

The Transitional Image

by **Laura Ruth Scholl**

Bachelor of Fine Arts, University of Massachusetts at Amherst, 1980

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning in Partial Fulfillment of the Requirements for the
Degree of

Master of Science in Visual Studies

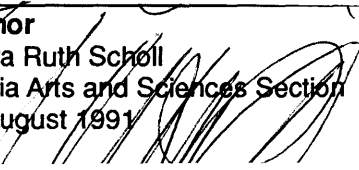
at the

Massachusetts Institute of Technology
September 1991

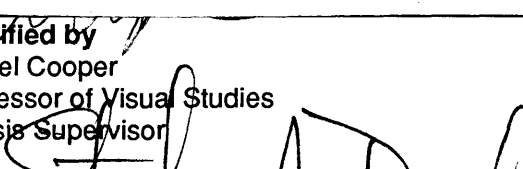
© Massachusetts Institute of Technology 1991, All Rights Reserved



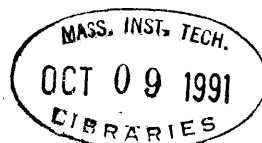
Author
Laura Ruth Scholl
Media Arts and Sciences Section
12 August 1991



Certified by
Muriel Cooper
Professor of Visual Studies
Thesis Supervisor



Accepted by ✓
Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students



Botch

The Transitional Image

by Laura Ruth Scholl

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning
in partial fulfillment of the requirements
for the Degree of Master of Science in Visual Studies
at the Massachusetts Institute of Technology, on
12 August 1991

Abstract

An interactive computer image can easily become complex and incomprehensible without the ability to isolate information of interest. A new type of image, the *transitional image*, is designed to assist us to this end. This thesis describes both a system of algorithms and the control over this system that generates such images. The system is based upon two hypotheses: (1) The perception of space helps us distinguish specific features within an image; and one way space can be perceived is from gradients. Focal, transparency, and color value gradients are chosen to exemplify this theory. (2) Our attention is drawn to change which becomes evident only over time. The control over the representation helps us begin to understand how perception reduces what we see to that which interests us. In this thesis, in order to grasp information an observer may select features in the image, which are organized spatially in a multi-layered object. For example, with a geographical map an observer can isolate feature layers concerned with air transportation (i.e.: airports, airspace, obstructions, etc.) and describe gradient values for transparency, focus and color value, and start and end points in time. The result is the *transitional image*, a continuous but changing image that enables us to apprehend information.

Thesis Supervisor: Muriel Cooper
Title: Professor of Visual Studies

This effort was sponsored in part by the Rome Air Development Center (RADC) of the Air Force System Command and Defense Advanced Research Projects Agency (DARPA) under contract No. P30602-89-C-0022. The views and conclusions in this document are those of the author, and should not be interpreted as necessarily representing the official policies, express or implied, of RADC of the Air Force System Command, DARPA, or the United States Government. This research was also sponsored in part by NYNEX.

Acknowledgments

It is with much gratitude and respect that I acknowledge the following people:

Muriel Cooper and Ron MacNeil for accepting me into the master's program, and allowing me to pursue research of my own interest.

My colleagues at the Media Lab for their help and ideas.

Alan Kay, who suggested I come to the Media Lab in the first place.

My friends in California, who were always there for me. Special thanks to the Conn family, Nancy Leiviska, Robin Meyers, and Traci Wheeler.

My friends Copper Giloth and Justin West, whose home served as a great escape from MIT.

My friend Ina Broida, who encouraged me to continue my academic studies.

And to my family, whose love and support knows no bounds.

Table of Contents

Abstract	2
Acknowledgments	3
Chapter 1	6
Introduction to Transitional Images	
Motivation	6
Definition	8
About this Thesis	9
Chapter 2	10
Approach:	
Design of the Representation	
The Multi-layered Structure	10
Why Gradients?	11
Focal Gradient	12
Color Value Gradient	12
Transparency Gradient	13
Chapter 3	14
The Interface:	
Studying Transitional Images	
Chapter 4	16
Background Theory	
Visual Perception	16
The Philosophy of Bergson	19
Cubist Art	20
Transitions in Motion Pictures	22
Chapter 5	25
Related Research	
in Computer Graphics	
Focus Pull	25
Depth from Focal Gradients	26
Pyramid Coding	28
Visual Dynamic Range	30
Image Compositing	32
Interactive Cinema and Video	33

Chapter 6	36
Program Design and Implementation	
Gradient of Focus	36
Gradient of Color Value	37
Gradient of Transparency	38
Image Compositing	38
Transitions	39
The Connection Machine	39
Interface on Hewlett-Packard	41
Chapter 7	44
Examples of Transitional Images	
A Geographical Map	44
An Image of a Three-dimensional Space	46
An Image Containing Multiple Perspectives	47
Chapter 8	49
Conclusions	
Results	49
Future Work	51
Bibliography	53

1

Introduction to Transitional Images

“The formulation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old problems from a new angle, requires creative imagination and marks real advance in science.”

EINSTEIN AND INFELD¹

1. Einstein, *The Evolution of Physics*, 92.

An interactive computer image can easily become complex and incomprehensible without the ability to isolate that which is of interest. This thesis describes a new type of image, the *transitional image*, which continuously adapts to assist us to this end. This thesis describes both a system of algorithms and the control over this system which generate such images. The system is based upon two hypotheses: (1) The perception of space helps us distinguish specific features within an image; and one way space can be perceived is from gradients. (2) Our attention is drawn to change which becomes evident only over time. Control over this representation helps us begin to understand how perception reduces what we see to what we want to see.

Motivation

Throughout history new types of images have emerged as a result of technological innovation. These images have helped us to learn more about how we perceive information and have actually changed our perceptual process. Medieval painters rendered all sides of an object, even the sides that were not within our viewpoint. Then, during the Renaissance, Brunelleschi invented a method for rendering perspective, and painters painted only the sides that were in their field of vision. Renaissance images indicated three-dimensional relationships from an illusionistic one-point perspective. Illusionism helped researchers and artists analyze fundamental properties of the concrete visual world.

With the invention of the camera in the nineteenth century, an image could be captured in a moment. This pushed images toward new levels of scientific realism, altered our sense of time, and enabled exploration

of the perception of light. Pre-mixed paints in tubes were created. Painters could leave the studio to paint, and Impressionism was born. Images were painted as patches of color, a method that echoed the ideas of the Structuralists who theorized that we perceive space from patches of light¹ and the technique of photography. Then in the twentieth century, motion pictures redefined the perceptual awareness of time and space. Now images from television are shaping this awareness. For example, we are able to follow an hour long television program even though it has breaks in continuity (commercials) at approximately every eight minutes.

1. Hochberg, *Perception*, 83.

Advances in computer technology have also brought about the invention of new images. These images describe data that were previously inaccessible, for example, weather from satellite data, fluid flow from simulation data, and molecules from three-dimensional models. In addition, for a number of years it has been possible for a computer to receive and send information in 'real time'. Consequently, we have interactive computer programs, such as computer-aided design packages, video games, and paint systems.

When we combine all of the information available to us with interactivity, we can easily create complex, cluttered, illegible images, densely packed with information (in the form of imagery, text and graphics). These images have extended the amount of information that we can take in at one glance. Although we are now able to absorb more visual information than we were before having been exposed to new forms of images, if too many objects of information change from image to image we have difficulty keeping track of what is happening. Therefore, our perception is limited by the number of objects of which we are consciously aware.

Three questions emerge. Just how many distinct objects of information can we apprehend from an image at one glance? How can we adapt an image to coincide with our perceptual limitations, so that we can obtain

all possible information from that single image? And how do we maintain constant perceptions over time, so as not to lose context?

In order to explore these questions, a new scheme must be invented. It must enable us both to emphasize and to de-emphasize differences and similarities. With it, we must be able to simulate the fading of information in a manner analogous to the fading of information from our own awareness. In addition, the newly invented representation must allow us to select the features which interest us in a manner analogous to the way in which we are predisposed to retain only what interests us in our surroundings. In other words, we need to be able to generate an image that is no longer a snapshot of a particular moment in time, but one that is an evolving visual description of information that exists over time. This image is called the *transitional image*.

Definition

The *transitional image* is a single, continuous, but changing image, whose representation combined with user interaction reduces a complex, informationally dense image to an image that changes based on the observer's interests. The resulting image is not a discrete state of the complex image, but is a series of qualitative transformations which flow into one another, and exist over time.

Each element of information contained within a transitional image is organized spatially as a layer in a multi-layered data object. An observer can select layers from the multi-layered object and can describe a change in focus, transparency, and color value, as well as the rhythms of these changes for each layer. These changes are what guide the observer's attention to the elements that have been selected.

A geographical map, for example, can be organized in layers of features. The observer may select layers related to a particular concern. For instance, a person curious about air transportation, may choose to view airports, airport labels (Figure 1), and airspace (Figure 2) in full focus

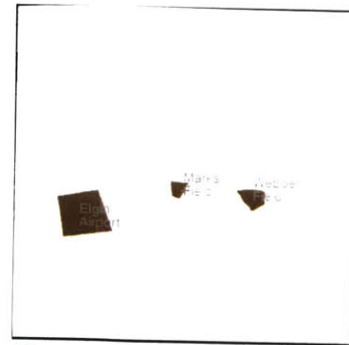


Figure 1. Airports and airport labels layers.

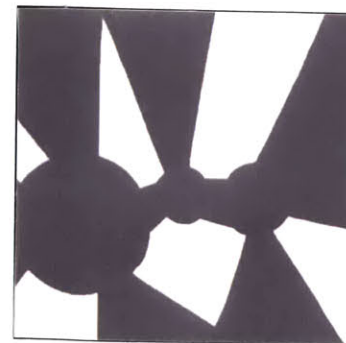


Figure 2. Airspace Layer.

and full brightness. Relevant layers such as the roads and waterways are also visible, but appear somewhat blurry, transparent, and darkened (Figure 3). In addition, timing descriptions are assigned to indicate when the selected layers will become most prominent. The user's attention is drawn to the changing features and come to rest on the features that are clearly in focus, opaque and bright; blurry, transparent, and dark information are perceived to be in the background.

Transitional images take advantage of the *a priori* existence of the entire complex image and the expected outcome of the observer's selection. Transitional images simulate both the way the mind fuses recollections with an awareness of the instant present and an anticipation of the immediate future, and the way it turns this fusion into a working perception of knowledge.

