

WASHINGTON UNIVERSITY

Sever Institute  
School of Engineering and Applied Science

Department of Computer Science and Engineering

Dissertation Examination Committee:

Cindy Grimm, Chair

Richard Abrams

Jeremy Buhler

Ann McNamara

Robert Pless

Gruia-Catalin Roman

PERCEPTION-GUIDED IMAGE MANIPULATION

by

Reynold Justin Bailey

A dissertation presented to the  
Graduate School of Arts and Sciences  
of Washington University in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

August 2007

Saint Louis, Missouri

copyright by  
Reynold Justin Bailey  
2007



# Acknowledgments

I would first like to thank my advisor, Dr. Cindy Grimm for her unwavering support and patience throughout my tenure at Washington University in St. Louis. In addition to providing invaluable input as I developed my research, she has also helped me to establish excellent collaborative relationships with other researchers. These relationships will no doubt continue to be fruitful throughout my academic career.

I would also like to express my gratitude to the other members of my dissertation committee: Dr. Gruia-Catalin Roman, Dr. Robert Pless, Dr. Jeremy Buhler, Dr. Richard Abrams, and Dr. Ann McNamara, for their guidance and willingness to work so closely with me.

Many thanks to my colleagues in the Media & Machines Lab for being available to discuss ideas, listen to practice talks, and help guide my research.

To my special friends Delvin, Donna, Jenner, Glynis, Nigel, and Shakir, thanks for just being there for me and my family. It was good to know that you were just a phone call away.

I must express my sincere gratitude to the Chancellor's Graduate Fellowship Program at Washington University in St. Louis for not only the financial support, but also for providing a community where I could interact socially and intellectually with Chancellor's Fellows from other disciplines. Thanks to Dean Sheri Notaro, Dean Robert Thach, and Amy Gassel for creating this nurturing environment.

My time at Wash-U would never have been a reality if it had not been for Dr. Randy Glean. Randy has not only been a very good friend, but an excellent mentor who saw my potential and encouraged me to pursue my doctorate at Wash-U. Randy, thanks for everything.

I wish to also thank my family, immediate and extended, for their love, prayers and support. I would especially like to say thanks to my parents, Reynold and Lorna Bailey, my sisters, Helen and Ronelle, and my brother, Larry, for their encouragement.

To my brothers and sisters-in-Christ at Oak Hill Baptist Church, thank you for embracing me. You became my family away from home. I truly appreciated your encouragement and prayers.

To my wife, Eucelia, thanks for your support, constant prayers, and always believing in me. This is as much your success as it is mine. I love you.

My little son Jonathan, I love you dearly. You have certainly made my life more interesting and fulfilling.

To my Savior and Lord, Jesus Christ, You have been my strength and my assurance when I was overwhelmed. Thank You for blessing me so abundantly and for seeing me through.

To God be the glory!

Reynold Justin Bailey

*Washington University in Saint Louis*

*August 2007*

To my wife Eucelia and my son Jonathan

# Contents

<b>Acknowledgments</b> . . . . .	<b>ii</b>
<b>List of Figures</b> . . . . .	<b>vi</b>
<b>Abstract</b> . . . . .	<b>vii</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Simulating Artistic Control of Apparent Depth . . . . .	3
1.1.1 Color as a Depth Cue . . . . .	4
1.1.2 Luminance as a Depth Cue . . . . .	6
1.1.3 Approach for Simulating Artistic Control of Apparent Depth . . . . .	6
1.2 The Effect of Object Color on Depth Ordering . . . . .	8
1.3 Conveying a Sense of Motion in Images . . . . .	9
1.4 Subtle Gaze Directing . . . . .	15
1.5 Layout of Dissertation . . . . .	16
<b>2 Background</b> . . . . .	<b>17</b>
2.1 Human Vision and Computer Graphics - Historical Perspective . . . . .	17
2.2 Human Visual System - Overview . . . . .	20
<b>3 Previous Work</b> . . . . .	<b>30</b>
3.1 Depth . . . . .	31
3.2 Motion . . . . .	32

3.3	Gaze Directing . . . . .	33
<b>4</b>	<b>Simulating Artistic Control of Apparent Depth . . . . .</b>	<b>35</b>
4.1	Notation . . . . .	36
4.2	Luminance Target Values . . . . .	37
4.3	Color Target Values . . . . .	38
4.4	Results . . . . .	39
4.5	Discussion . . . . .	44
<b>5</b>	<b>The Effect of Object Color on Depth Ordering . . . . .</b>	<b>46</b>
5.1	Method . . . . .	47
5.1.1	Participants . . . . .	47
5.1.2	Stimuli . . . . .	47
5.1.3	Procedure and Design . . . . .	48
5.1.4	Scoring . . . . .	49
5.2	Results . . . . .	49
5.3	Discussion . . . . .	55
<b>6</b>	<b>Conveying a Sense of Motion in Images . . . . .</b>	<b>60</b>
6.1	Segmentation . . . . .	60
6.2	Spatial Perturbation . . . . .	61
6.3	Results . . . . .	61
6.4	Discussion . . . . .	61
<b>7</b>	<b>Subtle Gaze Directing . . . . .</b>	<b>65</b>
7.1	Overview . . . . .	65
7.2	Description of Gaze Directing Technique . . . . .	67
7.3	Experimental Design . . . . .	70
7.3.1	Stimuli . . . . .	70
7.3.2	Participants . . . . .	70
7.3.3	Procedure . . . . .	72

7.4	Results . . . . .	73
7.4.1	Evaluation of Image Quality . . . . .	73
7.4.2	Effectiveness of Gaze Directing Technique . . . . .	75
7.5	Discussion . . . . .	83
<b>8</b>	<b>Conclusion . . . . .</b>	<b>86</b>
8.1	Simulating Artistic Control of Apparent Depth . . . . .	86
8.2	The Effect of Object Color on Depth Ordering . . . . .	87
8.3	Conveying a Sense of Motion in Images . . . . .	88
8.4	Subtle Gaze Directing . . . . .	88
<b>9</b>	<b>Future Work . . . . .</b>	<b>89</b>
9.1	Objective Tone-Mapping Evaluation . . . . .	90
9.2	Perception-Guided Image Compression . . . . .	97
	<b>Appendix A Eye-Tracking . . . . .</b>	<b>99</b>
	<b>Appendix B Visual Phenomena . . . . .</b>	<b>102</b>
B.1	Color Constancy . . . . .	102
B.2	Simultaneous Color Contrast . . . . .	105
	<b>Curriculum Vitae . . . . .</b>	<b>106</b>

# List of Figures

1.1	Photograph of macaque monkey . . . . .	2
1.2	Pictorial depth cues . . . . .	3
1.3	Examples of using color and luminance to create apparent depth in traditional art . . . . .	5
1.4	Possible physiological basis for color as a depth cue . . . . .	7
1.5	Example of a repeated asymmetric pattern . . . . .	10
1.6	Example of artwork that uses adjacent equiluminant colors . . . . .	11
1.7	Example of using motion blur in photography. . . . .	12
1.8	Example of using suggestive lines in cartoons to convey a sense of motion in a particular direction. . . . .	13
1.9	Example of using spatial imprecision to convey a sense of motion . . . . .	14
1.10	Example of a high detail image that lacks a sense of motion . . . . .	15
2.1	The earliest diagram of the human eye . . . . .	18
2.2	Timeline summarizing history of vision research . . . . .	19
2.3	Photograph of head-mounted display system and virtual reality driving simulator . . . . .	20
2.4	The human eye . . . . .	21
2.5	Cross-section of the retina . . . . .	22
2.6	Distribution of rods and cones in the retina . . . . .	23
2.7	Falloff in visual acuity . . . . .	24
2.8	The effect of visual acuity falloff and small fovea size on perception . . . . .	25

2.9	Processing that occurs at the ganglion cells . . . . .	27
2.10	Pathway of visual signals from the retina to the visual cortex . . . . .	28
2.11	Diagram of the lateral geniculate nucleus . . . . .	29
3.1	Example of creating the depth of field effect using image editing software .	31
3.2	Example of computer-generated cartoon-style renderings that suggest motion in a particular direction . . . . .	32
3.3	Stylized rendering of a photograph guided by gaze pattern of a single observer	34
4.1	Mimicking the artistic enhancement of apparent depth using luminance and color . . . . .	36
4.2	Target luminance and color values for a hypothetical input image . . . . .	40
4.3	Using technique to perform simple color replacement . . . . .	41
4.4	Mimicking the artistic reduction of apparent depth . . . . .	42
4.5	Attempting to change the perceived depth of individuals in a photograph by combining luminance-based manipulation with color-based manipulation in a very subtle manner . . . . .	43
5.1	Original teapot photograph, manually selected perceptually distinct colors, and adjusted equiluminant colors . . . . .	47
5.2	Stimuli used in depth ordering experiment . . . . .	48
5.3	Examples of a preference matrix and a cumulative preference matrix . . . .	50
5.4	Cumulative scores for each teapot and color patch against each of the 4 backgrounds . . . . .	51
5.5	Circular representation of color spectrum . . . . .	52
5.6	Results of a One-Sample T Test to establish if the mean subject score of any of the colored teapots varied significantly from the mean expected score . .	53
5.7	Results of a One-Sample T Test to establish if the mean subject score of any of the colored patches varied significantly from the mean expected score . .	54



5.8	Results of a One-Sample T Test to establish if the participants' depth preferences for a given colored teapot are significantly different than their depth preferences for the corresponding color patch . . . . .	56
5.9	Participants' response times for teapot data set and color patch data set against each of the 4 backgrounds . . . . .	57
6.1	Example of applying technique to convey a sense of motion over an entire image . . . . .	62
6.2	Example of applying technique to convey a sense of motion over an entire image . . . . .	63
6.3	Example of applying technique to convey a sense of motion in a specific region of an image . . . . .	64
7.1	Hypothetical image with current fixation region $F$ and predetermined region of interest $A$ . Inset illustrates saccadic masking. . . . .	68
7.2	Photograph of experiment setup and examples of luminance and warm-cool modulation . . . . .	69
7.3	The complete set of images used in gaze-directing study . . . . .	71
7.4	Analysis of quality scores for static and modulated images . . . . .	74
7.5	Images that received higher average quality scores under the modulated condition than under the static condition . . . . .	76
7.6	Quantifying gaze distribution . . . . .	77
7.7	Plot of all Pearson coefficient of correlation values obtained by comparing the gaze distribution of modulated images with the corresponding natural gaze distribution . . . . .	79
7.8	Gaze distributions for an image under static and modulated conditions . . .	81
7.9	Gaze distributions for an image under static and modulated conditions . . .	82
9.1	An example of tone-mapping . . . . .	91
9.2	Receptive field of a double-opponent cell in V1 . . . . .	95

9.3	Relationship between cones, pixels, and neurons . . . . .	96
9.4	Computing double-opponent response . . . . .	97
A.1	Overview of ViewPoint EyeTracker <sup>®</sup> system . . . . .	100
B.1	Example of color constancy . . . . .	103
B.2	Example of color constancy . . . . .	104
B.3	Example of simultaneous color contrast . . . . .	105

# ABSTRACT OF THE DISSERTATION

## Perception-Guided Image Manipulation

by

Reynold Justin Bailey

Doctor of Philosophy in Computer Science

Washington University in St. Louis, 2007

Cindy Grimm, Chairperson

This dissertation presents a novel approach to image editing and manipulation where the goal is to explicitly trigger certain visual cues. We refer to this as *perceptually meaningful image manipulation*. Existing image editing approaches typically do not take human visual perception into account, whereas our approach relies heavily on the principles of human visual perception. We present computer-based techniques for simulating artistic control of apparent depth in an image, conveying a sense of motion in an image, and subtly directing a viewer's gaze about an image. Our techniques combine elements of traditional art with research from various fields of science.

Our image-based technique for simulating artistic control of apparent depth works by automatically adjusting color or luminance in specific regions of an image. The color and luminance adjustment is based on several simple rules that artists use to convey and manipulate apparent depth in their work. Our technique for conveying a sense of motion in a visually static image introduces spatial imprecision in the image plane. This is also patterned after work done by traditional artists. It has been suggested that spatially

imprecise stimuli are processed differently by our foveal vision and our peripheral vision and that this difference in visual processing causes the image to appear to have a dynamic component. Finally, our subtle gaze directing technique exploits differences in visual acuity and processing speed between our foveal and peripheral vision. We present brief, subtle image-space modulations to the low acuity peripheral regions of the field of view. These stimuli, which are detected quickly, attract the slower, high acuity foveal vision to fixate on the modulated regions.

# Chapter 1

## Introduction

Traditional artists have developed numerous techniques for creating interesting visual effects. Although much of their work was done on flat, static surfaces, they were still able to create a sense of depth and motion in their paintings. Many of these artists had very little knowledge of the inner workings of the human visual system. Instead they viewed the human visual system as a black box and through experimentation, they learned to exploit its features. In a sense, they have reverse engineered the human visual system to learn what type of inputs elicit certain responses in the brain. Modern research from the fields of biology, physics, psychology, and neuroscience has given us better insight into the functioning of the human visual system. Additionally, there have been in-depth physiological studies of the visual system of the macaque monkey (see Figure 1.1), which is closely related to that of humans [?, ?, ?, ?, ?, ?]. Although the visual system is far from being fully understood, the knowledge we have gained, especially of the early stages of the visual pathway, is quite substantial. In fact, many of the visual effects created by artists can now be explained in terms of the features of the human visual system.

In this dissertation, we go beyond simply explaining these visual effects. We present a novel approach to computer-based image editing and manipulation where the goal is to explicitly trigger certain visual cues. We refer to this as *perceptually meaningful image manipulation*. Current image editing software, such as Adobe<sup>®</sup> Photoshop<sup>®</sup> [?] and The GNU Image Manipulation Program (The GIMP<sup>®</sup>) [?] are widely used by novices and



Figure 1.1: Photograph of macaque monkey. The visual system of the macaque is very similar to that of humans and has been studied extensively. Photograph courtesy of Viral Immunology Center, Georgia State University.

experts for tasks such as photo retouching, image authoring, and image composition. These packages provide powerful image editing and manipulation features that are based on sound mathematical principles along with a very basic knowledge of the human visual system. In contrast, our techniques rely heavily on the principles of human visual perception. We combine elements of traditional art with research from various fields of science to develop techniques for simulating artistic control of apparent depth in an image, conveying a sense of motion in an image, and subtly directing a viewer's gaze about an image. Additionally, this dissertation presents the results of an experiment that explores the relationship between color and perceived depth for realistic, colored objects. The remainder of this chapter presents introductory material for each of these contributions.

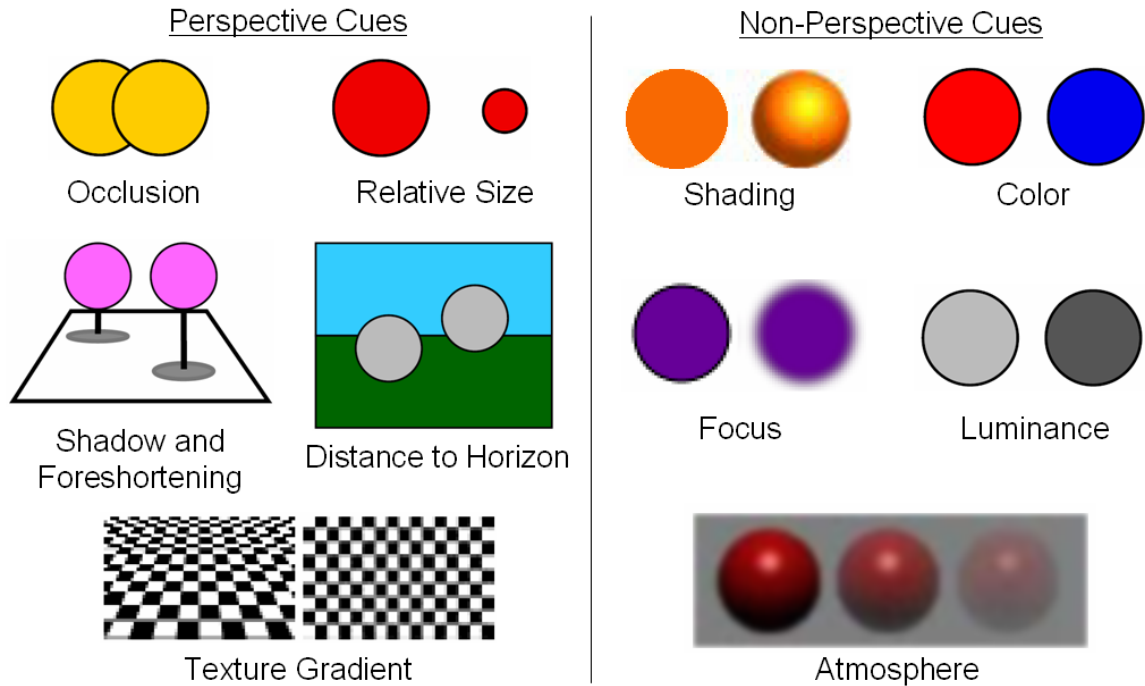


Figure 1.2: Pictorial depth cues. Perspective cues (left). Non-perspective cues (right). The perspective cues are directly related to the geometry of the scene and are generally more effective at conveying depth than the non-perspective cues. Figure adapted from Pfautz [?].

## 1.1 Simulating Artistic Control of Apparent Depth

Various sources of information in a scene contribute to our perception of depth. These are referred to as depth cues and are typically grouped into three categories: *pictorial*, *occulomotor*<sup>1</sup> and *stereo cues*<sup>2</sup> [?]. The pictorial depth cues are ones that we obtain from 2D images. These include perspective cues such as relative size, occlusion, distance to horizon, shadows, and texture gradient and non-perspective cues such as shading, color, focus, luminance, and atmospheric effects [?] (see Figure 1.2).

<sup>1</sup>Occulomotor depth cues result from changes to the physical configuration of the eyes. This includes information such as the angle of rotation of the eyes (convergence) and focal distance of the eyes (accommodation) [?].

<sup>2</sup>Stereo depth cues result from the angular disparity between the images in the left and right eye. The brain uses this disparity to compute the relative depth of points within the scene [?].

The perspective pictorial cues are generally more effective at conveying depth than the non-perspective cues [?, ?, ?]. In traditional art however, once the basic scene geometry has been established, the artist typically manipulates the subtler non-perspective cues (especially color and luminance) to change the apparent depth of objects in the scene. Figure 1.3 shows examples of traditional artwork where color and luminance have been used to create a sense of depth.

### 1.1.1 Color as a Depth Cue

Artists and designers refer to colors that appear closer to the red end of the visible spectrum as warm colors, and colors that appear closer to the blue end as cool colors<sup>3</sup> Early references to the color-depth relationship in literature refer to reds and oranges as *advancing colors* and blues as *receding colors* or *retiring colors* [?, ?]. The actual scientific study of the effect of color on perceived depth began in the latter half of the 19<sup>th</sup> century (see [?] and [?] for summaries). In most of these studies, human observers were asked to view physically equidistant, colored stimuli and compare them for relative depth. It was generally observed that warmer colors tend to appear closer to the viewer than cooler colors.

In order to account for these observed depth differences, several theories that are based on the physiology of the human visual system have been proposed. For example, it has been observed that the color sensitive cones in the retina exhibit a slight bias (stronger responses) to colors toward the warm end of the visible spectrum (see Figure 1.4(a)). Some physiologists have suggested that this bias may be strong enough to result in perceived depth differences between colors [?].

Another more accepted theory states that the difference in perceived depth is due to fact that shorter wavelengths of visible light are refracted more than longer wavelengths [?, ?]. As a result, equidistant sources of differing wavelengths cannot be simultaneously focused

---

<sup>3</sup>Physicists use the opposite terminology: on a black-body radiator, bluer colors are said to be hot while redder colors are said to be cool.





(a) Warm / cool colors. Renoir, Fille au chapeau de paille (1884).



(b) Luminance range. Seurat, Le Noeud Noir (1882).

Figure 1.3: Examples of using color and luminance to create apparent depth in traditional art.

onto the retina by the eye’s optical system. This is referred to as *chromatic aberration* and is illustrated in Figure 1.4(b).

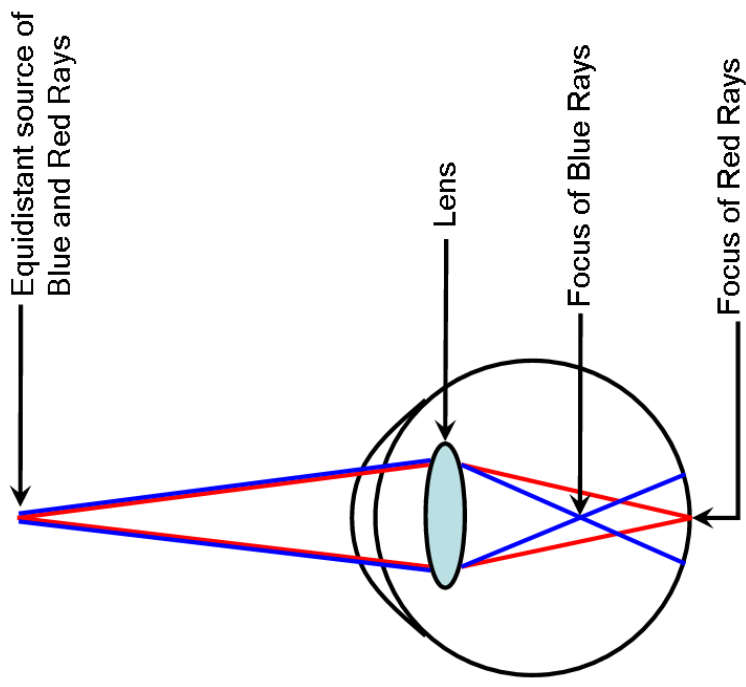
An example of the use of warm-cool colors to create a sense of depth is shown in Figure 1.3(a). The warm orange and red colors of the hat, hair and cheeks of the girl against the cooler light blue causes her face to appear more advanced than the rest of the painting. This enhances the sense of depth that we perceive. Compare this to her left elbow which appears flat because there is very little warm-cool color contrast in that area.

### 1.1.2 Luminance as a Depth Cue

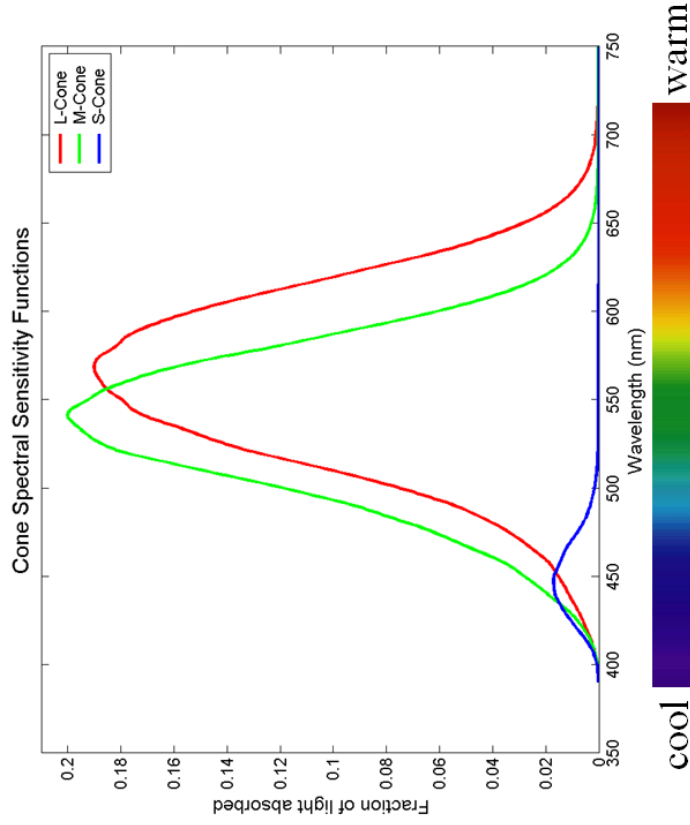
Artists also create apparent depth in their paintings by manipulating the luminance range (see Figure 1.3(b)). They do this by abruptly changing the luminance values across the boundary between two objects in specific regions of the painting. This is sometimes referred to as *haloing* or *counter-shading* and has the effect of increasing the local luminance contrast. O’Shea et al. [?] have shown that luminance contrast affects the perception of depth for both monocular and binocular viewing. They found that areas of lower contrast with the background appeared farther away while areas of higher contrast appeared closer. Guibal and Dresch [?] provide a brief summary of additional literature that describe the effect of luminance contrast on depth perception.

