THE TEMPORAL IMAGE MOSAIC AND ITS ARTISTIC APPLICATIONS IN FILMMAKING

by

CRAIG B. KNOWLES

A thesis submitted to the
School of Computing
in conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
December 2003

Abstract

The mosaic is relatively old form of art. The computer as an artistic device is relatively new. The recent combination of these two disciplines has furthered the evolution of the mosaic as an art form. The main goal of this thesis is to continue this evolution by extending the image mosaic into the third dimension, that of time, by melding the image mosaic with the discipline of the filmmaking. Specifically, this thesis provides a detailed background of the mosaic art form, then extends the image mosaic by adding the dimension of time. This results in a new mosaic form, the Temporal Image Mosaic(TIM). Respecting the artistic nature of this thesis, the use of the Temporal Image Mosaic in an artistic setting is demonstrated in the form of a short motion picture.

Acknowledgments

I would like to take this opportunity to thank my advisors. Thank you, James, for your grateful help over the coarse of this thesis, and especially for letting me take a break when I needed it. Thank you, Clarke, for helping to make this interdisciplinary thesis a reality, and for your help on all film aspects of the thesis.

Thanks also to my thesis committee for their help, time and valuable suggestions — from both the technical and artistic perspectives. The committee consisted of Malcolm Stott, James Stewart, Clarke MacKey, Tom Dean, David Rappaport and Francis Leeming.

Many thanks to my family — Laurie and Dora Mae Knowles, sisters Cindy and Crystal — and friends Melanie Kintner, Darren DeGrace, Andre LeBlanc and Doug Jagoe. Your love and support got me through the rough parts. I would not have been able to do it without any of you. Thanks also for keeping difficult questions like, "Why aren't you finished yet?" to a minimum.

Finally, I would like to dedicate this thesis to the memory of Blake Jagoe, who knew a long time before I did that making movies was something I should be doing.

Contents

Appendices

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Context

A large part of the study of computer science deals with the refinement of the computer itself — making algorithms run faster and more efficiently while minimizing the amount of resources used. The evolution of the computer to this point affords us a tool sufficiently powerful to have an impact on many diverse fields outside the realm of computer science.

The range of applications for the computer seems almost limitless, and new applications are found for the computer every day. Covering the broadest of spectrums, the computer is finding uses in disciplines from chess, cosmology, and agriculture to medical imaging, music and space travel. Another discipline which has found the computer a welcome tool, is that of Art.

The Art world is one of change and exploration, searching for new methods of expression, and new mediums in which to create. For the artist, the computer is yet another tool — another *brush* — with which to explore human expression. Artists spanning a wide variety of disciplines have embraced the computer. Writers, painters, illustrators, graphic designers, animators, sketch artists, and filmmakers now have the computer at their disposal as a creative tool.

The computer as an expressive device has brought forth new and exciting creative possibilities. The developments of computer graphics, once solely an academic endeavor, have been embraced by artists to create new means and methods of expression. Animated commercial movies, computer games and digital media are some of the many results of the computer as an artistic tool. Digital artists have a wide variety of high-end software products from which to choose, tailored to perform many standardized tasks.

The problem occurs when the artist desires to create something which is outside the scope of the software. A truly empowered artist requires the knowledge necessary to harness the underlying power of the computer outside the context of the commercial products. Today, such artists exist.

One such artist, benefiting from the power of the computer as an artistic tool, as facilitated by his underlying knowledge of the machine itself, is Robert Silvers. Silvers invented the *Photomosaic*^{TM} at MIT in 1997 as part of his Master's Thesis work in Computer Science. His art works (which are, curiously, *patented*) hang in art galleries and have been embraced by popular culture. Posters, magazine covers, puzzles and books are some of the many venues that display his creations. Silvers created a new means of expression using the power of the computer. He applied the brush to a new canvas, and created a new area of art in the process.

1.2 Motivation

One specific art form taking advantage of the computer is that of filmmaking. The discipline as a whole is embracing new digital formats, editing tools, film processes and work flows that are revolutionizing the craft in many areas. Along with many of the digital advents, comes increased flexibility and malleability of the filmed image, lending itself readily to image processing and manipulation techniques.

This flexibility is reflected in both the technical and artistic sides of the filmmaking process. However, since the end result of the overall process is artistic in nature, technical advances in image processing and manipulation, in this case, must be tempered by the goals and motivations of the artist. The technology must be used to further artistic expression — thought of as an addition to the grammar of the art, and not simply as a technical exercise.

This thesis will discuss the *Temporal Image Mosaic* (TIM) from the standpoints of both technical construction and artistic expression. The extensive technical considerations involved in the creation and refinement of the TIM will be detailed, and artistic considerations regarding the TIM as it relates to film grammar will be discussed.

Technically, the TIM is an extension of the static two-dimensional image mosaic, with time being the added dimension. Many of the problems and issues that arise during the creation of a two-dimensional image mosaic are also encountered with the TIM. Mosaic and image resolution, colour-space consideration, image-comparison technique, candidate pool size and distribution, and image formats are major factors which must be taken into account in order to create a coherent, visually-pleasing image mosaic in two dimensions.

The addition of the temporal dimension introduces a host of new issues and greatly increases the complexity of the task. The sheer size and color distribution of the candidate pool, and the increased matching complexity required to match the images in three dimensions are major considerations. The image matching technique itself, which is conceptually similar to the two-dimensional method, must be modified and extended to deal with matching in the additional dimension. Also as a result of the added dimension, temporal characteristics of the candidate clips themselves must be considered. Temporal continuity, overall colour distribution, and clip length all play important roles in producing a visually-pleasing resultant TIM.

As a primary goal of this thesis is to incorporate the effect into an artistic production, several constraints from the artistic aspect are additionally imposed on the technical creation of the TIM. One major artistic constraint is the desire to keep image colour constant throughout the process such that the colour palette of the resultant mosaic is as close to the original images as possible. Another major artistic constraint is maximizing the temporal length of the clips in the tiles of the mosaic to reduce the amount of *flicker*, while at the same time minimizing the colour difference between the base and candidate sequences.

Artistically, the TIM is previously unseen in the realm of film. Incorporating the effect into a film production in a coherent and expressive manner which abides by the rules of film grammar is a challenge. Although the TIM is new to film, it can be studied using related film concepts and forms which have been used in motion picture settings. As such, these related forms can be used to comment about the grammatical implications of the TIM, suggest possible uses for the effect, and possibly to help impose more artistic constraints on the technical construction of the effect.

In summary, this thesis is quite different in nature and approach from that of a standard computer science thesis. Although the end result of the thesis is a technical tool, it is produced from an artistic perspective, designed to be used in an artistic setting, and makes several contributions to the area of filmmaking itself. The thesis allows artistic constraints imposed by the needs of the film production itself to shape the technical aspect of the thesis. In the final analysis, it is felt that this thesis is more than simply the creation of an image manipulation tool. It is approached first and foremost from the perspective of a filmmaker whose primary concern is to achieve a defined artistic goal, and utilizes the technical tools to achieve that goal.

1.3 Contributions

As a result of the interdisciplinary aspect, there are several contributions this thesis makes to the research areas of Computer Science and Film. The contributions are in the form of both original ideas, and extensions of previous ideas in the areas.

A large contribution of this thesis is the research into the history of the mosaic arts. This thesis compiles an in-depth review of the mosaic as an art form, tracing its evolution from the beginnings up to the computing age. The influence of the mosaic on, and by, the computer are detailed, and the various related art movements which were influences on the mosaic, and those which were influenced by the mosaic.

By developing several key algorithms and a set of tools which enable the production of a temporal image mosaic, this thesis extended the ideas contained in a previous master's thesis in the field of Computer Science — that of Robert Silvers and the Photomosaic™.

By applying the image processing capabilities of the computer to the art of filmmaking, this thesis creates a new and novel means of expression — the temporal image mosaic.

By creating a short film which incorporates the temporal image mosaic — under co-advisor, Clarke MacKey, from Queen's Department of Film — this thesis demonstrated one possible use of the temporal image mosaic in an artistic setting.

By extending the previously two-dimensional mosaic art form into the temporal dimension, this thesis extends the mosaic arts themselves.

1.4 The Thesis

The structure of the thesis is as follows:

Chapter 2, the Background, will provide the reader a basis for the topics covered in the thesis. It will detail the history of the mosaic art form, the history of the tiled image as used in film, and give a general introduction to the image mosaic itself. The various techniques and issues involved in creating image mosaics will be discussed, as well as papers in this area of research. Peripheral work in the area will also be covered.

Chapter 3, the Technical Background, will provide the reader the technical detail necessary for further exploration into image mosaic construction. Details such as the image library structure, image formats, colour spaces considerations, and other concerns will be discussed.

Chapter 4, the Temporal Image Mosaic, will document the steps and processes used in creating the temporal image mosaic. The various methods and approaches employed and the reasoning behind each of the choices will be detailed. This chapter will also discuss the short film, and the use of the TIM.

Chapter 5, the Mosaic Results, will discuss the results of the technique as described in Chapter 4.

Chapter 6, the Conclusions and Recommendations, will discuss the overall results and lay out recommendations for future work.

Chapter 2

Background

2.1 Introduction to the Mosaic

The *temporal* image mosaic, as discussed in this thesis, is a combination of a number of ideas from a number of backgrounds. Decorative mosaics, as developed by the Greeks, were the genesis for all that followed in the world of mosaic art. Several artistic movements concerning the representation of a conceptual whole from a variety of related pieces — classical renaissance painting, *Impressionism* and *Pointillism* — also play a large part in the chain of artistic advancement that lead to the eventual creation of Silvers' Photomosaic$^{\text{TM}}$. These factors, in conjunction with the several key ideas from film — the multiple on-screen image, ideas developed concerning *mise-en-scène*, and the theory of *montage* — all lead to the creation of the temporal image mosaic.

Each of the contributing factors which culminate in this thesis will be discussed in the following section. The background will then take a look at the processes involved in the creation of the Photomosaic$^{\text{TM}}$, enhancements to Silvers' original technique, alternative mosaic generation techniques, and considerations involving colour spaces. Peripheral work in this area of research and art will also be discussed.

2.1.1 History of the Mosaic

The mosaic as an art form is defined in the Merriam-Webster dictionary as "a surface decoration made by inlaying small pieces of variously coloured material to form pictures or patterns." The mosaic dates to the 4th century B.C., and is generally associated with the Greeks. In fact, the word *mosaic* is of Greek origin, meaning "patient work of art, worthy of the muses" [47].

Figure 2.1: Seahorse Mosaic - Rome, 1st Century A.D.

The Greeks, and later the Romans, embraced the mosaic in many areas of architecture as a decorative element. During the first centuries A.D., mosaics were popular among the Romans, and could be found in many parts of Roman society — set in floors and walls of houses, temples, baths, columns and entrances to many residences and buildings [47]. The complexity of these mosaics ranged from random patterns, to simple decorative floor pieces, like the Seahorse shown in Figure 2.1, to the extremely elaborate wall mounted art pieces, like the Chariot Racers shown in Figure 2.2. The themes of the mosaic renderings during this era are varied, including mythology, sea and earth imagery, hunting scenes, and depictions of past and present emperors. With the fall of the Roman Empire and the rise of Christianity that marked the beginning of the *Byzantine era*, Christian imagery became the dominant theme [40].

Figure 2.2: Victorious Charioteer and Team - Rome, 3rd Century A.D.

During the Byzantine era, the mosaic reached its highest level of quality [40]. These high standards of the art form continued for several centuries increasing in artistic grandeur and decadence, where the mosaic attempted to imitate paintings and, as such, became thought of as a secondary art form [47]. The art form was reinvigorated in the modern age with the excavations of *Pompeii* beginning in the

18th century [47], and countless mosaic works from history remain preserved to this day.

2.1.2 Conceptual Contributions to the Image Mosaic

The image mosaic is directly related to the mosaic through the use of the grid-based compositional structure, but there are other conceptual factors which contribute to the visual impact of the image mosaic. Two additional visual components of that effect the impact of the image mosaic will also be discussed — the composition of a conceptual whole from a variety of pieces, and the perceptual blending of colours by the eye when viewed from a distance.

Image Composition

The idea of composition of a larger image from sub-components is not a new one. Giuseppe Arcimboldo, the renowned Italian 16th century painter, produced a series of portrait paintings in which theme-related sub-objects were used to form the larger compositions [42, 41]. Arcimboldo's most famous work, *Vertumnus*, is shown in Figure 2.3. It is a portrait of King Rudolf II, portraying the Roman God of Metamorphoses in nature and life. Rudolf's face is composed of fruit and flowers — his lips are cherries, his nose is a pear, his hair is made of flowers and wheat. All meant to symbolize the balance between nature and harmony that his reign allegedly represented [42]. Arcimboldo produced many of these composite portraits, including two famous four-work series — *the Seasons*(1573) and *the Elements*(1566-1570) — all exhibiting portraits composed from thematic sub-components.