About this Thesis

This thesis is divided into three main parts. The first part, Chapters 2 to 5, comprises a description of the approach and the interface that is proposed by this thesis, and the theoretical preliminaries and related work that led to the transitional image. The second part, Chapter 6, discusses the realization of the approach and controlling interface. The third part, Chapters 7 and 8, consists of three examples of transitional images, the results of these examples, and some suggestions for future direction.



Figure 3. An image showing air transportation heightened.

2

Approach: Design of the Representation

“In nature we never see anything isolated, but everything in connection with something else which is before it, beside it, under it, and over it.”

GOETHE¹

In this thesis, a *representation* is a system of transformational algorithms that produce a description. A user can interact with this system to generate that which I have termed the “transitional image.” The purpose of the representation outlined herein is to enable an observer to reduce a complex, informationally dense image into a ‘real time’, continuous, but changing image of the concern of any moment.

This system is based upon the following two hypotheses. The first hypothesis states that perception of space helps us distinguish specific features in relation to other features within an image. Space can be perceived from gradients. Pentland report evidence that spatial information can be derived from focal gradients.² Also included are gradients of transparency and color value. The second hypothesis suggests that we discern changes in the relationships between features as movement, which is a consequence of time. The first goal of this thesis to design and build such a representation is described in this chapter.

The Multi-layered Structure

The first consideration is how the features within complex image should be organized. The design of a data structure is often tied to the data at hand and assumptions about that data. This design assumes that the technology to extract objects from an image already exists and that additional data can be easily simulated.

The first data set used to experiment with the hypotheses of this research met both of these assumptions. They are geographical, land/satellite (LANDSAT) data and radar reflectivity, weather data. It is fortuitous that the technology for feature extraction from LANDSAT data is rather

1. Goethe, *Truth and fiction relating to my life*, 171.

2. Pentland reports the positive findings of two experiment's in *The Focal Gradient: Optics Ecologically Salient*, a supplement to *Investigate Opthomology and Visual Science* in April 1985.

advanced, and so some of the features were already extracted from the original data set. The weather data consist of slices of radar data at various altitudes above the earth's surface. Each slice exists as a complete, but partially transparent image in its own right. These data sets became the perfect makings of a map, a type of image known to be complex and difficult to read. Additional features could easily be simulated because maps consist of symbols and typographic information as well as image data.

These data sets led to the idea of a multi-layered structure (Figure 4), wherein each feature has its own layer. In this format the final image is a composite of all the layers. Another influencing factor is what is actually going to be done with this information.

Why Gradients?

The question, "how do we come to know the visual world?" plagues us through out this thesis. One way to begin to think about this question is to define some of the properties that we see in the visual world. For example, we see color, light and shade, and textures. The visual world is made up of shapes, surfaces and edges. It is right side up, goes on as far as the eye can see, and is modelled in three-dimensions.¹

Images, although they lack the three-dimensionality of the visual world, act as useful surrogates because they share many of the other properties such as color, light and shading, surface quality, distance etc. By examining these properties within an image we can easily create gradients. A gradient is the rate of change of some property over a continuous stimulus. Gradients can help us reconstruct the missing three-dimensionality. In addition, gradients are computable. The algorithms to generate focal, color value, and transparency gradients are three of the main transformational algorithms of the transitional image representation.

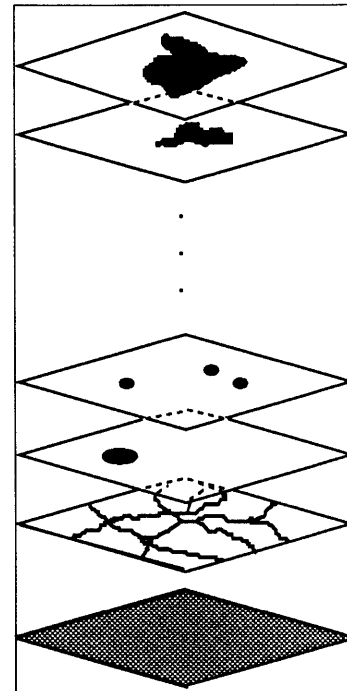


Figure 4. Multi-layered structure: layer 1, at the bottom, is the LANDSAT image, the middle layers 2-19 are features, and at the top are 9 layers of weather data.

1. Gibson, *The Perception of the Visual World*, 3.

Focal Gradient

The focal gradient is chosen for several reasons. One reason is based on the eye itself. The mechanics of the eye suggest that the effect of contrast on our visual attention is due to retinal lateral inhibition, which causes faster neural firings where the retinal image changes quickly. This means that the eye is most sensitive to sharp edges and less sensitive to surrounding regions.¹ By varying the depth of field of particular layers of the transitional image, we essentially create contrast between that which is sharply in focus (Figure 5) and that which is blurry (Figure 6).

Another reason is the effect of the *focus pull* which will be discussed in detail in Chapter 3 entitled, Related Research in Computer Graphics. A focus pull, in traditional cinematography, is described by the change of the camera's point of focus from either the foreground to the background, or the opposite. It is a known effect used to direct the viewer's attention to an area of interest inside of the framed image of a motion picture. A gradient of focus is what is necessary to simulate a focus pull.

The last reason, also discussed in the chapter on related research, is the discovery that depth information can actually be derived from the focal gradient. In Pentand's research, he presents the use of two images that vary only in focus to calculate depth. If we have the entire gradient, consisting of an infinite number of images which creates a smooth gradient, we can literally see information either come forward or recede in depth. The Gaussian pyramid encoding technique is chosen because of its near 'real time' computability and because the Gaussian simulates the appropriate lens for depth of field.

Color Value Gradient

The color value gradient is used because it too creates contrast in the image. It is well known that if we vary the color value of an area of color, it becomes either lighter or darker. If we keep that area constant, but vary the areas surrounding it, the surrounding areas change and so it seems

1. Gregory, *Eye and Brain*, 83.

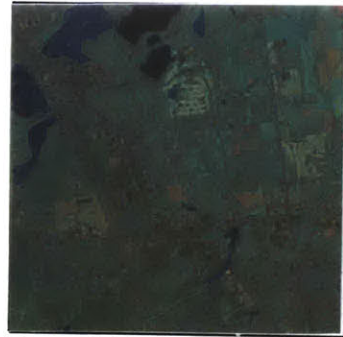


Figure 5. Background map in full focus.

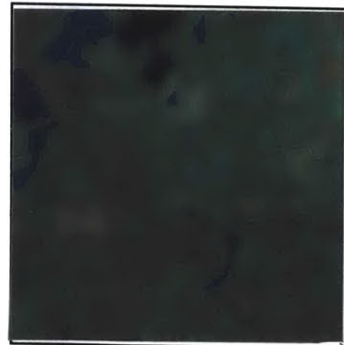


Figure 6. Background blurred.

does the specific area in question! What is important then is the relationship between colors. By varying the color value gradient we can experiment with these relationships. Also color value is used in traditional image making (i.e. painting and printmaking) to create a sense of space. The darker areas appear to recede into the background and the lighter areas come forward and are more easily noticed (Figure 7).

Transparency Gradient

The transparency gradient acts as an analogue to the abstract concept of information fading from our consciousness. By creating a gradient, fading can appear to be gradual. Used in isolation this transformation allows the observer to eliminate information completely from the image. When combined with the other gradients it can be used to emphasize and de-emphasize information.



Figure 7. Color value is used to isolate regions in the map, the background is shown at two depths.

3

The Interface: Studying Transitional Images

Control over the representation can help us understand how perception reduces what we see to that which interests us. A transparency gradient can be used to mirror the fading of an object from our consciousness. Focal and color value gradients can be used to heighten or diminish the amount of contrast in an image. Described in this chapter is the second goal: to control the representation.

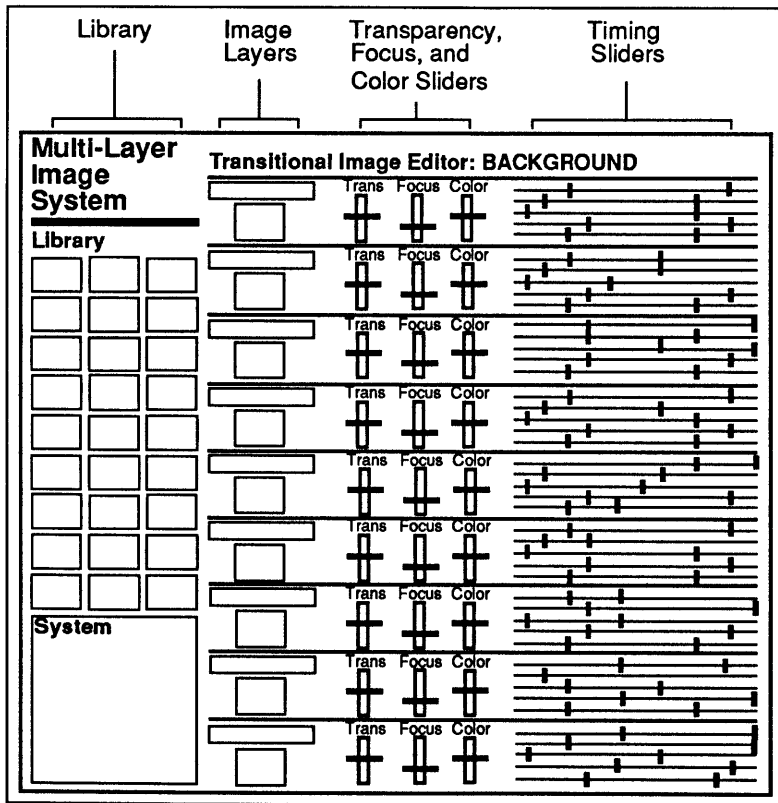


Figure 8. Schematic drawing of interface screen:
Library and Transition Sliders.

It is through the interface that the observer may affect the image (Figure 8). As a consequence of selecting features, the image becomes transitional. Ideally, the observer should be able physically to reach out, turn dials and mix in the information of greatest interest; but we do not have such an interface and its accompanying input device. Somehow,

therefore, we must precisely determine the settings of blur, transparency and color value for each layer with respect to what has been selected. First, we must study the effects of gradients; therefore, the current interface is designed to allow for exploration. This means that the interface must enable the observer to have 'real time' control at the most detailed level.

The observer is given the freedom to organize the multi-layered structure, by being allowed selection from a library of layers. The observer places the chosen layer into the structure at a specific position, determined by him/her. Once a layer is in the structure, the user can describe the desired amount of blur, transparency and color value by using virtual sliders on the control screen or with the knob box, or by typing in values.