### 1.1.3 Approach for Simulating Artistic Control of Apparent Depth

Techniques for manipulating luminance values and the warmth (or coolness) of colors are well established in existing editing systems. However, their use for adjusting apparent depth in an image is typically not a defined operation. Users are still left with the time-consuming task of manually selecting and manipulating groups of pixels. Additionally, users have to ensure that any edits they make are seamless. In chapter 4 of this dissertation, we present a technique that eliminates these drawbacks. The user simply selects an object and specifies whether it should appear closer or further away. Our technique attempts to



(b) Chromatic aberration. Assumes eye is red accommodated. Adapted from [?]. Due to the refractive properties of light, equidistant sources of differing wavelengths cannot be simultaneously focused onto the retina. This phenomenon may result in perceived depth differences between warm and cool colors.



(a) Response of long, medium, and short wavelength sensitive cones to different wavelengths of light. Data obtained from experiments conducted on human observers [?, ?]. The cones respond slightly more intensely to warmer colors than to cooler colors. This bias may be strong enough to result in perceived depth differences between warm and cool colors.

Figure 1.4: Possible physiological basis for color as a depth cue.

achieve the desired depth change by automatically generating color and luminance target values for the object and/or background. The method for determining these target values is based on approaches developed by traditional artists to create similar depth effects in their work. Distance-weighted blending is used to ensure that the appearance of false edges is minimized.

## 1.2 The Effect of Object Color on Depth Ordering

Although color is considered to be one of the weaker depth cues, its impact on depth perception has been studied extensively by psychologists and other vision researchers [?, ?, ?, ?, ?, ?]. Most of these studies have shown that warm-colored stimuli tend to appear nearer in depth than cool-colored stimuli. The majority of these studies asked human observers to view physically equidistant, colored stimuli and compare them for relative depth. However, in most cases, the stimuli presented were rather simple: straight colored lines, uniform color patches, point light sources, or symmetrical objects with uniform shading. Additionally, the colors used were typically highly saturated. There are two main reasons for the use of such simple stimuli. First, many early researchers used the *weak observer theory* [?] as a basis for their experiments. This theory suggests that the various depth cues are processed by the visual system independently of each other and our final overall depth perception is simply a weighted combination of the contribution of the individual depth cues. Additionally, while most modern researchers quickly dismiss this theory, it is difficult to design effective experiments for the alternative *strong observer theory* [?, ?, ?]. The strong observer theory suggests that there are complex interactions between the various depth cues. The presence of one or more depth cues may either enhance or suppress the effect of other cues. Secondly, the use of simple stimuli by early researchers can be attributed to the lack of computing power, display technology, and accurate measuring devices.

Although such simple stimuli are useful for isolating and studying depth cues in certain contexts, they leave open the question of whether the human visual system operates

similarly for realistic objects. In chapter 5 of this dissertation, we present the results of an experiment designed to explore the color-depth relationship for realistic, colored objects with varying shading and contours. The stimuli for this study were created using the technique introduced in section 1.1 and described in detail in chapter 4. The results of this experiment can be used to guide the selection of color to further enhance the perceived depth of objects presented on traditional display devices and newer immersive virtual environments.

### 1.3 Conveying a Sense of Motion in Images

The human visual system is well equipped to detect actual motion in a scene [?, ?]. Certain types of cells along the visual pathway are directionally selective - responding to motion in a preferred direction. This type of motion perception enables us to determine the speed and direction of objects in the environment.

It is also possible for the visual system to infer motion from a static scene. Stimuli such as *static repeated asymmetric patterns*, commonly used in optical illusions (see Figure 1.5), are known to generate spurious motion signals. There is still a lot of debate over the exact neural mechanisms that cause this effect [?]. *Adjacent equiluminant colors* also trigger spurious motion signals [?, ?]. Two colors are said to be equiluminant if they have the same luminance. Figure 1.6 shows an example of a work of art that uses adjacent equiluminant colors to create an illusion of motion. The illusion of motion occurs because the visual system relies on luminance contrast across boundaries to accurately detect edges and to assign locations to objects in the scene. When there is low or no luminance contrast, the adjacent colors appear to float ambiguously.

Finally, several approaches have been developed to convey a sense of motion. For example, in photography, the camera's shutter is open for some period of time to allow light to reach the image sensor or film. If an object in the scene, or the camera, moves during this time then the resulting image appears blurred along the direction of the motion. This is called motion blur. Photographers sometimes use long exposure times or rig their cameras



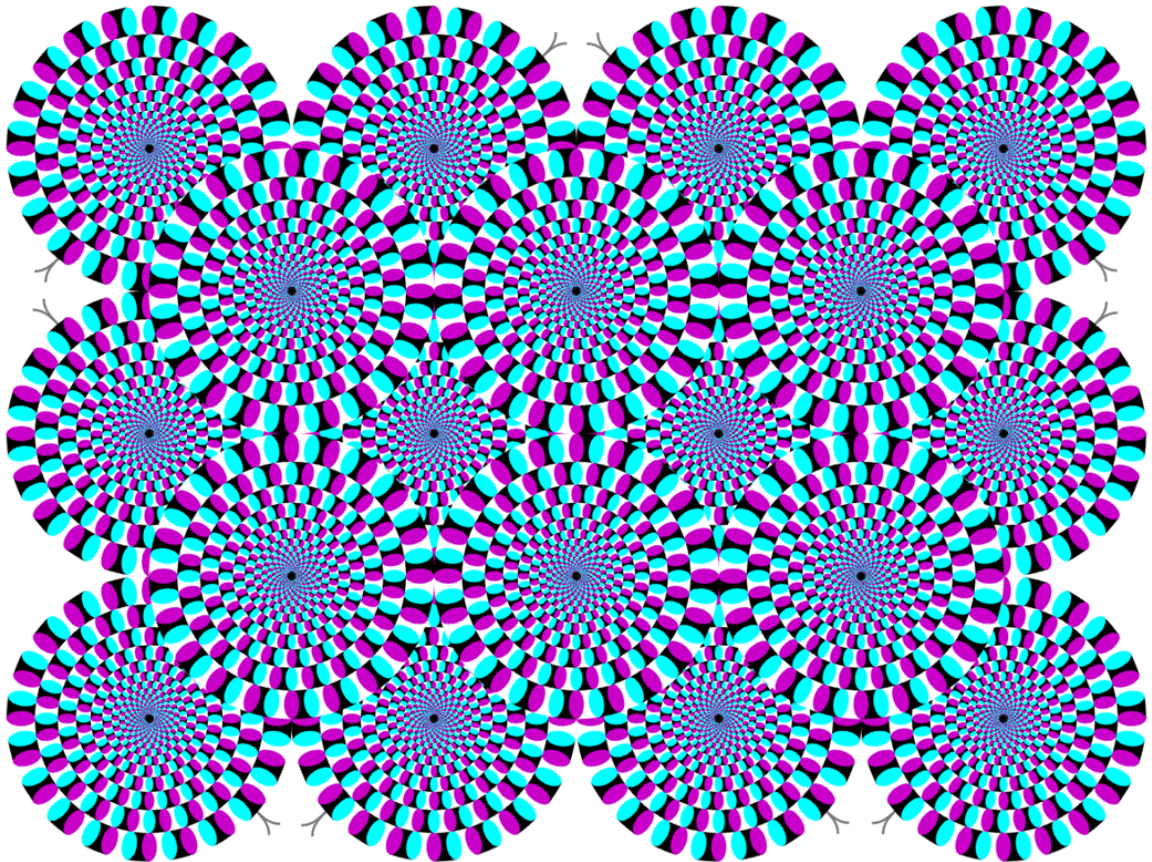
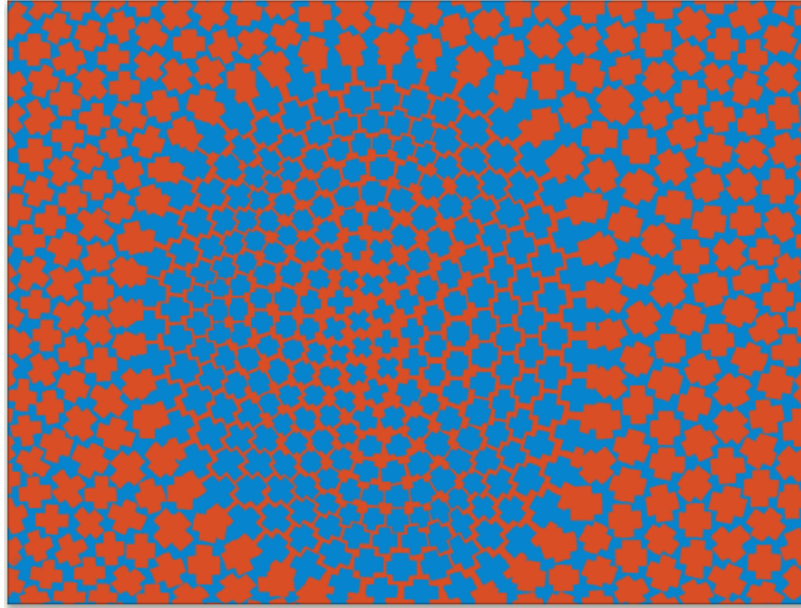
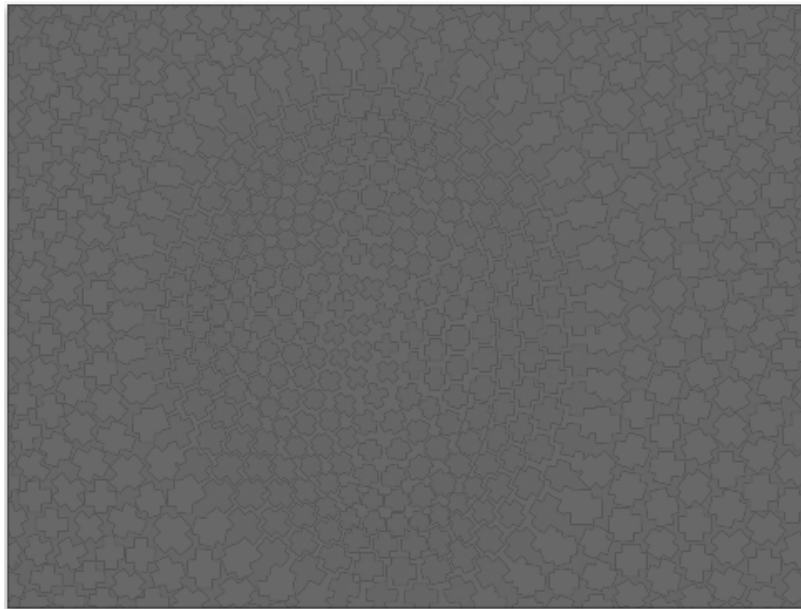


Figure 1.5: Repeated asymmetric patterns such as the one shown in this example are known to generate spurious motion signals along the visual pathway.



(a) Original



(b) Luminance

Figure 1.6: This example of artwork by Richard Anuszkiewicz, *Plus Reversed* (1960), uses adjacent equiluminant colors. (a) Original work: notice that there is a strong color contrast. (b) Corresponding luminance channel: notice that there is very little luminance contrast. The low luminance contrast makes it difficult for the visual system to detect the edges between the colors; hence they appear to float ambiguously.



Figure 1.7: The motion blur in this photograph conveys a sense of motion. Photographers create this effect by using long exposure times or by rigging their cameras to some moving device. Photograph courtesy of Bostan Alexander.

to some moving device to further enhance this effect (see Figure 1.7). Although motion blur has not been established as a cue in motion perception [?], it has been shown that it does provide significant information about the direction of motion [?].

Artists also seek to convey a sense of motion in their work. Cartoonists sometimes draw a few lines to suggest motion in a particular direction [?]. These are referred to as *action lines* or *speed lines*. Research has shown that humans are quite skilled at interpreting such lines [?, ?]. The use of suggestive lines in Figure 1.8 communicates the direction that the balls are moving. Similarly, traditional artists sometimes use *spatial imprecision* (misalignment of brush strokes) to convey a sense of motion [?] as illustrated in Figure 1.9.

It has been suggested that spatial imprecision conveys a sense of motion because of differences in our peripheral and foveal vision [?]. The fovea is a small indentation in the retina where we have the highest visual acuity. Our peripheral vision on the other hand,





Figure 1.8: The use of suggestive lines in cartoons convey a sense of motion. In this example, they suggest the direction in which the balls are moving. Cartoon courtesy of Royston Robertson.

which occupies the majority of our visual field, has much lower acuity. Additionally, research has shown that the peripheral vision responds (detects stimuli) faster than the foveal vision. This is due to the fact that the optic fibers that carry signals from the peripheral regions of the retina to the primary visual cortex for processing are fast-conducting, while the optic fibers that carry signals from the fovea are slower [?]. The theory is that when we first look at paintings where spatial imprecision is used, our fast acting, low acuity peripheral vision gives us a rough idea of where the brush strokes are in the scene. Mentally, we join these brush strokes together to form a complete picture. It is only upon closer scrutiny with our slower, high acuity foveal vision that we notice that the strokes are misaligned. The sense of motion is created because our visual system completes the picture differently with every glance (explanation adapted from *Vision and Art: The Biology of Seeing* [?]). Compare the painting in Figure 1.9 with the one in Figure 1.10. Although the scene depicted in Figure 1.10 has a lot of action, it lacks the sense of motion because it was painted in high detail.

In chapter 6 of this dissertation we present a simple two-step technique that conveys a sense of motion in a visually static image by introducing spatial imprecision to the image plane.



Figure 1.9: This painting demonstrates the use of misaligned brush strokes, also known as spatial imprecision, to convey a sense of motion. Claude Monet, *Rue Montorgueil - Paris, Festival of June 30, 1878* (1878).





Figure 1.10: Nicolas Poussin’s 1634 painting, *The Rape of the Sabine Women*, clearly portrays a lot of action; however it was done in high detail and lacks the strong sense of motion seen in Monet’s painting (Figure 1.9).

## 1.4 Subtle Gaze Directing

When viewing traditional images, the viewer’s gaze is guided solely by image content. For example, it is natural for humans to be drawn immediately to faces or other informative regions of an image [?]. Additionally, research has shown that our gaze is drawn to regions of high local contrast and high edge density [?, ?].

For digital images however, it is possible to develop display algorithms that provide a less passive approach to image viewing. Such algorithms can be used to dictate the pattern of eye-movements over an image. In chapter 7 of this dissertation, we demonstrate successful use of an algorithm we developed to accomplish gaze direction to specific image

locations, without substantially altering the experience of viewing an image. Our technique, which combines eye-tracking with subtle image-space modulation, exploits differences in visual acuity and stimulus response time (detection time) between the peripheral vision and the foveal vision. By presenting brief, subtle modulations to the peripheral regions of the field of view, the technique attracts the viewer’s foveal vision to the modulated region. Additionally, by monitoring saccadic velocity and exploiting the visual phenomenon of *saccadic masking*, modulation is automatically terminated before the viewer’s foveal vision enters the modulated region. Hence, the viewer is never actually allowed to scrutinize the stimulus that attracted her gaze. We call this paradigm *subtle gaze direction*.

## 1.5 Layout of Dissertation

The remainder of this dissertation is organized as follows: In chapter 2, we present a historical summary of the research on the human visual system and discuss its relationship to the field of computer graphics. Additionally, we provide an overview of the human visual system, highlighting several important features that we exploit to develop the perceptually meaningful image editing and manipulation techniques presented in this dissertation. Previous work by the computer graphics community in the areas of depth manipulation, creating a sense of motion, and gaze directing is presented in chapter 3. Our technique for simulating artistic control of apparent depth in an image is described in chapter 4. The results of our experiment to explore the color-depth relationship for realistic, colored objects with varying shading and contours are presented in chapter 5. Our technique that uses spatial imprecision to convey a sense of motion in visually static images is described in chapter 6, and our technique for subtle gaze direction is explained in chapter 7. We conclude in chapter 8 and discuss future work in chapter 9. Appendix A provides supplemental information about the eye-tracking system that we use and Appendix B describes two visual phenomena: *color constancy* and *simultaneous color contrast*.

## Chapter 2

# Background

In this chapter, we provide background information about the history of vision research and discuss its relationship to the field of computer graphics. Additionally, we present an overview of the human visual system.

### 2.1 Human Vision and Computer Graphics - Historical Perspective

The oldest theories about the nature of vision, light, and the eye date back to the early Greek philosophers<sup>1</sup> (see [?] for a review). Their speculation about vision tended to be egocentric. They believed that the eyes produced rays which interacted with objects in the environment to facilitate vision. These theories, though incorrect, persisted for several centuries, even spreading to the Arabic world. This is evidenced by the earliest known diagram of the eye which shows an area designated as the *visual spirit* (see Figure 2.1). This speculative and egocentric approach did not encourage the scientific study of vision. Additionally, with much of Europe being ravaged by poverty, plagues, and religious wars

---

<sup>1</sup>Empedocles (490 - 430 BC), Plato (428 - 348 BC), Aristotle (384 - 322 BC), Ptolemy (367 - 283 BC), and Euclid (325 - 265 BC).

during the Middle Ages (5<sup>th</sup> century - 15<sup>th</sup> century), very little was done to further our understanding of vision during that time [?].

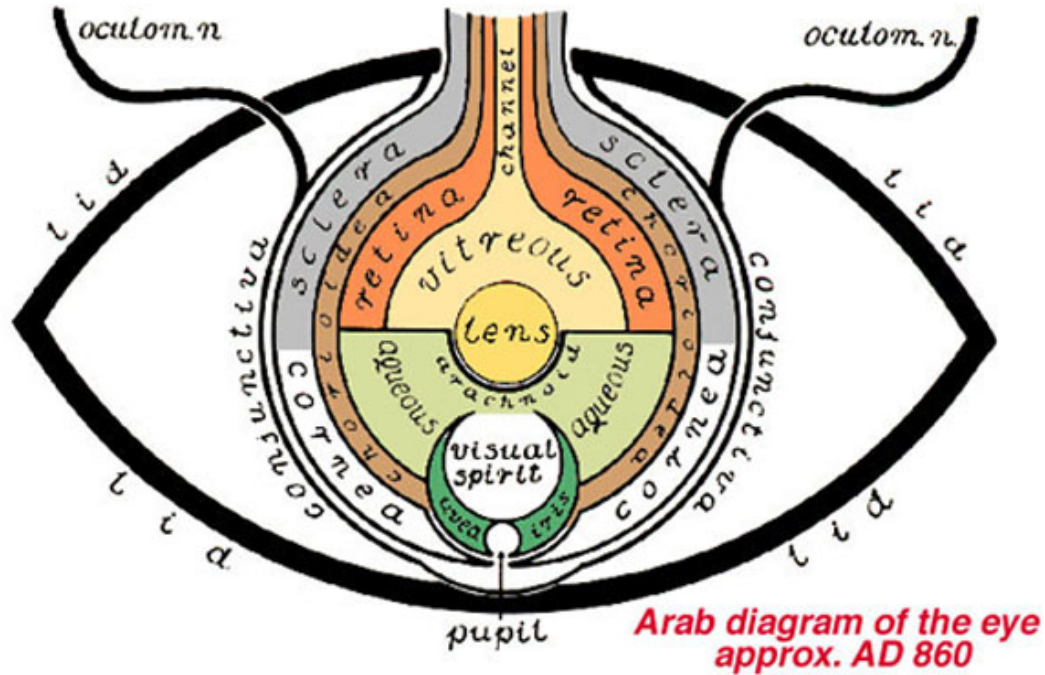


Figure 2.1: The earliest diagram of the human eye. The area designated as the *visual spirit* highlights the early egocentric theories concerning the human visual system.

Diagram courtesy of Polyak [?].

The actual scientific study of vision did not begin until the start of the Enlightenment Period in the early 17<sup>th</sup> century. In 1604, Johannes Kepler presented the first accurate explanation of the optics of the eye [?]. Later, in 1666, Isaac Newton conducted experiments with prisms to split white light into the colors of the rainbow [?]. Since that time, numerous researchers from the fields of biology, physics, psychology, physiology, and more recently neuroscience, have been involved in the scientific study of vision. The amount of research literature on the human visual system, and the rate at which new findings are published, is extraordinary. Figure 2.2 shows a timeline (not exhaustive) summarizing the history of vision research.

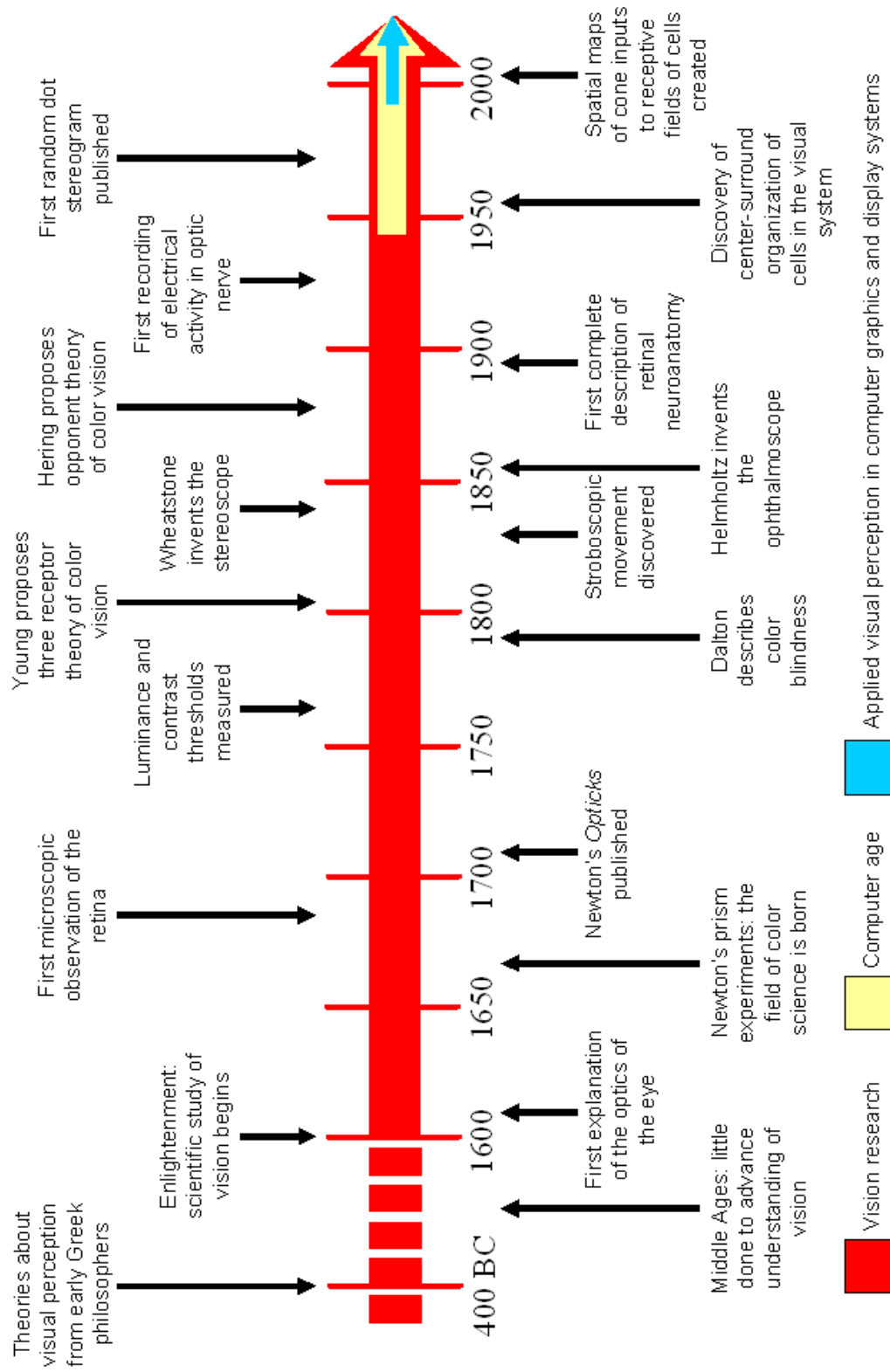


Figure 2.2: Timeline summarizing history of vision research. Data courtesy of Jack Yellott [?, ?], Brian Wagner and Donald Kline [?], and Wayne Carlson [?].



In 1946, the debut of the ENIAC [?] ushered in the age of the computer. Later, in 1960, William Fetter of the Boeing Company coined the term *computer graphics* [?]. It was not until the mid 1990's, however, that researchers became interested in the role that visual perception plays in display systems and computer graphics applications. Interest in this relatively new field (which has become known as *applied visual perception in computer graphics*) has grown rapidly due to advances in virtual reality, head mounted display systems and eye tracking technology (see Figure 2.3).



Figure 2.3: (a) Photograph of head-mounted display system. Courtesy of Broll et al. [?].  
 (b) Photograph of a virtual reality driving simulator. Courtesy of 9<sup>th</sup> International Conference on Virtual Reality.

## 2.2 Human Visual System - Overview

Much of the information about the physiology of the visual system presented in this section was gathered from *Fundamentals of Spatial Vision* [?] (see [?] and [?] for a more comprehensive discussion about the visual anatomy).