There are two important things to recognize about these works. First, the symbolic

Figure 2.3: Vertumnus - Arcimboldo, 1590

representation of the sub-components, in the way that they relate symbolically to the portraits they are used to create, is a commonality with the image mosaic. Secondly, the method in which the objects themselves are composited to create the overall portrait image produces a visual effect shared with the image mosaic. The sub-components are skillfully placed in such a way that a brief glance at the portrait does not immediately reveal the complexities of the compositional elements.

Perceptual Blending

The perceptual blending of the colour as witnessed in an image mosaic can be traced to the Impressionist art movement of the late 19th century. The Impressionist and post-Impressionist movements, considered the beginning of modern art [28, p. 415], began as a reaction to the formal style of painting supported by the rigid art establishments in France at the time [29, p. vii]. It is categorized by the use of swift brush strokes, swashes of colour and dominant primary colours used to simulate actual reflected light. The artists were concerned with the emotional representation, interpretation and thus *impression* of reality as opposed to strict photographic reproduction [36].

The blending of colour and freedom to explore outside the boundaries of strict photographic representation allowed Impressionists such as Claude Monet, Edouard Manet and the Post-impressionist Pointillist, Georges Seurat, to produce many distinct works of art which played with the use of colour as an approximation of the visual reality of their subjects. The Impressionist works convey the underlying feeling of the subjects, while at the same time, the works put the onus on the viewer's eye to blend the colours and suggested shapes in the scenes.

Seurat's Pointillist masterpiece, *Sunday Afternoon on the Isle of Grande Jatte*(1884-1886), shown in Figure 2.4, is composed of a complex myriad of colour, applied in hundreds of thousands of tiny points of individual colours. Seurat was an avid reader of scientific color theory, and thought that the perceptual blending as "mixed" by the eye made the resulting colours more luminous [28, p.114]. From a distance, the eye blends the colours together, giving the impression of complex colour patterns, and natural reflected light within the scene. This effect is not unlike like that of the pixilated representation of images on a computer screen, and edge-smoothing techniques like *anti-aliasing*, used to create the effect of smooth edges on pixilated images. This technique uses gradual variations in colour to represent smooth curves, lines and shapes. The subtle colour variations fool the eye into interpreting the shapes as being fully formed, but closer inspection reveals the shapes are merely alluded to.

Figure 2.4: Sunday Afternoon on the Isle of Grande Jatte - Seurat, 1884-1886

2.1.3 Art in the Computing Age

The mosaic art form, through its use of the grid-based compositional structure, finds itself naturally suited for development using the computer. A digital image, as viewed on a computer display screen, relies on the same visual representation technique as does a mosaic. This representation is accomplished using an array of small coloured squares, called *pixels*, to form an overall image. No matter how complex the image, it is built from smaller, simpler components — both digitally, and in the mosaic arts.

The development and subsequent availability of the computer in the 1960's and 1970's created interest in the digital display of images on a computer screen. With this newfound image representation came many new areas of research related to the functions and abilities that digital representation of images could offer, such as image comparison, image manipulation techniques and animations.

In 1973, Leon Harmon of Bell Laboratories wrote an article which was published in Scientific American, entitled "The Recognition of Faces" [11, 41]. The article discussed a minimum set of conditions required to recognize a human face, and what made some faces more recognizable than others [51]. A demonstration of his discussed conditions involved a pixelated portrait of Abraham Lincoln, a publicly recognizable face to Americans. The portrait, shown in Figure 2.5, would today be described as a low-*resolution* image — rendered down to 252 pixels from a much higher-resolution original image. Lowering image resolution can readily be accomplished today using an image manipulation program, but Harmon likely created the Lincoln portrait using a photographic process, as it is doubtful that image processing tools of the time could have accomplished the task [41].

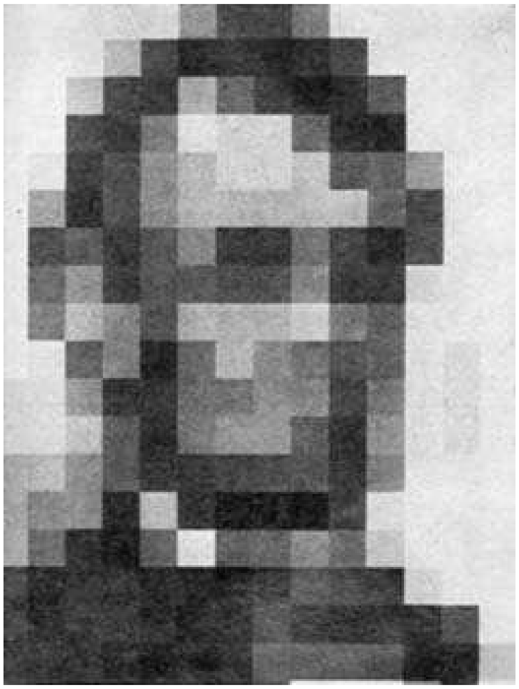


Figure 2.5: Portrait of Lincoln - Leon Harmon, 1973

This portrait of Lincoln is interesting in the history of the image mosaic for two reasons. First, when viewed in close proximity, it may not be immediately evident that it is Lincoln — or even a portrait, for that matter. When viewed from a distance, however, the eye blends the large pixels together, and the face becomes increasingly apparent. This visual blending, as discussed earlier, is a feature shared with the mosaic arts, Impressionism, and with the image mosaic. The second important consideration to recognize is the notion of redundant visual information, and image resolution — the balance between the size of the pixels and the desired rendering detail of the image. When an artist begins a mosaic composition, he or she must choose the size of the individual mosaic pieces and the overall composition size such that it will be possible to render the smallest desirable details. In the same way that an original image of Lincoln may be reduced in resolution to a point at which the image is still recognizable, below which point the image may become obscured, so shall we see that this is a major concern in developing a *good* image mosaic.

As the computer and monitor became more prevalent in modern culture, the effect of its pixelated visual representation of images became an influence to some in the art world. In the same manner that the computer was to become an important artistic tool, influencing the way in which art was created, artistic expression itself was influenced by the way in which digital images were displayed on a computer screen. Some artists began to make reference to and incorporate the digitized, pixelated look into their art work.

One such artist, working from technological ideas, was Salvador Dali — and in this case, the influence sprung directly from the Harmon's Lincoln portrait in Scientific American. In 1976, only three years after the appearance of Harmon's article, Dali produced a composition directly from Harmon's Lincoln portrait. Shown in Figure 2.6, "Gala Contemplating the Mediterranean Sea which at Twenty Metres Becomes the Portrait of Abraham Lincoln" was painted in colour, using oil paint, on large canvas of size 2.5 meters by 1.9 meters. The work is a combination of a number of ideas — the Lincoln portrait, a nude woman gazing away from the viewer, a window, a fiery horizon and the Mediterranean Sea.

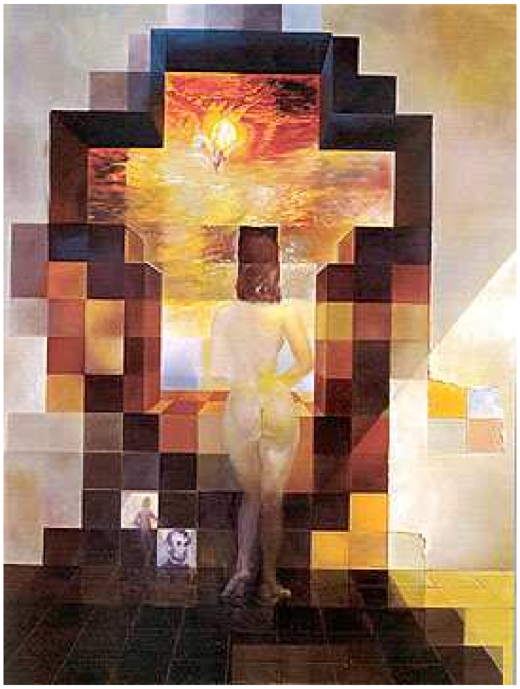


Figure 2.6: "Gala Contemplating the Mediterranean Sea which at Twenty Metres Becomes the Portrait of Abraham Lincoln" - Salvador Dali, 1976

Dali's painting is a notable step in the progression of art leading from the mosaic to the image mosaic for several reasons. It is one of the first clear pieces of artistic evidence reflecting the influence of the computer on art from the perspective of content. A computer did not create the art work, but the results of a computer-based study — the minimum characteristics required for facial recognition, resulting in the Lincoln portrait — was the influencing factor for the work, and was used as the basis by Dali to achieving his artistic goals.

Another interesting facet of the painting, an emphasis in the work overall, is the nature of viewer perspective. The painting contains multiple images at multiple scales, and depending on viewer proximity to the painting, these different images become the focus of viewer experience. That is, when viewed from afar (at 20 meters, as the title of the work suggests) the Lincoln portrait dominates the viewers perception. However, as the viewer moves closer, other details such as the nude woman, the sea and the sky, become more prominent and the Lincoln portrait becomes obscured. Closer inspection of the painting reveals more levels of detail, like a smaller Lincoln portrait and a smaller woman looking away from the viewer.

These ideas of changing viewer perspective and levels of detail according to viewer proximity are both conceptual and visual characteristics shared with the image mosaic. In fact, a deeper analysis of the subject of Dali's painting reveals references toward this very nature of the role of the viewer in perspective and in the actual work itself. That is, in viewing the painting, in addition to the multiple levels of detail based on viewer proximity, we the viewer see the viewpoint of the subject in the painting (the woman) from relatively the same perspective as does she, although she obscures some of our view. This complex notion of viewing a painting of a scene, composed of a woman viewing the same scene herself, makes reference to the nature of perspective and questions not only the actual subject of the painting and the role of the viewer in viewing the art, but makes reference to the idea of multiple layers of perspective in an art work.

Another notable artist to be mentioned in the evolution of the image mosaic is Chuck Close. Close, a leading figure in modern contemporary art since the early 1970's, is best known for his monumentally-sized portrait paintings [49]. Working in the style of *photorealism* — paintings so detailed that they look like photographs — Close's technique involves photographing his subjects with a camera, enlarging the photo, drawing a grid on the photo, then transferring each grid square one at a time to a larger canvas. Figure 2.7 shows a self-portrait from 1968, which was his first work using this style of painting from a photograph [37, 43]. The painting measuring almost 3 meters in height.

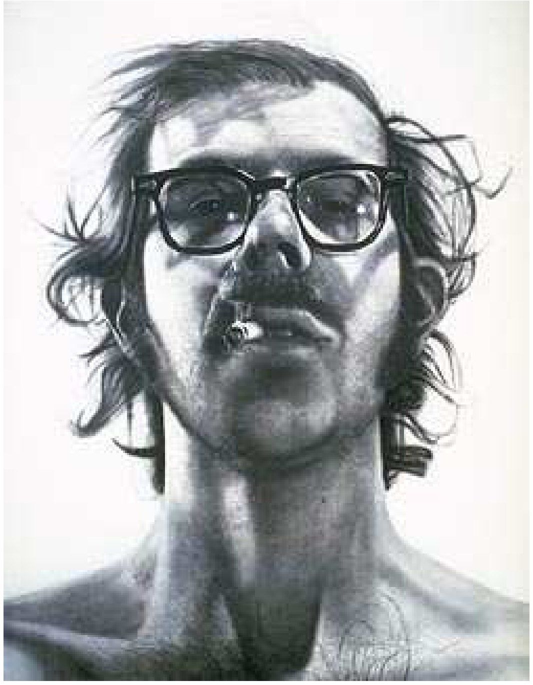


Figure 2.7: Self Portrait - Chuck Close, 1968

Close's grid-based technique evolved over time from the Self Portrait in 1968, a rigorously systematic approach with no visible grid structure, to grid-based experimentations in portraiture with distinctly visibly grid formats [49]. The works shown in Figure 2.8 are examples of his evolving style and grid-based experimentation. Both of the portraits were painted from photographs Close himself had taken. Although the use of subtle shades of colour and the meticulous grid-structure make these works look computer generated, or at least influenced by the digital age, Close has always denied any influence from the computer, or that of technology, on his work [6]. His style was a result of his overall grid process used to divide up photographs, and his own words best convey his feelings regarding any possible technological influence on his work:

> “Some people wonder whether what I do is inspired by a computer and whether or not that kind of imaging is a part of what makes this work contemporary. I absolutely hate technology, and I’m computer illiterate, and I never use any labor-saving devices, although I’m not convinced that a computer is a labor-saving device.” [49]

Close continued his experimentations using the grid-based approach, moving away from the harsh black and white renderings to works that display bright colour patterns and variations in the grid format itself, using circular and rotated grids. His use of colour became much more pronounced, sharing similarities to techniques used in impressionism and pointillism. Within each individual cell in the grid structure, the style became looser, bolder, more vibrant, full of abstract doodles [6]. When viewed in close proximity, the cells are *mini-paintings* in their own right [37, 6]. When the works are viewed from a distance, the grids converge into a single unified image - that of the overall portrait — a fusion of realism and abstraction [6]. Figure 2.9 shows

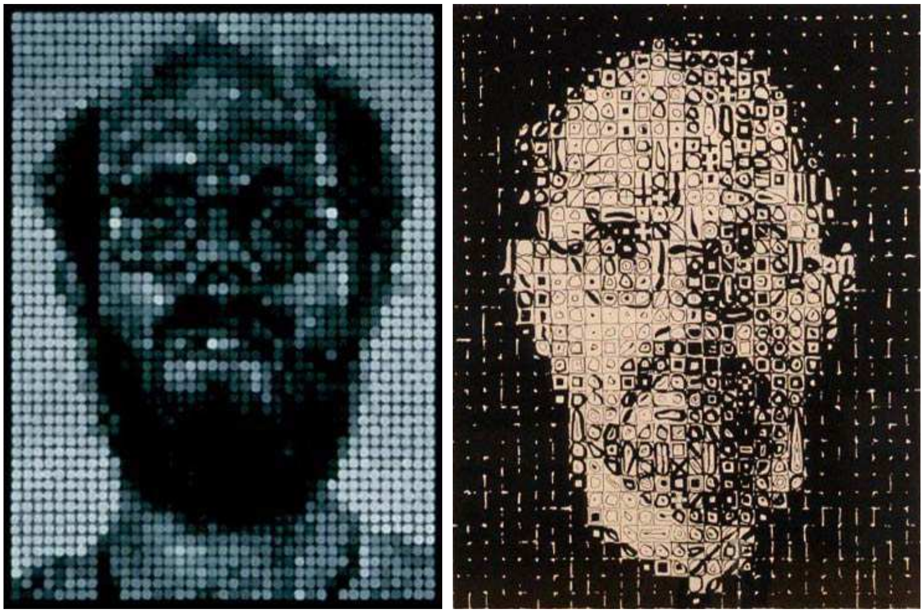


Figure 2.8: Two Self Portraits - Chuck Close

two of Close's most famous works of this nature, with (a) an example of a circular grid as the basis for the work, while (b) is an example of the use of the rotated, or skewed grid basis. Notice how the use of colour in both portraits approaches the style of impressionism in that the colours blend at a distance to form other colours.

Close's work, irrespective of technological influence, shares several key characteristics with the image mosaic. Viewer proximity plays a role in the viewing of Close's art, as it does with the image mosaic, via the different levels of detail based on viewer proximity. Close's impressionistic, pointillist use of colour, and the inherent perceptual blending of these colours at a distance is a dominant characteristic of the image mosaic. Close's use of the grid-structure, in which he treats each cell of the grid as a individual mini-painting, is a dominant concept present in the image mosaic. Finally, the concept of the individual cells of the work itself, forming a single unified image is another core concept shared with the image mosaic.

Figure 2.9: (a) Lucas, and (b) Self Portrait - Chuck Close

The computer evolved throughout the 1980s, increasing in processing speed and overall performance with each successive model. The development of a wide variety of productivity software established the computer as a prominent fixture in both the home and the office, capable of performing a variety of tasks. During this time period, the computer as artistic device was mainly driven by improvements in the display itself and by computer game artists. As a result of the limited colour palette and raster video display technology, image displayed on a computer screen are, in fact, a mosaics in their own right, composed of a grid of rectangular pixels. As the resolution of the computer display increased, and techniques like anti-aliasing were developed, pixels became smaller and jagged diagonal lines became less noticeable, yet the computer display still uses pixels to display images.

The 1990s ushered in an age of unprecedented advances in computing power, storage capacity which greatly improved computer display colour and resolution. As a result, software companies such as Adobe and Corel developed professional-quality image processing programs capable of performing complex image manipulation tasks.

In 1994, a comic-book artist named Dave McKean, produced an image using a program called Adobe PhotoShop, which is a notable work in both form and conceptual construct. Shown in Figure 2.10, McKean's composition is that of a grotesquely distorted face. It was formed by blending a large number of similarly-coloured smaller face images, likely also varying in *opacity*, over a base image. The smaller images are blended and incorporated in such a way that the entire composition serves to suggest one overall image.

Aside from the absence of a distinct grid-based structure, the McKean composition, published several years before the patented Photomosaics by Silvers, shares core similarities with the image mosaic representations. First and foremost, McKean's use of many smaller images, composed in such a way as to suggest a greater whole, is the basis behind the image mosaic concept. Secondly, the notion of changing content based on viewer proximity is also a shared characteristic with the image mosaic. A closer inspection of the McKean composition reveals the multitude of smaller face images, which at that point become the subject of the piece. When viewed from afar, the overall face becomes apparent, as a result of the perceptual blending smaller images, thus effectively changing the subject of the work.

One detail about this piece must be noted, however. Since the smaller images in this piece seem to be composited on top of the larger, preformed image of the face, (i.e. the face image the smaller images are attempting to suggest seems to be present even if the smaller images were to be removed from the composition) the overall composition does not function in the same manner as that of a true image mosaic.

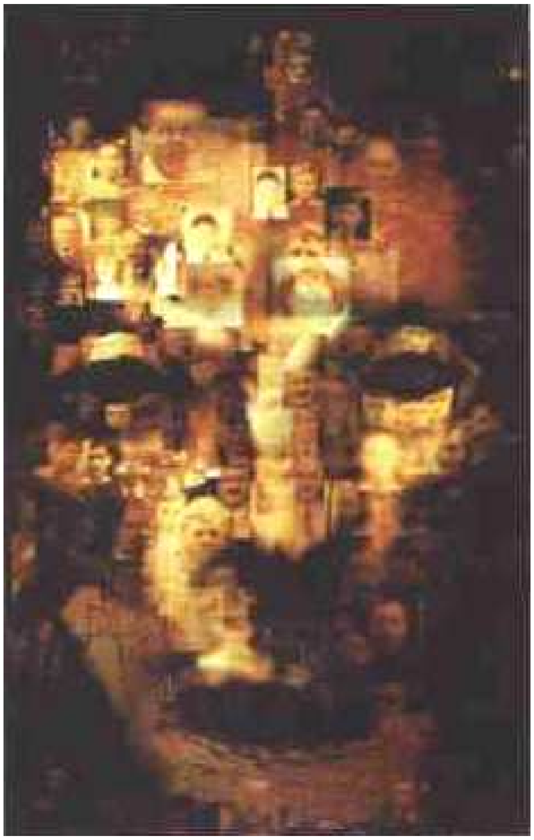


Figure 2.10: Face - Dave McKean, 1994

A image mosaic functions as a result of the sum of the pieces creating a visual whole that is not actually there to begin with. In this particular case the work must be discounted, since the overall face is on some level already present, thus lessening the impact of this work from the perspective of comparison to an image mosaic in terms of true suggestion of a visual whole from sub-pieces.

The early 1990s witnessed the computer evolve into a powerful tool in many areas of media development as the digital image became paramount. Illustration software packages gave artists the flexibility to design digitally from concept to creation. Digital image processing and creation tools, adopted by magazines, newspapers, television stations, advertising companies, and certainly web content creators, became an crucial new tool and source of power, and changed the work-flow of many of these industries.

Widespread changes at the development level in many industries, coinciding with increased interest in the internet design and advertising, created new job titles for artists who chose to adopt the new computer-based image tools. Industry positions such as Digital Artist, Web Media Specialist, Digital Illustrator, Computer Animator, Non-Linear Editor, and Digital Art Director became commonplace in a relatively short period of time. Many educational institutions changed curriculums in an attempt to provide industry with artists for this new digital age of creation.

As the digital arts became established as a new medium in the declared Digital Age, the artists themselves worked towards mastering this new digital craft. As the demands of the industry and the artists themselves evolved, so did the software packages, becoming increasingly powerful and somewhat standardized in their evolution. As an artist, it became important to avail oneself with the latest techniques and software, as to keep pace with the latest tools and technology of the trade.

A product of this evolution was the complexity of tasks the digital tools were capable of performing, and as a result, the learning curve for many of the tools became a concern. Many digital artists found the need to be educated in both art and computing in order to take full advantage of their tool-set. An illustrator, attempting to make the switch from a traditional medium to a digital one, may be faced not only with the task of learning the illustration tool — confronted with vectors, image resolution, anti-aliasing, importing, exporting, file formats, layers and the undo command – but would also be required to learn to use the computer itself. File format portability, the mouse as an input device, file systems, printers, hardware and software conflicts, colour spaces and operating systems are concepts above and beyond those posed by the tools the artist is required to deal with.

As artists become familiarized with the computer in order to master their craft, many computer scientists were extending the field of computer graphics. The artistic software tools themselves had to be created by programmers, but from an artistic insight into what tasks and abilities the artist desired the tools to perform. The process of research and advancement became a loop which drove itself. Research developments in computer graphics were reflected new and updated software tools. The new tools were in turn used by the artists who pushed the technology to its limits in creating new works, and in the process, placing new demands on the software. These new demands again drove research and software development.

In the digital age, a digital artist, well educated in both disciplines — the artistic and the technical sides of their art form — has complete creative freedom. An artist with a programming background and in-depth computer science knowledge has not only the ability to use the computer as an artistic tool, but are also afforded the freedom to create and modify artistic software to achieve any artistic goals they desire. Whether it be extending the functionality of existing tools, modifying existing tools to bypass limitations, creating completely new tools, or experimenting with completely new ideas, the freedom to push the boundaries of their art was attractive to at least one artist with computer science background, and lead directly to the development of the image mosaic.

2.1.4 Image Mosaics and the Photomosaic$^{\mathrm{TM}}$

In 1995, Robert Silvers, working as a graduate student at MIT's Media Lab under Michael Hawley, produced a work that many regard as the first image mosaic. With access to a digital image database at MIT, Silvers developed a software tool which used the contents of this image database to create a mosaic of a given base image, using the library images as smaller tiles of the overall mosaic.

In general terms, Silvers process works in the following manner. The base image is divided up into a grid-based structure, with individual parts of the grid being referred to as tiles. Then, for each tile in the base image, the software tool searches the image database to find the image which most closely matches the tile in question. Each image in the database is given a rank based on metrics such as "tone, brilliance, texture, shape, (and) colour" [41]. The best matching *sub-images* for each tile are then composited to form the resultant image mosaic. Figure 2.11 shows of one of Silvers first commercially available mosaics.

Silvers image mosaics gained public recognition and became a pop-cultural phenomenon in the late 1990s and early 2000s. Appearances on many prominent magazine covers, television commercials, talk shows (including Oprah), puzzles and games in some ways reinvigorated the mosaic art form in the public eye.