The activity of moving through the gradients to select a desired percentage of blur, transparency and color in itself defines a transition. However, since the controls are virtual, the user has to pay attention to the input devices to be certain that the cursor continues to point to the selected feature. Despite efforts to develop a representation that guides the user's attention to features within the transitional image, they may be missed entirely because the observer's attention is elsewhere.

In order to study the effects of the selections being made, the observer is given control over both when the selections begin and end in time and how the rhythm changes. This enables the user both to see the effects of the choices made, and to integrate the changes of any number of selections. This would be impossible if the user depended solely on adjusting the input devices one at a time. The user is able to integrate and see changes by defining a length of time in abstract units and then setting start and end points and the motion type for each layer. At any point the observer may send a "playback" message and review the transitional image.

4

Background Theory

“The whole difficulty of the problem that occupies us comes from the fact that we imagine perception to be a kind of photographic view of things, taken from a fixed point by that special apparatus which is called an organ of perception – a photograph which would then be developed in the brain-matter by some unknown chemical and psychical process of elaboration.”

BERGSON¹

1. Bergson, *Matter and Memory*, 40.

How much information can we grasp from an image at one time? How can we adapt an image to the limits of our perceptual system so that it describes the amount of information that we can apprehend? How do we maintain continuity in continually changing image? Before we can answer such questions, it is helpful to examine existing theories that are concerned with perception. This chapter concentrates on four areas:

(1) visual perception which is directly concerned with the act of seeing, (2) the philosophical discourse of Bergson, because of its direct relationship to the changing nature of transitional images, and because of its historical relationship to the following two areas, (3) Cubist theory which developed a new system for seeing three-dimensions in a two-dimensional image, and (4) the comparison of transitions in motion pictures to the transitional image.

Visual Perception

Germane to this thesis is the question of how we perceive space; or more generally, of how we come to know the visual world. Theories developed centuries ago continue to influence the way in which researchers think about these questions today. However, in the twentieth century, James J. Gibson developed a theory that influenced this research.

In the seventeenth century, Kepler hypothesized that sight is based on the retinal image. He believed that the eyeball is like a lens that produces a copy of an image inside our heads. A generation later, Descartes challenged Kepler's idea and developed a theory in which the retinal image is decomposed into points on the retinal plane, which stimulate

fibers in the optic nerve. According to this doctrine, seeing is the process of interpretation (made by the mind) of the subjective effects of these “stimulations”, later to be termed “sensations.” The Cartesian theory of sensations is based on Descartes’ beliefs that we are aware of our own cerebral excitations and that perception is “a result of our imagination.”¹

1. **Reed**, *James J. Gibson and the Psychology of Perception*, 31.

In 1867, Helmholtz, in his treatise on physiology of optics, constructed a model of seeing, based on Cartesian theory. He first analyzed the physics of light, then reduced sensations to points of light, finally defined perception as an “intellectual combination and interpretation of the sensory data.”²

2. *ibid.*, 32.

The first scholars to challenge Cartesian theory were the Gestalt psychologists. Experiments (carried out by Wertheimer in 1912) with the phi-phenomenon disproved the theory that perception is a strict by-product of sensations. The phi-phenomenon occurs when two dots of light, projected in rapid succession in two positions on a blank background, appear to be one moving dot. Wertheimer, by combining two displays of apparent motion in opposite directions, proved that there is no evidence that apparent movement is caused by spatial sensations.

James J. Gibson was the first to formulate an actual theory in complete rejection of the Cartesian tradition. He based his theory on the idea that vision cannot be explained by retinal images because optical information exists not on the retina, but in the environment. The retinal image is flat, yet the physical environment is three-dimensional. If perception depends solely on the stimulations of retinal images on the optic nerve, how is the third-dimension restored? Gibson’s important contribution redirects research away from a concentration on sense-data by maintaining that the perception of space is based upon the variations in stimulation that correspond to the continuous, visible surface of an object.

The traditional cues for sensing depth usually include the following: linear perspective, size constancy, motion parallax, change in color, changes in relative brightness, shading, and the apparent overlapping of objects seen as edges when one object is in front or behind another.

Gibson considered these cues to be variables which create gradients of stimulation over the physical surfaces of objects, not to be facts about how we see the geometrical shapes of objects. A gradient is the rate of change of some property over a continuous stimulus. For instance, edges may be seen as an interruption in two otherwise uninterrupted gradients; motion parallax may be thought of as a deformation gradient; and brightness is a gradient of color value. Gibson concentrated specifically on texture gradients.¹

Gibson's theory, called the "ground theory," is constructed from five main principles. The first states that an edge may account for a figure against the ground. The second states that a stimulus is discoverable for the quality of depth over a continuous surface. This stimulus might be a gradual increase or decrease in the texture density of the retinal image. Similarly, an edge could have a stimulus for the impression of depth. The third principle refutes Kepler's theory. There is no copy of the outer world in the eye, however, surfaces, slopes and edges may very well have correlates in the retinal image. Under the fourth principle, the pattern of the retinal image can also be considered as a stimulus. Under the fifth, perception is divided into two types, literal and schematic. Literal perception, which is more or less constant, includes colors, textures, surfaces, edges, slopes and interspaces. Schematic perception, which is selective, is made up of objects, places, people, signals, and written symbols.²

This thesis draws upon Gibson's theory that gradients aid us in perceiving space. In so doing, it uses gradients both to distinguish objects or features within an image, and to heighten the sense of space in an image that is nevertheless two-dimensional.

1. Gibson, *The Perception of the Visual World*, 78.

2. *ibid.*, 10.

The Philosophy of Bergson

The effect of time on perception is another important element of this research. Bergson's philosophy on the duration of perception was chosen because it closely matches the ideas which were conceived in the design of the transitional image. In addition, there exists an established relationship between his philosophy and both Cubism and motion pictures.

In *Matter and Memory* (1896), Bergson states, "However brief we suppose any perception to be, it always occupies a certain duration, and involves, consequently, an effort of memory which prolongs, one into another, a plurality of moment."¹ Bergson believed that experience gives us a composite of movements through space and in time. This composite serves to unite movement in such a way that it can no longer be grasped as separate positions in space.

1. Bergson, *Matter and Memory*, 1991.

Movement is a translation in space. Every time something is translated in space, there is a corresponding qualitative change in a whole. When thinking about movement as juxtaposed events, or a sequence of positions, we can only perceive differences in 'degree'. Bergson's philosophy proposes that we think instead in terms of time, where we can perceive differences in 'kind'. In *Creative Evolution*, Bergson exemplified his point with the physical experiment of putting a cube of sugar in water. He says, 'I must, willy-nilly, wait until the sugar melts.'² What does he mean? As the sugar particles detach and change position in the water, there is a qualitative transition from the cube of sugar in water to sugared water. A rhythm of duration is revealed in the process of dissolving, which is ultimately a change in the whole.

2. Bergson, *Creative Evolution*, 10.

Movement, then, has two aspects. One aspect is that which happens between objects; the other aspect is that which expresses duration. Through movement the whole is divided up into objects, and objects are re-composed in the whole, and, as a result of this division and re-composition, 'the whole' changes.

Perception reconstitutes movement. It is only when the objects within an image begin to move that the image changes. Transitional images use this movement to direct the observer's attention within the image, because we are conditioned to notice change.

Cubist Art

The question "how do we see?" directly effects this thesis. It is useful to take a look at how others, with similar goals, have studied this problem.

The Cubist's were concerned with the way we perceive both space and time. Challenging illusionistic one-point perspective, they created a new system to describe three-dimensional relationships. They broke up objects in the image into parts and then re-combined discrete viewpoints of these parts to formulate a composite object. They also brought together objects that would normally lie at different points in space, in order to imply a sense of simultaneity.

Cubism was an expression of the scientific discoveries of its time. It was the pictorial expression of the theory of relativity, which assimilates space to time in order to express the invariance of distance.¹ The philosophy of relativity maintains that man does not objectively perceive the world, but invents reality while he perceives it.² The Cubists created a pictorial view of this reality by expressing a fusion of space, time and form on a two-dimensional surface.

Cubism is thought to be primarily an outgrowth of Cezanne's experiments with diminishing the division between form and space.³ [FRY, 66]. He developed a way of denying illusionism by means of integrating the surfaces of objects with the surrounding space. These experiment relate to Gibson's theory of perception (1950), which states that visual space should be conceived as a continuous surface, or an array of adjoining surfaces, instead of as objects surrounded by air.

1. Deleuze, *Bergsonism*, 86.

2. Glover, *The Cubist Theatre*, 11.

3. Fry, *Cubism*, 66.

The Cubists dismantled the objects in an image, thus forcing the viewer to scan the image for structural parts. Continuity emerged through the movement of the viewer's eyes and memory. Cubism transformed the dynamic nature of perception into the static, two-dimensional medium of painting. Objects became located in time, and pictorial space was not just a container, but part of an event. The outcome was a new manner of seeing, guided by a new compositional order. This order was a network of transparent, overlapping planes that fade in and out of each other over time.

The experience of the viewer, assimilating accumulated parts of the image with memory, parallels the philosophy of Bergson. The Cubists were aware of Bergson's theories. Gleizes and Metzinger paraphrase Bergson's cinematographic model for the "mechanism of our ordinary knowledge," in their explanation of movement in Cubist painting, "around an object to seize several successive appearances, which, fused in a single image, reconstitute it in time."¹

1. Lawder, *The Cubist Cinema*, 20.

By 1913, the Cubists used collage as a way to escape illusionism altogether. Ready-made facsimiles were incorporated into the painted image to act as symbols of themselves. Words, letters and numbers were also included to help orient the viewer. Spatial ambiguity could be achieved by overlapping the integrated materials in one order at a specific point in the painting and in the opposite order at another point. Lines and surfaces were painted over, under, or separate from the material, and so used to describe further the color and form of an object.

The Cubists pushed the limits of painting in order to push the limits of our perception. Their paintings serve as experiments for the analysis of vision and time, and for understanding the role of memory in perceiving images.

Transitions In Motion Pictures

The techniques for articulating the sense of a continuous flow of time, discovered with motion pictures, serve as a starting place for developing a representation for transitional images. The path from photography and Cubist painting to motion pictures is significant, and so it is outlined briefly below, before delving into the use of transitions in motion pictures.

