graph.*, 28:23:1–23:11. → pages 65

Samadani, R., Mauer, T. A., Berfanger, D. M., and Clark, J. H. (2010). Image thumbnails that represent blur and noise. *IEEE Trans. Img. Proc.*, 19(2):363–373. → pages xi, xii, 53, 55, 57, 58, 68, 69, 70, 71, 72, 73, 82, 95, 96, 98, 99

Santella, A., Agrawala, M., DeCarlo, D., Salesin, D., and Cohen, M. (2006). Gaze-based interaction for semi-automatic photo cropping. In *Proc. CHI*, pages 771–780. → pages 65

Shan, Q., Jia, J., and Agarwala, A. (2008a). High-quality motion deblurring from a single image. *ACM Trans. Graph.*, 27:73:1–73:10. → pages 45

Shan, Q., Li, Z., Jia, J., and Tang, C.-K. (2008b). Fast image/video upsampling. volume 27, pages 153:1–153:7, New York, NY, USA. ACM. → pages 64, 65

Shapley, R. M. and Tolhurst, D. J. (1973). Edge detectors in human vision. *J. Physiol. (Lond.)*, 229:165–183. → pages 32, 34

Shin, D.-H., Park, R.-H., Yang, S., and Jung, J.-H. (2005). Block-based noise estimation using adaptive gaussian filtering. *IEEE Trans. Consumer Electronics*, 51(1):218–226. → pages 66

Simoncelli, E. P. and Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *Proc. ICIP*, volume 1, pages 379–382. → pages 66

Smith, K. (2008). *Contours and Contrast*. PhD thesis, Max-Planck-Institut fur Informatik. → pages 33, 35

Smith, K., Krawczyk, G., Myszkowski, K., and Seidel, H. (2006). Beyond tone mapping: Enhanced depiction of tone mapped HDR images. *Proceedings of Eurographics*, 25(3):427–438. → pages 63

Smith, K., Landes, P., Thollot, J., and Myszkowski, K. (2008). Apparent greyscale: A simple and fast conversion to perceptually accurate images and video. *Proceedings of Eurographics*. → pages 63

Subbarao, M. (1992). Radiometry. chapter Parallel depth recovery by changing camera parameters, pages 340–346. Jones and Bartlett Publishers, Inc., , USA. → pages 79

Suh, B., Ling, H., Bederson, B. B., and Jacobs, D. W. (2003). Automatic thumbnail cropping and its effectiveness. In *Proc. UIST*, pages 95–104. → pages 65

Sullivan, G. and Georgeson, M. (1977). The missing fundamental illusion: Variation of spatio-temporal characteristics with dark adaptation. *Vision Research*, 17(8):977–981. → pages 33, 34, 110, 119

Thomos, N., Boulgouris, N. V., and Strintzis, M. G. (2006). Optimized transmission of JPEG2000 streams over wireless channels. *IEEE Transactions on image processing*, 15(1). → pages 130

Tiippana, K. and Näsänen, R. (1999). Spatial-frequency bandwidth of perceived contrast. *Vision Resarch*, 39(20):3399–3403. → pages 30

Tolhurst, D. J. (1972). On the possible existance of edge detector neurones in the human visual system. *Vision Res.*, 12:797–804. → pages 34

Tong, H., Li, M., Zhang, H., and Zhang, C. (2004). Blur detection for digital images using wavelet transform. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 17–20. → pages 48

Trentacoste, M., Lau, C., Rouf, M., Mantiuk, R., and Heidrich, W. (2010a). Defocus techniques for camera dynamic range expansion. In *Proceedings of Human Vision and Electronic Imaging XXI*, page 75370. → pages iv

Trentacoste, M., Mantiuk, R., and Heidrich, W. (2010b). Quality-preserving image downsizing. SIGGRAPH Student Research Competition Poster. → pages iv

Trentacoste, M., Mantiuk, R., and Heidrich, W. (2011a). Blur-aware image downsampling. *Computer Graphics Forum (Eurographics)*. → pages iii, iv

Trentacoste, M., Mantiuk, R., and Heidrich, W. (2011b). Scale-dependent perception of countershading. In *To Appear*. → pages iii

Trentacoste, M., Mantiuk, R., and Heidrich, W. (2011c). Synthetic depth-of-field for mobile devices. In *To Appear*. → pages iii, iv

Trussell, H. J. and Fogel, S. (1992). Identification and restoration of spatially variant motion blurs in sequential images. *IEEE Transactions on Image Processing*, 1:123–126. → pages 49

van den Brink, G. and Kleemink, C. J. (1976). Luminance gradients and edge effects. *Vision Research*, 16:155–159. → pages 35

Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., and Tumblin, J. (2007). Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*. → pages 42, 127, 129, 131, 134

Vishwanath, D. and Blaser, E. (2010). Retinal blur and the perception of egocentric distance. *J Vis*, 10:26. → pages 24

Wachtler, T. and Wehrhahn, C. (1997). The Craik-O'Brien-Cornsweet illusion in colour: Quantitative characterisation and comparison with luminance. *Perception*, 26:1423–1430. → pages 36

Walker, P. and Powell, D. J. (1974). Lateral interaction between neural channels sensitive to velocity in the human visual system. *Nature*, 252:732–733. → pages 36

Wang, B. and Ciuffreda, K. J. (2005a). Blur discrimination of the human eye in the near retinal periphery. *Optom Vis Sci*, 82:52–58. → pages 21

Wang, B. and Ciuffreda, K. J. (2005b). Foveal blur discrimination of the human eye. *Ophthalmic Physiol Opt*, 25:45–51. → pages 20, 21

Wang, B., Ciuffreda, K. J., and Irish, T. (2006a). Equiblur zones at the fovea and near retinal periphery. *Vision Res.*, 46:3690–3698. → pages 21, 22

Wang, B., Ciuffreda, K. J., and Vasudevan, B. (2006b). Effect of blur adaptation on blur sensitivity in myopes. *Vision Research*, 46(21):3634 – 3641. → pages 22

Wang, Y.-P., Wu, Q., Castleman, K. R., and Xiong, Z. (2001). Image enhancement using multiscale differential operators. In *IEEE Proc. of the Acoustics, Speech, and Signal Processing*, pages 1853–1856. → pages 62

Ware, C. and Cowan, W. (1983). The chromatic Cornsweet effect. *Vision Res*, 23(10):1075. → pages 35

Watson, A. (1987). The cortex transform: Rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing*, 39:311–327. → pages 16

Webster, M. A., Georgeson, M. A., and Webster, S. M. (2002). Neural adjustments to image blur. *Nat. Neurosci.*, 5:839–840. → pages 23

Webster, M. A., Webster, S. M., MacDonald, J., and Bahradwadj, S. R. (2001a). Adaptation to blur. In B. E. Rogowitz and T. N. Pappas, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4299 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 69–78. → pages 23

Webster, S. M., Webster, M. A., Taylor, J., Jaikumar, J., and Verma, R. (2001b). Simultaneous blur contrast. In B. E. Rogowitz and T. N. Pappas, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4299 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 414–422. → pages 24

Wiener, N. (1964). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge, Mass. → pages 45, 129

Witkin, A. (1987). Scale-space filtering. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*. → pages 51

Wuerger, S. M., Owens, H., and Westland, S. (2001). Blur tolerance for luminance and chromatic stimuli. *J Opt Soc Am A Opt Image Sci Vis*, 18:1231–1239. → pages 22

Xiangjun, L. and Jianfei, C. (2007). Robust transmission of JPEG2000 encoded images over packet loss channels. In *IEEE International Conference on Multimedia*, pages 947–950. School of Computer Engineering, Nanyang Technological University. → pages 130

Yang, D., Gamal, A., Fowler, B., and Tian, H. (1999). *IEEE Journal of Solid-State Circuits*, 34:1821–1834. → pages 122, 123

Yi, F., Iskander, D. R., and Collins, M. J. (2010). Estimation of the depth of focus from wavefront measurements. *Journal of Vision*, 10(4). → pages 86

Yoshida, A., Ihrke, M., Mantiuk, R., and Seidel, H. (2008). Brightness of the glare illusion. In *Proc. of Aymposium on Applied Perception in Graphics and Visualization*, pages 83–90. ACM. → pages 86

Young, I. T. and van Vliet, L. J. (1995). Recursive implementation of the gaussian filter. *Signal Process.*, 44:139–151. → pages 76

Yuan, L., Sun, J., Quan, L., and Shum, H.-Y. (2008). Progressive inter-scale and intra-scale non-blind image deconvolution. *ACM Trans. Graph.*, 27:74:1–74:10. → pages xi, 45, 46

Zhang, X. M. and Wandell, B. (1996). A spatial extension to CIELAB for digital color image reproductions. In *Proceedings of the SID Symposium Digest*, volume 27, pages 731–734. → pages 40

Zhou, C. and Nayar, S. (2009). What are good apertures for defocus deblurring? *IEEE Computational Photography*. → pages 44, 127, 129, 131, 132, 133, 134