Silvers submitted his technique and refinements as his Master's thesis, terming his creations 'Photomosaics.' In a somewhat curious move from an artistic perspective, Silvers went on to apply for a *patent* for his image mosaic creation method and for the apparent "look and feel" of the image mosaic. On November 28th, 2001, a patent was granted to Silvers for "digital composition of a mosaic image." It is interesting to note the terse patent statement on his company website, Runaway Technology. It

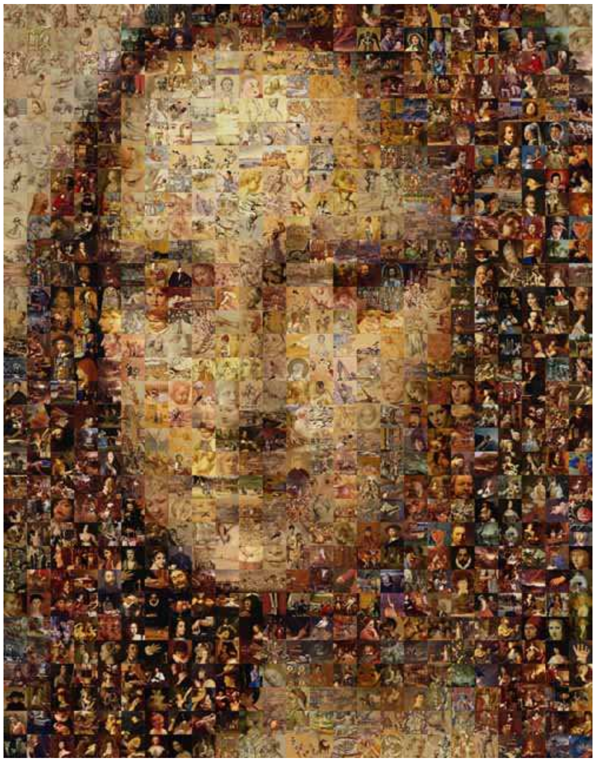


Figure 2.11: Mona Lisa Photomosaic$^{\mathrm{TM}}$- Robert Silvers

reads as follows:

> “Photomosaic$^{\mathrm{TM}}$is a trademark of Runaway Technology. The Photomosaics software is patented (US Patent No. 6,137,498) and the Photomosaic “look and feel” are protected by the patent, copyright, and other intellectual property laws of the United States and other major countries. We protect these rights vigilantly. [46]”

Although beyond the scope of this thesis, this patent raises a number of very interesting questions, from both a technical and artistic standpoint. If an artist or company is granted a patent for a means of artistic expression, especially for something as broad as the ‘digital composition of a mosaic image,’ consider how a company choosing to enforce such a patent would effectively be able to limit both technical research and artistic advancement in the given area. By extension, consider the potential artistic repercussions if the Greeks had imposed a patent on the original mosaic art form, or perhaps Leonardo da Vinci (although not its inventor) being granted a patent for the paintbrush.

Nonetheless, interest in the image mosaic has spawned many progress-friendly academic research papers. Covering many issues associated with image mosaic creation, some including speculation into Silver’s actual (unknown) methodologies, the academic research has put forth a host of refinements, additional problems, variations, modifications and proposed extensions to the form, as well as mutations of the form itself into such things as non-grid based compositions, directionally-oriented mosaics and jigsaw image mosaics. These related forms will be discussed at the end of this chapter.

As well the academic research, many commercial programs have since become available for producing image mosaics. The more robust commercial products include pre-packaged image libraries, exhibiting a diverse colour distribution, which is a crucial in creating a visually-pleasing mosaic. PhotoMontage, Retriever, PictureImage, PhotoSuiteIII [41] and Juggle [50], produce results of various quality - although some programs produce “better” end results than others [41].

This concludes our look at the evolution of the mosaic art form, from its origins in the early Greek period, in combination with related conceptual ideas from other areas of art, involving the construction and representation of a conceptual whole from a variety of sub-pieces, culminating in the image mosaic.

2.2 Motion Picture and the Multiple Image

Although the temporal image mosaic is a new concept in the area of the motion picture arts, its contributions can be examined from the perspective of several stylistic and conceptually similar ideas and previous achievements in the form. Thus, with the history of the mosaic arts and the conceptual contributions leading to the development of the image mosaic firmly in mind, this section will explore the issues and implications of a temporal image mosaic as reflected in the established grammatical[1] language and discourse of the motion picture.

Motion pictures, like photography, comprise an interesting art form that is, in its most basic live-action form, a *capturing* of physical reality. That is, a live-action motion picture is an actual visual sampling of the physical world from the perspective of the camera. In contrast to painting and the mosaic arts, forms in which the artist renders a *representation* of the physical world using a particular medium, the motion picture not only captures discrete instances of time from the viewpoint of the camera, but captures a temporal epoch of time as a series of discrete photographs. If captured at a rate upwards of 12-15 [24, p. 91] images per second, when these discrete images are played back in succession at the captured frequency, the human visual phenomenon of *persistence of vision* allows the brain to interpret the discrete images as being one continuous motion picture. Oddly enough, persistence of vision was discovered by Ptolemy of Alexandria in 130 A.D. [24, p. 570] — at a time when the mosaic was a popular art form and decorative element of society.

---

[1] Film is a visual language. Film grammar is the set of techniques used to convey meaning to the viewer. Unlike the strict grammars of natural and programming languages, the grammar of film is based on historical convention, and these conventions continue to evolve. Determining the meaning of a new expressive device in the context of this historical grammar is what is referred to over the course of this thesis as "grammatical implication."

As compared to the lengthy history of the mosaic arts, the on-screen image is in its relative infancy. The birth of the motion picture occurred in the late 19th century, and like the mosaic, although it has become a well-defined form in certain settings, it simultaneously continues to evolve with each new idea, achievement and discovery.

2.2.1 The Multiple Exposure

The basic unit of film, so to speak, is the image. Most *standard*[2] commercial motion pictures and television programs, whether exhibited in a movie theater or on a television, are displayed as a single moving image, from a single viewpoint, displaying a given point in time. Cuts and juxtapositions may occur between different images during the course of a motion picture, the image itself may be formed from several composited layers of images, such as during a special effect sequence, and the temporal flow of the images need not be chronological, but for the large part of a motion picture, the viewer is watching a single moving image which comprises the majority of the viewer's screen.

Although much of a standard motion picture is composed of a single motion image on the screen at a given time, there can be many instances throughout the duration of a movie that there is more than one image on the screen. Many types of transitions from one cut to another, several cinematic devices, and cinematic special effects — camera-based effects, optical/processing effects and computer generated effects — are all cases where more than one image could be visible on the screen at one time. As one defining characteristic of the TIM is the exhibition of many images on the screen

---

[2] The term 'standard' in this instance is used to draw a distinction between the commercial North American and Western commercial motion picture form, and that of the many experimental and true art-based motion pictures, which are less easily generalized.

simultaneously, these cinematic cases and their uses will provide insights into the motivation and effect of the use of multiple on-screen images.

One of the simplest and earliest forms of representing multiple images in a single frame is the *multiple exposure*. Conceptually, it is like taking more than one picture on the same piece of photographic film. The technique is used in a wide variety of instances to achieve varying artistic goals.

Multiple exposure is a technique native to the photographic arts whereby the film is exposed to light, registering an image, then this same piece of film is exposed again, registering yet another image. These two images are both registered on the film, and, depending on the desired effect, can have an layered and semi-transparent look if the two images overlap, can also lay side by side, or can be used for a more complex composition effect by matting off specific parts of the film or field of view, exposing targeted parts of the film, one at a time. The film can be exposed multiple times, as much as necessary to achieve the desired end results.

The multiple exposure technique is also used in motion pictures, in similar fashion, where the film is exposed multiple times in overlapping or targeted areas to achieve the desired artistic effect. Multiple exposures can be achieved both *in camera*, by rewinding the film each time, or in a photographic laboratory using a special device such as an optical printer, whereby the exposures are under a greater degree of control. Digital composite and layering techniques developed in the late 20th-century have the flexibility to produce simulated multiple exposures, although some filmmakers prefer the visual characteristics inherent in certain multiple exposures on film.

Two images, spanning seventy-two years of film, show two possible incantations of the multiple exposure in an artistic setting. Figure 2.12 is a screen-capture from the 1927 film *Sunrise*, by F.W. Murnau. It depicts a young couple sitting in the grass in the foreground, sharing a daydream, as the female tells the young man about the wonders of big city life. The bottom half of the film was exposed to the young couple on the grass, while the top half of the film was exposed to the depiction of the city.

Figure 2.12: Sunrise - Murnau, 1927.

Multiple exposure, varying in intent depending on artistic requirements for the film, shares a conceptual commonality with the TIM in that it is a method of displaying more than one image on the screen simultaneously. It is a tool which allows a filmmaker to draw visual and thematic relationships between images, temporal points in time, and suggest multiple layers of reality. Although the multiple exposure is one of the most unnatural codes, or logical relationships, in cinema, it can also one of the most meaningful [24, p. 196].

Figure 2.13 is a screen-capture from the 2000 film *Requiem for a Dream*, by Darren Aronofsky. This particular shot is one frame from a 30-second segment depicting a haunting hallucinatory scene in which the main character, Ellen Burstyn, slowly loses her grip on reality as the result of an addiction to prescription medication. In this particular shot, there are three images of the main actor, overlapping. This effect was created by multiply exposing the film the actress while she engaged in a euphoric dance. The result is a layered, multiple view of her character, which coupled with the story itself, serves as a visual representation of her physical and mental breakdown, and also serves as a reference to her drug-induced mental state.

Figure 2.13: Requiem for a Dream - Aronofsky, 2000.

2.2.2 Variations of the Multiple Exposure

One distinctive variation of the multiple exposure is the ***dissolve***. Used widely as a transition effect, it is an important effect to mention, in that it is also a method at the disposal of a filmmaker who desires to display multiple images on the screen simultaneously. Dissolves were originally employed by theatrical film productions to indicate time or space transition, a change in a mental condition, or used as a means of blending the shots of a montage sequence. It also often used in contemporary cinema and television to soften cuts between badly-staged or poorly-edited material, which often can be prevented by thoughtful camera work and direction [22].

The dissolve is produced in the laboratory by an optical printer. This device consists of an optical apparatus which allows multiple images, contained on film strips, to be combined optically and exposed together onto a new film strip. The results of a dissolve can be similar to that of the multiple exposure, as discussed previously in this section, when the dissolve is of long duration.

One of the most famous cinematic examples of a dissolve is shown in Figure 2.14. The shot is from opening of the 1940 film, *Citizen Kane*, directed by Orson Welles. Depicted is the second dissolve from a sequence of ten consecutive dissolves which, shot by shot, visually transports the viewer from the front gates of massive estate up a hill and into a lone, lit window in a mansion. The sequence is highly original in that throughout each successive dissolve, the lit window in the mansion remains in static position throughout each successive dissolve, enlarging slightly each time as each shot is progressively closer to the mansion. The sequence establishes a number of very important thematic ideas with relatively few shots; the isolation of the lonely window; the subjectivity of the films presentation by making the viewer a trespasser, the grandeur, security and also isolation of the estate — while at the same time the duality of this privacy — both to keep out, and to keep in.

In this case, the dissolve helps to aid the economical flow of the overall presentation. The slow pace of the dissolves gives the sequence dream-like quality, and also serves to enhance the etherial impact of the estate in its isolated, “castle in the

Figure 2.14: Citizen Kane - Orson Welles, 1940.

clouds" existence. The linearity and grid-like overlap of many the images during the transitional dissolves drives home the security/imprisonment duality. It should be noted, however, that there is more at work from a filmic standpoint in this particular transition. It is, in addition to simply being a series of dissolves, a montage. This introduces a number of other functioning mechanisms, which will be covered later in this section.

The dissolve, as a cinematic device, is of interest to this thesis as it is yet another method at the disposal of the filmmaker which allows the use of multiple images on the screen simultaneously. In general, as with the multiple exposure, it is often the artistic requirements of the film and the context of the effect in the film which determines the meaning and expressive intent of the dissolve. In its most basic case, as mentioned earlier, it often serves as a transitional effect between two shots to suggest the passing of time. In more artistic settings, the partial dissolve — one that is held for a long period of time — is an very effective device used to make dramatic or dream-like connections between shots and groups of shots. In this case, the dissolve is used in similar manner to a multiple exposure shot, and is also as a form of montage, in which several shots can be on screen at the same time, overlapping and fading into one another.

Two great film examples of this use of the dissolve can be seen in *The Birdman of Alcatraz*(1962) by John Frankenheimer, and *Apocalypse Now*(1979) by Francis Ford Coppola. Frankenheimer uses a long, slow dissolve across several shots, holding many shots for long periods of time, which give a multiply-exposed look as result. The entire sequence is used to symbolize the monotony of prison life for the main character, the slowness of passing time, and the character's inner struggle to cope with his own dream-like existence in captivity. In *Apocalypse Now*, Coppola uses the dissolve in a similar fashion, employing slow dissolves between several shots simultaneously to represent a dream-like states, hint at multiple layers of reality, and to visually portray the character's thematic descent into madness.

Another cinematic device worthy of note to this thesis is the ***superimposition***. A specialized case of the multiple exposure, the superimposition, also referred to as a composite, is used as a method of layering two frames or sequences on top of one another, normally as an attempt to place the elements of one sequence into another. The superimposition has a wide range of possible uses, but one of the most common uses is superimposing a character into another location, creating a composite shot which would otherwise be too costly, too dangerous, or simply impossible to attempt to capture in actuality. One great example of this can be seen in Figure 2.15.

Taken from the 1927 film *Sunrise*, by F.W. Murnau, the shot is a superimposition

Figure 2.15: Sunrise - F.W. Murnau, 1927.

of a young couple over the background image of a busy street. Cars pass back and forth as the couple attempts to cross a dangerous street. The cinematic motivation to construct this particular scene as a superimposition may have been because of the danger imposed upon the actors by the busy street, but the scene is also used to great artistic effect. As the couple continues to walk, the background of the scene changes several times, to visually represent the couple walking through a shared daydream. In all likelihood, this is the reason for the superimposition in this case — so the background could be changed seamlessly as the characters continue to walk uninterrupted on their dreamful excursion.