With the advent of photography, images of the world could be recorded directly. By the mid-nineteenth century, still photographers were already responding to the inflexibility of the medium and began to think about motion, or the time element.

Oscar G. Rejlander and Henry Peach Robinson, in England, experimented with merging portrait and landscape photographs by compositing images together. They created photographic dramas using actors and photographic effects achieved by constructing collages with film negatives.¹ Cubism can be seen as a reaction to photography, although it is not known whether or not the Cubists had been aware of these photographic collages.

The experiments of Eadweard Muybridge stand as precursors to motion pictures. In 1887, Muybridge was particularly interested in human and animal motion. In order to study their motion, he devised a setup using three cameras, each with twelve lenses and all connected to an electrical clock mechanism switch. With this setup he could photograph his subject at the same time from three different views. Jules-Etienne Marely also performed experiments in recording movement. The Cubists were aware of these investigations. In fact, Marcel Duchamp acknowledged his debt to Marley.² This is exemplified by Duchamp's famous painting *Nude Descending the Staircase*, which is compared with Marley's multiply exposed chronophotographs of men in motion. These chronophotographs are simultaneous images, unlike motion pictures which are sequential. Nevertheless, motion pictures are an extension of the Cubist

1. Monaco, *How to Read a Film*, 26.

2. Lawder, *The Cubist Cinema*, 7.

image into the dimensions of movement through time.

At the outset motion pictures imitated natural perception. At first, the view point of the shot (which consists of a succession of instances extracted from a spatial and temporal continuum¹) was immobile. It was defined by its location in space. Motion pictures evolved through montage and the mobile camera. They were no longer restricted by a single viewpoint and our perceptual abilities were changed. With multiple viewpoints, the shot was no longer defined by its spatial characteristics, but instead by time. Motion was not added to the series of fragments of space and time; it was part and parcel.

1. Burch, *Theory of Film Practice*, 4.

Motion pictures served as a model for Bergson's concepts of time and perception. He states, "what is real is the continual change of form: form is only a snapshot view of a transition".² For Bergson, motion pictures and, in this thesis, time is contained in the transition. In filmmaking the definition of transition, however, is different by degree. A transition is defined as an indirect image of time. It produces the appearance of continuity by joining two different spatial/temporal situations, or shots.

2. Bergson, *Creative Evolution*, 302.

There are five types of temporal articulation and three types of spatial articulation that are used to define different transitions.³ The first temporal type is known as the *reverse angle shot*. An example of this is a cut from one person who is speaking to another person who is listening, or from one side of a door and to the other. The second type is a gap or time abridgement that occurs over some measurable amount of time, this is called a *temporal ellipsis*. In this type, a part of the action is omitted. This can be done with successive close ups, or with cross cutting, which is an alternation between two actions that take place in two distinct locations, or two different camera angles. The third, is the *indefinite ellipsis*. This type is only measurable by something external, such as a change in clothing style, or from daylight to darkness. A *time reversal* is the term for the fourth type. Here part of the action is repeated in an artificial way. An example of this is a cut from someone walking through

3. Burch, *Theory of Film Practice*, 8.

the door and to the moment when the door was opened. The last type is the *infinite time reversal*, such as a flashback into the past or a flashforward into the future.

There are three types of articulation between spaces that are used as transitions. The first is similar to the reverse angle shot and is called the *straight match cut*. It is used to preserve spatial continuity. This shot is taken from the same camera angle, but is either closer or farther away. The last two types are the extremes of discontinuity: The second is a cut to another shot of close *proximity*, such as another area of the same room. The third type is a *radical spatial discontinuity*, which is a cut to a completely different location.

All of these transitions involve “cutting” both literally and figuratively. Literally, the filmstock itself is “cut” and spliced back together. Figuratively, the apparent movement is no longer strictly continuous, but “cuts” to a different image. The concept of cutting is a key difference between the transitions of film and transitional images. For the transitional image, which is always a description of the objects of interest within one image (which is a composite of any number of images contained in the multi-layered structure), there is no cutting to different images.

Another major distinction is the concept of a shot, or a linear succession of instances. Transitional images are not made up of a sequence of fragments of time and space. They are the fusion of simultaneous elements in space, and they only succeed each other in so far as they are remembered by a conscious observer. If nothing changes, the image continues to appear. In film (or video), in order for the image to continue over time, the frame must be duplicated. With transitional images, change, not frames, denote the passage of time.

5

Related Research in Computer Graphics

This research produces a 'real time' representation that creates transitional images with which we may explore visual perception. I have not come upon any other computer graphics research that focuses on understanding the limits of our perception by generating a single, continuous, but changing image. However, some researchers and artists have developed techniques that have either become a part of this research or make useful analogies.

Focus Pull

The cinematographic technique *focus pull* plays a significant role in the design of the transitional image representation. Filmmakers use *focus pull*, or the changing of the camera's point of focus from the foreground to the background and vice versa, for the purpose of directing our attention within the scene.

Sturman

David Sturman has developed a technique to simulate the focus pull in computer graphics. In order to pull focus we must first compute the depth of field, which is the range of depth for which objects will be in focus at a particular exposure plane and then blur the detail that is outside this range. Sturman uses similar triangles to compute the depth of field. The amount of blurring at a given depth, known as the blur circle, is derived as a function of the center of (1) focus, (2) object depth, (3) focal point and (4) the f-stop.¹

Sturman's approach is based on the camera model and has proven to be extremely effective in computer animation, in which the camera is simulated. Transitional images are based on a model of our perceptual system. Perception includes understanding optics, which share many characteristics with the camera, but is much more complex than that

1. Sturman, "Motion Picture Cameras and Computer Graphics."

which can be described by a mathematical model. In addition, the transitional image does not restrict blurring to a three-dimensional model in space. Both precise calculation of depth of field and precise derivation of a blur circle, therefore, are not necessary. What is needed is a smooth blurring effect on each layer that is computationally efficient.

Johnson

Michael Johnson has developed a simulation tool that takes RGB/Alpha¹ images and traditional camera parameters (f-stop, focal length, focal point, and film width), blurs them according to Sturman's camera model and mattes them together to make a final image. He chose to implement this tool on the Connection Machine 2 System (CM-2) because the design of the hardware is inherently appropriate for filtering images. He chose the Fast Fourier Transform (FFT) for its efficiency, because the number of RGB/Alpha images needed to make the final composite could be arbitrary and the size of the blur circle could be as much as 75% of the image.

The choice of filter used for blurring depends on the type of lens being simulated and the degree of accuracy needed to model the optical properties of the lens. Johnson chose the Gaussian filter for this application. To filter each image it takes about 8 seconds. Although his results create beautiful images that are extremely accurate to the camera model and the eye, his method is not fast enough to generate these images in 'real time'. Johnson's insight about the appropriateness of the CM-2, is useful to this research, and is one of the reasons it was chosen as a platform (see Chapter 5).

Depth from Focal Gradients

Pentland

Researchers in computer vision are interested in developing mathematical models for apprehending information from pictures. These models can be extremely useful for understanding the mechanics of our

1. Porter, "Compositing Digital Image."

perceptual system. Understanding how we see is vital to the development of transitional images. Sandy Pentland discovered that focal gradients may be used to recover depth information, which he defined as the distances between the viewer and points in a scene. By measuring the amount of blur, one can reliably estimate depth, simultaneously at all points, from one or two images.¹

1. Pentland, "A New Sense for Depth of Field."

Previously, an autofocus method was used to derive depth from depth of field by searching for the lens setting that is clearly in focus at a particular point. With this method depth can only be measured at one point at a time. In contrast, Pentland's method required no search; instead it measured the amount a given point is blurred by the imaging optics, or the "blur circle," and combined this with the focal length of the lens system to compute distance at that point. Pentland also found that a two-dimensional Gaussian had the same general shape as the blur circle.

Pentland developed two methods that use focal gradients. The first method measured apparent blur near sharp discontinuities. This required calculating the blur circle by measuring the rate at which image intensity changes. He used the slope of the Laplacian across the zero-crossing to measure the rate of change. This method required that the scene characteristics be known before focus could be measured.

The second method used two or more images of exactly the same scene, but which were taken with different aperture sizes to vary the depth of field. The distance could be recovered by comparing corresponding points in the two images and determining the change in focus. This method is comparable to deriving spatial information from motion parallax, but without the image matching problem.

The algorithm he used was very simple, and inexpensive enough for 'real time' implementation with 256 x 256 images. First, the two images were convolved with a small Laplacian filter to estimate the local high-frequency content. Next the output of the Laplacian filters were summed

and normalized by dividing them with the output from convolving the images with a Gaussian filter. Finally, the “focal disparity” map was produced by dividing the estimated high-frequency content of the blurriest image by the sharper image.

Pentland’s research was one factor in leading me to the idea of using Gaussian pyramids to create a focal gradient that could in turn act like the effect of blur in the focus pull. First, the Gaussian is chosen because it approximates the type of blur necessary to convince the eye of depth, and second pyramid coding is chosen for its computational efficiency.

Pyramid Coding

The format of the image data for transitional images is a crucial consideration. The ‘real time’ nature of the transitional image representation requires that the data be easily accessible in order for it to be adapted to the observer’s interactions. It would also be useful if the data format mirrored what we know about the external visual world. Image decomposition by two-dimensional binary subdivision was a pioneering strategy for visible surface recognition.¹ This method has been subsequently customized to many other image processing problems including texture mapping, pattern recognition, image transmission and spatial data structures. The image pyramid is an efficient multi-resolution format that is thought to be analogous to the multiple scale analysis that is achieved with the human visual system.

Adelson

Ted Adelson has pointed out in his research² that in image processing the format of the data is as critical as are the techniques applied. The raw format of a digital image is simply an array of pixel intensities.

Generally, this array is converted to another format for processing. The FFT is a widely used format because operations can be applied to the transform coefficients while preserving the original pixel values. FFT’s, however, are unsuitable for retaining both the spatial location of pattern

1. Warnock, “A Hidden-line Algorithm for Halftone Picture Representation.”

2. Adelson, “Pyramid methods in image processing.”

elements and the localization at different scales. A multi-resolution scheme, such as the image pyramid, which decomposes the image into a set of spatial frequency bandpass component images, retains information localized in space and spatial frequency.

Adelson's research describes the Gaussian pyramid. The Gaussian pyramid is designed to support image convolution (a basic operation of image processing) in an efficient manner. The pyramid comprises a sequence of copies of the original image, wherein sample density and resolution are decreased in regular interval. Each level is low-pass filtered and subsampled, or decreased in size by a factor of two in order to create the next pyramid level. It is called a Gaussian pyramid because the low-pass filter used is a Gaussian-like weighting function.