Figure 2.4(a) shows an external view of a human eye. The pupil is the central aperture that allows light to enter the eye. Its size is controlled by a pigmented ring of muscle called the iris. Figure 2.4(b) is a simplified diagram of the human eye. The eye has a roughly spherical shape with an average diameter of approximately 25mm in adults. It



has three basic layers. The sclera is a tough outer layer that helps to maintain the shape of the eyeball and also protects the internal structures. The middle layer is the choroid which provides the blood supply for the eye. The retina is the innermost layer and is lined with photosensitive cells (photoreceptors). The cornea, lens, and iris are the optical structures of the eye. They work together to focus light onto the retina.

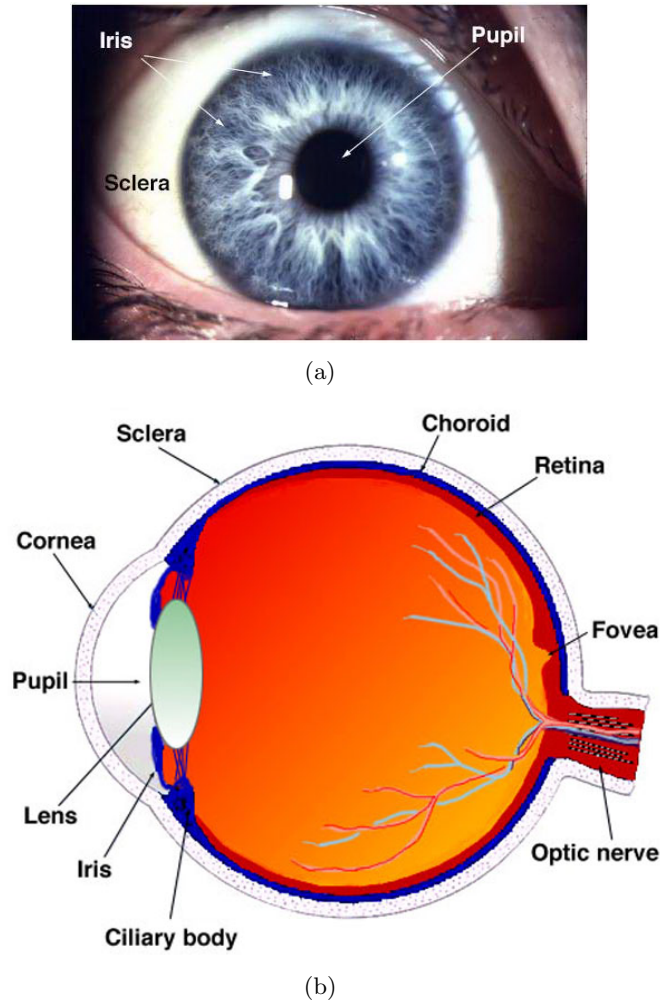


Figure 2.4: (a) External view of the human eye. (b) Diagram of the human eye. Images courtesy of John Moran Eye Center, University of Utah [?].

There are two classes of photoreceptors known as rods and cones. When photons from the visible spectrum are absorbed by these photoreceptors, a biochemical cascade is

triggered which results in an electrical signal (see [?] for a review). There are about 120 million rods in the human eye, which are very sensitive to light and provide achromatic vision in dim lighting conditions. On the other hand, there are about 8 million cones which provide color vision at higher (daylight) levels of illumination. There are three types of cones with peak sensitivities corresponding to different wavelengths of light. These three types of cones are commonly referred to as red, green and blue cones. This naming system is misleading because all cones are in fact sensitive to a wide range of wavelengths of light (see Figure 1.4(a)). Therefore, many researchers refer to them as long wavelength sensitive cones (L-cones), middle (or medium) wavelength sensitive cones (M-cones), and short wavelength sensitive cones (S-cones) respectively.

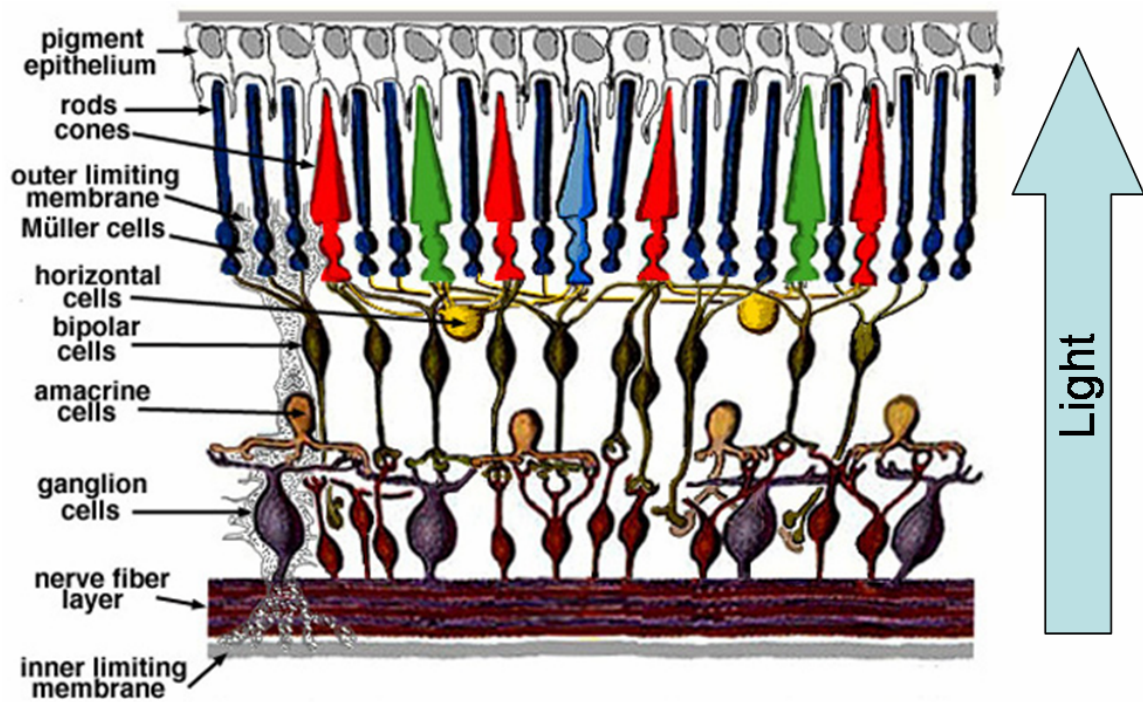


Figure 2.5: Cross-section of the retina. Light must first pass through several layers of cells before interacting with the rods and cones located toward the back of the retina. Diagram courtesy of John Moran Eye Center, University of Utah [?].

Figure 2.5 shows a cross-section of the retina. The rods and cones are located toward the back of the retina. Light must first pass through several layers of cells before interacting

with them. There is a small indentation in the retina called the fovea (see Figure 2.4(b)) where these layers of cells are pushed aside, allowing the photoreceptors to be directly exposed to the light. It is in this area that we have the highest visual acuity. The fovea has a diameter of  $1.5\text{mm}$  and subtends an angle of only 2 degrees of the visual field.

The photoreceptors are not distributed evenly across the retina [?]. Figure 2.6 shows the distribution of rods and cones as a function of angle (relative to the center of gaze). The density of cones is very high in the fovea (0 degrees) and falls off rapidly as the angle increases. The distribution of rods, on the other hand, is most dense around 20 degrees and there are no rods in the center of the fovea. There are also no photoreceptors on the optic disk. This is the area where the optic nerve is connected to the eye and is commonly referred to as the blind spot. The response of the rods is completely saturated for any level of illumination brighter than  $10\text{cd}/\text{m}^2$  (moonlight)<sup>2</sup> and can be ignored for the purposes of this research.

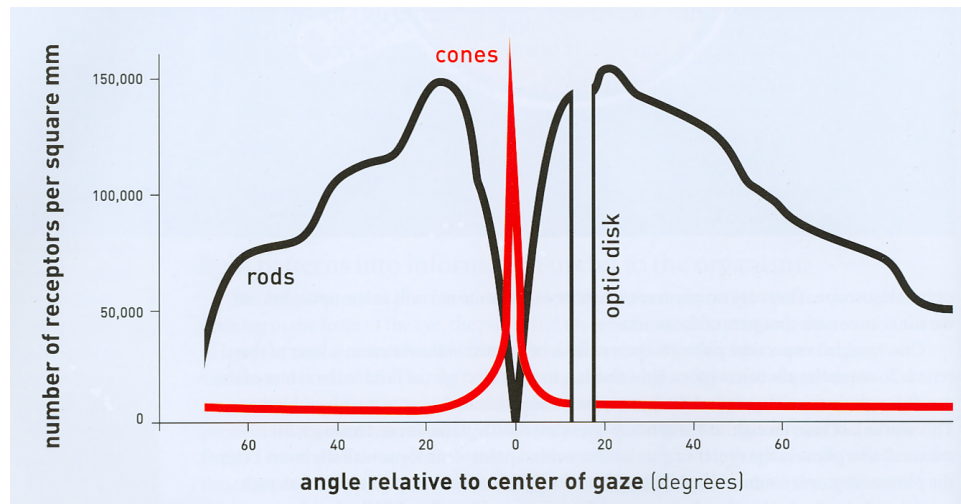


Figure 2.6: Distribution of rods and cones in the retina. The density of cones is very high in the fovea and falls off rapidly as the angle increases. The distribution of rods, on the other hand, is most dense around 20 degrees and there are no rods in the center of the fovea. Diagram courtesy of Margaret Livingstone, 2002 [?].

<sup>2</sup>The candela (cd) is the SI (International System of Units) unit of luminous intensity.

The distribution of cones directly affects how visual acuity varies with distance from the fovea (see Figure 2.7). This rapid falloff, coupled with the fact that the fovea subtends such a small visual angle, means that at any instant, only a small area in our field of view is seen in high resolution. We overcome this limitation by quickly and involuntarily scanning about the scene. This creates the impression that we are actually seeing the entire scene in high resolution (see Figure 2.8).

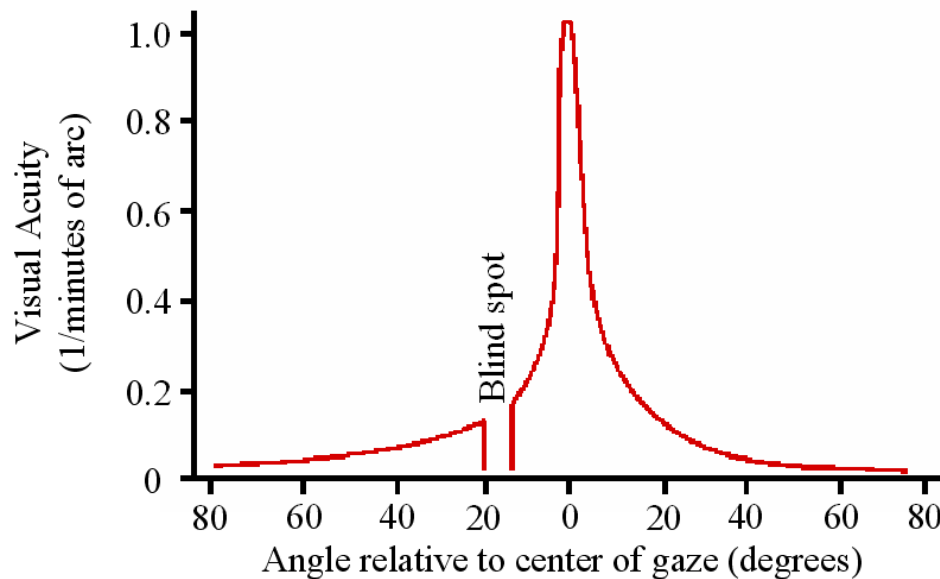


Figure 2.7: This graph shows the falloff in visual acuity as distance from the center of gaze increases. This falloff is directly related to the distribution of cones in the retina shown in Figure 2.6.

The signals from the cones are passed down to the ganglion cells. A single ganglion cell may be connected to one or more of the cones. These cones make up the receptive field of that ganglion cell. The receptive fields of the ganglion cells exhibit a center-surround organization [?]. A center-surround receptive field consists of a central disk called the *center* and a concentric ring around the center called the *surround*. The center and surround regions typically respond oppositely when exposed to light<sup>3</sup>. Center-surround cells are

<sup>3</sup>Light in the center of the receptive field either increases the rate of the cell response while light in the surround decreases the rate of cell response (called on-center cells) or light in the center of the receptive field decreases the rate of cell response while light in the surround increases the rate of cell response (called off-center cells).



(a)



(b)

Figure 2.8: The rapid falloff in visual acuity coupled with the small size of the fovea means that only a small region of a scene is seen in high resolution at any instant. We overcome this limitation by quickly and involuntarily scanning about the scene, which creates the impression that we are actually seeing the entire scene in high resolution.

(a) Scene shown in high detail. (b) Perception of scene at one instant.

often illustrated in literature by a small circle inside of a larger circle. The actual shape of the centers and surrounds are, however, less structured and vary greatly between different ganglion cells. In some cases the centers are even comprised of multiple disjoint regions. Despite this seemingly random structure of the receptive fields, the ganglion cells generally respond to the stimuli from the cones in one of two ways:

1. Color-opponent responses:- Color-opponent ganglion cells are typically found in vertebrates that can discriminate color. The color-opponent responses are generated by taking the difference of the signals from the various classes of cones. There are two types of color-opponent responses:
  - The red-green opponent response is formed by taking the difference of the signals from the L- and M- cones.
  - The blue-yellow opponent response is formed by taking the difference of the sum of the signals from the L- and M- cones and the S- cones.
2. Luminance response:- The luminance response is generated by adding the signals from the L- and M- cones.

Hence, the net effect of the processing that occurs at the ganglion cells is to combine the signals from the L-, M-, and S- cones to form a red-green opponent channel, a blue-yellow opponent channel, and a luminance channel. Figure 2.9 illustrates this process. This processing minimizes the correlation between the original cone signals and allows for efficient transmission of information to the brain [?, ?, ?].

The outputs of the ganglion cells form the individual fibers of the optic nerve. The optic nerves from each eye meet in the optic chiasm where some of the fibers cross each other. Research has shown that the fibers that receive signals from the left half of each retina (right visual field) are sent to a structure called the left lateral geniculate nucleus (LGN) and the fibers that receive signals from the right half of each retina (left visual field) are sent to the right LGN [?]. From there, the optic fibers carry the signals to the left and

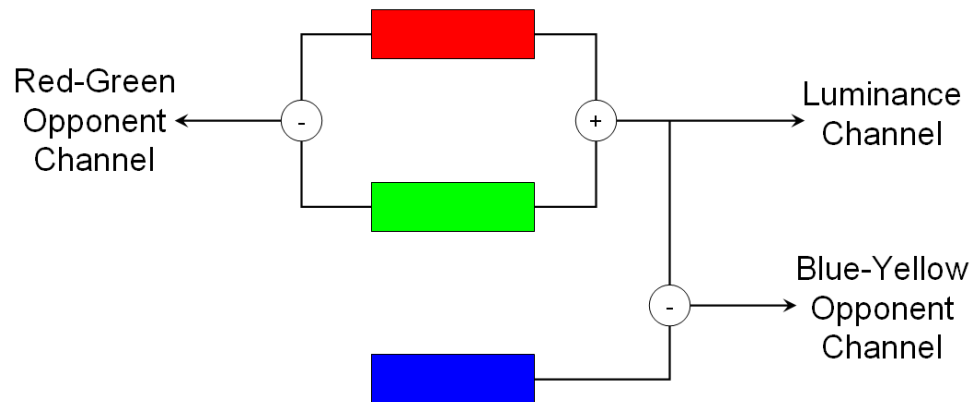


Figure 2.9: Processing that occurs at the ganglion cells. The signals from the three types of cones are combined into different channels. Red, green, and blue are used to represent the L-,M-, and S-cones respectively.

right visual cortex. Figure 2.10 shows the pathway that our visual signals take from the retina to the visual cortex.

The lateral geniculate nucleus (Figure 2.11) is a six layered structure. The two lowest layers are called the magnocellular layers. These layers receive signals from photoreceptors in the periphery of the retina. The next four layers are called the parvocellular layers. These layers receive signals from the foveal photoreceptors. It was discovered in 1966 that the optic fibers that carry signals from the periphery to the magnocellular layers are fast-conducting while the optic fibers that carry signals from the fovea to the parvocellular layers are slower [?]. These observations suggest that our visual system is comprised of two subsystems:

- A fast achromatic system with low visual acuity. This system processes achromatic (luminance) information from our peripheral vision. This system is sensitive to motion and is common among most mammals. Livingstone [?] refers to this system as the *where* system since it carries enough information for us to determine where an object is and how fast it is moving.
- A slower chromatic system with high visual acuity. This system processes color information from our fovea and is only well developed in higher primates. Livingstone [?]



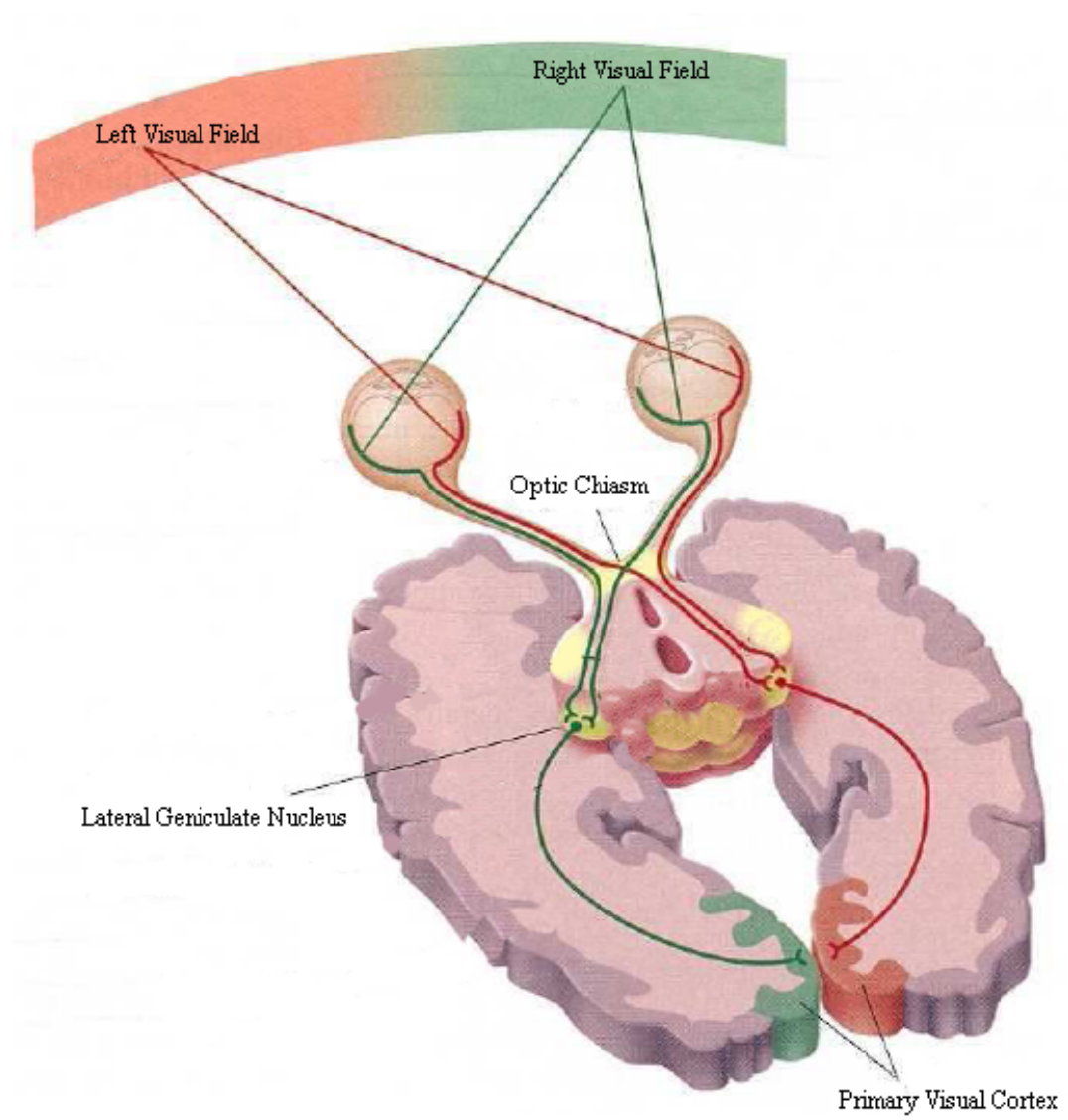


Figure 2.10: Pathway of visual signals from the retina to the visual cortex.



refers to this system as the *what* system since it processes high detail color information that allows us to determine what an object is.

Further evidence to support the existence of these two separate systems comes from lesion studies on monkeys and also from case studies of stroke victims, who have experienced loss of some or all functionality of one system while the other remains intact [?, ?].

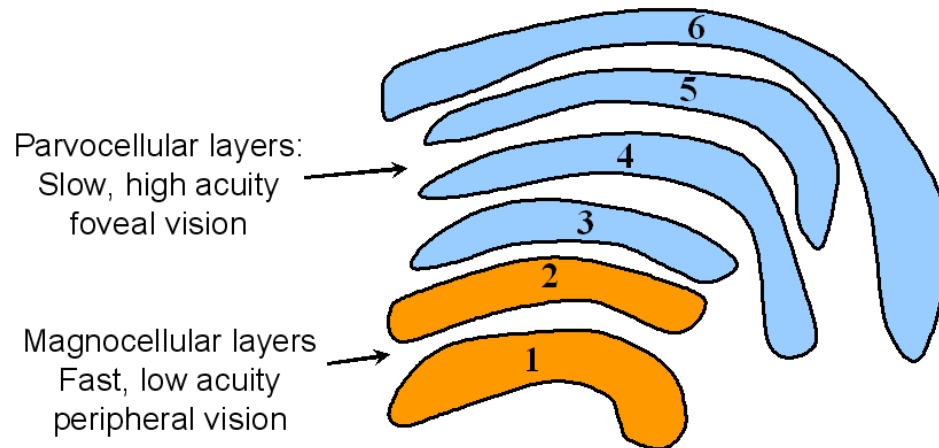


Figure 2.11: Diagram of the lateral geniculate nucleus. The two lowest layers process signals from the peripheral vision while the next four layers process signals from the foveal vision.

The two systems are not completely independent, but exactly how information is shared between them is not yet fully understood. Signals from the LGN are sent to the primary visual cortex where they are processed to extract information about low level visual features such as the orientation of edges in a scene. These processed signals are then sent on to increasingly complex areas toward the front of the brain. Eventually they combine with processed signals from other sensory organs in polymodal areas of the brain. This gives us significant information about our environment. The 20 billion neurons of the brain correlate this sensory information with our long- and short-term memories to allow us to react to a given situation.

## Chapter 3

# Previous Work

The field of photography predates the age of the computer by more than a century<sup>1</sup>. The early photographers edited their photographs by piecing together negatives in a dark room or by using airbrushes, paint, or ink. Even without the aid of computers and image editing software, skilled photographers were able to create images that deceived the viewer. Stalin's regime (1922 - 1953) is noted for falsifying photographs for propaganda purposes [?] (see [?] for a discussion on the ethics of image manipulation).

Today, most photographers rely on computer-based editing systems to manipulate their images. As discussed in chapter 1, modern image editing software provides a wide range of editing and manipulation features. However, these features typically do not take human visual perception into account. Over the last decade, there has been some work in the computer graphics community to address this need. This chapter discusses previous computer graphics approaches for manipulating apparent depth, creating a sense of motion, and directing viewer gaze.

A discussion of the previous work from the vision research community that focused on depth cues, motion cues, and visual attention cues has been incorporated into the introduction of this dissertation (see chapter 1).

---

<sup>1</sup>The first permanent photograph was created by French inventor, Joseph Nicéphore Niépce, in 1826.



Figure 3.1: An example of creating the depth of field effect using image editing software (a) Input image. (b) Output image. Images courtesy of Richard Rosenman-Advertising and Design.

### 3.1 Depth

The knowledge gained from the study of human depth perception has been used by computer graphics researchers to develop techniques for manipulating the apparent depth of objects in computer-generated scenes. These approaches are typically scene-based (relying on 3D representations where depth information is readily available). Examples include the introduction of atmospheric effects like haze or fog [?, ?], perspective changes such as the shape, size, or position of objects (or camera) [?], and level of detail (LOD) changes [?, ?].

Very little work has been done in the area of image-based depth manipulation. The most commonly used approach is to simulate the depth-of-field effect from traditional photography. This effect can be achieved in commercially available image editing packages by applying a sharpening filter to the foreground and a blurring filter to the background [?]. This has the effect of bringing different areas of an image in or out of focus (see Figure 3.1). There has also been some recent work by Gooch and Gooch which suggests that the perceived depth in an image can be enhanced by adding an artistic matting [?].