The superimposition, as with the previous variation of the multiple exposure, is important to this thesis more so from a conceptual perspective than from a strictly visual one. The superimposition allows images captured at different times and places to be combined into a single image. The contextual use of the effect determines its cinematic goal and meaning, but it is often used in an attempt to trick the viewer into believing the shot occurred as shown on the screen.

2.2.3 Split-Screen and Beyond

The next cinematic device of interest that bears both a conceptual and visual relation to the image mosaic is the split-screen. Visually, the split-screen and its extensions will form the basis of the discussion of the temporal image mosaic as it relates to the motion picture. The split-screen is an effect in which the screen is divided into two or more parts of varying dimension, and separate shots are shown in each of the divisions. The reader will likely be most familiar seeing the split-screen shot cinematically represent two sides of a telephone conversation, as has been a cliche of this effect since the silent film era [15].

Outside of a few prime examples, narrative film has seen little use of the split-screen effect. In 1927, French director Abel Gance used the effect in the film Napoleon, where he divided the screen into three parts, a triptych, where then three projectors projected images into each of the three parts of the screen to show the film. Dubbed Polyvision by Gance, these three parts sometimes combined to form wide panoramic shots, sometimes exhibited three separate images, and sometimes showed one central shot flanked by two identical outside shots [15]. Other films which use forms of split-screen effects are the 1966 film *Grand Prix* by John Frankenheimer, *Woodstock* (1970) by Michael Wadleigh, and *More American Graffiti* (1979) by Bill Norton.

In 2000, the movie *Requiem for a Dream* exhibited several examples of the split-screen effect in a narrative setting. Figure 2.16 shows a split-screen in which the screen is divided vertically, with one image on each side of center. In this particular scene, a mother behind a door watches through a keyhole as her son steals her television set from her living room, which he will pawn to support his drug habit. The keyhole shape on the left splits the screen on the left even further.

Figure 2.16: Requiem for a Dream

The shots cinematically represent two views on the same point in time, although closer inspection of the segment reveals that the action does not exactly coincide between each of the views (the two *takes* must be different) and produces an interesting temporally *out-of-sync* feeling, because what is presented as being two viewpoints of the same point in time, do not match on action, nor on time. This produces a perceptual dissonance in this particular scene, as the split-screen has been historically coded to represent two simultaneous events. That is, prior motion picture experiences on the part of audience have taught them to expect that most times, two sides of a split-screen presentation occur in simultaneous cinematic time.

This leads to the second point of interest for this split-screen shot — it is the combination of two distinct points of view. The image of the keyhole on the left represents the subjective point of view of one character — the mother — while the image on the right is the relatively objective viewpoint of the camera/viewer, fixed on the son stealing the television. Thus, this scene, in its combination of these two points of view, results in an overall presentation of ambiguous point of view, forcing each to be considered separately, but together forming one temporally dissonant event. This subjective impact of this visual effect on the viewer serves the films narrative by cinematically representing the emotional distance between the two characters in the film.

A second example of the split-screen in *Requiem for a Dream* is shown in Figure 2.17. Similarly to the first example, the screen is split vertically down the center. This particular scene displays two young lovers who lay next to one another in bed. The screen is split down the middle between the two, with each occupying their own half of the screen. The two images represent the same point in time, as the characters are interacting, and are physically co-located in the same bed, but again, the images are temporally dissonant. When the young man on the left screen raises his hand to touch the young woman's face, his hand does not appear in the right image until several seconds after the viewer would expect it to. In some shots, the action does not match at all — his hand is simultaneously on her face and by his side.

Historically, the split-screen has been used to relate distance. In the instance of showing a telephone conversation using a split-screen, the viewer assumes and expects the events to be happening simultaneously, but over a physical distance, and in such cases, the split-screen has become *coded*, or associated with that feeling. The split-screen in this instance is no different. Narratively, the young couple are having problems in their relationship and the split-screen is used to convey the emotional

Figure 2.17: Requiem for a Dream

distance between the two characters. Again, the nature of subjectivity is present, as even though the point of view in both screens is objective, the split-screen in this instance represents the distance the two characters are feeling from one another, and the overall presentation conveys that distance visually to the viewer.

The last example from *Requiem for a Dream* is shown in Figure 2.18. This split-screen is interesting because it is a horizontal split. In this instance, the woman in the scene is laying out her diet pills for the week. On the top screen, we see an objective shot of the character looking down at the pills. On the bottom, we see her point of view on the pills. Narratively, the woman feels compelled to diet so she can fit into her new red dress, but feels somewhat apprehensive about taking the pills. Visually, it may be somewhat difficult to determine the meaning of this particular presentation, but narratively this split-screen follows the other examples, having been coded earlier in the film as representing emotional distance and separation. In this case, it conveys the disconnection between the woman and the medication which she does, apprehensively at first, become addicted to.

Figure 2.18: Requiem for a Dream

The year 2000 saw the release of another important film to note in this exploration of the mosaic as it applies to the motion picture. *Timecode*, by Mike Figgis, is unique in the history of film. It is an experimental Hollywood film in which the screen is split into four equally sized quadrants for the whole course of the film, with each quadrant containing a separate character storyline. The events and characters in the four quadrants exist in the same cinematic time frame, and over the course of the film, the characters interact, culminating in the four storylines merging at the climax of the film, where the four quadrants then becoming four different points of view on the same overall scene. The film is experimental for a number aspects beyond the quadrant presentation — each of the quadrants are filmed from beginning to end without any cuts, and the film is shot completely using four digital cameras, one for each quadrant. Figure 2.19 is a screen-shot from the film, from a point in the film where two characters in the film are talking to one another via cell phone, narratively connecting two of the quadrants over the period of the conversation.

Figure 2.19: Timecode - Mike Figgis

Timecode's quadrant presentation is of interest to this thesis as the image mosaic, in several aspects, is a direct extension. In viewing *Timecode*, several poignant issues and concepts becomes apparent in relation to the cinematic implications of a temporal image mosaic as a motion picture device. One such issue is that of viewer attention. In the film, with four storylines on-screen at any given time, it can be difficult to take in everything at once. In fact, the onus is placed squarely on the viewer to choose which of the storylines to pay attention to. Through the use of sound editing, the director attempts to guide user attention by manipulating the audio levels of certain quadrants over time, bringing the audio level of certain quadrants to the forefront of the mix, while pushing others to the background. However, the choice is ultimately left to the viewer as to which storyline to attend visually.

With four storylines to attend over the course of the film, information overload becomes an issue. Traditionally, motion picture present narrative storylines one at a time, cross-cutting between multiple storylines when desired or necessary. As such, viewers are accustomed to viewing and processing one storyline at any given time. To display four storylines simultaneously, expecting the viewer to keep pace, may pose a problem. Figgis, the director of the film, addresses this issue:

> “I wondered if it was going to be too complicated, and too much information for an audience to absorb. I thought not, because audiences have become incredibly sophisticated in terms of [assimilating] visual information. Rather than bombarding them with an MTV editing style, where you never get to settle on one image, [I wanted] to go in a complex but alternate route by running four images without any editing at all — in parallel, in real time. [3]”

Another interesting aspect of the quadrant presentation is quadrant association. At times during the film, narrative connections are made between certain quadrants — a telephone call links the characters of two quadrants at one point, two characters in separate quadrants meet face to face and have a conversation, at one point, two characters actually switch quadrants, and at the climax of the film, all four characters are in the same room and the quadrants display four viewpoints on the same scene.

As in *Requiem for a Dream*, the quadrants isolate their characters from one another, but characters can overlap and cross into each others worlds. Narratively, the meanings generated by the connections between quadrants are context-based, but on the whole, separating the characters into their own quadrants helps to emphasize the points of view of each character — resulting in an overall ambiguous point of view — but allowing the director to exhibit the story from four points of view simultaneously.

When two characters meet or when narrative connections are made between quadrants, the viewer, due to the presentation of the story from the viewpoints of both characters at once, objectively sees the scene from both points of view. This is quite different from the traditional cinematic presentation of one viewpoint at a time. In choosing to present the scene from a particular viewpoint, traditionally, this meant the exclusion of all other viewpoints. Figgis comments:

> "What really fascinated me was the fact that with two angles [quadrants], you seem to psychologically see more that twice the amount of information than you do with one. [3]"

In 2002, Figgis released another experimental film, entitled *Hotel*, which saw him continue his experiments with multiple on-screen images, multiple points of view and digital motion picture production. Figure 2.20, a still from the picture, exhibits an exterior shot of a hotel which looks somewhat distorted and off-kilter. Closer inspection of the shot reveals the shot is composed of four different shots, arranged in quadrants, whose frames are overlap slightly [2], forming the overall, almost Cubist, image of the hotel. This conceptual step, the creation of an overall image from sub-components, is a conceptual commonality shared by the temporal image mosaic.

Another interesting shot from the movie can be seen in Figure 2.21. It is a superimposition of four shots of varying size and transparency over a larger background shot. One of the smaller shots is identical to the background shot. All of the shots are of the same scene, they are thematically linked, but the time frames are slightly dissonant. The shots are placed over the background in such a way that they do not obscure the background shot, but in fact blend in with the background. As with the past examples, this is does not exhibit all the characteristics of a temporal image

Figure 2.20: Hotel - Mike Figgis

mosaic, but what it does show are several key conceptual characteristics, that when extended, will result in the temporal image mosaic.

This brings us well into the current age of motion pictures — the digital age. It should be noted at this point that many of the film exposure techniques discussed above, developed since the beginning of the 20th century, now have counterparts in the digital age. The digital influence on the motion picture world at the professional level has afforded a director the option to perform several stages to film production digitally, one of these being the option to perform many of the effects mentioned above digitally. In such cases, the term "multiple exposure" becomes non-applicable since computer images do not function like film which has to be exposed, but more like a painting in that new layers and pieces can be added on top of one another.

For the effects discussed above, when the motion picture is shot entirely on film and then edited on film, the effects are done as discussed previously — some in camera for lower budget productions, and many in processing laboratories. For directors wanting

Figure 2.21: Hotel - Mike Figgis

to perform the effects digitally, and additional step must be performed. If the motion picture is shot on film, the film is then digitized and processed on a computer. After the desired effects are completed, the footage is then printed back to film for projection in movie theaters. If the motion picture is shot digitally, the footage can be loaded directly into the computer. It should be noted that many large-budget Hollywood films are done by this process — where the picture is shot on film, digitized, edited digitally, then printed back to film for distribution.

In all cases, whether the result of an in-camera exposure, the product of optical effect in a laboratory, or aided by the use of a computer, the effects and devices discussed above — the multiple exposure, the dissolve, the split-screen — have associated cinematic meanings from the standpoint of motion picture grammar and from past and historical usage in motion pictures. However, of equally importance is the consideration of context of use, which also a prime factor in determining their grammatical meaning. Many of the examples of effects shown above have a strong contextual component, and this should be of no surprise considering the visual basis of the motion picture. In artistic settings, effects are chosen to aid the director in stating, underscoring and emphasizing narrative and thematic issues visually.

2.2.4 Montage and Mise-en-Scène

Beyond the presentation format of the motion picture image are two important cinema construction concepts which can further our understanding of the grammatical implications of a temporal image mosaic. As noted earlier, narrative context plays a key role in determining the meaning of a particular shot, but so does the context of the shot in relation to the shots before and after it — the editorial process by which the overall segment is constructed. This is known as montage. Similarly, meaning is also generated from the visual content of a shot or scene — the filmic space of the shot itself — which is referred to by the French term *mise-en-scène*.

Mise-en-scène holds that filmic meaning can be generated from inside the frame itself, irrespective of the chronological axis of a film, and is influenced by qualities of the composition such as balance, shape, space, light, colour, movement and tension [24, p. 183]. These visual qualities in relation to meaning are shared among the many pictorial arts, such as painting, photography and the original image mosaic.

The relation of mise-en-scène to the TIM is dual-layered — from the standpoint of overall composition, as well as from the perspective of the individual mosaic pieces. Considering the overall image as a result of the TIM and the visual contents of the entire frame is the first layer one would use in determining the cinematic meaning of the TIM in the context of a motion picture. The question becomes “What is the overall presentation in the frame suggesting?” The second layer of the TIM to consider is that of the individual mosaic tiles themselves, and their relationship to the first layer — the overall image. In the same way mise-en-scène describes the relationship of the content of a frame to the overall presentation, the term can be used to relate the content and conceptual relation of the tiles themselves to the overall composition.