With this method, each level of the image appears as sharp as the original. When an image at a particular level is expanded back to original size, by using the same low-pass filter with a gain modification of 50%, it appears blurry. This blurriness simulates the blur circle created by Johnson using a Gaussian filter in the frequency domain, but here converting from spatial to frequency domain and then back again is unnecessary. The pyramid is efficient to compute and is also faster to compute than the equivalent filtering done with an FFT. In addition, the information is available in a format that is convenient to use.

Demos

Gary Demos has compared a variety of schemes for hierarchical resolution pyramids with respect to its efficiency and the image fidelity that results in image decompression.¹ First, he discusses Jones' and Rabbani's subsampling pyramid. In this pyramid technique, the reduced resolution image is created by a factor of two reduction, by deriving the mean average of four pixels. Although the reduced resolution images look good in most cases, the lower resolution pictures are actually undersampled. This causes apparent aliasing that cannot be corrected in some images. In addition, reconstructing the original image requires 33%

1. Demos, "A Comparison of Hierarchical High Definition Imagery Coding Schemes."

extra data. This technique is similar to Lance Williams' MIP maps developed in 1983.¹

1. Williams, "Pyramidal Parametrics."

A variation of this subsampling pyramid technique is termed the reduced sum pyramid, which is similar to Frank Crow's Summed Area Table.² With this method, the additional overhead is reduced to 8.3%.

2. Crow, "Summed Area Tables."

A second variation is the reduced difference pyramid, wherein the reduced images are expanded by using the deltas to calculate the original four pixels from one pixel. An additional 12.5% of data is added to the original high resolution image.

Demos believes a better solution is to combine this delta picture (like the one used in the reduced difference pyramid) with a better filtration method than the mean average. He considers Gaussian filtering to be such a method. The combination of the Gaussian pyramid and the residual differences create what is known as the Laplacian pyramid. The Laplacian pyramid requires little overhead because it can be easily compressed using entropy coding.

The Gaussian and Laplacian pyramids are discussed in detail by Adelson in the research mentioned above. The only difference between the technique proposed by Adelson and that by Demos is in the image expansion of the Gaussian pyramid. Instead of reversing the reduce function, he uses pixel replication.³ This method is extremely efficient when the image is used for the purpose of encoding. It does not create an appropriately blurred image for the purposes of the this thesis.

3. Demos, personal interview on July 3, 1991.

Visual Dynamic Range

Demos

Visual dynamic range is the range of brightness between no light (black) and the brightest light we can see without damaging our eyes (white). At the lowest level of this range we see only greyscale. When a certain threshold of brightness is reached we begin to perceive color. This range

affects some of the algorithms associated with the transitional image representation including compositing and color value. It also had an effect on image scanning, which was used to convert 35mm slides to digital data for this thesis. Generally, this range is represented linearly; however, the effect is information loss in the low light areas of an image. Gary Demos proposes a logarithmic representation.¹

Dynamic range of pixel brightness on a CRT is described by digital values for the three color primaries of red, green, and blue. For “full-color” we use 8 bits for each primary which allows 256 levels of brightness. These levels, however, have not been standardized. This lack of standardization becomes apparent, for example, when the displayed output from scanning an image with one film scanner differs from the output of another film scanner.

It is well known that the eye is sensitive to dynamic range by percentages of brightness in relation to the surrounding environment.² Therefore, we must use some metric to perceive a grey value properly. Generally, brightness is established in reference to white. On a CRT white is created by assigning a value of 256 for each of the three primaries. If we use a linear representation, the difference in brightness between the range of 128 and 256 will be perceived as the same as the difference between the range of 64 and 128, or 32 and 64, and so on. Each of these ranges will be seen as twice as bright as the next range. But there is obviously more steps between 128 and 256 than there are between 32 and 64. Consequently some detail will be completely lost in the dark areas and the image will look crudely digitized.

Since the eye perceives brightness in percentages, a logarithmic representation is optimal. For example, the number of values that can be represented is doubled each time a bit is added to a given number of bits; 7 bits represents 128 values, and 8 bits represents 256 values. When each step is made in a logarithmic scale, each step will be of equal significance. Then the values in the range between 0 and 128 and the

1. Demos, “The use of Logarithmic and Density Units for Pixels.”

2. Hochberg, *Perception*, 118.

range between 128 and 256 will have equal significance.

Normal 32-bit floating point numbers are capable of representing brightness in the linear space with relatively constant precision, because the range is embodied in the exponent. The trade-off is that floating point calculation is expensive. If pixel brightness must be represented in a small number of bits per color, such as the 8 bits per color that is generally used for “full-color,” rather than the luxury of 32 bits per color, a logarithmic representation is a useful solution.

Since the CM-2 is the platform used for this thesis, optimization became a non-issue. Pixel values are converted from fixed point to floating point once for all operations, and are then converted back to fixed point prior to displaying the image on the CRT. Nevertheless the dynamic range is important to consider because the concept is applicable to determining a range of the acceptable amount of black and white that can be added to the image for color value gradient (see Chapter 5, Gradient of Color Value). A change in the color value affects the image legibility, under some circumstances to an even greater degree than blurring.

Image Compositing

Porter and Duff

Image compositing is an essential aspect of transitional images. Each layer of the multi-layered object is composited at some percentage of transparency to the previous layer to produce a single, transitional image, displayed in ‘real time’ on the monitor. Tom Porter and Tom Duff proposed an image compositing technique that uses the alpha channel.¹

First, Porter and Duff separate the RGB image they are trying to create into elements for independent rendering. In order to place one image of partial coverage over another with an arbitrary background, a mixing factor is required at each pixel to control the linear mixing of foreground and background colors. Each of these elements have an associated matte which describes this mixing factor. They treated the mixing factor, or

1. Porter, “Compositing Digital Images.”

alpha, of 0 as no image coverage and alpha of 1 as complete coverage. They associate this matte to each element by adding a fourth component to the existing RGB component information. This additional component is the alpha channel.

For efficiency they stored the *pre-multiplied* value in the color component so that for example, (.5,0,0,.5) indicated a full red object half covering a pixel. The quadruple (r,g,b, alpha) indicates that the pixel is alpha covered by the color (r/alpha, g/alpha, b/alpha). The RGB channels record the true colors where alpha is 1, and linearly darkened colors for fractional alpha's along edges, and black where alpha is 0. This maintained the anti-aliased edges of the elements.

Porter and Duff outlined thirteen distinct compositing operations including: **A over B**, **A in B**, **A held out by B**, **A atop B**, and **just A**, the opposite of these five operations, clear where both alpha values are zero, **xor**, and the **plus** operator. The plus operator is not concerned with precedence in any area covered by both elements. The components are simply added together.

In this thesis the need for the mixing factor, or alpha, information is present. Transitional images are unlike the images that Porter and Duff decompose into elements in that they are continuously changing in the percentage of transparency, blur, and color value. This means that the precision of the original RGB values must be maintained; therefore, any efficiency gained by pre-computing the color component is lost in degradation of the color. Nevertheless a separate matte "channel" is associated with each layer.

Interactive Cinema and Video

Filmmakers and video artists are also concerned with how we perceive information. Some useful analogies can be made with both the work of Davenport, Aguierre Smith and Pincever, and Viola.

Davenport, Aguierre Smith, and Pincever

The transitional image defined in this thesis has a parallel relationship to some new research in the field of motion pictures. Glorianna Davenport, Thomas Aguierre Smith and Natalio Pincever propose a new framework for managing story elements (shots and sound) and the relationships between these elements for creating multiple contexts in multimedia.¹

The shot is defined as one or more contiguous frames, that represent a continuous action in time and space. Their idea is to exploit the contiguous (temporal) nature of the shot by tracking the changes in the state of the camera and the images being recorded. Tracking the changes, instead of describing every detail of every frame, will greatly reduce the amount of data necessary to describe a shot.

The model they use is called *Stratification*² which enables them to describe information about a shot verbally in layers or *strata*. A layer can contain start and end frames and environmental information (who, what, when, where, etc.). Layers of sequences can be built from the layers that describe a shot. In this way, shot descriptions will propagate in an upward direction.

It is easy to see that entering all of these data will quickly become an overwhelming task. Since the framework described relies on a situated camera, Davenport is focusing her research on the design of the *data camera*, a camera that collects some of this descriptive information. The type of information that will be collected includes: time code (SMPTE), camera position, focal length, focus settings, and voice annotation.³

A parallel can be made between this framework and this thesis, because both time is thought to be continuous and data are segmented into a layered structure. However, their research describes a method for verbal annotation for motion picture footage as it is recorded, whereas a transitional image is the description itself. Also a transitional image does not exist as a pre-recorded medium. It is a 'real time' image that exists

1. Davenport, "Cinematic Primitives for Multimedia."

2. Aguierre Smith, "Stratification: Toward a Computer Representation of the Moving Image."

3. The data camera does not collect three-dimensional data as does the range camera developed by Michael Bove at MIT.

while it is observed.

Viola

The works of artist Bill Viola are concerned with the limitations of visual perception. In his video “Ancient of Days” (1980), he establishes the overwhelming presence of Mount Rainer for several minutes.¹ As we observe Mount Rainer, the only perceptible changes are in the cloud cover. Our attention shifts to a young boy in the foreground. It is not until the cloud cover, which has been completely obscuring Mount Rainer, begins to clear that we discover it has been blocked from our view since the boy faded into the picture. Nevertheless, the presence of Mount Rainer has been established in our memory. It is through the change in the cloud cover that we realize the mountain has not literally been in our field of view for some time.

In “The Reflecting Pool” (1981), the image is divided in half horizontally. Action happens either in the entire picture, in the top above the pool, or in the bottom where the pool is located. Viola suspends our belief by freeze-framing, in mid-air, a man jumping into the pool. Slowly, very slowly, the man fades away. While he is fading, however, his reflection is not in the pool. Instead we see the rippling effect of something being dropped into the water, in addition to other peoples reflections.

Eventually, a man emerges from the pool and fades off only to reappear again walking away through the woods.