By automatically adjusting color and luminance in specific regions of images, our technique, described in chapter 4, simulates the artistic manipulation of apparent depth.

## 3.2 Motion

Computer graphics techniques for manipulating static images to trigger a motion response from the visual system typically rely on animation (changing the image over time). Color table animation [?], one of the earliest techniques developed, worked by swapping or cycling through colors in the frame buffer’s color table. Freeman et al. [?] later demonstrated that applying local filters to images and continuously varying their phase over time creates the illusion of motion. More recently, Chuang et al. [?] created animated clips from single images by applying time varying 2D displacement maps to various segmented regions of the original scene.

Several static techniques that do not rely on animation to convey a sense of motion have also been developed. The most common are non-photorealistic techniques for creating cartoon-style effects that suggest motion in a particular direction [?, ?, ?] (see Figure 3.2). Our technique for creating a sense of motion in images, described in chapter 6, can also be categorized as a non-photorealistic technique. We introduce spatial imprecision to the image plane. This gives the image a dynamic effect which may be explained by the differences in the foveal and peripheral vision as described in chapter 1.



Figure 3.2: Example of computer-generated cartoon-style renderings that suggest motion in a particular direction. Images courtesy of Kawagishi et al. [?].

### 3.3 Gaze Directing

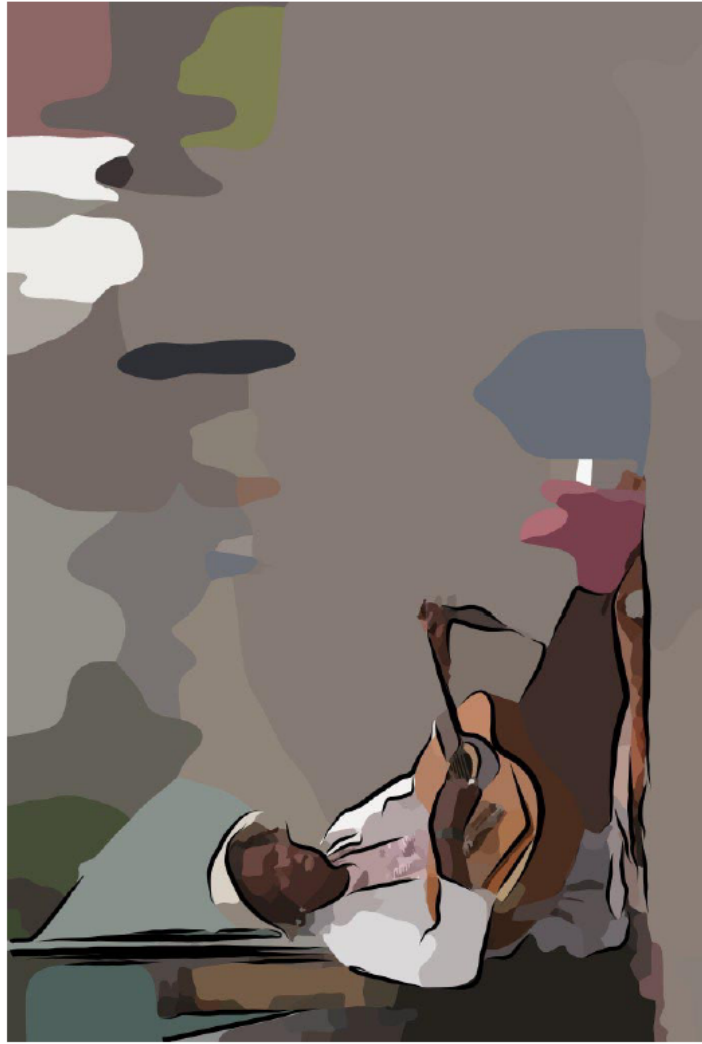
Eye tracking systems first emerged in the early 1900s [?, ?] (see [?] for a review of the history of eye-tracking). Until the 1980s, eye trackers were primarily used to collect eye movement data during psychophysical experiments. This data was typically analyzed after the completion of the experiments. During the 1980s, the benefits of real-time analysis of eye movement data were realized as eye-trackers evolved as a channel for human-computer interaction [?, ?]. More recently, real-time eye tracking has been used in interactive graphics applications [?, ?, ?] and large scale display systems [?] to improve computational efficiency and perceived quality. These systems use real-time eye-tracking to *follow* the viewer’s gaze. The technique described in chapter 7 combines real-time analysis of eye movement data with subtle image space modulation to *direct* the viewer’s gaze about a scene. Appendix A provides an overview of the eye-tracking system that we use for this research.

Previous computer graphics approaches for directing a viewer’s gaze have typically relied on the fact that the foveal vision is naturally drawn to regions of sharp focus or high detail [?]. The most commonly used approach to direct gaze in 2D images is to manipulate the depth-of-field of images using sharpening and blurring filters as described in section 3.1. The depth-of-field concept has also been applied to 3D scenes [?]. DeCarlo and Santella [?] used a different approach to direct viewer gaze. They recorded the gaze pattern of a single observer over an image and used it to direct the gaze of others. This was achieved by creating a stylized rendering of the image where only the areas attended to by the first observer are shown in high detail (see Figure 3.3). Cole et al. [?] applied similar stylized rendering techniques to direct gaze in 3D scenes.

Unlike these approaches which draw the viewer’s gaze by manipulating focus or level of detail, our technique, uses brief luminance or warm-cool image-space modulations presented to the peripheral regions of the field of view. Additionally, unlike the previous approaches, our technique attempts to preserve the viewing experience by not changing the overall appearance of the image.



(a)



(b)

Figure 3.3: Stylized rendering of a photograph guided by gaze pattern of a single observer. (a) Input image and recorded gaze data. (b) Output image. Images courtesy of DeCarlo and Santella [?].

## Chapter 4

# Simulating Artistic Control of Apparent Depth

In this chapter, we present a technique for simulating the artistic control of apparent depth of objects in an image. Our technique (see Figure 4.1) is based on the fact that traditional artists generally manipulate the apparent depth of objects in a scene, without changing the scene geometry, either by changing the warmth or coolness of colors or the luminance values in particular regions of the scene<sup>1</sup>. Our approach eliminates many of the drawbacks associated with traditional approaches to manipulating color and luminance (outlined in section 1.1.3). The user loads an image, selects an object<sup>2</sup>, and specifies whether the object should appear closer or further away. In an attempt to achieve the desired depth change, our system automatically determines luminance or color target values for the object and/or background. The method for determining these target values is based on approaches developed by traditional artists to create similar depth effects in their work (see section 1.1). Target values are assigned to the pixels that lie on the object boundary (i.e., object pixels

---

<sup>1</sup>Previous research that establishes color and luminance as depth cues is described in section 1.1.

<sup>2</sup>Any existing object selection (or (more generally) image segmentation) technique can be used (see [?] for a review). We use an implementation of the *Lazy Snapping* method [?].



Figure 4.1: Mimicking the artistic enhancement of apparent depth using luminance (left) and color (right). Input image (middle).

that lie closest to the edge) and to pixels that lie on the background boundary (i.e., background pixels that lie closest to the edge). These values are then propagated to every pixel in the image.

## 4.1 Notation

Let  $\mathbf{B}$  be the set of background pixels,  $\mathbf{O}$  be the set of object pixels,  $\mathbf{I} = \mathbf{O} \cup \mathbf{B}$  be the set of all image pixels,  $T_{lum}$  be the set of luminance target values, and  $T_{col}$  be the set of color target values. Let  $p_o \in \mathbf{O}$  be an object boundary pixel and  $p_b \in \mathbf{B}$  be the closest corresponding background boundary pixel. Let  $n$  be a user specified neighborhood size ( $n = 9$  for the images in this chapter) and let  $N_{p_o}$  and  $N_{p_b}$  be the  $n \times n$  neighborhoods about  $p_o$  and  $p_b$  respectively.



## 4.2 Luminance Target Values

Let

$$Avg_o = \frac{1}{sizeof\{N_{p_o} \cap O\}} \sum_{p_i \in \{N_{p_o} \cap O\}} p_i$$

and

$$Avg_b = \frac{1}{sizeof\{N_{p_b} \cap B\}} \sum_{p_i \in \{N_{p_b} \cap B\}} p_i$$

be the average neighborhood pixel values of  $p_o$  and  $p_b$ , respectively. Furthermore let

$$AvgLum_o = luminance(Avg_o)$$

and

$$AvgLum_b = luminance(Avg_b)$$

be the average neighborhood luminance values of  $p_o$  and  $p_b$ , respectively. Target luminance values  $T_{lum}$  are assigned to  $p_o$  and  $p_b$  using the following rules:

- Case 1 - Desired reduction of apparent depth by manipulating both object and background:

$$- T_{lum}(p_o) = T_{lum}(p_b) = \frac{1}{2}(AvgLum_o + AvgLum_b)$$

- Case 2 - Desired reduction of apparent depth by manipulating the object only:

$$- T_{lum}(p_o) = AvgLum_b$$

- Case 3 - Desired reduction of apparent depth by manipulating the background only:

$$- T_{lum}(p_b) = AvgLum_o$$

- Case 4 - Desired enhancement of apparent depth by manipulating either the object and/or background:

$$- \text{if } AvgLum_b \leq AvgLum_o$$

- \*  $T_{lum}(p_o) = 1.0$  (white)
- \*  $T_{lum}(p_b) = 0.0$  (black)
- if  $AvgLum_o < AvgLum_b$ 
  - \*  $T_{lum}(p_o) = 0.0$  (black)
  - \*  $T_{lum}(p_b) = 1.0$  (white)

Once boundary target values have been assigned, they can be propagated to all pixels in the image using any flood-fill algorithm. We use a fast breadth-first approach. A Gaussian blur filter is applied to the resulting target image. This gives smoothly varying target values and hence a more natural looking result once the operation is complete. Target values for all cases can be precomputed and stored as images. This reduces lag time, especially when processing large images.

When assigning target values for a desired reduction in apparent depth, it should be noted that we can substitute  $Avg_o$  and  $Avg_b$  for  $AvgLum_o$  and  $AvgLum_b$ , respectively. This has the same effect of reducing luminance contrast between the object and the background, however it does not limit us to grayscale target values. We found that this generally gives more aesthetically pleasing results and is the approach used for the images in this chapter. The top row of Figure 4.2 shows the target luminance values generated for a hypothetical image.

### 4.3 Color Target Values

In order to assign target color values, we simply modify the target luminance values  $T_{lum}$ . The user specifies a warm color  $w$  and a cool color  $c$ . Target color values  $T_{col}$  are assigned to  $p_o$  and  $p_b$  using the following rules:

- Case 1 - Desired reduction of apparent depth by manipulating both object and background:

$$- T_{col}(p_o) = T_{col}(p_b) = \frac{1}{2}(T_{lum}(p_o) + c)$$

- Case 2 - Desired reduction of apparent depth by manipulating the object only:

$$- T_{col}(p_o) = \frac{1}{2}(T_{lum}(p_o) + c)$$

- Case 3 - Desired reduction of apparent depth by manipulating the background only:

$$- T_{col}(p_b) = \frac{1}{2}(T_{lum}(p_b) + w)$$

- Case 4 - Desired enhancement of apparent depth by manipulating either the object and/or background:

$$- \text{if } AvgLum_b \leq AvgLum_o$$

$$* T_{col}(p_o) = \frac{1}{2}(T_{lum}(p_o) + w)$$

$$* T_{col}(p_b) = \frac{1}{2}(T_{lum}(p_b) + c)$$

$$- \text{if } AvgLum_o < AvgLum_b$$

$$* T_{col}(p_o) = \frac{1}{2}(T_{lum}(p_o) + c)$$

$$* T_{col}(p_b) = \frac{1}{2}(T_{lum}(p_b) + w)$$

We experimented with other approaches for assigning target color values including estimating the color temperature (position along the spectrum) of values in the image then traversing the spectrum to the right or left to make the target values warmer or cooler. Such approaches are generally computationally expensive and do not offer any significant advantage over the approach presented here. The bottom row of Figure 4.2 shows the target color values generated for a hypothetical image.

## 4.4 Results

The actual value  $V$  assigned to a pixel during this operation is a weighted combination of the target pixel value  $T_{lum}(p)$  (or  $T_{col}(p)$ ) and the original pixel value  $\mathbf{I}(p)$  at that point:

$$V(p) = (1 - f(d))\mathbf{I}(p) + f(d)T_{lum}(p)$$

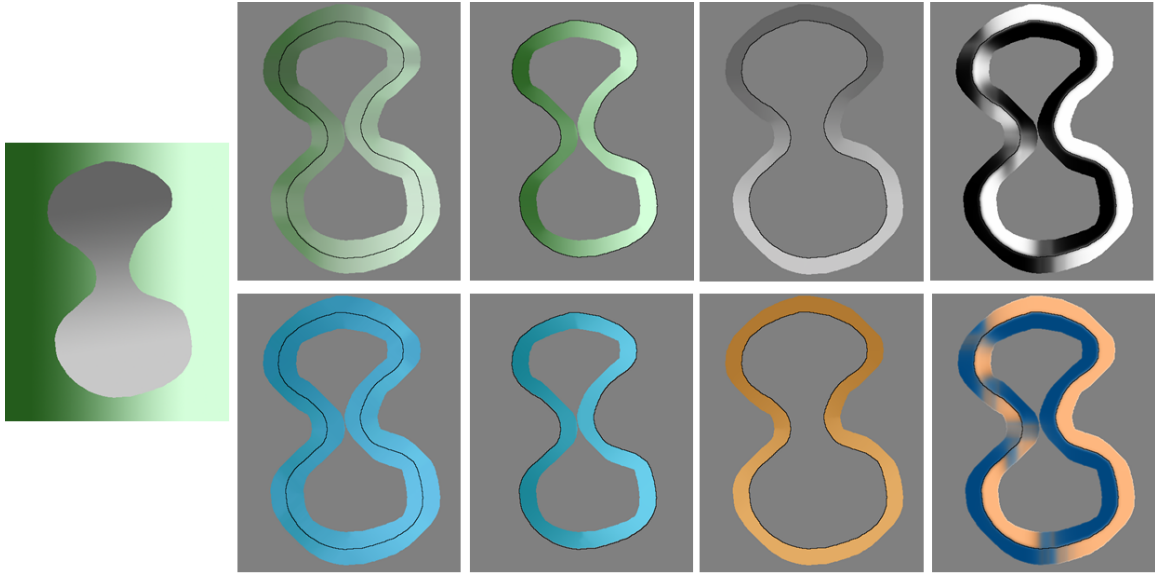


Figure 4.2: Hypothetical input image (left). Target luminance values (top row). Target color values (bottom row). Case 1, 2, 3, and 4 (left to right). The original edge is shown by a black line. Only a 60 pixel wide area of the target values are shown to better illustrate the relationship with the original image.

where the parameter  $d$ , is a user specified value representing desired depth, and  $f$  is some user specified falloff function of distance from every pixel to the edge between the object and background. We experimented with Gaussian, quadratic, and linear functions, all with pleasing results. Additionally, the user is allowed to select the distance over which blending occurs or have the system generate variable distances based on distance to the object skeleton.

With blending turned off and a zero falloff function, adjusting the parameter  $d$  simply results in color replacement as illustrated in Figure 4.3. Notice that the effect of luminance on perceived depth (middle column) is immediately obvious (brighter objects appear closer), however, the relationship between color and perceived depth (right column) is more subtle. In chapter 5 we describe a psychophysical study that we conducted using stimuli generated by this technique to explore the relationship between color and perceived depth for realistic colored objects with varying shading and contours.



Figure 4.3: Using technique to perform simple color replacement. Input image (left). Luminance manipulation (middle column). Color manipulation (right column).

By varying the distances and falloff function, we mimic the artistic manipulation of perceived depth of objects by adding undertones and halos. Figure 4.1 shows an example of enhancing perceived depth and Figure 4.4 shows an example of reducing perceived depth. Notice that drastically reducing perceived depth results in an interesting ghosting effect. This occurs because target values are propagated from the boundary to every pixel in the image. In this manner, we obtain an estimate of the pixel values behind the object. A similar effect can be achieved by performing a cutout of the object, using an image inpainting [?] or texture synthesis [?] approach to estimate the missing pixel values, and finally blending the object with the new pixels.

In the case of photorealistic editing, especially of human subjects, this type of image manipulation needs to be more subtle. Figure 4.5 shows an attempt to change the perceived depth of individuals in a photograph by combining luminance-based manipulation with color-based manipulation in a very subtle manner. This type of manipulation can be used to draw the viewer's focus to the individual who is furthest away from the camera [?]. The difference image (bottom left) highlights just how subtle these manipulations can be to result in a perceptual change.

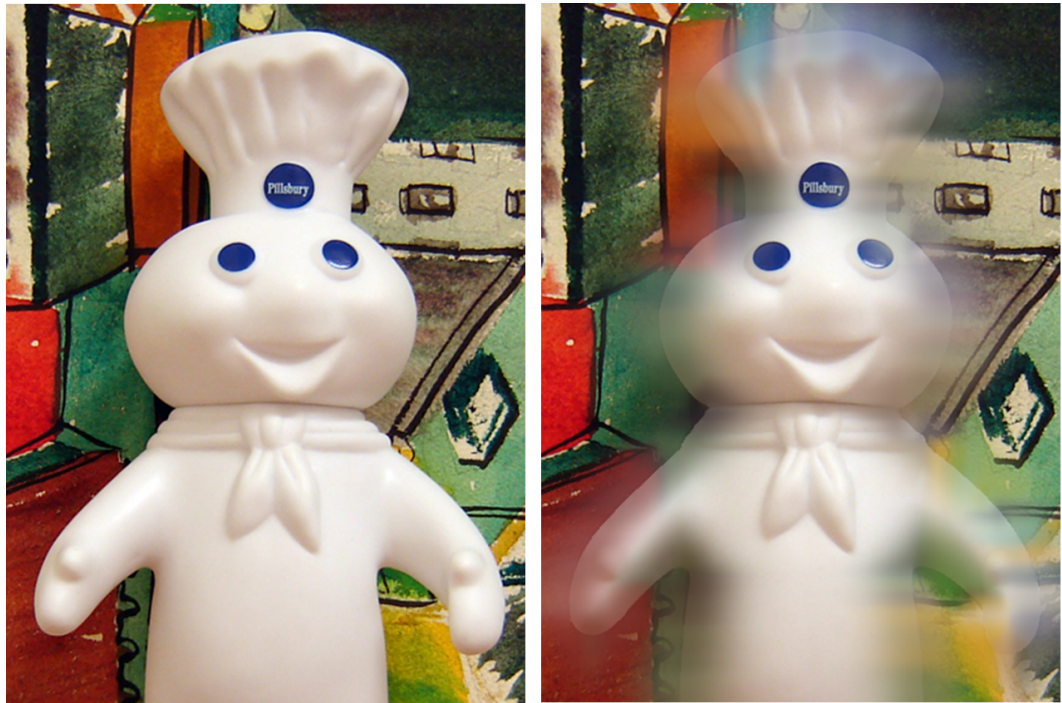


Figure 4.4: Mimicking the artistic reduction of apparent depth. Drastically performing this type of manipulation results in a ghosting effect. Input image (left). Output image (right).





Figure 4.5: Attempting to change the perceived depth of individuals in a photograph by combining luminance-based manipulation with color-based manipulation in a very subtle manner. This type of manipulation can be used to draw the viewer's focus to the individual who is furthest away from the camera. Input image (top left). Output image (top right). Difference image (lower left) shows that a warm reddish undertone was added around the face of the girl on the left and that the luminance of her face was increased slightly. A cooler gray-green undertone was added around the face of the girl on the right. Scaled difference image (lower right) emphasizes these changes for display and printing purposes.

## 4.5 Discussion

The technique presented in this chapter can be useful in post-production film editing as part of a suite of digital grading tools. The digital grading process involves adding subtle color tones and smoothing effects to various scenes to further enhance the emotion that is being conveyed and also to ensure that there is coherence between shots taken from multiple cameras at different times. Grading techniques can also be applied to specific regions of a scene to emphasize/deemphasize certain objects or characters. Consider a case where the character closest to the camera is not the one performing some important action. By manipulating the luminance and, to a lesser extent, the color across the boundary of each character, our technique can be used to add visual emphasis to the active character [?]. The traditional approach for accomplishing this involved having an artist manually adjust the luminance range and paint subtle undertones and halos around the characters.

One of the limitations of our technique is that poor selection of the warm and cool colors by the user may give rise to cue conflicts when performing color-based manipulation. Colors have an inherent luminance component, hence it is possible to select pairs of colors where the cooler color is brighter than the warmer color. In this case, since luminance is a stronger cue, the cooler color will appear closer to the viewer [?]. This phenomenon known as *depth reversal* may lead to unexpected results. One way to reduce the occurrence of this type of cue conflict is to limit the selection of warm and cool colors to only include equiluminant pairs. This, however, will severely limit the artistic freedom afforded to the user and will also eliminate pairs of colors where the color cue and luminance cue actually reinforce each other. A better approach and one possible avenue for future research is to develop a more robust method for color selection that eliminates color pairs that lead to cue conflicts. Ideally this method should also take into account the content of the scene since features such as gradients, the orientation of edges, or the presence (or absence) of a horizon also contribute to cue conflicts when 3D scenes are displayed on flat screens [?].



Another important area of future research is to develop and conduct experiments to test the effectiveness of this technique at manipulating perceived depth.

## Chapter 5

# The Effect of Object Color on Depth Ordering

In this chapter we describe an experiment that was conducted to explore the relationship between color and perceived depth for realistic, colored objects with varying shading and contours. Participants were presented with all possible pairs from a set of images of differently colored objects and were asked to select the object in each pair that appears closest to them. This approach is called a paired comparison test.

Ledda et al. [?] present a strong case for the use of this type of testing. Although the number of test pairs, and hence experiment time, grows quickly as the number of objects in the test set increases, they point out that the results of a paired comparison test do not suffer from the distortion that is common in ranking or rating approaches [?]. The objects were presented against 4 different uniform backgrounds of varying intensity levels. This allows us to analyze the impact of luminance contrast on the depth preferences recorded.



Figure 5.1: (a) Original teapot photograph courtesy of Jack Tumblin and Ankit Mohan. (b) Manually selected perceptually distinct colors. (c) Adjusted equiluminant colors.

## 5.1 Method

### 5.1.1 Participants

Fifteen students (4 females, 11 males), between the ages of 18 and 40 volunteered to participate in this study. They all reported normal or corrected-to-normal vision with no color vision abnormalities.

### 5.1.2 Stimuli

Stimuli were presented on a 17 inch monitor, operating at 60 Hz with a resolution of 1280 x 1024. The stimuli used in this experiment consisted of images of 7 colored teapots and 7 uniform colored patches. To create the stimuli, we began with a photograph of a teapot<sup>1</sup> (see Figure 5.1(a)) and performed segmentation using an implementation of the *Lazy Snapping* method [?]. Seven perceptually distinct colors were manually selected (see Figure 5.1(b)). The luminance of each of these colors was adjusted to match the average luminance of the original teapot. The resulting equiluminant colors are shown in Figure 5.1(c).

<sup>1</sup>This particular object was chosen because it appeared natural regardless of its color.

Each of these target colors was then applied to the teapot. This was done using our color replacement technique, described in chapter 4. This approach, which scales the target colors by the individual pixel luminance values of the original teapot, preserves the shading and specular highlights of the original teapot (see Figure 5.2(a)). The 7 uniform colored patches were simply images of each of the equiluminant target colors (see Figure 5.2(b)). Each teapot image and each color patch image subtended a visual angle of  $5^\circ$  (i.e.  $5^\circ$  high and  $5^\circ$  wide).



Figure 5.2: Stimuli used in experiment. (a) Teapots (b) Uniform color patches.

### 5.1.3 Procedure and Design

Participants were seated in front of a computer screen in a dimly lit room. Two differently colored objects of the same type (teapot or color patch) were presented on screen for each trial. The objects were displayed on either side of an imaginary central, vertical line on the screen. The background for both objects in a single trial could be one of the four different uniform levels of gray, evenly spaced from black to white. Participants were instructed to decide which object appeared nearer and responded by moving the mouse over their choice and clicking on it. If participants were unsure, they were instructed to choose based on their initial impulse. Participant response time for each trial was also recorded. Between each trial, a blank screen was displayed for 2 seconds. This was done because it has been shown that our perception of depth diminishes with viewing time and that a 2 second interval prevents total adaptation, thereby ensuring that depth perception remains stable [?].

Each participant performed 21 comparisons for each type of object against each of the 4 possible backgrounds, for a total of 168 experimental trials. An object of a particular color was never compared against itself, but all other within-object color comparisons occurred. All trials that fell under one type of object and one type of background comprised a block, for a total of eight blocks (i.e. teapots presented against each of the 4 backgrounds, and color patches presented against each of the 4 backgrounds).