This type of relation can be viewed in many of Robert Silver's Photomosaic™creations, where the smaller images bear a relation to the image they are working to compose. Examples include a composition of the character *Yoda* from the film *Star Wars*, composed purely of frames taken from the movie itself, and a portrait of the late animator Walt Disney composed completely of frames taken from his famous animated features. In both instances, the sum of the whole is created from conceptually related pieces.

The second important motion picture concept which can be applied to furthering the understanding of a temporal image mosaic is *montage*. Montage, one of the primary methods of generating meaning in the motion picture, describes the process by which a third meaning is created from the original two meanings of two adjacent shots [24, p. 216][3]. That is, a particular shot with an inherent narrative meaning, followed by another shot, also with some narrative meaning, is able to generate a third meaning not contained in either of the two shots themselves.

Russian film teacher Lev Kuleshov demonstrated this in the early 1920s, as a result of his research into the theory of montage. In one experiment, Kuleshov displayed a sequence of three shots to an audience, and found that depending on the order of the shots, audiences attributed emotional content to the clips based solely on editing. For example, in one experiment, Kuleshov constructed a sequence using three close-ups

---

[3] To the French, montage refers the editing of the film, which is the form referred to in this thesis. Montage is also used to describe a process in which several short shots are placed in sequence, usually with dissolves between shots, communicating a large amount of information in a short time [24, p. 216].

— a man with a neutral expression, a bowl of soup, and the same man with a neutral expression. When presented to the audience in that order, the audience commented on the actors great ability to convey hunger. Then Kuleshov replaced the clip of the bowl of soup with that of an injured girl laying on the ground. This time the audience indicated they felt the actor was expressing pity. The important point of this experiment is that the actors expression is constant, but the viewer attributes changes his emotional state on the basis of editing.

In several cases of the multiple on-screen image discussed in the previous section, most prominently in the dissolve, montage is the dominant grammatical factor at work. In the dissolve examples showcased in *Citizen Kane*, *Apocalypse Now* and especially *The Birdman of Alcatraz*, the dissolve is used to great effect, but it is predominantly used as the visual glue that holds the sequence together, allowing images to be displayed and cut together in a seamless fashion, softening the cuts between each image into more visually pleasant presentation.

The montage is an important concept in relation to understanding the implications of a TIM because of the many ways which cuts can relate to one another with a multitude of images on-screen at one time, and the many possible relations between these images. Considered at the sub-image level, the TIM is composed of many grid-based image clips, each of which contains its own cutting pattern. Considered at the overall image level, the image suggested by the TIM in turn contain cuts of its own. Finally, considered at a sum of these two pieces, the smaller images can relate to the large image, and vice versa. Thus, the multitude of potential relations between images, governed by the theory of montage, becomes an important tool in determining the cinematic meaning of the TIM in a motion picture setting.

This ends the look into the mosaic and the multiple on-screen image as it relates to the motion picture[4]. Covered have been the many variations of the multiple exposure, and of the split screen and its extensions. Examples of their use in practice have been shown, and their narrative and visual implications have been discussed. The relation of montage theory and mise-en-scène to the tiled image has also been explored. All these ideas provide a foundation from which to gauge the grammatical implications of the temporal image mosaic in a motion picture setting.

The next section will cover peripheral work in the realm of the image mosaic, as well as related art forms, whether they be as a result of strictly computer science development or from the world of art.

2.3 Related Work and Art Forms

Many new visual discoveries and areas of art have come as a result of the digital age. Digital techniques, tool sets, means of display and image manipulation abilities have allowed artists and computer scientists alike to produce many new forms of art and to venture into areas of expression that would have been difficult if not impossible without the aid of a the computer as an artistic device. What follows is a look at forms of art conceptually related to the image mosaic, research into the mosaic, and extensions of the image mosaic idea.

---

[4] Coverage of the mosaic as it applies to the motion picture was strictly limited to the distributed, Western commercial variety. Notably absent from this discussion are the abundance of experimental, non-commercial, and undistributed films, which are difficult to find and view.

2.3.1 ASCII Art

ASCII art[5], or text art, is an art form in which pictures are produced using only text-based characters as pieces. ASCII art became popular in the late 1970s and 1980s, and gained great popularity with the rising use of the internet, the World Wide Web and electronic mail. Though ASCII art was commonplace throughout the 1990s with the widespread use of the internet, text-based art is over one hundred years old, tracing its origins back to the invention of the typewriter in 1867. Figure 2.22(a) displays the earliest known work of typewriter art, a depiction of a butterfly, produced in 1898 by a woman named Flora Stacey. Taken from the October 15th, 1898 version of *the Phonetic Journal*, the entire work was created using a typewriter, and the butterfly itself is comprised of brackets, hyphens, points, oblique strokes, a single asterisk, and several ‘o’s [48].

Figure 2.22(b), artist unknown, is a modern ASCII art depiction, circa 1990, of Leonardo DiVinci’s *Mona Lisa*. This work differs from the previous example of the butterfly in that the spacing between the characters is pre-determined by the computer itself. In the case of the butterfly, done with a typewriter, the composition seems smoother and more linear. This was a result of the paper on which the butterfly was created being turned and twisted such that each character and letter could be placed precisely in the desired location [48].

Text-based art is traditionally created by hand, with the aid of a typewriter or a computer, although examples do exist of text-based art being created completely by hand with only a pencil, a paintbrush or a printing press. Lewis Carol’s 1865 book, *Alice in Wonderland*, which predates the typewriter, contained several examples of

---

[5] The term ASCII is an acronym, standing for American Standard Code for Information Interchange, and refers to the predominant 7-bit character set encoding of present-day computers.

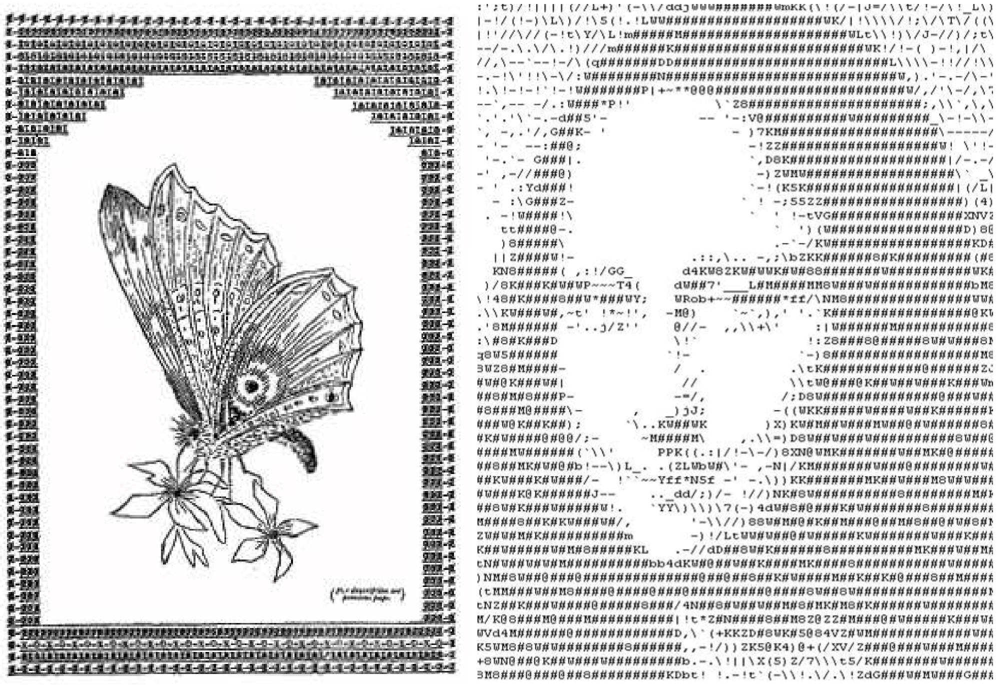


Figure 2.22: (a) Typewriter Butterfly, and (b) ASCII Mona Lisa

text-based art of a similar variety in which some sentences were molded to form certain shapes, such as the tail of a mouse. Today, many computer utilities exist which can convert images directly into ASCII art.

The major conceptual relation of ASCII art to the image mosaic is the idea of composition of a whole from a variety of sub-pieces. In this case, the parts are the set of ASCII characters. ASCII art also shares the idea of user proximity with regard to level of detail. When viewed from afar, perceptual blending helps to unify the image, but when viewed from very close, the characters themselves become the subject and the suggestion of the overall image becomes degraded.

2.3.2 Bar Code Art

Bar code art is an interesting variety of computer art produced by Scott Blake. Shown in Figure 2.23(a) is a work by Blake entitled *Bar Code Jesus*. The overall image is that of a portrait of Jesus, composed of 27 vertical lines of varying width, which itself suggest the parallel lines of a bar code. Detailed inspection reveals the depth in the work, as the 27 lines, in turn, are themselves composed by rows of bar codes. The visual density and overall length of each bar code is placed in such a way as to form the changes in width of each of the 27 lines from which the overall portrait is composed. On a deeper level yet again, Figure 2.23(b) depicts a sampling of the bar codes themselves, which all represent words from the Biblical Book of Revelations. These words, in bar code form, bear an overall conceptual relation to the portrait based on artistic intent.

Blake's own artist statement best describes his work, and his artistic intent:

> "My bar code art explores the process of making art with information and technology. As a computer artist I am in the business of selling 0's and 1's. The bar code represents automation, efficiency, and commodities. It is the universal icon for the computer revolution.
>
> I started working on my first bar code portrait in 1998. Since then I have experimented with bar codes in many different forms, both on and off the computer. I have made videos, paintings, and gallery installations all based on bar codes. My bar code portraits have remained the main focus of this body of work. The people I choose to depict in my unique style of digital pointillism are very recognizable faces. I am interested in exposing what goes into the commercialization of individuals. My bar

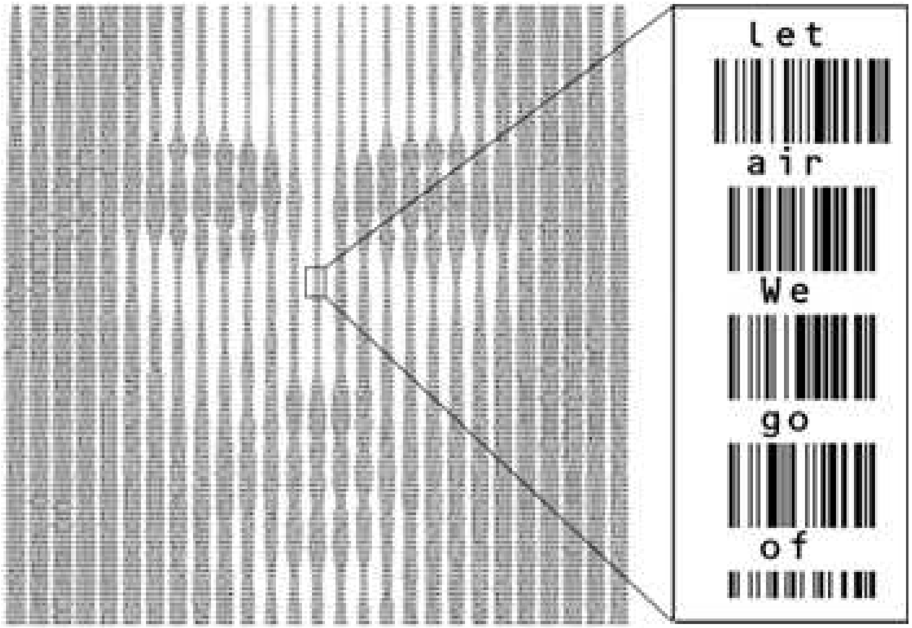


Figure 2.23: (a) Bar Code Jesus, and (b) Enlarged section

> code portraits show what these people are made of. [38]"

Blake's bar code art shares several commonalities with the image mosaic, the largest of which being that of the multiple layers of the composition based on viewer proximity. When viewed from afar, the bar code art portraits suggest a larger overall composition. Closer inspection of the portrait reveals a deeper layer which can be interpreted on its own merit. In the case of the *Bar Code Jesus*, the depth of the portrait continues to the conceptual level by the bar codes themselves being representative of words which bear some relation to the overall work.

A interesting aspect of art in the digital age is that most artists have their own web sites, and Blake is no exception. In the Frequently Asked Questions section of his site, Blake answers the question, "What is the difference between your work and

Rob Silvers's Photomosaics?"