Although Viola’s perceptual experiments are achieved completely with analog video techniques, the results are similar to those which can be achieved with gradients and digital compositing. Viola seems to share some of the concerns of this thesis. He experiments with shifting our attention, fading images from our view, and our awareness of that which is in the image (even if it is not explicitly visible). Also, Viola’s work is about duration, because the video tapes are meant to be watched for a period of time. The changes in the images are subtle and would be missed if the observer only watched for a moment.¹

1. Personal interview with Bill Viola in Long Beach, California on July 8, 1991.

6

Program Design and Implementation

This chapter describes in detail (1) the algorithms used to create gradients and image compositing, (2) the method used to implement control over the duration of transitions, (3) the use of the Connection Machine 2, and (4) the control interface on the Hewlett-Packard.

Gradient of Focus

The first step in creating the focal gradient is to construct a Gaussian pyramid¹ for each layer of the multi-layered image. This results in a cascaded sequence of reduced resolution images that decrease in bandwidth at one-octave intervals for each layer. The first level of the pyramid is constructed by low pass-filtering the original image layer with a 5-tap weighting function, that resembles a Gaussian filter, and subsampling by a factor of two. This process is repeated with each resulting pyramid level to generate the remaining levels. This process is referred to as a standard REDUCE operation.¹

For $0 < l < N$:

$$G_l(ij) = \sum_m \sum_n w(m, n) G_{l-1}(2i + m, 2j + n) \quad (1)$$

Figure 9 shows an original image and two levels of its corresponding pyramid which decrease in size by a factor of 2. Once the pyramid is created, the next step is to expand each level of the pyramid to the size of the original by reversing the reduction process.¹

$G_{l,0} = G_l$, and for $k > 0$:

$$G_{l,k}(ij) = 4 \sum_m \sum_n G_{l,k-1}\left(\frac{2i+n}{2}, \frac{2j+m}{2}\right) \quad (2)$$

This produces a sequence of images that have the visual effect of increasing exponentially in blurriness. Figure 10 shows an example of an expanded pyramid, the upper left image G_0 is the original image, and G_1 through G_3 are the expanded pyramid levels. Simply displaying them in series as they are would be jarring to the eye. The last step, in order to

1. Adelson, "Pyramid methods in image processing."



Figure 9. An original image and two levels of its corresponding pyramid, decreasing in size by a factor of two.



Figure 10. An example of an expanded pyramid. G_0 is the original, G_1 - G_3 show the expanded pyramid levels.

create a smooth effect of continuous blurring, is to add inbetween images. Inbetweens are computed by linear interpolating between pyramid levels. For efficiency we want to create as few images as possible. The minimum number of images that must be computed between each level to give a convincing continuous blur is described by the exponential function where n is the pyramid level number:

$$2^n - 1. \quad (3)$$

Gradient of Color Value

The color value is computed by first determining the percentage of either white or black that will be mixed into the original image. This percentage is derived from the total possible percentage that can be mixed, which is a range of about 15% white and 25% black, and a value from 0 to 255 that is sent by the user. For black the equations:

$$R = \left(\frac{1.0}{127.0} \right) \times V \quad (4)$$

$$P = ((0.25 \times R) \times R) + 0.75 \quad (5)$$

and for white the equations (which differ slightly since it is at the opposite end of the scale):

$$R = \left(1.0 - \frac{1.0}{127.0} \times (V - 127.0) \right) \quad (6)$$

$$P = ((0.15 \times R) \times R) + 0.85 \quad (7)$$

(where R is a ratio computed from the acceptable range, P is the percentage, and V is the input value from the user) are used to determine the percentage. Then a variation of standard linear color mixture equation is used:

$$D = (S \times P) + (V \times (1.0 - P)) \quad (8)$$

(where D is the destination image, S is the source image, P is the percentage calculated above, and V is the input value from the user).

Gradient of Transparency

Three variations of Talbot's law of color mixture¹ are used to compute the transparency. Each variation is based on the type of image. The first function:

$$D = ((S \times P) + (D \times (1.0 - P))) \quad (9)$$

(where D is the destination image, S is the source image, and P is the percentage of transparency) is used for opaque images that become transparent uniformly. An example of this is an RGB image. The second:

$$D = (S \times \frac{P}{255.0} \times C) + (D \times (1.0 - (S \times \frac{P}{255.0}))) \quad (10)$$

(where C is the color) is used for images that are partially opaque and partially transparent, such as feature layers. In this case, black is considered to be transparent. The third function is used for RGB images that have associated mattes. The transparency is factored into the alpha information:

$$D = (S \times \frac{A}{255.0} \times P) + (D \times (1.0 - (\frac{A}{255.0} \times P))) \quad (11)$$

(where A is the alpha information).

Image Compositing

The transparency and image compositing basically happen in one step. The destination image is also the running total for the final image. Each layer is treated as a separate image. While the list of layers is traversed it is determined whether or not it is actually visible (images that are 100% transparent are not visible). If it is, then the transparency is computed and the image is added to the existing state of the final image. For RGB images without mattes and greyscale images the entire image is added. For RGB images with mattes only the matte information was added. The black areas, which are considered to be completely transparent under all circumstances was masked by setting the context flag of each corresponding processor. In other words the processors that contained no matte information were simply turned off, which is a distinct advantage

1. Weitzman, "Digital Transparency Applied to Interactive Mapping."

to using the CM-2 (see the section describing the Connection Machine below). Transitional images use only the “plus” operator.¹ The observer may determine the order in which each layer is composited, since different orders make sense under different circumstances. For example, our geographical map also has associated weather layers. If the multi-layered image is organized such that the weather is on top, we can not see the underlying information. The observer may want to see the airports in relation to the weather. It is useful to be able to reorganize the layers with the airport layers on top of the weather.

1. Porter, “Digital Image Compositing.”

Transitions

In order to study transitions I have implemented the basic functionality of an animation scripting system. The observer is able to mark start and end points for each translation over some duration of abstract time. The observer can also describe the rhythm of movement (i.e.: linear, slowin, slowout, slowin/out, etc.). When the user sends a “play” message, the procedure will loop and calculate the percentage of blur, transparency, and color value based on the combination of these start and end points and the rhythm.

The Connection Machine

Gradients of focus and transparency were first implemented as a three level Gaussian pyramid on the Hewlett-Packard (HP) 9000 series 835 workstation. The pyramid method and interpolating between pyramid levels proved to be both visually effective and computationally efficient. Pyramid levels, however, could not be computed in ‘real time’. It turned out that even by precomputing the blurred images and loading them at run-time, ‘real time’ interpolating, calculating of transparency and image compositing, was simply not realizable on the HP.

The Connection Machine-2 (CM-2) was chosen as an alternate platform, because the parallel nature of the architecture lends itself to bit-mapped image manipulation. The CM-2 is a massively parallel SIMD¹ machine,

1. Single Instruction Multiple Data

which supports memory and time sharing of the physical processors designed to simulate virtual processors. In this thesis, each pixel is assigned to one virtual processor. This means that each pixel is treated as though it has its own serial computer. This thesis used 16K processors.

Initially, the precomputed blurred images (the original and its corresponding three blurred images) were loaded into the frame-buffer memory of the CM-2 and all image computation was done in floating point with 'real time' results. Eventually, the pyramid encoding was reimplemented directly on the CM-2 as well, because the precomputed images required a huge quantity of disk space, and loading the images was slow.

To implement the pyramid technique described above on the CM-2, separate virtual processor sets (Vp-sets) must be created for each level with geometries that decrease in size by a factor of two. This means that to compute each level, both neighbor communication within Vp-sets and communication across Vp-sets is required. While neighbor communication is extremely efficient, sending messages across Vp-sets is much less efficient. As a matter of course, I opted to forego the compactness of convolving smaller and smaller images and use just a single-sized Vp-set to represent all levels.

The pyramid technique is revised to take advantage of organizing the Vp-set as a logical grid, where each processor can be thought of as corresponding to each pixel on the CRT. The revision also takes advantage of nearest-neighbor communication in regular patterns based on powers-of-two. This is extremely efficient because each processor can access the memory of another processor without computing the other's address.² Only the grid axis and the pattern of selected processors must be specified. Subsampling is then implemented by selecting processors by a factor of two and writing black to the inverse of the selected processors. Patterns for the selected processors for each pyramid level are shown in Figure 11. These patterns are used to reduce the image. G_0

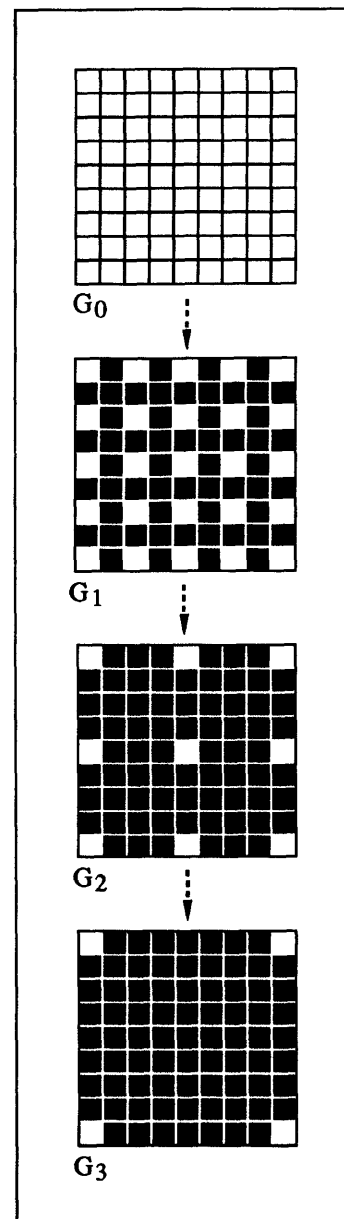


Figure 11. These patterns represent the selected processors at each level of the pyramid. G_0 is used to convolve the first level. G_1 is used both to subsample the first level and convolve the second level. G_2 is the subsampling pattern for level two and the convolution pattern for level three. G_3 is used to subsample the third level. Expanding uses these patterns in reverse order.

is the pattern used to convolve the first level of the pyramid. G_1 represents both the pattern used to “subsample” the first level and the pattern used to convolve the second level. G_2 is the subsampling pattern for the second level and the pattern for convolving the third level. G_3 is the pattern used for subsampling the third level. Expanding uses these patterns in reverse order.

The architecture of the CM-2 also lends itself to creating the inbetween images, computing the transparency and color value, and compositing each separate layer into one final image in ‘real time’. To create the inbetween images the average between two successively blurred levels is computed on a pixel-by-pixel basis. On the CM-2 one set of machine-level instructions can be sent to each processor in parallel. Similarly, another set of instructions is sent to calculate the transparency, and another for the color value. Compositing, which is essentially summing the image layers together requires only one instruction. Another advantage of the CM-2, with the use of two sequencers (16K processors), is that the multi-layered image system can contain over one hundred 512×512 resolution images. The CM-2 allowed me to visualize my hypothesis of using gradients of focus and transparency to direct an observer’s attention to specific information of interest with ‘real time’ feedback.