The presentation order of the blocks was randomly determined and trials within each block were randomly ordered. It was randomly determined for each trial which object would be on the left and which would be on the right. Participants were given breaks after every 7 minutes of testing.

#### 5.1.4 Scoring

Participant responses were stored in 7 x 7 preference matrices (see Figure 5.3(a)). This approach, patterned after Ledda et al. [?], allows for easy analysis of the results. Eight preference matrices were generated for each participant: four for the teapot data set presented against each of the 4 possible backgrounds and likewise four for the color patch data set. A preference matrix is best interpreted in a row-wise fashion. For example in Figure 5.3(a), the 1's in the first row specify that the purple teapot appears closer than the orange and red teapots.

## 5.2 Results

Corresponding preference matrices of all participants were summed together to form eight cumulative preference matrices. Figure 5.3(b) shows the cumulative matrix for the 15 participants for the teapot data set presented against the black background. Summing across each row produces a score for each colored teapot (shown in the last column). This score tells how many times a given teapot was chosen as appearing nearer relative to all other teapots. The maximum possible cumulative score for a given teapot (or color patch) is 90.

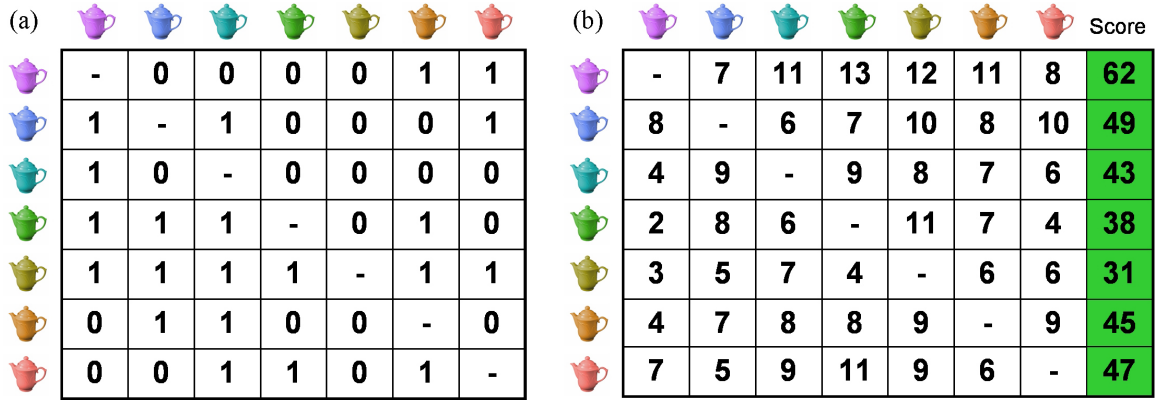
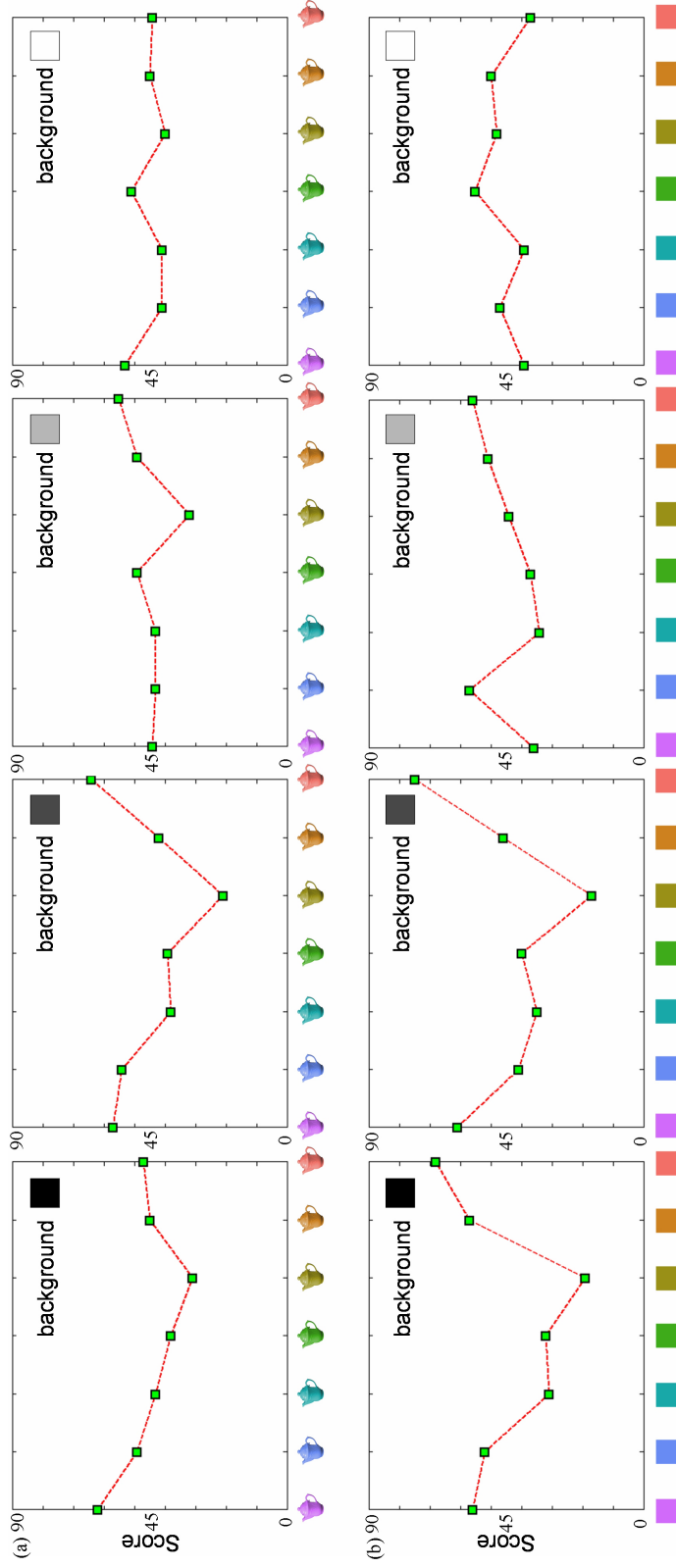


Figure 5.3: (a) Preference matrix. A 1 in row  $i$ , column  $j$  signifies that stimulus  $i$  appeared closer than stimulus  $j$ . (b) Cumulative matrix for teapot data set against black background.

This occurs if each of the 15 participants selected that teapot as appearing nearer for each of the 6 trials involving that teapot against a particular background. Figure 5.4 shows plots of the cumulative scores for each teapot and color patch against each of the 4 backgrounds. The two darker uniform backgrounds were darker than the average luminance of the stimuli while the two lighter uniform backgrounds were lighter than the average luminance of the stimuli. Notice that when stimuli were presented against the darker backgrounds that a clear cyclic trend emerged in the scores from cool to warm. This cyclic behavior correlates nicely with the familiar circular representation of the color spectrum<sup>2</sup> (see Figure 5.5). Notice however that as the background luminance increased that this cyclic relationship was lost. This observation suggests that the relative strength of color as a depth cue is enhanced against darker backgrounds.

Inferential statistics were used to determine if there is indeed a depth effect associated with the colors of the stimuli chosen. For a given preference matrix, the mean of each row was computed. This gives the mean subject response for trials involving each teapot (or color patch) for a given background. These mean scores lie between 0 and 1. This computation was performed for all preference matrices created during the experiment. Using this data,

<sup>2</sup>The color spectrum is often presented as a continuous color wheel with red blending into purple. The original color wheel is credited to Sir Isaac Newton [?].



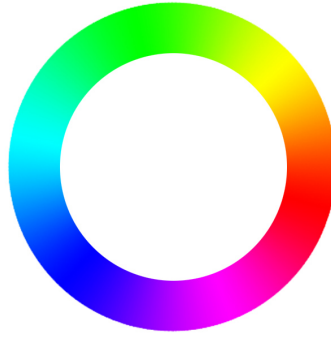


Figure 5.5: Circular representation of color spectrum with red blending into purple.

a One-Sample T Test was conducted to establish if the mean scores for the teapots or the color patches varied significantly from the expected mean score of 0.5. The results of this test for the teapot data set are shown in Figure 5.6. Similar results for the color patch data set are shown in Figure 5.7.

Notice in Figure 5.7 that against the black background, 6 of the 7 color patches had scores that differed significantly from the expected mean score. There were 3 such patches for the second background, 2 for the third background, and finally 1 for the white background. This observation supports the theory that the relative strength of color as a depth cue is enhanced against darker backgrounds. For the teapot data set, on the other hand, there were only 2 significant results against the black and second backgrounds and 1 for the third background. None of the teapots had mean scores that differed significantly from the expected mean score for the white background. This suggests that when complex images are viewed, such as the teapots, that color is not necessarily the overriding depth cue.

Additional statistical tests were performed to determine if the participants' depth preferences for the 6 comparisons involving a given colored teapot were significantly different than their depth preferences for the 6 comparisons involving the corresponding colored patch. To set up this test for a given color and background, the 6 entries in the preference matrix for the color patch were subtracted from the 6 corresponding entries for the teapot.



Background		Stimuli					
							
	$t(14) = 3.523;$ $p < .05$ <b>Significant</b>	$t(14) = .654;$ $p > .05$	$t(14) = -.367;$ $p > .05$	$t(14) = -1.451;$ $p > .05$	$t(14) = -2.357;$ $p < .05$ <b>Significant</b>	$t(14) = .000;$ $p > .05$	$t(14) = .487;$ $p > .05$
	$t(14) = 1.871;$ $p > .05$	$t(14) = 2.073;$ $p > .05$	$t(14) = -1.825;$ $p > .05$	$t(14) = -1.031;$ $p > .05$	$t(14) = -5.870;$ $p < .001$ <b>Significant</b>	$t(14) = -.676;$ $p > .05$	$t(14) = 4.219;$ $p < .001$ <b>Significant</b>
	$t(14) = -.202;$ $p > .05$	$t(14) = -.343;$ $p > .05$	$t(14) = -.435;$ $p > .05$	$t(14) = 1.000;$ $p > .05$	$t(14) = -2.982;$ $p < .05$ <b>Significant</b>	$t(14) = .807;$ $p > .05$	$t(14) = 1.625;$ $p > .05$
	$t(14) = 1.948;$ $p > .05$	$t(14) = -1.169;$ $p > .05$	$t(14) = -.888;$ $p > .05$	$t(14) = 1.871;$ $p > .05$	$t(14) = -.717;$ $p > .05$	$t(14) = .000;$ $p > .05$	$t(14) = -.174;$ $p > .05$

Figure 5.6: Results of a One-Sample T Test to establish if the mean subject score of any of the colored teapots varied significantly from the mean expected score.



Background		Stimuli						
								
	$t(14) = 2.442$ ; $p < .05$ <b>Significant</b>	$t(14) = 1.240$ ; $p > .05$	$t(14) = -3.500$ ; $p < .05$ <b>Significant</b>	$t(14) = -2.578$ ; $p < .05$ <b>Significant</b>	$t(14) = -5.030$ ; $p < .001$ <b>Significant</b>	$t(14) = 3.055$ ; $p < .05$ <b>Significant</b>	$t(14) = 4.766$ ; $p < .001$ <b>Significant</b>	
	$t(14) = 3.756$ ; $p < .001$ <b>Significant</b>	$t(14) = -.888$ ; $p > .05$	$t(14) = -2.000$ ; $p > .05$	$t(14) = -1.581$ ; $p > .05$	$t(14) = -7.299$ ; $p < .001$ <b>Significant</b>	$t(14) = -.211$ ; $p > .05$	$t(14) = 7.246$ ; $p < .001$ <b>Significant</b>	
	$t(14) = -1.288$ ; $p > .05$	$t(14) = 3.292$ ; $p < .05$ <b>Significant</b>	$t(14) = -2.750$ ; $p < .05$ <b>Significant</b>	$t(14) = -2.086$ ; $p > .05$	$t(14) = -.135$ ; $p > .05$	$t(14) = 1.146$ ; $p > .05$	$t(14) = 1.703$ ; $p > .05$	
	$t(14) = -.972$ ; $p > .05$	$t(14) = .354$ ; $p > .05$	$t(14) = -1.247$ ; $p > .05$	$t(14) = 2.197$ ; $p < .05$ <b>Significant</b>	$t(14) = .544$ ; $p > .05$	$t(14) = .892$ ; $p > .05$	$t(14) = -1.524$ ; $p > .05$	

Figure 5.7: Results of a One-Sample T Test to establish if the mean subject score of any of the colored patches varied significantly from the mean expected score.

The average of these 6 differences was then computed. This was done for each participant. A One-Sample T Test was then conducted to determine if these average differences significantly varied from the expected mean difference of zero. Figure 5.8 shows the results of this test. Negative T values occur when a color patch was chosen as appearing nearer more often than the corresponding colored teapot. Positive T values occur when the colored teapot was chosen as appearing nearer more often than the corresponding color patch. Notice in Figure 5.8 that there are only 3 cases where the differences in depth preferences between the teapots and color patches was significant. Most notably, against the black background, the differences between the depth preferences for the red color patch and the red teapot were found to be very significant. This is an interesting observation since red has been established in literature as a color that appears nearer to the viewer [?, ?, ?, ?, ?, ?]. This is indeed the case for the red color patch; however, this behavior was not observed for the red teapot. This observation further supports the notion that color is a relatively weak depth cue and is not necessarily the overriding cue in complex stimuli.

Figure 5.9 shows plots of all the recorded response times as a function of variance from the expected mean probability (0.5). The red line in each plot illustrates the best linear fit of the data. Guibal and Dresp [?] observed a correlation between the probability of a *near* response and the associated response time (the higher the probability, the shorter the response time) for simple highly saturated stimuli. Such observations are often used to help validate experimental setup. We observed similar trends in our experiment as illustrated in Figure 5.9.

### 5.3 Discussion

This study was conducted to explore the effect of color on the perceived depth of realistic objects. Previous studies using simple stimuli have been useful in isolating and studying depth cues in certain contexts, but have left open the question of whether the human visual system operates similarly for realistic objects. The results presented here suggest that color















Background	Comparison					
	 - 	 - 	 - 	 - 	 - 	 - 
	$t(14) = 1.309;$ $p > .05$	$t(14) = -.435;$ $p > .05$	$t(14) = 2.175;$ $p < .05$ <b>Significant</b>	$t(14) = .727;$ $p > .05$	$t(14) = 1.633;$ $p > .05$	$t(14) = -1.977;$ $p > .05$
	$t(14) = -.636;$ $p > .05$	$t(14) = 1.713;$ $p > .05$	$t(14) = .445;$ $p > .05$	$t(14) = -.186;$ $p > .05$	$t(14) = .745;$ $p > .05$	$t(14) = -.654;$ $p > .05$
	$t(14) = 1.017;$ $p > .05$	$t(14) = -1.793;$ $p < .05$	$t(14) = 1.871;$ $p > .05$	$t(14) = 3.055;$ $p < .05$ <b>Significant</b>	$t(14) = -1.702;$ $p > .05$	$t(14) = -.242;$ $p > .05$
	$t(14) = 1.974;$ $p > .05$	$t(14) = -.945;$ $p > .05$	$t(14) = .280;$ $p > .05$	$t(14) = -.590;$ $p > .05$	$t(14) = -.877;$ $p > .05$	$t(14) = -.892;$ $p > .05$
						$t(14) = -.168;$ $p > .05$
						$t(14) = .861;$ $p > .05$

Figure 5.8: Results of a One-Sample T Test to establish if the participants' depth preferences for a given colored teapot were significantly different than their depth preferences for the corresponding color patch.

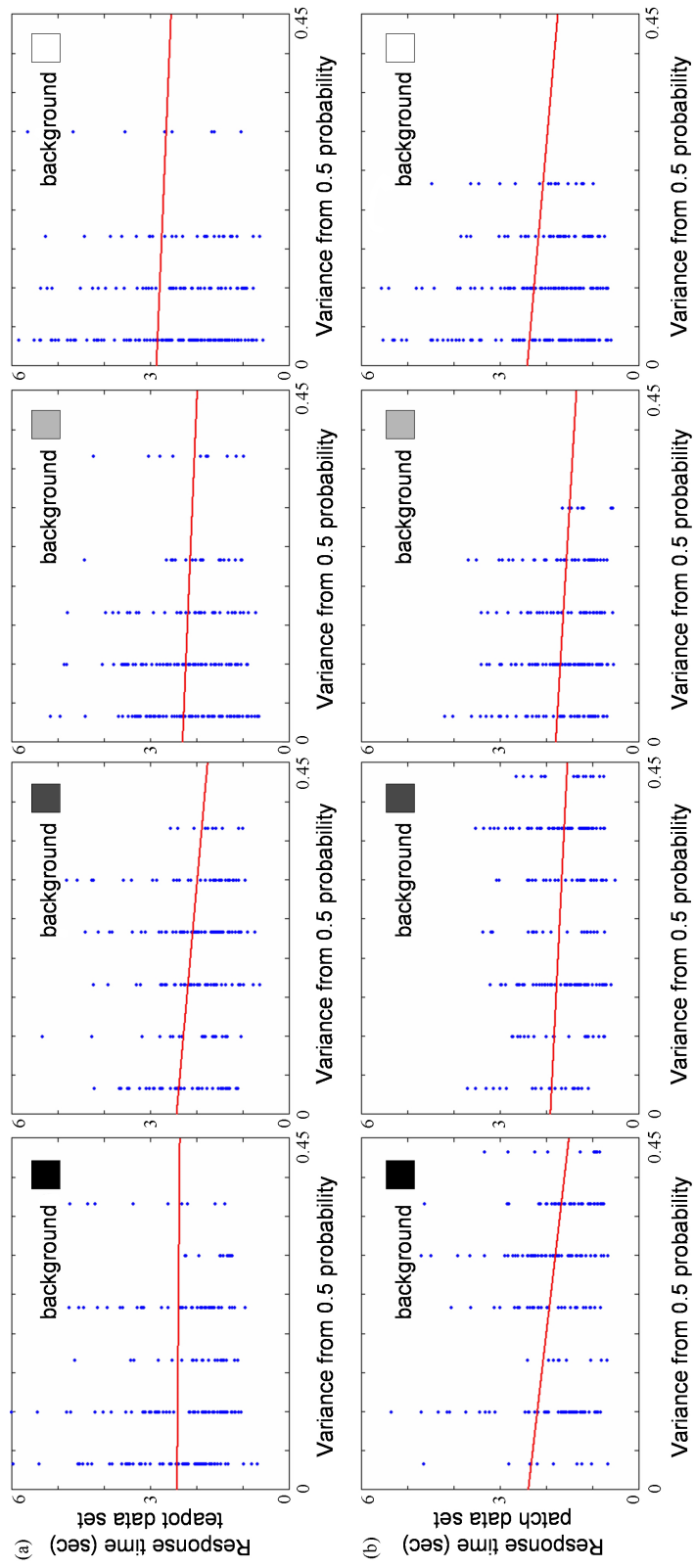


Figure 5.9: (a) Participants' response times for teapot data set against each of the 4 backgrounds as a function of variance from the expected mean probability (0.5). (b) Participant response times for color patch data set against each of the 4 backgrounds as a function of variance from the expected mean probability (0.5). The red line in each plot illustrates the best linear fit of the data.

is not necessarily the overriding depth cue for complex stimuli. Additionally, the results suggest that there are circumstances under which the relative strength of color as a depth cue can be enhanced. For the simple patches, and to a lesser extent the more complex teapots, color appeared to function as a more powerful depth cue against darker backgrounds than against lighter backgrounds. These observation provides valuable insight into the impact of luminance contrast on our depth perception.

The observations made in this experiment have several useful applications to the field of computer graphics. First, many modern software systems such as 3D-modeling programs, paint programs, or even word processors and spreadsheets have a plethora of user options available. This often leads to complex user interfaces with numerous buttons, sliders, and drop-down menus in multiple panels. The clever use of non-perspective depth cues such as color may further enhance the readability and hence user-friendliness of such complex user-interfaces. By incorporating our observations or by conducting similar experiments during the user-interface design stage of the software development process, suitable colors can be selected to ensure that these goals are met. Second, several recent studies have shown that humans generally underestimate distance and depth<sup>3</sup> in virtual environments [?, ?, ?]. Our observations can be used to guide the selection of color to help improve the perceived depth of objects displayed on such systems. Finally, consider an electronic visual alert or warning system that demands quick attention. Our results, in addition to that of Guibal and Dresch [?], show that certain foreground/background color combinations result in stronger depth cues and that participants responded quicker to these cues. By making informed color selection decisions, an additional depth cue can be introduced which might bring the alert into awareness quicker than an otherwise equal alert that did not capitalize on this cue.

There are two main limitations to this study. First, for the paired comparison test, subjects were only asked to identify the stimulus that appeared closest to them. If the

---

<sup>3</sup>The terms distance and depth are often incorrectly used interchangeably. In the context of virtual environments, distance can either be defined to be egocentric (distance from self) or exocentric (relative distance between objects). Depth refers to egocentric distance.

participants did not follow the instructions and instead made their decision based on some other criteria, such as what appears nicest to them, then these choices could masquerade as apparent depth effects which would be incorrectly attributed to the color of the stimuli. A more robust approach would be to also have a group of subjects identify the stimulus that appears farther away. If the participants are following the instructions correctly, then we would expect the depth preferences to change thereby providing an additional source of validation for the experiment. Second, while our observations suggest that the relative strength of color as a depth cue is enhanced against darker backgrounds, additional ANOVA-based statistical tests are needed to confirm if this is indeed the case. However, since the colored patches and colored teapots were not compared against themselves, the resulting data does not facilitate ANOVA testing since not all cross comparisons were performed. A follow-up experiment with the necessary design changes would help to address these limitations.

During the course of this study, it was observed that there is a lack of standardized experiment methodology for studying the complex interactions between the various depth cues. Additional efforts from the research community are necessary to address this need.

## Chapter 6

# Conveying a Sense of Motion in Images

In this chapter, we present a technique for conveying a sense of motion in images. This technique is based on the fact that traditional artists sometimes use spatial imprecision to convey a sense on motion in their work as described in section 1.3. We use a a simple two-step process to introduce spatial imprecision to the image plane. The input image is first segmented into regions of roughly uniform color and the resulting segments are then spatially perturbed. This technique can be applied over an entire image or to specific regions of an image.

### 6.1 Segmentation

Any existing image segmentation technique can be used (see [?] for a review). However, since no one segmentation technique works well in all cases, we recommend techniques where the user at least has some control over the degree of segmentation. We use an approach proposed by Comaniciu and Meer [?].



## 6.2 Spatial Perturbation

Each segment is assigned a unique ID and a local Cartesian coordinate system whose origin was arbitrarily chosen to correspond to the top-left corner of the bounding box of the segment. Different 2D spatial transformations are then applied to each segment with respect to its local coordinate system. For the images shown in this chapter, the 2D spatial transformations comprise of a small random rotation (between -10 degrees and 10 degrees) and a small random translation (at most 20 pixels).

Let  $p \in \mathbf{S}$  be a pixel in a segment  $\mathbf{S}$ . After undergoing the 2D spatial transformation, its new location  $p'$  is given by the standard composite transformation expression:

$$p' = \mathbf{T} * \mathbf{R} * p$$

where  $p$  and  $p'$  are expressed in homogeneous coordinates,  $\mathbf{T}$  represents the translation matrix and  $\mathbf{R}$  represents the rotation matrix.

Perturbing segments in this manner introduces holes in the image plane. We experimented with various *inpainting* [?] approaches for filling these holes, but found that simply pasting the perturbed segments onto the original image gave better results.

## 6.3 Results

Figure 6.1 and Figure 6.2 show examples of applying our technique over entire images and Figure 6.3 shows an example of applying the technique to a specific region of an image.

## 6.4 Discussion

The main limitation of this technique is that performing random perturbations may cause perceptually relevant segments, such as faces, to become (partially) occluded by other segments. One possible solution for this problem, will be to extend our framework to allow



Figure 6.1: Example of applying technique to convey a sense of motion over an entire image. Input image (left). Output image (right).

users to specify the relative importance of the various segments. By processing the most important segments last, we can then ensure that they are not occluded.

Another downside to performing random perturbations is that it may not necessarily be suitable for all scenes. Consider, for example an image depicting a windy scene. In such cases it will be better to organize the perturbations to follow a more structured spatial pattern (along the direction that the wind is blowing). One way to accomplish this is to allow the user to provide a flow field which specifies the desired spatial pattern. Segments can then be aligned to this field using an optimization approach.

Finally, since little research has been conducted to study the impact of spatially imprecise stimuli on our perception, it would be fruitful to conduct experiments using stimuli created by our technique.



Figure 6.2: Example of applying technique to convey a sense of motion over an entire image. Input image (top). Output image (bottom).