> "We have both created a new gimmick using an old idea. The mosaic has been in art since the eighth century BC. Some of the early floor pattern designs where made using only white and black tiles, like bitmap mode. Other artist(s) have experimented with this idea such as Chuck Close, Roy Liechtenstein, and Georges Seurat. Chuck Close has been a big influence on my art. His process of building a painting one block at a time encouraged me to work on my own designer pixels. Roy Liechtenstein's examination of Benday dots opened my eyes to the science of halftone printing. Photomosaics use tiles that are made out of photographs to create a mosaic image. My bar code portraits are made using bar code tiles. Each bar code in my portraits is nothing better than a fancy pixel. One tile in a Photomosaic represents 'much more than a pixel'. [38]"

2.3.3 Chuck Close Filter

Another interesting digital creation from the mind of Scott Blake which bears mention in this thesis is the Chuck Close filter. This 'filter' or process takes a given image and produces a tiled resultant image in the style of artist Chuck Close (mentioned earlier). Figure 2.24(a) depicts an original Chuck Close self-portrait, painted from an original photograph, while Figure 2.24(b) shows the result of Blake's Chuck Close filter on this same original photograph.

To make the filter, Blake took an original Chuck Close painting and broke it down into its basis tiles. Blake then assigned a density value to each tile, grouping them into 52 possible values ranging from dark to light, with 10 possible variations

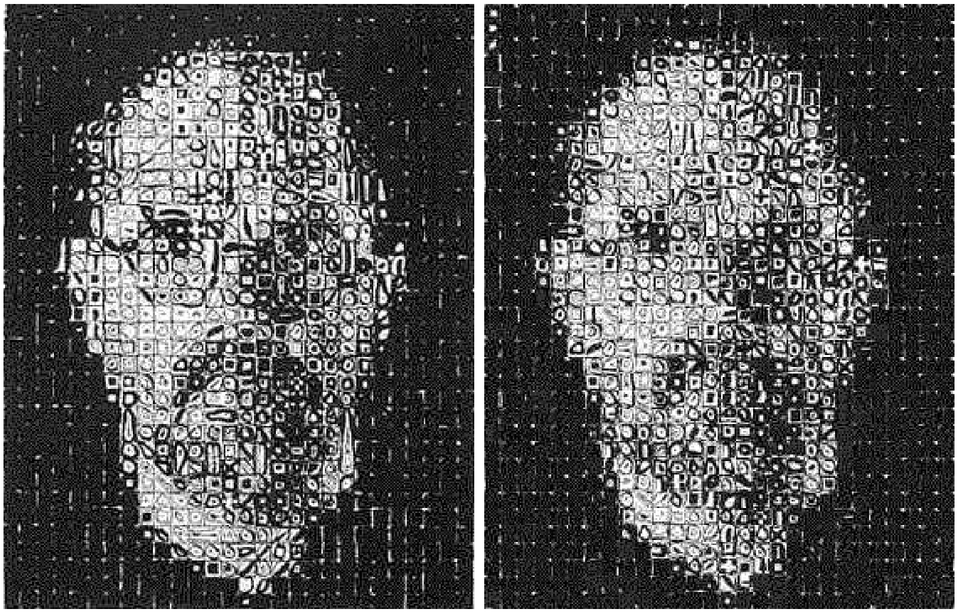


Figure 2.24: (a) Chuck Close, and (b) Chuck Close filter

of tiles within each of the 52 values. Figure 2.25 depicts the value grouping of the tiles extracted from the original Chuck Close portrait, arranged from dark to light, with 10 tiles corresponding to each value. Then, with a given base image, Blake's algorithm constructs the filtered image using these value-sorted tiles as a palette from which draw.

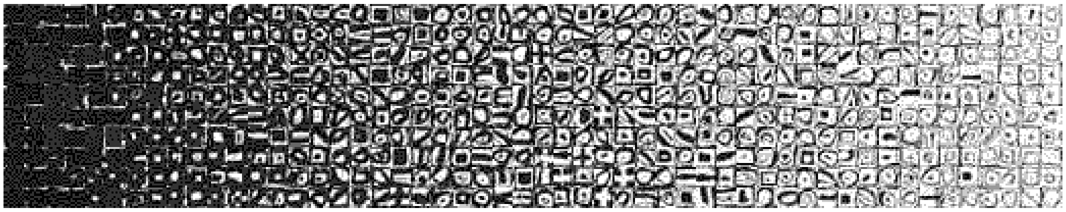


Figure 2.25: Chuck Close value scale

The Chuck Close filter is very similar to the image mosaic. Both involve a strict grid-based structure, both replace parts of this grid with other images, based on which part most closely approximates the original image to form a suggestion of the original image, and both exhibit visual depth in that the pieces themselves are artistic pieces in their own right. The only differences between the two are the way in which the tiles of the Chuck Close filter are graded based on shading value, and the non-temporal nature of the tiles, which would make a 'temporal Chuck Close filter' rather unappealing due to tile discontinuity between successive frames.

2.3.4 Fresnel Art

It is often the case in the world of art, and similarly in the digital world, that not everyone wishes to elaborate or have known how a particular piece of art (or technology) was created. Such is the case with Glenn Zucman and his "MagnaDott" creations — at least, this author was unable to find conclusive evidence as to how the works were created, although clues do exist which shed some light on the potential process behind his intriguing creations. Shown in Figure 2.26 is a work by Zucman entitled "Fresnel Jodie."

The portrait is that of Hollywood actress Jodie Foster. It is a six colour composition (black, white, green, blue, yellow and red), arranged in circular, moire-like fashion, which at a distance blends to form a fully coloured image. Many overlapping concentric circles of colour, whose widths are not constant, serve to distribute colour over the image in specific quantities in certain areas, which then become blended by the viewer eye.

Although the work itself is painted in canvas and is done in acrylics, the artist refers to the creation technique as a "MagnaDott" process. An internet search reveals a Magnadot (only one 't') process to be associated with that of laser image-setting

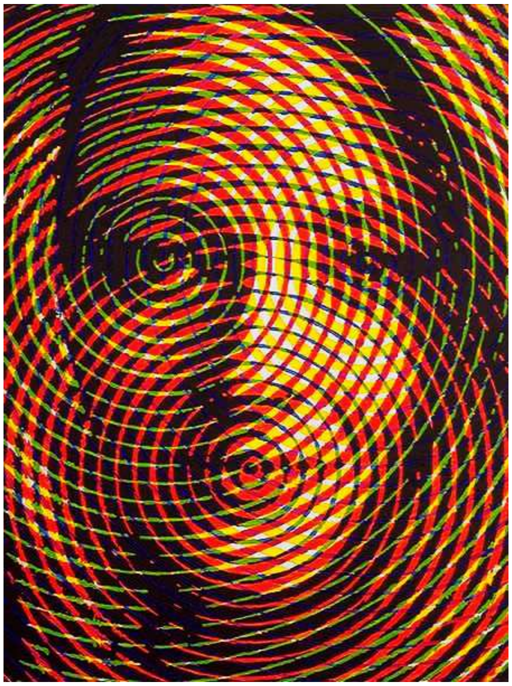


Figure 2.26: Fresnel Jodie - Glenn Zucman/Artboy$^{\text{TM}}$

devices, involved in the halftone printing process [35]. Although the details of the creation process are less important, it does not appear that it is done explicitly with the aid of computer — that is, by an image processing algorithm.

The similarity between this work and the image mosaic is the perceptual blending of colour at a distance. The colour blending is striking in *Fresnel Jodie*, forming a full-colour portrait — complete with correct shades of skin-tone, red lipstick and blonde hair. The colour blending is very similar to Impressionistic painting.

2.3.5 Impressionistic Video

To this point, many of the related and peripheral art forms discussed have been two-dimensional in nature, with no temporal component. Impressionistic Video [20] is a technique which exploits the temporal nature of video. By processing a given video using the technique, a video which has hand-painted, Impressionistic qualities can be produced.

Several previous papers have covered the topic of painterly rendering, including techniques for computer-assisted transformation of images [10], interactive techniques for producing pen-and-ink representations of images [25, 26], and systems for transforming 3D geometry into animations with a hand-painted look [23]. Although the Impressionistic Video technique is not the first to produce Impressionistic images, it is the first which tracks pixel motion, resulting in a temporally coherent, painterly styled end result [20]. Figure 2.27 depicts a field of flowers processed using the Impressionistic technique.

The overall technique works by using optical flow fields to "push" brush strokes from one frame to the next, in the direction of pixel movement, while randomly perturbing each brush stroke in terms of length, color and orientation to generate the "hand-painted" look [20]. Edges are detected in the original frames, and brush stroke lengths are determined by clipping against these edges. In this manner, fine details such as linear features and object silhouettes are preserved.

Impressionistic Video as a whole shares little in similarity with the image mosaic, other than the desire for a temporally coherent result. Though Impressionistic Video is concerned with the re-representation of a given original image or video, it does not rely on a comparison or image replacement process. Instead, it concerns itself

Figure 2.27: Impressionistic Effect

predominantly with producing a modified representation of the original video which adheres to the visual qualities of an Impressionistic painting in terms of color, edge representation and painterly brush strokes.

2.3.6 Video Cubism

Video Cubism [17, 18] is another intriguing computer graphic research work which also uses the temporal aspect of video as the basis of the work. It is the product of a Microsoft Research project with Princeton University. The result is a new set of non-photorealistic (NPR) rendering tools which exploit the temporal aspect of video, providing a variety of interactive aesthetic controls which allow an artist to produce resultant works from the video itself [17, 18]. The types of works possible with the tools approximate the spirit of several modern art movements, notably Cubism, Impressionism, Abstraction — and even the Mosaic.

The heart of the video cubism tool set works by processing the desired video footage to produce a set of *stroke solids* [17, 18]. A stroke solid is a three-dimensional parameterized slice through the space-time volume of video, which allows the technique to track certain optical flows through consecutive frames of the video. The process then acts on these stroke solids instead of the individual frames, to form the desired artistic result, allowing the video cubism process to extend beyond static frame processing tools, which, when applied to a temporal image on a frame-by-frame basis, often produce undesirable temporal artifacts [17, 18].

Figure 2.28(a) depicts one frame from a unique video produced using the video cubism "shard" rendering style. The frame is Cubist in appearance, exhibiting many slices, each a different viewpoint on the same woman's face. The process involves the use of an authoring tool to subdivide areas of the screen referred to as shards. This tool then allows placement of these subdividing lines, in a key-frame manner, throughout the video. These subdividing lines are then linearly interpolated between key-frames, allowing for movement of the shard boundaries in the final resulting rendered output. Shards can also be processed in certain manners such as colouring, zooming, modifying time, and using two video streams as input.

Figure 2.28(b) is a single frame from the video cubism variation on the image mosaic. The fractured depiction is that of a young girl's face. The tiles, actually stroke solids through the video space, are arranged in a grid structure, with individual tiles possessing characteristics such as scaling, offset, rotation, zoom and time-offset [17, 18]. The tool set allows the artist to specify an average value and a variance

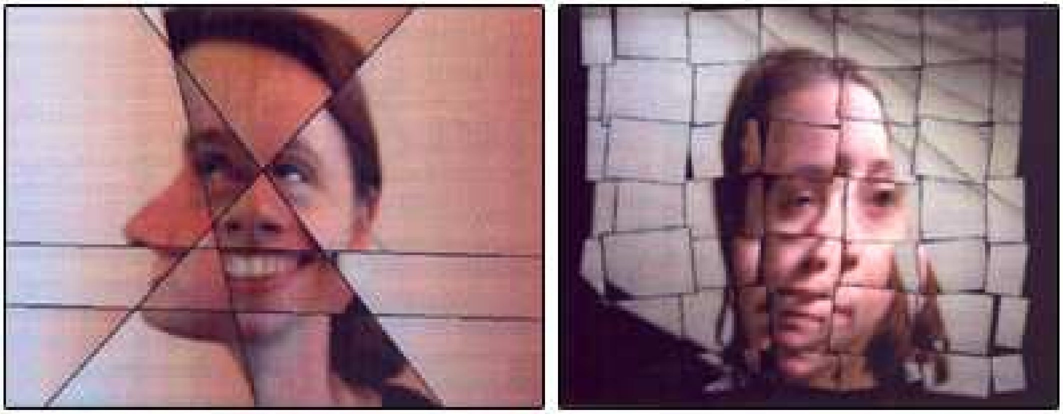


Figure 2.28: (a) Video Cubism, and (b) Video Cubism Photo mosaic

by which each of these characteristics can be randomly effected. Tile borders are distinctly visible, and tiles are permitted to overlap, which, along with time offset, produce a very Cubist mosaic effect.