Interface on Hewlett-Packard

The controls for transitional images are implemented in C on the HP 9000 series 835 workstation. This workstation has a RISC-based processor and two 24-bit high resolution framebuffers. The controls were built on top of “Bad Windows,” a window system developed primarily by Bob Sabiston in the Visible Language Workshop (VLW), that takes advantage of the graphics capabilities of the HP.

The interface consists of three main parts: (1) the library, (2) the multi-layered structure, and (3) the transitional editor. The library contains all

of the layers at extremely reduced resolution (Figure 12). These thumbnail sized images are not always legible at such a small size, so a title accompanies each image. These library images can be selected by clicking on them with the mouse or pen stylus. Once an image is selected it is in a sense attached to the mouse, so the user can then assign it to the multi-layered structure.

The multi-layered structure is designed to be seen as a vertical slice of all the layers, with the bottom most layer generally being the background. All of the layers cannot fit onto the screen at once, so the user is encouraged to organize the layers in chunks (Figure13). These chunks help the observer manage the information. The user has the ability to organize and label these groups anyway desired. Once the layers are assigned to the structure, the observer can begin to manipulate the various settings in the transitional editor.

In the transitional editor (Figure 14), the user can set focus, transparency and color value by either moving the virtual slider with the mouse, typing into the register window, or selecting the layer and using the knobs. Although typing is the most accurate input, the user may not always know the desired percentage of the gradient. The knobs are more accurate than the slider because the slider is limited by its size. The user may also set markers on a “time-line” to indicate the start and end points of each gradient. These markers are like a slider with two markers. They can be accessed through the mouse, knobs or by typing into the register window.

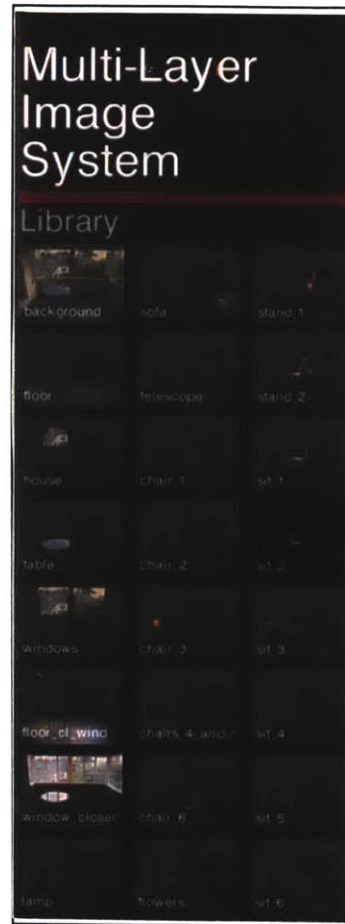


Figure 12. Detail of Library.

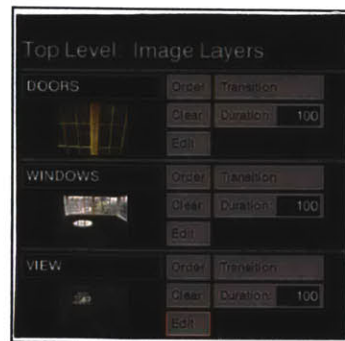


Figure 13. Detail of Multi-Image Structure.

Transitional Image Editor: BACKGROUND									
Object	Order	Start	Time	Phase	Color	Value	Value	Value	Value
stand	Order	Start							
	Clear	Move X	10						
	Close	Move Y	100	0.57	1.00	0			
sofa	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	0.59	0.00	190			
face of wind	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	1.00	0.00	127			
doors	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	1.00	0.80	127			
window closed	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	1.00	1.00	127			
window closed	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	0.00	0.00	0			
floor	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	1.00	0.70	250			
background	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	0.55	1.00	0			
background	Order	Start							
	Clear	Move X	0						
	Close	Move Y	0	0.00	0.30	127			

Figure 14. Transitional Image Editor.

7

Examples of Transitional Images

In this chapter three examples are described that have used this technology. They are a geographical map, an image of a three-dimensional space and an image of multiple perspectives. Each example differs in the type of information it contains, which is useful for understanding the effects of gradients and time.

A Geographical Map

A geographical map is chosen as one subject area to explore the influence of gradients and time for three main reasons: (1) Maps can easily be broken down into features, which lend themselves to the multi-layer image structure. (2) They are abstract descriptions of real information. (3) The data are readily at hand.

The map is constructed using a LANDSAT image as the background. Roads and waterways are extracted from the image and additional features such as metropolitan areas and airports are included to simulate a realistic map. In addition, nine layers of weather data are superimposed over the image because one may want the capability to combine onto one map information that is normally available only on separate maps (e.g. road maps and weather maps).

The image is organized as individual layers of information in order to control focus, transparency and color value on the selected layers within the image. In this example, there are a total of twenty-nine layers. The resolution of all feature layers is 512 x 512 pixels by 8 bits. They exist as greyscale images with color added to them when they are composited. The LANDSAT image is a 512 x 512 color image, twenty-four bits deep. Figure 4 shows the initial order of the layers for the map. The LANDSAT image, which is always the bottom-most layer, serves as the background. The next nineteen layers are feature data (roadways, metropolitan areas, labels, etc., see Figures 15 and 16). Weather data comprise the top nine



Figure 15. Roads Layer.

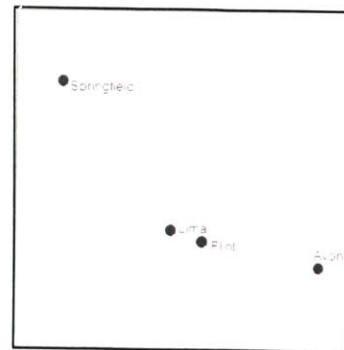


Figure 16. Towns Layer.

layers in the form of radar reflectivity data. Each weather layer represents a horizontal cut through the atmosphere, termed Constant Altitude Planned Position Indicator (CAPPI). The cuts range from 200 meters above ground to 8200 meters (Figure 17).

From these layers the user may chose information. Perhaps the observer is interested in available transportation. The waterways, the roads or the airports may be chosen. For example, first the user selects the waterways by setting its transparency to “opaque,” its focus to “full focus,” and its color value to “bright” (or the most white that can be added). Next the user sets start and end points for each gradient denoting the amount of time (in abstract integer units) and the order in which each gradient will reach the settings that have been described. In this way, the user can watch the waterways fade into view, come into focus, and brighten in relation to the surrounding information. Through observation the observer can begin to design the transformations that accentuate the waterway information and at the same time maintain context with related transportation information (Figure 18).

To continue this example, the user must determine the percentage transparency, blur and color value for related features (i.e. roads and airports). This is done the same way, by manipulating the sliders or knobs to determine the appropriate settings for each gradient. The observer may decide that the related information should fade away before attention to the waterways is heightened.

Exploration with setting the gradients is necessary to discover formal relationships between types of data. For instance, text blurred by 100% over a high frequency image cannot even be seen. Over a solid color background, which differs in color and intensity, it is somewhat more discernible. This example is very useful because the information is abstract. It has one major drawback, however, which is, since most of the information is symbolic the features are small in scale. This makes it more difficult to see the effects of blur on anything but the LANDSAT



Figure 17. An example of weather data at 200 and 1200 meters above the map.

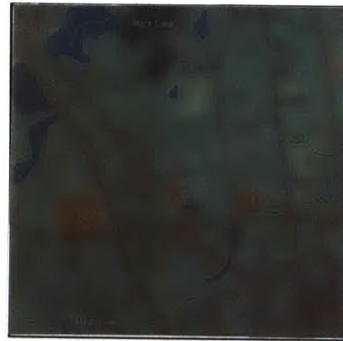


Figure 18. An example of the waterways heightened and other related transportation.

map used as the background. Also, there is very little overlap of the different features, this makes it difficult to see the effects of partial transparency.

Experiments with time enable the user to come to know how many changes can be made to the image at the same time and still be noticed. If the changes are too fast or too slow the observer will lose context.

An Image of a Three-dimensional Space

A second example is an image of a three-dimensional space. This example is chosen because it is literal rather than abstract and because most of the objects are larger, although they still do not overlap. With this example the observer can continue to experiment with gradients and time, but in another context.

The three-dimensional space is an image of a room (Figure 19), the objects contained therein (Figure 20), and a human actor (Figure 21) in various positions within the room. These data came from 35mm slides which were taken with the system in mind. Each slide was shot from a single position with a camera mounted on a tripod. The room was emptied and then one by one objects were placed back inside the room as each picture was taken.

Once the data were scanned and segmented they could be organized as a multi-layered object. In this example there are twelve, twenty-four bit images used in conjunction with twenty-six mattes to make a total of twenty-seven layers. The background, which is the room empty of objects, exists as the only layer with out an accompanying matte. The objects in the room include: a sofa, a lamp, a telescope, a bunch of hanging peppers, the table, six chairs, flowers on the table and a chess board. These layers are made from two slides, or two bit-mapped images with six matte channels each. The actor is posed in each chair and stands looking at the table and also out the window. A final layer is a set of french doors that close off the room.



Figure 19. An image of a room and a table. The room serves as the background image.



Figure 20. This example shows the windows, the table and the floor without the room itself.



Figure 21. This image show a scene in the room, with various objects and one actor.

With this example the user may select objects of interest, in the same way as in the map example, and examine the relationships between them.

Experimentation with this image exemplifies the difference between film transitions and transitional images. Although the *a priori* existence of the room is captured with twelve slides, we do not “cut” from one still slide to another to look at what is contained in the room. Instead the image seems to evolve as the observer selects and reveals the various objects.

By introducing the actor the desire for Cartesian three-dimensional data to describe both the room and the actor becomes evident. If only the actor could move within the space!

An Image Containing Multiple Perspectives

The third example describes a time and place, an afternoon in January 1989 at Hermosa Beach, California. It was chosen to explore the limits of the representation by combining images that were not precomposed with either a multi-layered structure or the gradients of blur, transparency and color value in mind. There are multiple backgrounds; the previous two examples had only one background. And the features segmented from the images overlap.