Figure 6.3: Example of applying technique to convey a sense of motion in a specific region (foreground) of an image. Input image (top): Nicolas Poussin, *The Rape of the Sabine Women* (1634). Output image (bottom).

## Chapter 7

# Subtle Gaze Directing

In this chapter we present a novel technique that combines eye-tracking with subtle image-space modulations to direct viewer gaze about an image. A description of the eye-tracking system that we use is presented in Appendix A.

### 7.1 Overview

When viewing a scene for the first time, the low acuity peripheral vision of the human visual system locates areas of interest<sup>1</sup>. The slower, high acuity foveal vision is then directed to fixate on these regions. The technique presented in this chapter operates by modulating regions of the scene that appear only to the peripheral vision. In an attempt to identify the stimulus detected in the periphery, the eyes move (saccade) to focus the foveal vision on the modulated region. Both luminance modulation and warm-cool modulation were studied. These modulations were chosen because the human visual system is very sensitive to luminance changes [?], and research has shown that the responses of the photoreceptors in the retina are combined into color-opponent channels that differentiate warm and cool colors [?].

---

<sup>1</sup>The areas of interest chosen vary depending on the visual task being undertaken.

Previous experiments have shown that the abrupt onset of salient stimuli is an effective approach for capturing viewer attention [?]. Ideally however, gaze directing algorithms should preserve the experience of natural image viewing. To that end, we chose to use very subtle modulations as opposed to salient modulations. To achieve subtlety in the gaze directing technique, a small study was conducted to establish thresholds for each type of modulation so that they were just intense enough to be detected by the peripheral vision. Additionally, the viewer’s foveal vision is never allowed to fixate on the modulated region. This is achieved by monitoring the direction component of the saccade velocity vector, to determine if the foveal vision is about to enter the modulated region. If this is the case, the modulation is immediately terminated. Taking advantage of the visual phenomenon known as *saccadic masking*<sup>2</sup> allows sufficient time to terminate the modulation before it can be scrutinized by the foveal vision. Saccadic masking, first described by Dodge [?], is the temporary suppression of visual processing during eye movements (i.e. between fixations (see Figure 7.1 inset)).

The results of a larger psychophysical experiment reveal that both luminance modulation and warm-cool modulation are effective at directing gaze. Subjective evaluations of the quality of the modulated test images tended to be slightly lower than the corresponding static images. One possible reason for this is that the subtle gaze direction technique forces viewers to violate their natural gaze pattern for a given image.

The remainder of this chapter is organized as follows: In section 7.2, the subtle gaze direction technique is described in greater detail. The design of an experiment to test the effectiveness of this technique and to study its impact on perceived image quality is presented in section 7.3. Analysis and discussion of the experimental results are presented in section 7.4. In section 7.5, a summary of the implications of the new technique along with potential avenues of future research is presented.

---

<sup>2</sup>Saccadic masking is also called saccadic suppression. Saccadic masking prevents perception of the blur of the retinal image caused by the ballistic movement of the eye. Complex neurological processes ensure that the resulting gaps in the visual signal are not perceived.

## 7.2 Description of Gaze Directing Technique

Consider the hypothetical image shown in Figure 7.1. Suppose that the goal is to direct the viewer's gaze to some predetermined area of interest  $A$ . Let  $F$  be the position of the last recorded fixation, let  $\vec{v}$  be the velocity of the current saccade, let  $\vec{w}$  be the vector from  $F$  to  $A$ , and let  $\theta$  be the angle between  $\vec{v}$  and  $\vec{w}$ . Either luminance modulation or warm-cool modulation is performed over the pixels in region  $A$ . Once the modulation commences, saccadic velocity is monitored using feedback from a real-time eye-tracking device and the angle  $\theta$  is continually updated<sup>3</sup> using the geometric interpretation of the dot product:

$$\theta = \arccos \left( \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| |\vec{w}|} \right) \quad (7.1)$$

A small value of  $\theta$  ( $\leq 10^\circ$ ) indicates that the center of gaze is moving toward the modulated region. In such cases, modulation is terminated immediately. It is important to note that the modulation is terminated *during* the saccade to take advantage of the gap in our perception caused by saccadic masking (see Figure 7.1 inset). This contributes to the overall subtlety of the technique. By repeating this process for other predetermined areas of interest, the viewer's gaze is directed about the scene.

The modulations are simply alternating interpolations of the pixels in  $A$  with black and white, in the case of luminance modulation, or with a warm and a cool color, in the case of warm-cool modulation (see Figure 7.2). Consider a pixel  $p = (x, y)$  in  $A$  with color  $col(p)$ . For warm-cool modulation, the color of the resulting pixel  $col'(p)$  for the warm interpolation is given by:

$$col'(p) = ((w * i) + col(p) * (1 - i)) * f(p) + col(p) * (1 - f(p)) \quad (7.2)$$

---

<sup>3</sup>Our software polls the eye-tracker at a rate of 20Hz. This relatively low sampling rate works well in practice because research has shown that only 3-4 fixations occur per second [?].

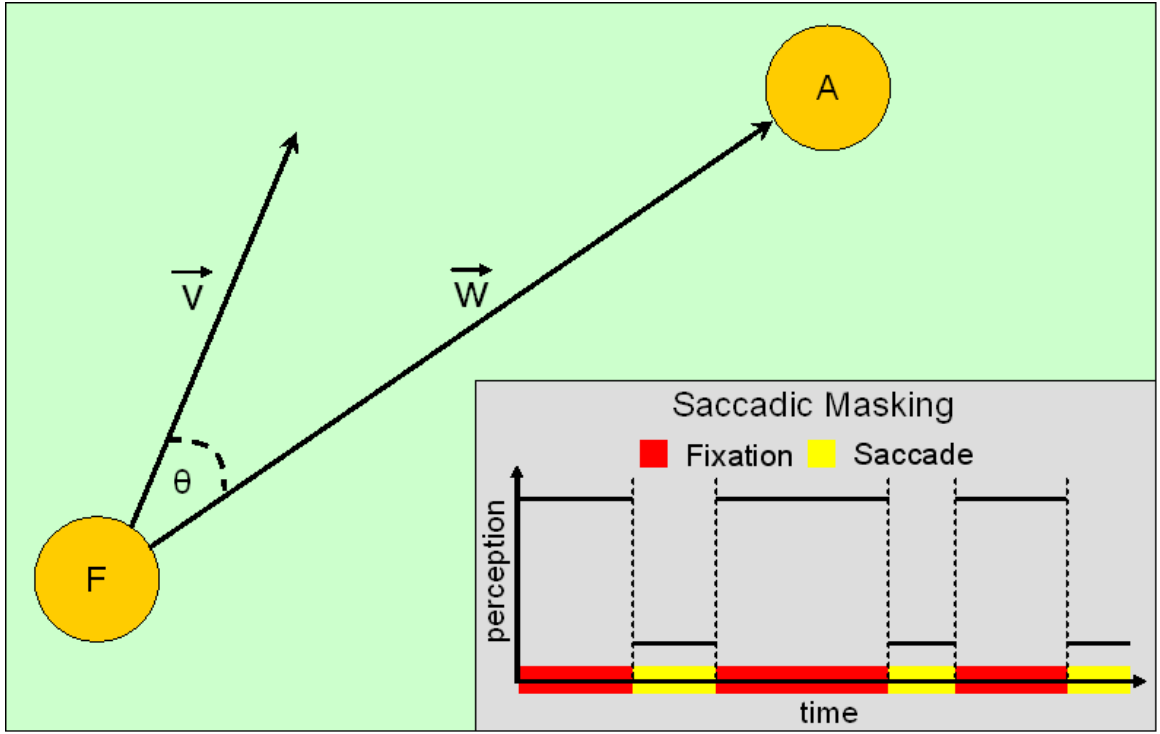


Figure 7.1: Hypothetical image with current fixation region  $F$  and predetermined region of interest  $A$ . Inset illustrates saccadic masking.

where  $w$  is a warm color<sup>4</sup>,  $f(p)$  is a falloff function based on the distance of  $p$  from the center of region  $A$ <sup>5</sup> and  $i$  is some scalar value in the range  $[0,1]$  which controls the intensity of the modulations. The cool interpolation of the warm-cool modulation cycle, and the black and white interpolations of the luminance modulation cycle are analogous.

A small pilot study was conducted to determine the value of  $i$  for which the modulations are just intense enough to be detected by the peripheral vision. Three participants were involved in this pilot study. They were each presented with five randomly selected images from the complete test set (see Figure 7.3). They were instructed to fixate on a cross in the center of the image while modulations were presented in random peripheral regions. Using the keyboard (+/-), they adjusted the value  $i$  in step sizes of 0.005 until

<sup>4</sup>For our experiment we use red as the warm color and blue as the cool color.

<sup>5</sup>We use a Gaussian falloff function with a radius of 32 pixels. This corresponds to approximately a 1cm diameter circular region on screen.





Figure 7.2: Photograph of experiment setup (left). Small patch from a test image (center). Example of luminance modulation (third column). Example of warm-cool modulation (right column).

the modulations were just noticeable. The final value for  $i$  was obtained by averaging the results of the three participants. For luminance modulation between black and white with a Gaussian falloff,  $i = 0.095$  and for warm-cool modulation between red and blue with a Gaussian falloff  $i = 0.105$ . It should be noted that  $i$  needs to be recomputed if any changes are made to the falloff function or to the colors used for the interpolations.

## 7.3 Experimental Design

This section describes an experiment that was conducted to test the effectiveness of the gaze-directing technique and to study its impact on perceived image quality.

### 7.3.1 Stimuli

Stimuli were presented on a 20 inch monitor, operating at 75Hz with a resolution of 1280 x 1024. The stimuli used in this experiment consisted of forty 1280 x 1024 images compiled from various sources. These images were chosen with no particular criteria and are shown in Figure 7.3.

### 7.3.2 Participants

Ten participants (5 females, 5 males), between the ages of 18 and 45 volunteered to participate in this study. All participants reported normal or corrected-to-normal vision with no color vision abnormalities. Participants were randomly assigned to one of two groups:

- **Static group:** Participants (five) were presented with a randomized sequence of the 40 images with no modulation. This group served as the control group for the experiment.
- **Modulated group:** Participants (five) were presented with a randomized sequence of the 40 images with modulations at pre-selected image locations. These locations, manually selected by the researchers, included areas that are not visually significant

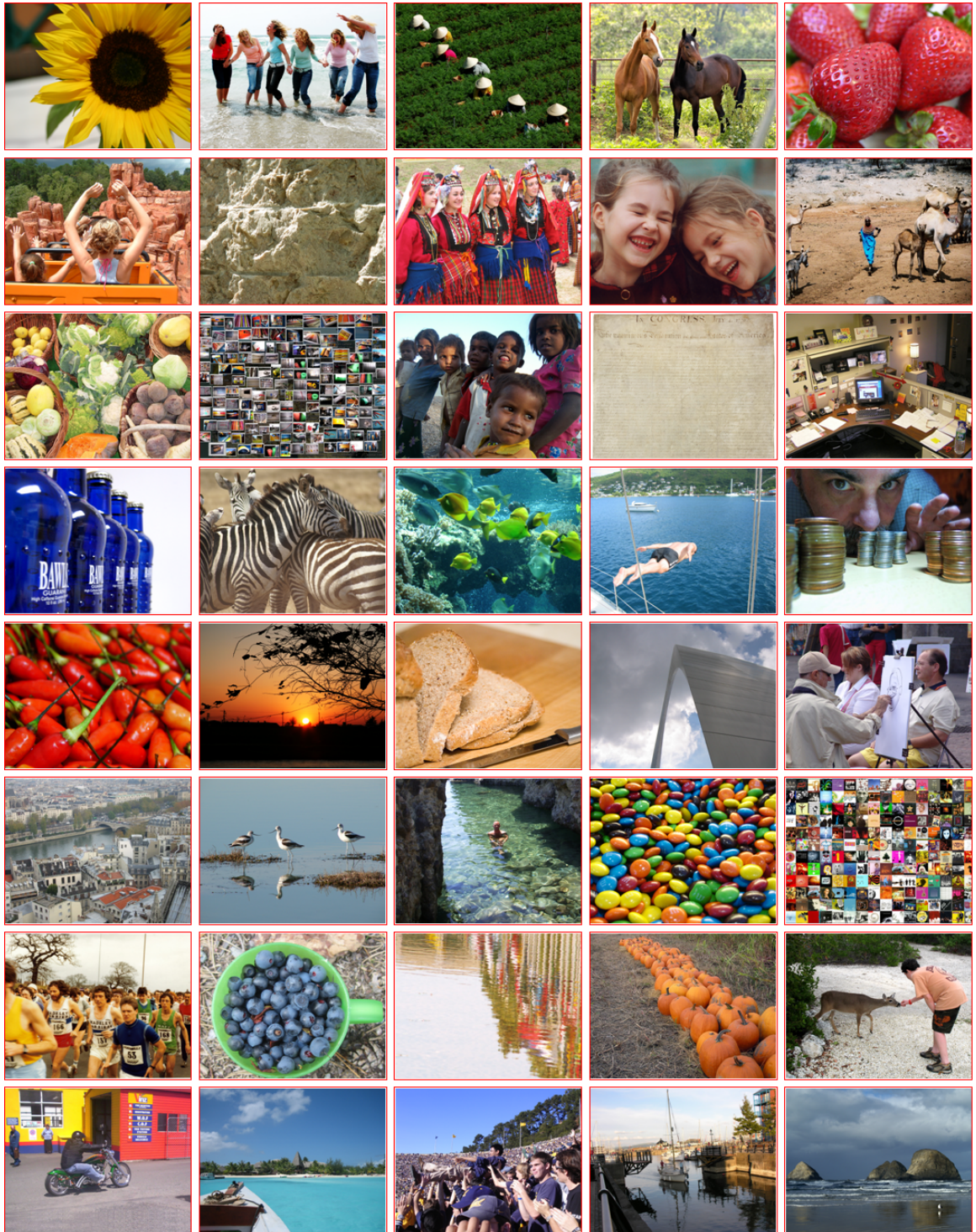


Figure 7.3: The complete set of images used in gaze-directing study. Images were chosen with no particular criteria and gathered from various sources on the web [?, ?].

- low contrast, low detail, low color saturation, and uninteresting objects. These are areas that an observer would not ordinarily attend to<sup>6</sup>. The type of modulation for a given image was randomly selected to be either luminance or warm-cool.

For both groups, the images were presented for a duration of 8 seconds. Between each image, a black screen with a small white cross displayed at the center was presented for 2 seconds. This allows participants to rest briefly between images.

### 7.3.3 Procedure

Participants were seated in front of the computer screen in a well lit room with their chin comfortably resting on a chin-rest to reduce head movement. Using an infrared camera-based eye-tracking system<sup>7</sup>, data pertaining to fixation position and saccades were recorded for the dominant eye of each participant<sup>8</sup> (see Figure 7.2). Participants were instructed to remain as still as possible while the eye-tracker was calibrated and the experiment was conducted. The chin-rest was positioned 75cm from the screen. At this distance, the actual perceptual span (area of high acuity) of the observer occupies a circular region of diameter 5cm on the screen [?]. To further promote subtlety, modulations were presented in a smaller (1cm diameter) circular region.

The participants in each group were asked to assess the quality of each image on a scale from 1 (low) to 10 (high) and to report it verbally during the 2 second period between images. These scores were recorded by the researchers. The term *quality* was not defined by the researchers. Instead it was left up to the participants to formulate their own notion

---

<sup>6</sup>The intuition behind this decision is that if we are able to direct the viewer to look at regions that they would not ordinarily attend to, then we can make strong claims about the effectiveness of our technique.

<sup>7</sup>ViewPoint EyeTracker<sup>®</sup> by Arrington Research, Inc.

<sup>8</sup>The dominant eye was established by asking participants to stare at an object a few feet away from them and to point at that object with their index finger. With their eyes focused on the object their index finger would appear blurred in the line of sight. They were asked to close one eye and then the other. The dominant eye is the eye with which the index finger appears to be pointing directly at the object.



of image quality. For participants in the modulated group, real-time analysis of the eye-tracking data was performed (as described in section 7.2) to determine when to terminate the modulations.

Data files for all participants were compiled during the experiment. The data files contain the following information: time elapsed, filename of image being displayed, position of modulated region, type of modulation, location of current fixation, and corresponding delta change from last fixation. Entries to the data files occur at a rate of 30Hz. Testing for each participant (including calibration) lasted approximately 15 minutes.

## 7.4 Results

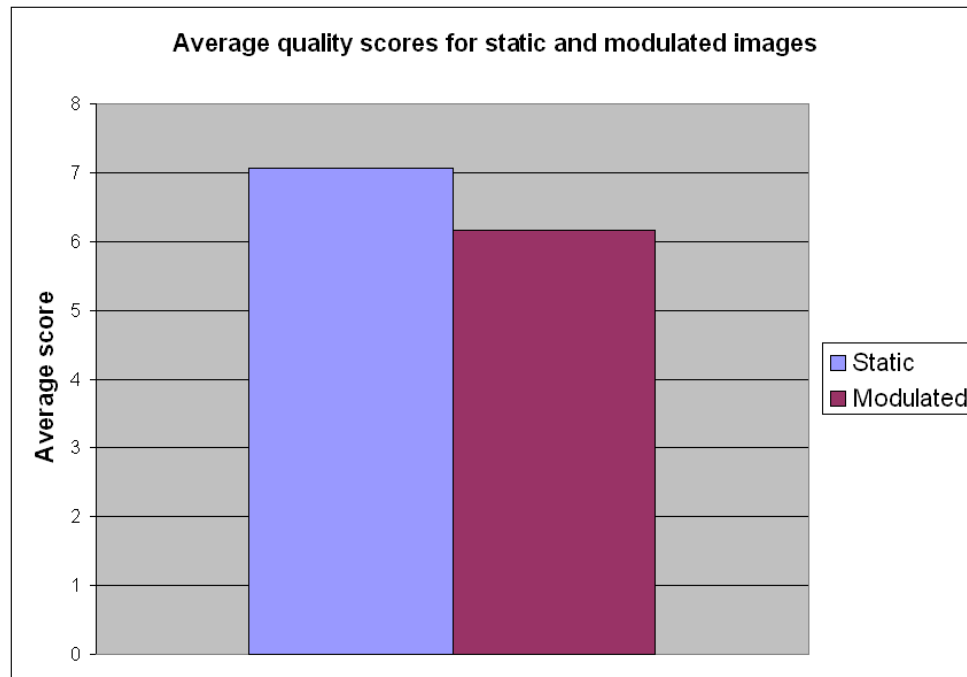
### 7.4.1 Evaluation of Image Quality

One way to evaluate the effect of gaze-directing modulations on perceived image quality is to compare the mean quality scores for the static and modulated images gathered from the eye-tracking experiment. Figure 7.4(a) summarizes these findings. Quality ratings for the static images average 7.07, while the modulated images receive a mean quality rating of 6.17. These mean scores were obtained by averaging the individual scores of the 5 participants over the forty images in each group. This observation suggests that introducing modulations, even subtle ones presented to the peripheral regions of the field of view, results in a reduction in the perceived quality of the image being viewed.

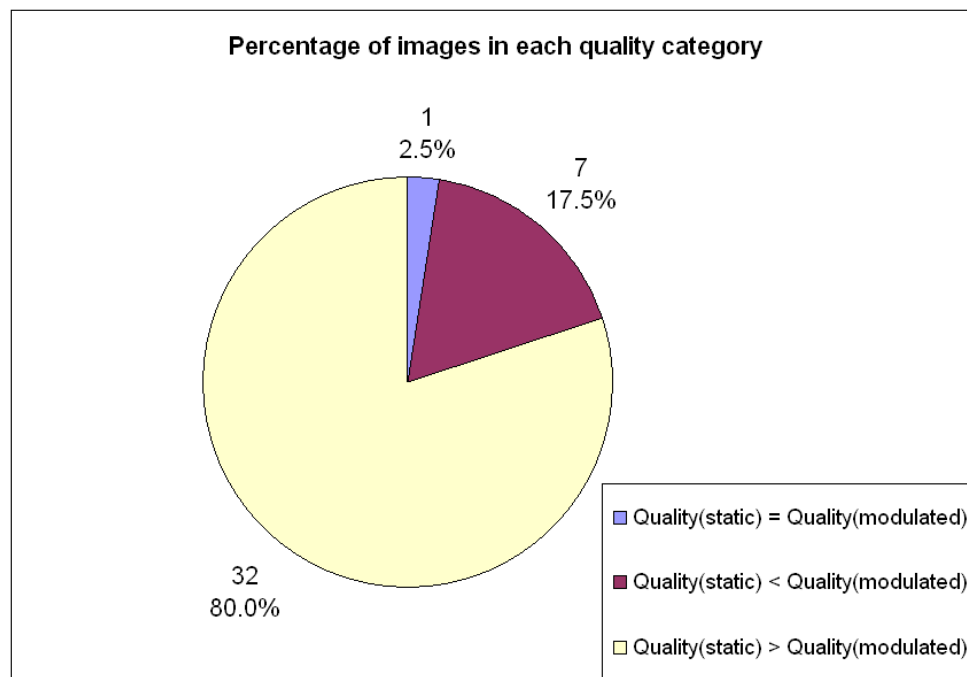
To substantiate this theory, an independent-samples  $t$ -test was performed to compare the average rating for each of the 40 images for the two groups, to determine if this pattern of results did not simply occur by chance. The results of the  $t$ -test show the following significant  $t$ -values:

$$t(78) = 5.790; p < .0005$$

The  $t$  statistic equals 5.790 with 78 degrees of freedom (80 average quality ratings). The probability  $p$ , of this distribution happening by chance is less than 0.005. This means that



(a)



(b)

Figure 7.4: (a) Average quality scores for static and modulated images. (b) Percentage of images that fall into each quality category.

the difference in quality judgments between the static and modulated images is due to the presence (or lack) of modulation and cannot be attributed to chance or other factors. We believe that the presence of the modulations causes viewers to violate their natural gaze pattern for a given image. This interruption of the natural scan path and the fact that the viewer is never allowed to scrutinize the modulated regions, leads to a lower quality rating<sup>9</sup>.

For the purpose of further analysis, we categorize the quality ratings of each static image with respect to the corresponding modulated image. A static image can fall into one of three categories:

- higher perceived quality
- equal perceived quality
- lower perceived quality

Figure 7.4(b) shows the percentage of images that fall into each of these categories. It is interesting to note that 80% of the static images received higher quality ratings than the corresponding modulated images. On the other hand, only 17.5% of the static images received lower quality scores than the corresponding modulated images. Again, we speculate this is due to deliberate disruption of natural viewing caused by the subtle modulations. Figure 7.5 shows the seven images (17.5%) that received higher average quality scores under the modulated condition than under the static condition. These images appear to be representative of the entire image set in terms of image features and image content, hence we are unable to speculate or draw conclusions about the specific image characteristics that contribute to this phenomenon.

#### 7.4.2 Effectiveness of Gaze Directing Technique

Since only areas of low visual significance were selected as regions of interest by the researchers, it would be expected that an effective gaze directing technique would result in

---

<sup>9</sup>The lower quality scores may also be due to the decision to only direct the viewer to visually uninteresting regions of an image. Additional testing will be necessary establish if this is indeed the case.

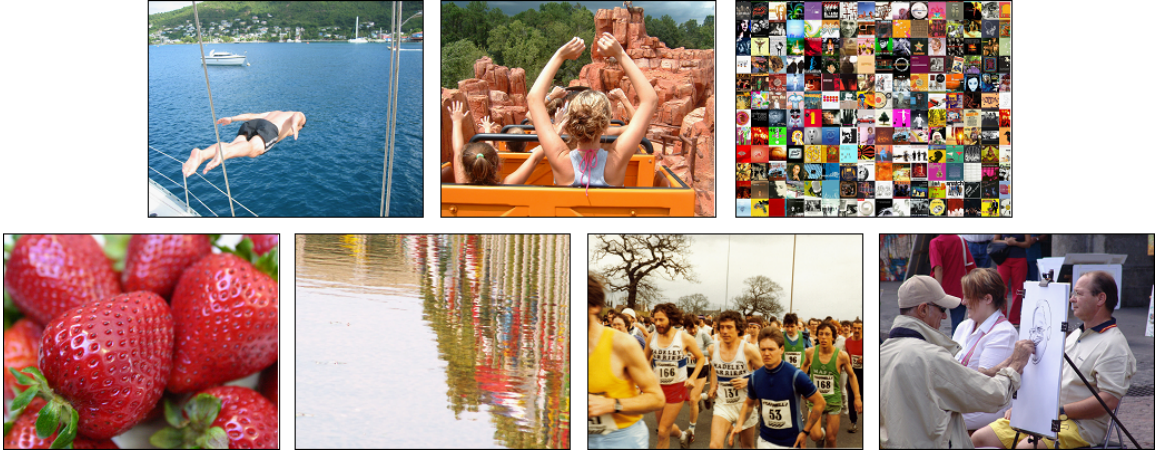


Figure 7.5: The set of images that received higher average quality scores under the modulated condition than under the static condition. We are unable to speculate about specific image characteristics that contribute to this phenomenon.

a gaze pattern that is significantly different than one obtain by natural viewing. By comparing fixations in the static scene to those in the luminance modulated and warm-cool modulated scenes, it is possible to evaluate the effectiveness of our gaze directing technique. Fixations denote regions of the image where viewers rest their gaze, indicating a feature of interest in the scene. To facilitate this comparison we determine the total fixation duration in various regions of an image in both the static condition and the modulated condition. The regions were specified by partitioning each image into a 5 x 5 regular grid.