Here again the images are taken from 35mm slides. This time, however, the slides were not taken with *a priori* knowledge of the representation. There are three slides that constitute the background, a picture of tire tracks and foot prints in the sand, the Hermosa Beach pier, and the skyline of the city of Hermosa Beach as seen from the pier. The features comprise, text describing the date and location, and three head shots of the same person on the pier (Figures 22-24).

One the one hand, the concept of multiple perspectives is similar to the discontinuous articulations between spaces discussed in Chapter 4, in the section entitled “Transitions in Motion Pictures.” The backgrounds resemble “radical spatial discontinuity” in that they are radically different images, although they describe the same place. The pictures of



Figure 22. Two views superimposed of Hermosa Beach, California.

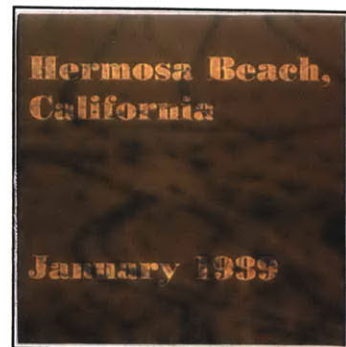


Figure 23. Blurred background of sand with a reversed matte of the text brightened and in focus.

the person are like shots of close “proximity,” because the camera angle is only slightly different, although the object of interest is the same. With this example the observer can experiment with changing the point of view by varying the background and moving between different perspective shots of the person. On the other hand, once again, we do not “cut” from one slide to another. The image is one multi-layered object comprising the various perspectives. In order to move from one perspective to another we adjust the gradient values. There are no frames of film or video to cut.



Figure 24. Two perspective views of the same person. The background is blurred and darkened, one head shot is transparent and blurred, the top is in focus, fully opaque and brightened.

8

Conclusions

In this thesis I have described the representation of, and user interactions with, the transitional image which enables an observer to adapt an image to the amount of information that can be perceived at one time. This is accomplished both by experimenting with gradients, which produce spatial information, as a way to increase the amount of information apprehended; and by then further exploring the limitations of what can actually be perceived, by describing a number of gradient changes at the same time. The results of this research and some future directions follow.

Results

The three examples have produced various results. The majority of the results come from the map example because the data set was available for over a year, where as the other data sets have only been available for a short time.

One of the first observations was that background image had a greater range of blur than text and iconic symbols over the background. Even if text and symbols were viewed over a solid color background of a different color and intensity, the range was less than that of image data. Two additional observations were found in the differences between single color layers (greyscale with color added) and RGB layers. The color value range of single color images appeared to be linear, while RGB images seemed to have a non-linear dynamic range. Also, the range of blur was about 15% greater in RGB images than greyscale images.

These observations led toward formalizing the amount of blur, transparency and color value that is useful in certain situations. For example, I was able to visually derive generalizations from the geographical map example. The LANDSAT map is the only RGB image and it is also the background image. If we assume that all features of interest will be in relation to the background (unless the observer wants to isolate

a layer completely), the background should never become transparent. With that assumption out of the way, I could generalize the percentages of blur and color value for RGB images, greyscale images, text and symbols. In addition I could generalize the gradient percentage of transparency for greyscale, text and symbols. When an observer is highly interested in any of the various types of data, the selected layer is completely opaque, fully in focus, and at the brightest end of the dynamic range. If the layer is only somewhat important, an RGB image will change only in color value to its normal color. A greyscale image will become 30% transparent, but will remain in full focus, and be in the middle of the range between the normal color value and the darkest end of the value range. Text and symbols, on the other hand, will remain opaque, but will become blurry by 44%. They will also have the same color value as greyscale images at this point. When the selected layer is only visible to maintain context, the RGB layer will be 100% blurry, and about 2/3 toward the darkest state. A grey scale image will still be 30% transparent, but it will be 83% blurry and at the darkest point in the color value range. Text and symbols will be 13% transparent, 66% blurry, and also at the darkest value of the color range.

As a result of these generalizations, there was much discussion about using the darker end of the color value range to diminish information. One observation was that adding black to the image took away too much information and that the observer could lose context. Still others thought that adding black did indeed heighten the information of interest and that blocking out some of the other detail was useful at times, especially at low light levels.

At first, each change to the image was made by selecting one gradient slider at a time. Once the scripting system was written we could begin to study the limits of the perceptual system by describing values for more than one gradient at a time and reviewing the changes in 'real time'. The user could set the begin and end points of any change in gradient value,

in any layer, relative to any other gradient change or any other layer. Although there was no strict testing, it seemed that an observer could notice between one and two changes at a time. If the changes were offset in relative time, even by a small amount an observer could notice between three and five changes. If the observer was actually making the selections the amount that could be noticed was even greater. But this of course takes into account the observer's expectations of the choices that were made. Naturally, an uninvolved observer would be less aware of these selections.

Another observation was made with the examples that rely heavily on the use of mattes. Blur and transparency must be used together, otherwise the edge of the mattes made the objects look unnatural. This may be a result of the crude method (by hand) in which the mattes were made. Further investigation in image segmentation may eliminate this problem altogether.

Future Work

The thesis is a beginning in the use of transitional images to understand more about our perceptual system. It has great potential for future research. There are six areas that instantly come to mind; they are: (1) A formal system can be developed that describes the relationships between types of data by experimenting with the representation described herein. (2) Research in image segmentation will be of tremendous benefit, and is a necessary extension to this system. (3) This thesis can be applied to the problems of information compression and progressive transmission in multimedia. (4) Algorithms can be developed that improve the efficiency including, non-linear color transformations and alternative blurring techniques. This research could even be designed to run on a serial computer, although the image processing algorithms used in this research lend themselves to parallel architectures. With hardware prices dropping, it is likely that massively parallel computers will become more mainstream. It is important that research remain in the avant garde so that

there are examples of how new hardware technologies can be used.

(5) The data can be extended to include three-dimensional coordinate data, for example, from range cameras. (6) This research can be expanded to work with video technology so that “cutting” is no longer the only solution for transitions.

Bibliography

Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., Ogden, J.M., "Pyramid methods in image processing", *RCA Engineer*, vol. 26-6, pp. 33-41, Nov/Dec 1984.

Aguierre Smith, T., "Stratification: Toward a Computer Representation of the Moving Image", a working paper, MIT/Media Lab Memo, June 1991

Bergson, H., *Matter and Memory*, translated by Paul, N.M. and Palmer, W.S., Zone Books, New York, 1991, (1911), *Matiere at Memoire*, 1896.

Bergson, H., *Creative Evolution*, translated by Mitchell, A., New York: Henry Holt & Co., 1911 (New York: Macmillian & Co., 1944), *L'Evolution Creatrice*, 1907.

Burch, N., *Theory of Film Practice*, translated by Lane, H. R., Praeger Publishers, New York, 1973.

Colby, G. and Scholl, L., "Transparency and Blur as Selective Cues for Complex Visual Information", SPIE Proceedings, March 1991.

Crow, F., "Summed Area Tables", *Computer Graphics*, vol.19 no.3, July 1985.

Davenport, G., Aguiere Smith, T., and Pinciver, N., "Cinematic Primitives for Multimedia", *IEEE Computer Graphics and Applications*, pp 67-74 July 1991.

Deleuze, G., *Cinema1: the Movement-Image*, translated by Tomlinson, H. and Habberjam, B., University of Minnesota Press, Minneapolis, 1986.

Deleuze, G., *Bergsonism*, translated by Tomlinson, H. and Habberjam, B., Zone Books, New York, 1988.

Deleuze, G., *Cinema2: the Time-Image*, translated by Tomlinson, H. and Habberjam, B., University of Minnesota Press, Minneapolis, 1989.

Demos, G., "The Use of Logarithmic and Density Units for Pixels", *SMPTE Journal*, Preprint No. 132-85, Oct 1990.

Demos, G., "A Comparison of Hierarchical High Definition Imagery Coding Schemes", unpublished paper, April 1991.

Eisenstein, S., *Film Form, Essays in Film Theory*, edited and translated by Leyda, J., Harcourt, Brace & Co., New York, 1949.

Einstein, A. and Infeld, L., *The Evolution of Physics*, Simon and Schuster, New York, 1938.

Fry, E.F., *Cubism*, Oxford University Press New York 1978, (1966).

Gibson, J.J., *The Perception of the Visual World*, Houghton Mifflin Co., Boston, 1950.

Glover, J.G., *The Cubist Theatre*, UMI Research Press, Ann Arbor, 1983.

Goethe, J.W., *Truth and fiction relating to my life*, translated by Oxenford, J., C.C. Brainard Publishing Company, Boston, 1982.

Gregory, R.L., *Eye and Brain, the Psychology of Seeing*, McGraw-Hill Book Company, New York, 1966.

Hochberg, J.E., *Perception*, Prentice-Hall Inc. Englewood Cliffs, New Jersey, 1964.

Johnson, M.B., "A Massively Parallel Simulation of Depth of Field for Cel Animation", *Internal MIT/Media Lab Memo*, Jan 1990.

Jones, P.W., and Rabbani, M., "Digital Image Compression Techniques", SPIE Optical Engineering Press, 1991.

Lawder, S.D., *The Cubist Cinema*, New York University Press, New York, 1975.

Marr, D., *Vision*, W.H. Freeman and Company, New York, 1982.

Monaco, J., *How to Read a Film*, Oxford University Press, New York, 1981.

Reed, E.S., *James J. Gibson and the Psychology of Perception*, Yale University Press, New Haven, 1988.

Pentland, A.P., "A New Sense for Depth of Field", *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. PAMI-9. no. 4, July 1987.

Pilkington, A.E., *Bergson and his Influences*, Cambridge University Press, Cambridge, 1976.

Porter, T., and Duff, T., "Compositing Digital Images", vol. 18 no. 32, July 1984.

Sturman, D. J., "Motion Picture Cameras and Computer Graphics", *Internal MIT Media Lab Memo*, June 1989.

Thinking Machines Corporation, *Introduction to Programming in C/Paris*, Version5, Thinking Machines Corporation, Cambridge, MA 1989.

Warnock, J.E., "A Hidden-line Algorithm for Halftone Picture Representation", Department of Computer Science, University of Utah TR 4-15, 1969.

Weitzman, L., "Digital Transparency Applied to Interactive Mapping", *Master's Thesis*, MIT Department of Architecture, Cambridge, MA, May 1978.

Williams, L., "Pyramidal Parametrics," *Computer Graphics*, vol.17 no. 3, pp. 1-11, July 83.