For each static image, the percentage of fixation time, averaged across the 5 participants, was determined for each grid cell. This averaged fixation time quantifies the natural gaze distribution for a given static image (see Figure 7.6). Essentially this data serves as the ground truth for a given image and we would expect any viewer’s gaze distribution to correlate highly under similar viewing conditions and assigned task. In fact, to verify this ground truth, a fresh participant was asked to complete the experiment. As expected, we observed a high degree of correlation between her gaze distribution and that of the averaged distribution. This is especially true for the images with strong salient features where we





(a)

0.14%	2.99%	2.28%	1.59%	0.00%
1.27%	9.60%	6.07%	11.88%	5.98%
0.71%	7.49%	14.98%	4.67%	3.12%
0.21%	1.13%	5.39%	2.19%	2.77%
0.38%	2.84%	4.55%	5.09%	2.68%

(b)

Figure 7.6: Quantifying gaze distribution. (a) Source image. (b) Average fixation time for each grid cell for participants viewing image without modulation. This quantifies the natural gaze distribution. Notice that regions containing visually significant features such as faces receive greater fixation time.

observed a Pearson coefficient of correlation as high as  $r = 0.96$ . For images with no dominant salient features, the minimum Pearson coefficient of correlation was still a respectable value of  $r = 0.67$ . This validates our supposition that averaging over the five participants for the static group would give us a ground truth or normal (average) gaze pattern for free viewing of each image. The challenge now was to demonstrate that gaze patterns over the modulated images did not correlate well with the ground truth indicating that the presence of modulations had caused some upset to the expected natural gaze pattern for a given image.

The participants in the modulated group viewed a total of 200 images (5 participants, 40 images). Of these, 107 were luminance modulated, and 93 were warm-cool modulated (the type of modulation was chosen randomly). Figure 7.7 plots of all the correlation values obtained by comparing the gaze distribution of the modulated images with the corresponding ground truth gaze distribution. The results indicate a poor correlation between the gaze distribution of the two groups. This is to be expected as the technique deliberately attempts to drive the viewer's scan path off their natural viewing path. The poor correlations indicate that viewing the modulated images yields very different results than viewing the static images. This provides an element of confidence that our technique does indeed function correctly and draws the viewer's gaze to pre-selected regions of the image, and prevents them from following their instinctive gaze configuration.

Further statistical analysis is necessary of course to validate that this variation in gaze patterns did not happen by chance. We conducted a repeated measures between subjects analysis of variance (ANOVA) test on the coefficients of correlation using the participants as the factor for the test. The following results were obtained:

$$F(39, 1.903) = 21.034; p < 0.001$$

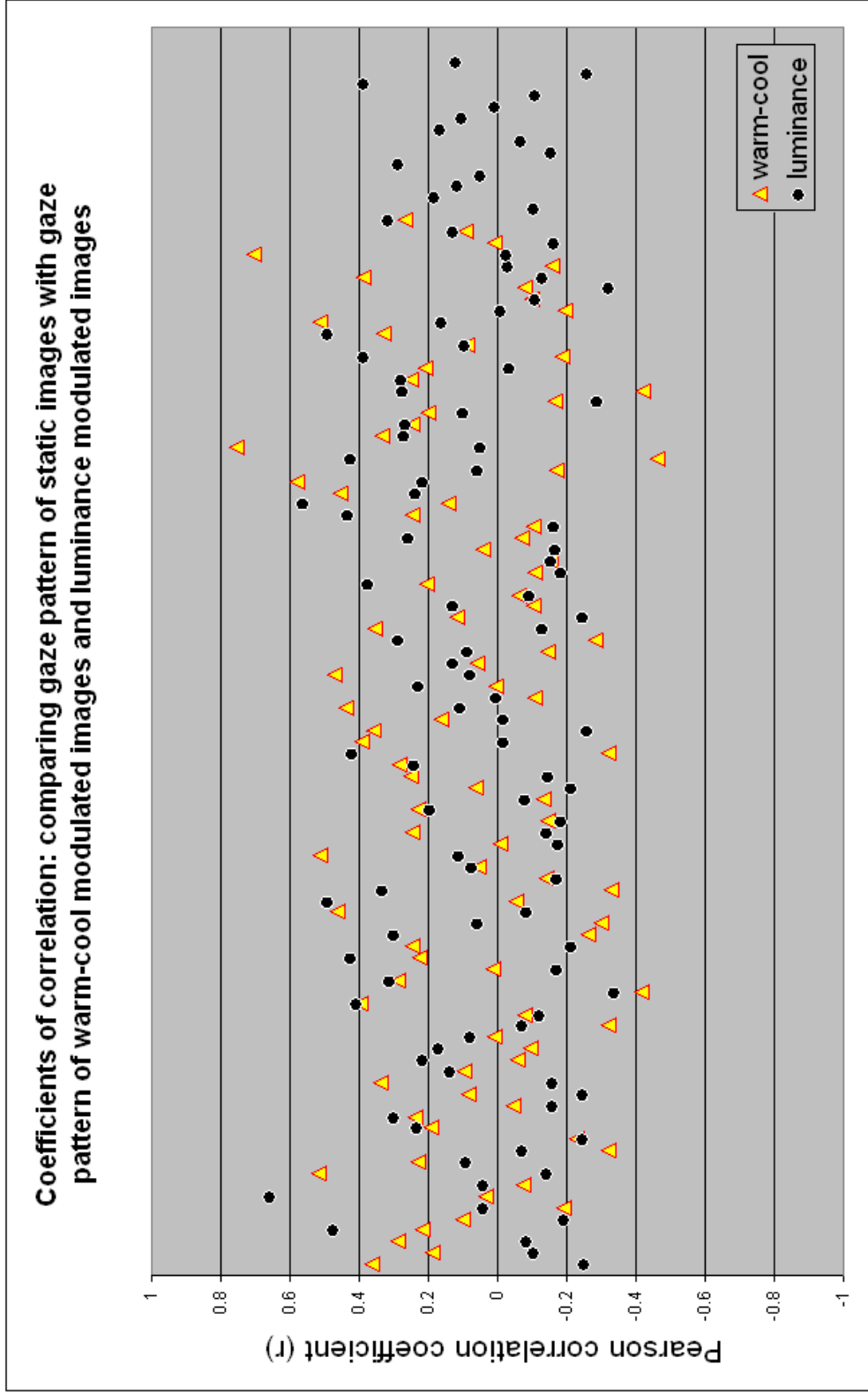


Figure 7.7: Plot of all Pearson coefficient of correlation values obtained by comparing the gaze distribution of modulated images with the corresponding natural gaze distribution. Note that most of the coefficients fall in a very narrow range close to zero ( $-0.2 \leq r \leq 0.4$ ). These small values show that there is poor correlation between the gaze patterns of the static group and modulated group.

This shows that the poor correlation values did not happen by chance. Hence we can conclude that the presence of gaze directing modulations results in a gaze distribution that is significantly different than the natural gaze distribution.

Notice in Figure 7.7 that the coefficients of correlation for the luminance modulated images occupy a narrower range ( $-0.34 \leq r \leq 0.66$ ) than the coefficients of correlation for the warm-cool modulated images ( $-0.46 \leq r \leq 0.75$ ). This suggests that the luminance modulation is more effective at directing gaze than the warm-cool modulation. An independent-samples *t*-test, however, revealed that this difference was not significant. The results of the *t*-test show the following *t*-values:

$$t(198) = .654; p > .05$$

Figure 7.8 and Figure 7.9 show examples of the averaged gaze distributions for the participants in the static group and the modulated group for two of the images used in the experiment. These figures demonstrate that our gaze directing technique does indeed result in significantly different gaze distributions, compared to the gaze distributions for the corresponding static images.

Notice that the overlaid gaze distributions are not necessarily well aligned with the researcher pre-selected locations for modulation. There are three possible causes for this inaccuracy. First, the simple chin-rest that we use does not completely eliminate head movement. Second, since the subtle modulations are terminated as soon as the eye starts moving toward them, it may be that the presentation time of the stimulus is too brief for the brain to accurately determine its location. Finally, some of the overlaid gaze distributions suggest that the fixations are actually occurring at regions of visual significance that lie closest to the modulated regions.



Figure 7.8: Gaze distributions for an image under static and modulated conditions. Input image (top). Gaze distribution for static image (bottom left). Gaze distribution for modulated image (bottom right). White crosses indicate locations preselected by researchers for modulation.





Figure 7.9: Gaze distributions for an image under static and modulated conditions. Input image (top). Gaze distribution for static image (bottom left). Gaze distribution for modulated image (bottom right). White crosses indicate locations preselected by researchers for modulation.

## 7.5 Discussion

We have presented a method capable of directing an observer’s gaze to chosen regions of a display. Using custom built software, we introduce brief subtle luminance or warm-cool modulations in still images. By monitoring the saccadic velocity of the dominant eye of the observer, we are able to terminate these modulations before the observer has an opportunity to scrutinize them. The results of an experiment show that this approach to gaze directing is highly effective. Subjective evaluations of image quality, however, tended to be lower for the modulated images compared to the corresponding static images.

Our technique has potential applications in a number of research and commercial areas:

- **Perceptually Adaptive Rendering:** By dictating where in a scene a viewer looks, we can reap enormous benefits in the rendering domain. If a particular region of a scene requires more rendering time (due to the complexity of the model or the need for higher resolution), our technique can be used to direct users to look at other regions of the scene that required less time to render, thus distracting their view from the more complex areas that are progressively updating.
- **Flight/Driving Simulation and Training:** In flight and driving simulators, the goal is to instill good navigation habits, such as checking certain information on the cockpit equipment, or routinely checking rear-view and side mirrors. Training simulators could be equipped with subtle gaze direction techniques to encourage users to frequently look at selected regions, such as mirrors, in the hope that the habit would transfer to the real world situation.
- **On-line Training and Distance Education:** Educators employing on-line technology in distance learning courses could use subtle gaze direction to encourage student viewing of relevant sections of the on-line course display. For example, while reading

through some text the reader’s gaze could be directed to an animation or figure illustrating the practicalities of the topic in question. Viewer’s gaze could also be directed to a sidebar which elaborates on the text, or to pertinent links to other web pages, or to coursework relevant to the readings.

- **Pervasive Advertising:** On large single-screen displays, advertisers could use subtle gaze directing to quickly guide the viewer to the important product information. This could potentially be a cost saving approach especially for high viewer volume segments such as Super Bowl commercials which currently cost about \$2.4M for 30 seconds of air time.

There are also several potential avenues of future research. We believe that the effectiveness of the gaze directing technique would be improved by making it more adaptive to the observer’s viewing configuration. For example, by adjusting the intensity of the modulation and/or the size of the modulated area based on the distance between the current fixation point and the desired fixation point, the modulations can be made more salient for larger distances or more subtle for smaller distances. For larger distances, adapting the modulations in this manner will lead to quicker detection by the peripheral vision and also a greater degree of accuracy when the eye saccades to fixate on the modulated region.

Additionally, an emerging area of research that explores the use of subliminal cues to trigger unconscious neural processing may offer a solution to the problem of reduced perceived quality that was observed for the modulated images. The approach that is currently being used presents a subtle stimulus to the peripheral regions of one eye while a continuously flashing mask is presented to the other eye. The presence of the flashing mask completely suppresses the conscious perception of the subtle stimulus. However, the subtle stimulus still results in electrical signals that trigger neural processing [?]. While this particular approach is impractical for our work, because we would like to facilitate natural binocular viewing of the display, it does encourage the exploration of subliminal approaches for gaze-directing.



Finally, a natural extension of the research presented in this chapter is the application of the gaze-directing technique to video. We believe that the presence of motion in the video will further suppress the subtle modulations. Hence, the modulated videos may not suffer from the perceived quality degradation that was noted in the modulated images. There are however, several challenges such as maintaining frame-to-frame coherency of the modulations and changing the position of the modulations to follow moving regions of interest, that will need to be overcome for this to be feasible.

As display technology advances and screen sizes continue to increase, there is a danger that viewers will be overwhelmed by the amount of visual information available at any one time. If their gaze is subtly directed as they navigate through the vast amount of information, the effects will be positive for both viewer and presenter.

## Chapter 8

# Conclusion

This dissertation introduced a novel approach to computer-based image editing and manipulation where the goal is to explicitly trigger certain visual cues. We refer to this as *perceptually meaningful image manipulation*.

This dissertation also presented three examples of perceptually meaningful image manipulation techniques:

- Simulating the artistic control of apparent depth of objects in an image
- Conveying a sense of motion in images
- Subtly directing a viewer’s gaze about an image

Additionally, the results of an experiment that explores the relationship between color and perceived depth for realistic, colored objects was presented. The remainder of this chapter provides a brief summary of each of these contributions.

### 8.1 Simulating Artistic Control of Apparent Depth

Our technique for simulating the artistic control of apparent depth is described in chapter 4. This technique, which attempts to mimic the artistic use of color and luminance to

convey and manipulate perceived depth in images, overcomes many of the time-consuming limitations associated with manually selecting and editing groups of pixels. The technique works by automatically generating luminance or color target values for the object and/or background. The method for determining these target values is based on approaches developed by traditional artists to create similar depth effects in their work. For large images, these target values can be precomputed, allowing for editing to be performed at interactive rates.

## 8.2 The Effect of Object Color on Depth Ordering

Our experiment to explore the effect of color on perceived depth for realistic, colored objects with varying shading and contours was presented in chapter 5. This study was motivated by the fact that much of the existing research literature described experiments where very simple colored stimuli were presented to human subjects and their depth judgments were recorded. The question of whether the human visual system operated similarly for more realistic stimuli was not extensively researched.

In our experiment, subjects were asked to compare equidistant pairs of differently colored realistic objects for differences in apparent depth. Our observations, which support the results of previous studies, show that in the presence of other cues, that color is not necessarily the dominant cue. Additionally, we observed that the relative strength of color as a depth cue can be enhanced by making the background darker. These observations can be used to guide the selection of color for computer graphics applications where depth perception is important. This is especially true for virtual environments where studies have shown that depth is typically underestimated.

### 8.3 Conveying a Sense of Motion in Images

Our approach for conveying a sense of motion in images, described in chapter 6, is patterned after a technique developed by traditional artists. Traditional artists often used misaligned brush strokes (spatial imprecision) to create a sense of motion in their paintings. Our technique used a simple two-step process to introduce spatial imprecision to the image plane. The input image is first segmented into regions of roughly uniform color and the individual segments are then spatially perturbed. The resulting spatially imprecise image appears to have a dynamic component. This sense of motion may be due to differences in the behavior of the foveal vision and the peripheral vision. This technique can be applied over an entire image or to specific regions of an image.

### 8.4 Subtle Gaze Directing

Our technique for subtly directing a viewer’s gaze about an image is described in chapter 7. This technique works by presenting brief modulations to the peripheral regions of the field of view. These modulations attract the viewer’s foveal vision. Additionally, we monitor saccadic behavior with an eye-tracker, and are able to terminate the modulation before the viewer’s foveal vision enters the modulated region. Hence the viewer is never allowed to scrutinize the stimulus that attracted their gaze. The results of an experiment show that this approach to gaze directing is highly effective. However, subjective evaluations of quality of the modulated images tended to be lower than that of the corresponding static image.

Our gaze directing technique has potential application in several commercial and research areas including large scale display systems, perceptually adaptive rendering, and complex visual search tasks.

## Chapter 9

# Future Work

Future work relating to, or extending, the contributions made in this dissertation has been presented in the discussion sections of chapters 4, 5, 6, and 7. For convenience, these are summarized below:

- Simulating Artistic Control of Apparent Depth:
  - Develop a more robust method for warm-cool color selection that avoids cue conflicts.
  - Develop and conduct experiments to test the effectiveness of this technique at manipulating perceived depth..
- The Effect of Object Color on Depth Ordering:
  - Conduct additional testing where some participants are asked to select the stimulus that appears farther away.
  - Modify experiment design to facilitate ANOVA-based testing.
- Conveying a Sense of Motion in Images:
  - Develop a priority-driven approach to ensure that perceptually relevant segments, such as faces, do not become occluded.

- Provide users the ability to organize the perturbations to follow a more structured spatial pattern.
- Develop and conduct experiments to evaluate the impact of spatially imprecise stimuli on perception.
- Subtle Gaze Directing:
  - Improve gaze directing technique by making the size and intensity of the modulations dynamically adaptive to the observer’s viewing configuration.
  - Explore the use of subliminal approaches for gaze-directing.
  - Extend gaze directing framework to video.

In addition to these opportunities for future work, the following sections describe two open problems that should also be addressed.

## 9.1 Objective Tone-Mapping Evaluation

The goal of this proposed area of research is to develop an objective approach for evaluating how well existing tone-mapping techniques preserve our visual perception of a scene. The range of luminance values that can be displayed on a typical monitor is only a fraction of the range that exists in the real world. As such, in order to display real world images on a monitor, this wide luminance range must be mapped to the smaller displayable range. This is known as *tone-mapping*. Numerous tone-mapping techniques exist [?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?]. These techniques are based on the fact that absolute brightness of visual stimuli is discounted in early stages of the visual pathway [?, ?]. Instead, our visual system extracts more information from the local variation of brightness within the stimuli.

Figure 9.1 illustrates how tone-mapping can be used to compress a large range of luminance values while preserving visual content. These images are from a recent paper by Fattal et al. [?]. For convenience their explanation is repeated here:



Figure 9.1: An example of tone-mapping. Courtesy of Fattal et al. [?]. The five smaller images are photographs of a scene taken using different exposure times. The larger image shows the tone mapped result.

These photographs (small images) were taken using a digital camera with exposure times ranging from 1/1000 to 1/4 of a second (at f/8) from inside a lobby of a building facing glass doors leading into a sunlit inner courtyard. Note that each exposure reveals some features that are not visible in the other photographs. For example, the true color of the areas directly illuminated by the sun can be reliably assessed only in the least exposed image (top left), since these areas become over-exposed in the remainder of the sequence. The color and texture of the stone tiles just outside the door are best captured in the middle image (second row left), while the green color and the texture of the ficus plant leaves becomes visible only in the very last image in the sequence. All of these features, however, are simultaneously clearly visible to a human observer standing in the same location, because of adaptation that takes place as our eyes scan the scene [?]. Using Debevec and Malik's method [?], we can compile these 8-bit images into a single high dynamic range (HDR) radiance map with dynamic range of about 25,000:1<sup>1</sup>. However, it is not at all clear how to display such an image on a CRT monitor whose dynamic range is typically below 100:1! (The large image shows the result of their technique).

There have been recent advances in the area of high dynamic range display systems [?]. Such displays are not yet practical for average consumers because they are still quite expensive, require significant power and generate a lot of heat. As such, tone-mapping continues to be an area of active research.

Researchers have become interested in evaluating how well these tone-mapping techniques preserve our visual perception of a scene. Early work in this area involved psychophysical experiments where participants compared tone-mapped images obtained using different techniques [?, ?]. They were presented with pairs of these images and asked to determine which one appeared more natural and also which one they preferred. In this

---

<sup>1</sup>The dynamic range is often expressed as the ratio between the highest and lowest luminance values.



manner, the researchers were able to rank the performance of several tone-mapping techniques for various scenes. In 2005, Yoshida et al. [?] conducted a study where participants did similar comparisons between the tone-mapped image and the actual scene. However, Ledda et al. [?] noted that comparing a 2D image directly with a 3D scene introduces many uncontrolled variables and instead opted to have participants compare the tone-mapped image with the scene shown on a high dynamic range display system. In all of these cases, the evaluation of the tone-mapping techniques is based on human preferences.

The remainder of this section proposes an objective approach for performing these evaluations that involves comparing the response of double-opponent cells (introduced below) to the tone-mapped image with the response of double-opponent cells to the original scene. Such objective approaches for evaluating tone-mapping techniques will eliminate the need for time consuming user-studies.

In section 2.2, color-opponent ganglion cells were discussed. These cells play a major role in our color perception. However, there are certain visual phenomena that cannot be attributed solely to color-opponent ganglion cells. Two of these phenomena, *color constancy* and *simultaneous color contrast* are presented in Appendix B.

For such phenomena to occur, cells must not only exhibit color-opponency but also spatial-opponency [?, ?, ?]<sup>2</sup>. These cells are referred to as *double-opponent* cells. Researchers initially expected to find double-opponent cells either in the retina or in the lateral geniculate nucleus (LGN) of primates (none have been found) [?]. Several researchers presented indirect evidence that double-opponent cells were actually located in the primary visual cortex (V1) [?, ?, ?]. In 2001, Conway [?] showed that the majority of the cells in V1 are in fact double-opponent cells.

Conway’s work is part of a larger effort in the Neuroscience community to develop techniques to create spatial maps of cone inputs to the receptive field of neurons in the LGN and also in the primary visual cortex [?, ?, ?, ?, ?, ?]. The following explanation of the process involved in creating these maps was adapted from [?]. In general, these maps

---

<sup>2</sup>There are several other opposing theories [?].

are created by presenting cone-isolating stimuli to the subject (macaque) and plotting a point when a response is detected in the neuron being studied. In the case of alert subjects, eye trackers are used along with methods to compensate for residual eye movements [?]. The cone-isolating stimuli can be created using the *silent substitution method* [?, ?]. A *plus stimulus* is one that selectively increases the response of one class of cones while maintaining the responses of the other two classes of cones. Similarly a *minus stimulus* is one that selectively decreases the response of one class of cones while maintaining the responses of the other two classes of cones.

Figure 9.2 shows the spatial maps of the L-, M-, and S- cone inputs for a typical double-opponent cell in the primary visual cortex. Figure 9.3 illustrates the relationship between cones in the retina, pixels on the display, and individual neurons in the primary visual cortex. The  $n \times n$  array of pixels corresponds to the size of the receptive field of the neuron.

We can think of the overlays of the plus and minus maps as spatial domain filters to the L-, M-, and S- cone responses (which I refer to as the L-, M-, and S- channels)<sup>3</sup>. Figure 9.4 illustrates how these filters can be used to compute the double-opponent response for a given cell over an entire image. Using this approach, we can evaluate how well existing tone-mapping techniques preserve the double-opponent response to the original scene. This can be done by comparing the double-opponent response to the original HDR image with the double-opponent response to the resulting tone-mapped image. One possible metric of response similarity is a sum of the response differences between the high dynamic range image and the corresponding tone-mapped image. This can be expressed as follows:

$$\sum_{x,y} |H_{R_{x,y}} - T_{R_{x,y}}| \quad (9.1)$$

where

---

<sup>3</sup>Golz and MacLeod [?] provide a method to determine the L-, M-, and S- cone responses (channels) for an image displayed on a CRT.

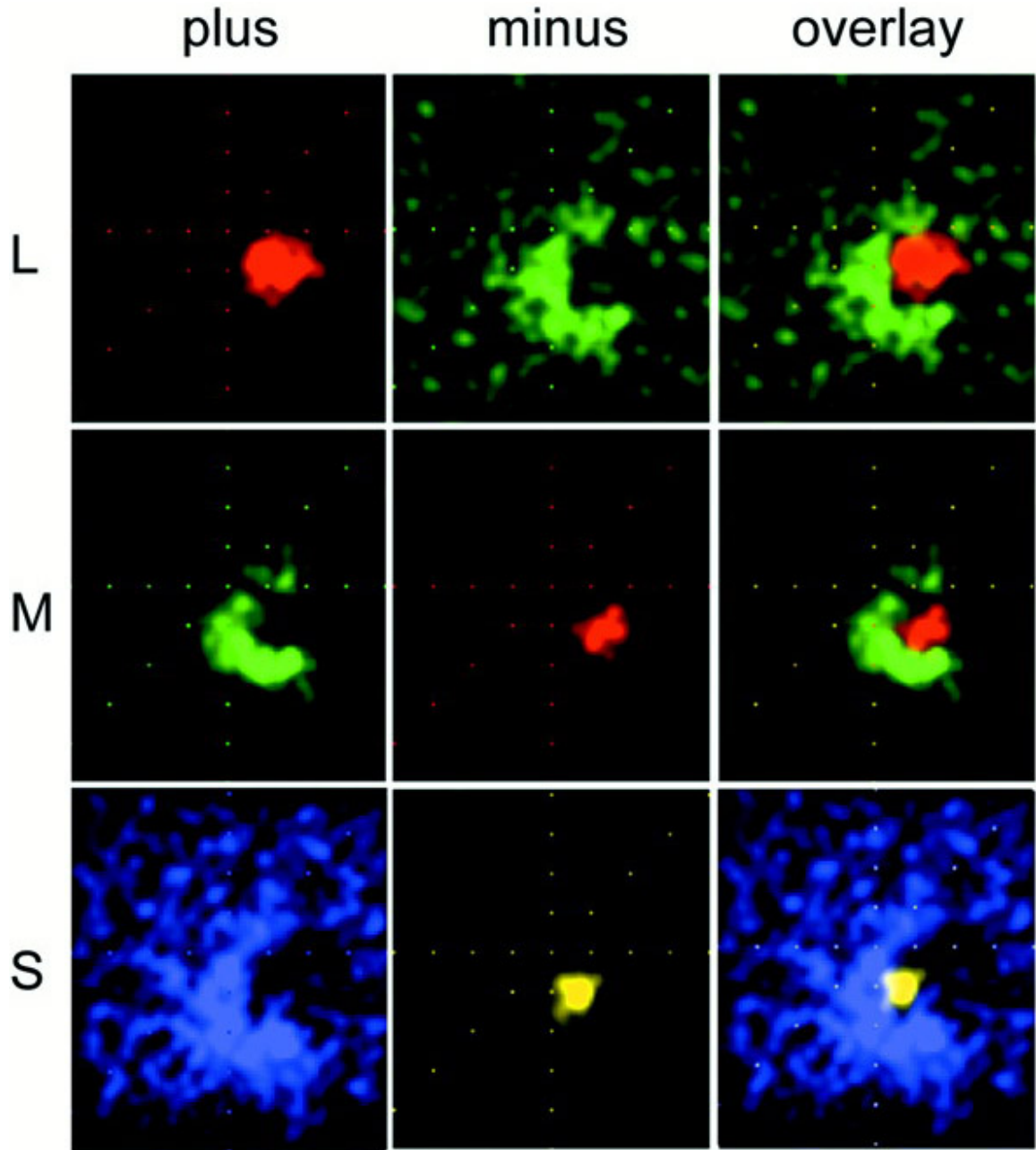


Figure 9.2: Receptive field of a double-opponent cell in V1. The images in the left column show the spatial maps of L-, M- and S- cones created using L-, M-, and S-cone isolating plus stimuli. The middle column contain maps obtained using minus stimuli. Finally, the last column shows an overlay of the images in the first two columns. Higher cell responses are indicated by more saturated colors. Black represents a cell response of zero spikes per second. Courtesy of B. R. Conway [?].

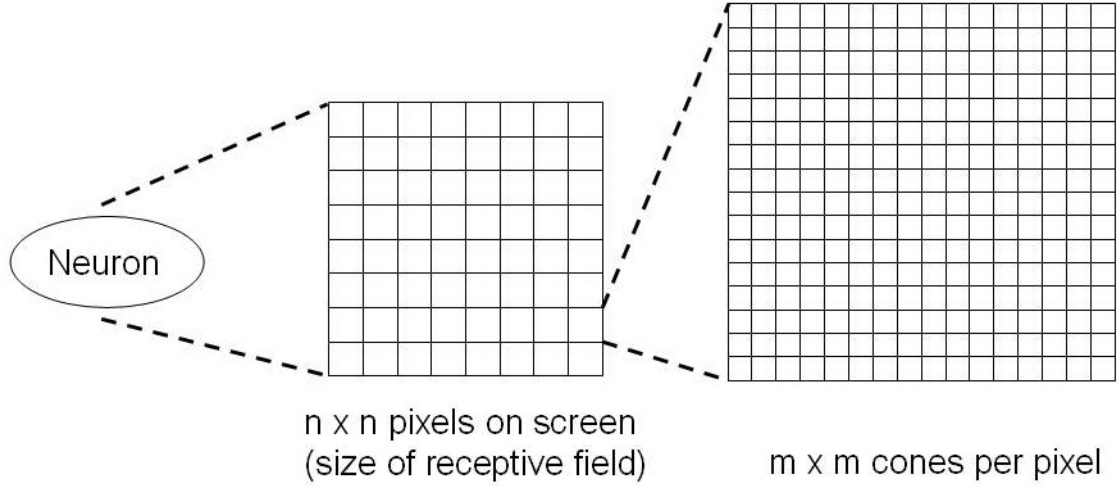


Figure 9.3: Relationship between cones in the retina, pixels on the display, and neurons in the primary visual cortex. The  $n \times n$  array of pixels corresponds to the size of the receptive field of the neuron

- $H_{R_{x,y}}$  is the double-opponent response to the high dynamic range image at location  $(x, y)$ .
- $T_{R_{x,y}}$  is the double-opponent response to the tone-mapped image at location  $(x, y)$ .

This gives a single number that describes how close the double-opponent cell responses are between the HDR image and the corresponding tone-mapped image. Consider ranking various tone-mapping techniques (for a particular scene) based on double-opponent response similarity. If this ranking is consistent with the results of previous preference based evaluations of tone-mapping techniques [?, ?, ?, ?], then we can conclude that double-opponent cells do in fact play an important role in our judgment of perceptual similarity. This in turn will encourage the pursuit of a double-opponent response preserving tone-mapping technique. Such a technique will be an important step toward the ultimate goal of a perception-preserving tone mapping technique.

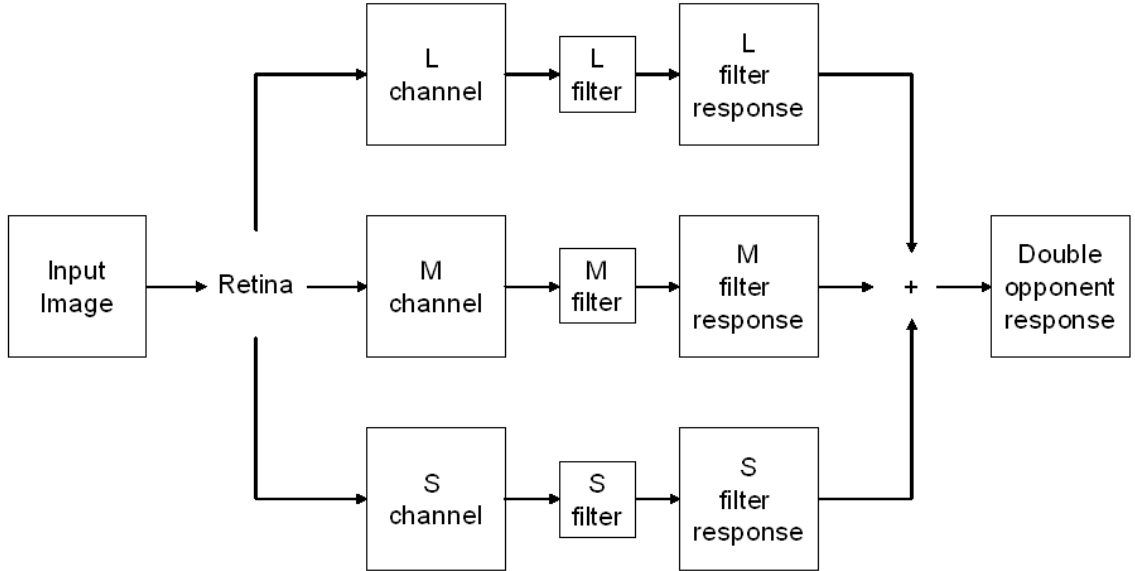


Figure 9.4: Computing double-opponent response over an entire image. The L-, M-, and S-channels of the input image can be extracted using a method proposed by Golz and MacLeod [?]. These channels are then filtered using the L, M, and S overlays of a particular cell. The resulting filter responses are summed to give the double-opponent response of that particular cell over the entire image.

## 9.2 Perception-Guided Image Compression

Recall from the discussion in section 2.2 that eye has approximately 130 million photosensitive cells. These cells generate enormous amounts of information (cell responses) for every retinal image. The visual pathway does not have enough capacity to allow for the direct transmission of this large amount of information to the brain in a timely manner. Instead, the human visual system performs extensive data reduction to extract only the salient features of a scene. This data reduction does not occur at a single physiological location, but rather, is the result of processing that occurs at successive stages along the visual pathway.

In the early stages of the visual pathway, this data reduction is possible largely because of redundancy in the incoming cone responses. Notice, for example, that there is considerable overlap in the absorption spectra for the various types of cones (see Figure 1.4(a)). This is especially true for the L- and M- cones. The processing that occurs at

the ganglion cells minimizes the correlation between these cone signals [?, ?, ?]. Subsequent processing further reduces data by extracting only the essential attributes of objects and surfaces that are necessary for the brain to categorize or identify them.

Researchers have been actively studying the stages along the visual pathway and in some cases are even able to describe the processing that occurs as mathematical transformations [?, ?, ?, ?]. An interesting extension of their work would be to apply these findings to the field of image compression with the goal of developing more efficient algorithms.

## Appendix A

# Eye-Tracking

For our gaze directing technique described in chapter 7, we use the infrared camera-based ViewPoint EyeTracker<sup>®</sup> developed by Arrington Research, Inc. Our custom-built gaze directing software uses the software developer's kit (SDK) provided by Arrington Research, Inc. to interface with the eye-tracker in real-time.

Figure A.1 illustrates how the eye-tracker works. An infrared light source is used to illuminate the eye. When viewed with an infrared sensitive camera, the pupil, which acts as an infrared sink, appears as a filled black circle. This method is preferred as it allows for the easy discrimination of the pupil.

Individual frames from the video signal of the camera are captured at a rate of 60Hz and a resolution of 640 x 240. Fast image segmentation algorithms are used to separate the pixels that are part of the pupil. The resulting segmented image is then analyzed to determine the center of the pupil.

During calibration, stimuli are presented at different regions of the display, called stimulus points, and the corresponding location of the center of the pupil is determined. This information is used to compute a function that maps the pupil coordinates to display coordinates. It should be noted that the best algorithms for performing this mapping are

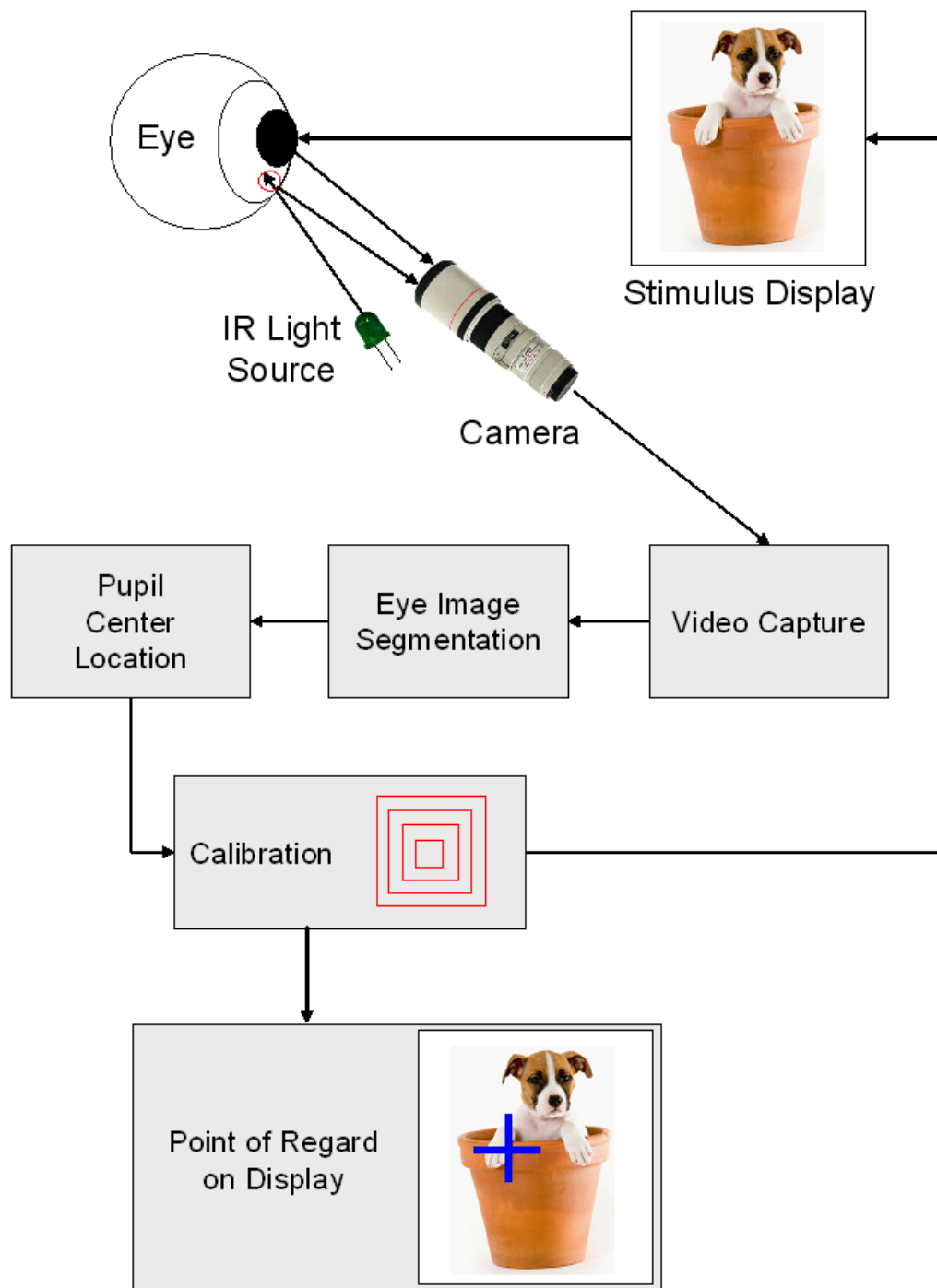


Figure A.1: Overview of ViewPoint EyeTracker® system. Adapted from ViewPoint EyeTracker® Software User Guide.



non-linear since eye movement has a rotational component. The algorithms used by the ViewPoint EyeTracker<sup>®</sup> are company proprietary.

## Appendix B

# Visual Phenomena

### B.1 Color Constancy

Humans have the ability to accurately perceive the color of an object or a surface under different illumination conditions. This phenomenon is known as *color constancy*. Figure B.1 shows a scene under two different lighting conditions. The colors that we perceive as blue, yellow, and red are shown below each scene. Notice that the color that we perceive as blue in the yellow-illuminated image is the same as the color we perceive as yellow in the blue-illuminated image. Figure B.2(a) shows a scene under normal daylight conditions. Figure B.2(b) shows the same scene with the color of the banana modified to a dark green color. The banana no longer looks natural. Figure B.2(c) shows the same scene under a green illuminant. The color of the pixels of the banana in Figure B.2(c) is identical to those in Figure B.2(b), however, we perceive them as being much lighter. Such images provide evidence that our visual system somehow discounts the contribution of the illuminant. Several researchers suggest that double-opponent cells are responsible for this phenomenon [?, ?, ?]. Opposing theories suggest that color constancy is caused by poor color memory or by a process called *chromatic adaptation* [?].

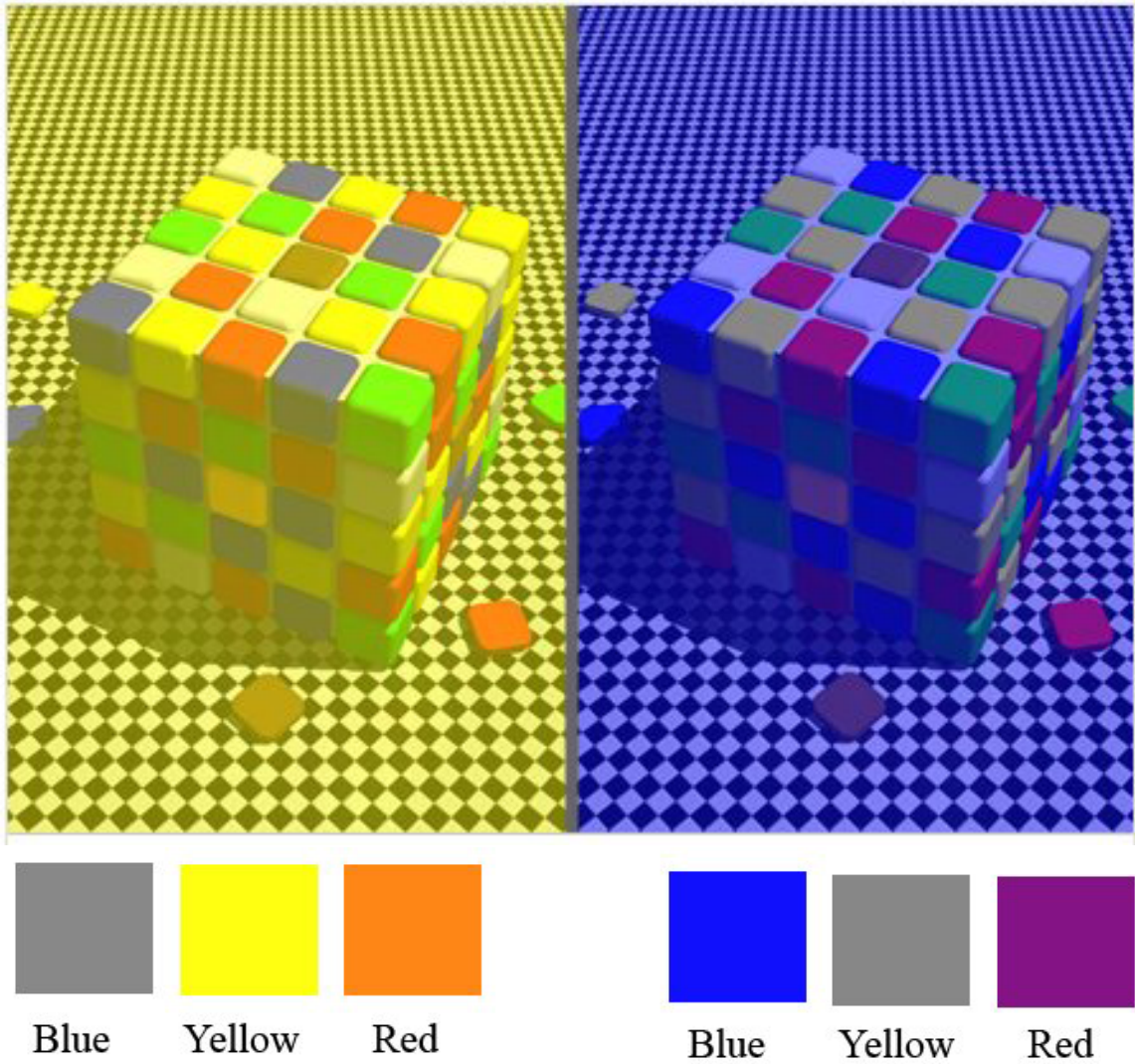


Figure B.1: Example of color constancy. A scene under different lighting conditions. The colors that we perceive as blue, yellow, and red are shown below each scene. Courtesy of Laboratory of Dale Purves, M.D., Center for Cognitive Neuroscience, Duke University [?].



(a) Daylight illumination



(b) Modified banana color



(c) Green illuminant

Figure B.2: Example of color constancy. The banana in (c) is identical to the banana in (b) but is perceived as being lighter. Courtesy of G. M. Johnson [?].

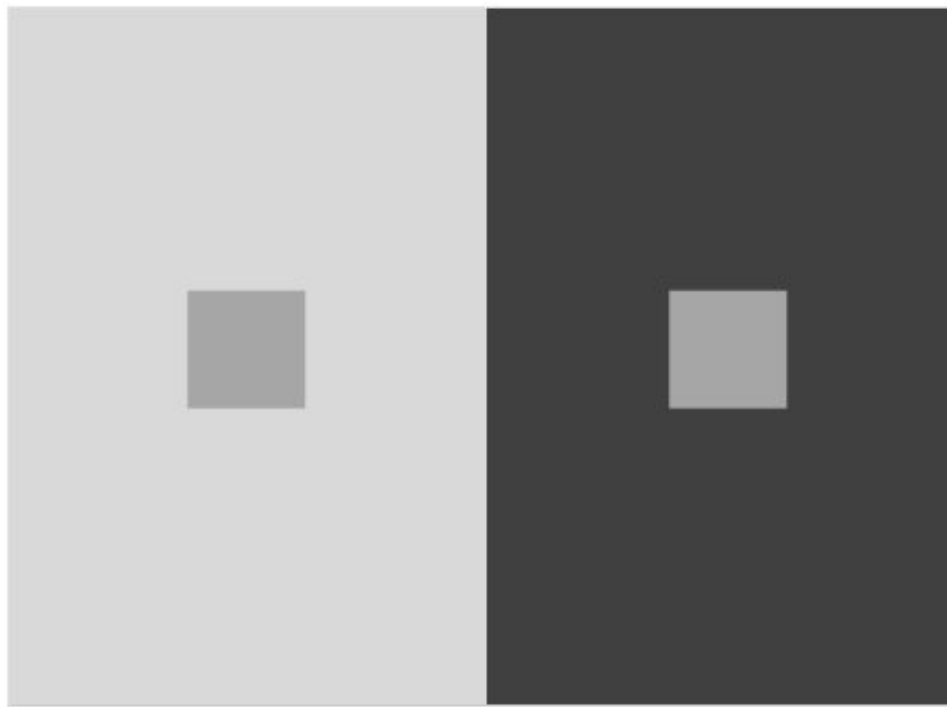


Figure B.3: Example of simultaneous color contrast. The small squares are identical, however the square surrounded by the darker background (on the right) appears lighter the square surrounded the lighter background (on the left). Courtesy of Laboratory of Dale Purves, M.D., Center for Cognitive Neuroscience, Duke University [?].

## B.2 Simultaneous Color Contrast

The perceived color of an object or a surface is influenced by surrounding colors. This phenomenon is known as *simultaneous color contrast* [?, ?]. Simultaneous color contrast is a corollary to the color constancy phenomenon [?] and can also be explained in terms of double-opponent cells [?, ?, ?]. Figure B.3 shows an example of simultaneous color contrast.

# Curriculum Vitae

Reynold Justin Bailey

**Date of Birth** April 23, 1977

**Place of Birth** Kingstown, Saint Vincent & the Grenadines

**Degrees** M.S. Computer Science, May 2004  
B.S. Magna Cum Laude, Computer Science, May 2001  
B.S. Magna Cum Laude, Mathematics, May 2001

**Publications** Reynold Bailey and Cindy Grimm, “*Perceptually Meaningful Image Editing: Manipulating Perceived Depth and Creating the Illusion of Motion in 2D Images*”, in The Visual Computer: International Journal of Computer Graphics, 2007.

Ankit Mohan, Reynold Bailey, Jonathan Waite, Jack Tumblin, Cindy Grimm, and Bobby Bodenheimer. “*Table-Top Computed Lighting for Practical Digital Photography*”, in IEEE Transactions on Visualization and Computer Graphics, July 2007.

Reynold Bailey, Cindy Grimm, and Christopher Davoli. “*The Effect of Warm and Cool Object Colors on Depth Ordering*”, in Proceedings of Symposium on Applied Perception in Graphics and Visualization, Boston, MA, July 2006.

Ankit Mohan, Jack Tumblin, Bobby Bodenheimer, Cindy Grimm, and Reynold Bailey. “*Table-Top Computed Lighting for Practical Digital Photography*”, in Proceedings of Eurographics Symposium on Rendering, Konstanz, Germany, June 2005.

Reynold Bailey. “*A Thesis on Techniques for Non-Photorealistic Shading Using Real Paint*”, Master’s Thesis, Department of Computer Science, Washington University in St. Louis, May 2004.

Christopher Kulla, James Tucek, Reynold Bailey, and Cindy Grimm. “*Using Texture Synthesis for Non-Photorealistic Shading from Paint Samples*”, in Proceedings of the 11th Pacific Conference on Computer Graphics and Applications, Alberta, Canada, October 2003.

Nelson Passos, Delvin Defoe, Reynold Bailey, Richard Simpson, and Ranette Halverson. “*Theoretical Constraints on Multidimensional Retiming Design Techniques*”, in Proceedings of AeroSense Aerospace/Defense Sensing, Simulation and Controls, Orlando, FL, April 2001.

Reynold Bailey, Delvin Defoe, Ranette Halverson, Richard Simpson, and Nelson Passos. “*A Study of Software Pipelining for Multidimensional Problems*”, in Proceedings of 13th International Conference on Parallel and Distributed Computing Systems, Las Vegas, NV, August 2000.

August 2007

Short Title: Perception-Guided Image Manipulation

Bailey, Ph.D. 2007