An interactive Video Cubism approach by Fels *et al* [13] at the University of British Columbia, produces similar visual results by treating the video as a three-dimensional solid, and allowing the user to manipulate cut planes or cut spheres through the video data. The manipulation of the planes can occur in real-time, allowing the user to explore the video data throughout its temporal component, from a variety of perspectives. Images can be abstract or concrete, depending on the position and orientation chosen by the user [13].

There is relatively little similarity between the Video Cubism mosaic effect and the temporal image mosaic, beyond the temporal aspects. The Video Cubism mosaic effect, though it also produces a tile-based mosaic as a result, does not concern itself with image replacement in any fashion similar to that of an image mosaic, but rather endeavors to produce an abstracted representation of the video — a simulation of

Cubism. As future work, it would be interesting to apply the idea of the stroke solid to the TIM as a method for reducing temporal artifacts.

2.3.7 Simulated Decorative Mosaics

One interesting aspect of computer science research is that considerable time is spent, especially in the computer graphic areas, on attempts at simulating the world right in front of our eyes. Large parts of computer graphics — animation, rendering, shading, lighting, modelling, movement — are judged by how close the end result of the computing process simulates, represents or duplicates the look and movement of our physical world. Such is the case with work done by Alejo Hausner at the University of Toronto in attempting to simulate decorative mosaics.

Several commercial products are available which can produce a mosaic from a given image file, but none take into account directionality of tiling, nor seek to emphasize edges. Adobe Photoshop contains functionality for producing a tiled mosaic of a given image, but it functions in a very basic manner — it uses square tiles and merely reduces spatial resolution of the original image to produce the mosaic [12].

Hausner's approach is an extension of two specific attempts at producing tiled mosaics — one which uses Voronoi diagrams, and another which uses oriented particles to interpolate surfaces and to control surface deformations [12]. Given a starting image, and user-selected edge features, Hausner's process produces a mosaic which reproduces original image colours as well as orienting tiles in such a way as to emphasize edges. Figure 2.29 depicts a mosaic generated from Michelangelo's *Lybian Sibyl*, created using 2000 tiles.

Figure 2.29: Lybian Sibyl mosaic - Hausner

In regard to this thesis, the simulated decorative mosaic bears only a casual relation to the TIM, as beyond the fact that both techniques aim to produce a mosaic from a given image, there is little technical or functional similarity between the two. The image mosaic is concerned with searching a database to find the best possible match for a given section of the base image; the decorative mosaic simulation is concerned with manufacturing its own tiles, colouring and orienting them in such a way as to emphasize edges in the original image, with the end result of simulating a mosaic of the Greek and Roman period. Though there is little similarity between the two approaches, Hausner's technique is of interest and produces visually pleasing results. It would make for interesting future work to explore the possibility of producing an animated decorative mosaic, with some temporal component to the image.

2.3.8 Variations on the Image Mosaic

Since the appearance of Silver's first Photomosaic$^{\text{TM}}$, several interesting varieties of the image mosaic have appeared that bear discussion. The first of these variations is the variable tile-size mosaic, which moves beyond the limitation of the uniform grid-structure and fixed tile aspect ratio. The second variation, the halftone mosaic, moves away from the necessity of a large candidate image library. The third variation, the jigsaw image mosaic, break away completely from the grid structure and the fixed, rectangular image size requirements. At this point, is should also be noted that the terms 'video mosaic' and 'image mosaic' bear no relation to the video mosaics as studied in the field of computer vision. These types of mosaics are concerned with stitching together component video segments of a single scene in order to form a resulting video with a larger field of view [31, 32].

Variable Tile Size

One limiting aspect of the Silvers Photomosaic$^{\text{TM}}$is the grid structure which necessitates uniform tile size. The grid is imposed on the image one wishes to mosaic, then the candidate database is searched for images which best match the tile areas of this predefined grid. Additionally, every image in the candidate database must be of the same aspect ratio as the grid tiles.

One technique which moves beyond these two limitation — uniform grid structure and fixed tile size — results in an image mosaic like that shown in Figure 2.30(a). The work was created by William L. Hunt, and it is a self-portrait composed of variable aspect size postage stamps [41]. The composition still relies on rectangular images, but frees itself from the grid-based structure. Closer inspection of the work shown in

Figure 2.30(b) reveals that images are allowed to overlap and to be positioned at any point on the image where best fit occurs. This is most likely the result of a sliding match approach (no details are given as to how Hunt's process functions) in which the images in the library are compared to the target image on a pixel by pixel basis, sliding each image in the library over the entire image to find the best possible match. On the basis of this mosaic, a maximum overlap amount may come into play, since the majority of images overlap other images, but by only a small margin.

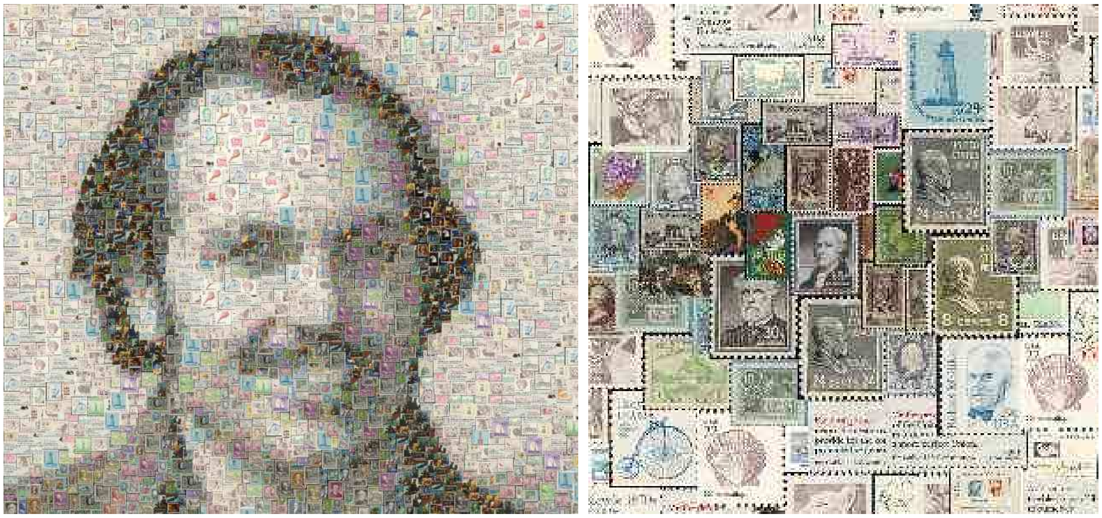


Figure 2.30: (a) Stamp Mosaic, and (b) detail

Hunt's mosaic is very clean and the tiles are well distributed in terms of proximity of repeated tile images. Identical images do appear in the mosaic (some algorithms prevent the use of the same image more than once in a mosaic), but these occurrences appear to be randomized in their proximity placement as no visible banding is present in the large areas of uniform colour. Another interesting point to note about this mosaic, although as much artistic as technical, is that the work is a black and white picture generated from colour photos.

Halftone Mosaic

A crucial aspect of creating the image mosaic is the library of candidate images. This library must be diverse and well distributed, containing a wide variety of colour and shape configurations. The better this library, the closer the matches will become when creating the image mosaic, and thus, the better the resulting mosaic will be. However, simply increasing the size of the Candidate Pool is not guaranteed to result in a more visually-pleasing mosaic [34]. Building a sufficiently large and well-distributed library can be a difficult task, and determining what constitutes a “diverse” Candidate Pool is an important concern [33].

Show in Figure 2.31 is a halftone mosaic as created by Finkelstein and Range. It moves away from the limitation of a large candidate library by creating the mosaic from a single repeated image, and adjusting the colour and brightness of this particular image accordingly in a technique which is a generalization of the halftone printing process [9].

The halftone process was developed as a method of printing on paper that used black dots of variable size, arranged in a grid-structure, to produce the effect of varying shades of gray . This process originated as a purely optical phenomenon, the result of photographing an image through a transparent screen, using high-contrast film [9], but has since moved to the domain of the computer. Shown in Figure 2.32(a) is one specific instance of a halftone value scale from black to white. Figure 2.32(b) shows the Finkelstein and Range generalized halftone value scale, created by varying the brightness of the image across the scale.

There are several interesting points to note about the Mona Lisa halftone mosaic. Close inspection of the work reveals the mosaic to be three levels deep. The overall

Figure 2.31: Mona Lisa Image Mosaic - Finkelstein and Range

image of Mona Lisa is composed of a tiled layer of smaller Mona Lisa's, as a result of Finkelstein and Range's generalized halftone pattern. In turn, these smaller Mona's are themselves composed of a third layer of halftone Mona Lisa's. This depth could conceptually infinite, reminiscent of the structure of fractals [9]. A second interesting point to note is that Finkelstein and Range's tile arrangement tool allows for either automatic or manual tile placement. In the case of their Mona Lisa mosaic, the arrangement is likely automatic since the halftone pattern is composed of a simple side-by-side tiling of the image. Another possible tile arrangement pattern is shown in the halftone value scale, shown in Figure 2.32(b).

The final point that bears reiteration is that no candidate library was necessary to create the halftone mosaic — no database, no intensive searching methods, no disk

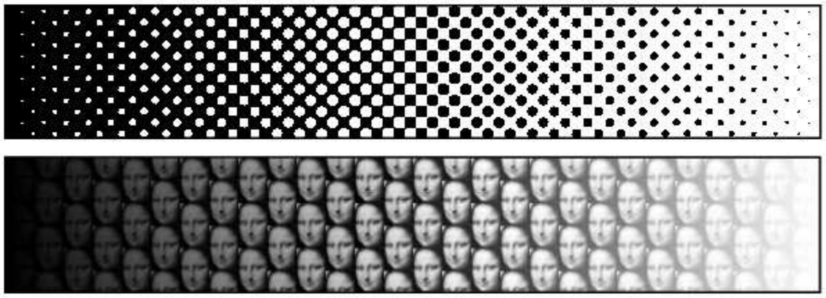


Figure 2.32: Halftone Value Scales - (a) Dots, and (b) Mona Lisa

space constraints. The entire mosaic is composed using only one image, which in the case of the Mona Lisa mosaic, is both subject and sub-component. As a result, the creation time of the halftone mosaic is substantially less than of the library-based image mosaic. Creation time of several hours or days in the case of the image mosaic is far beyond the few minutes processing time [9] required to produce a halftone mosaic.

Jigsaw Image Mosaics

The early computer-based mosaic effects were based on the use of rectangular tiles. The reason for this, both in the original analogue sense as well as digitally, is the same reason why bricks are rectangular. Rectangular tiles have straight edges, and straight edged tiles are easier to fit together with other straight edged tiles when constructing a mosaic (or brick wall). The Jigsaw Image Mosaic (JIM) breaks away entirely from this rectangular tile restriction by creating a mosaic with arbitrary shaped tiles [16].

The JIM process works by breaking the original image down into a number of similarly-coloured regions, or *containers*. Each of these containers is then packed with images from the image database, based on the underlying colour of the original image for each container. Figure 2.33 depicts a jigsaw image mosaic of a parrot. The individual containers of the parrot mosaic can be seen in the upper-left corner of the image. In an attempt to achieve the best packing results possible, the images are allowed to be deformed by up to a pre-specified amount. Although the containers are generated by the algorithm, the process also allows for the creation of user defined containers. This provides the user with a degree of artistic freedom in the containerization process. In the case of the parrot, detailed feather shapes were added to enhance the visual impact of the final result.

Figure 2.33: Jigsaw Image Mosaic

The jigsaw image mosaic is a visually interesting art form. The works are layered in that closer inspection of the overall work reveals the smaller images that comprise the image. Edges and coloured containers within each of the works are well-defined, making the overall subject of a JIM often more readily distinguishable than an image mosaic, in which edges and coloured areas tend to be less distinct.

There are a few possible limitations of the jigsaw image mosaic from an artistic perspective, however. Because only the colour of the image tiles is considered when filling a container, representation of fine detail in an image becomes problematic. That is to say, a container with an average underlying colour of yellow in the original image must then be packed with yellow images, irrespective of subtle shade changes or gradation in the underlying image. A second potential limitation of the JIM is due to the orientation of the tile images. In its attempt to pack a container, tile images may be placed in any orientation, including images which are upside down in relation to the viewer. This may obscure the identity of such images in the resulting mosaic.

Video Mosaics

During the course of the TIM implementation for this thesis, a closely-related attempt at producing an image mosaic with a temporal component appeared in a research paper entitled *Video Mosaics* [19]. Implemented by Allison Klein and Adam Finkelstein at the University of Princeton, in conjunction with Michael Cohen of Microsoft Research, their approach is similar to that taken in this thesis, in terms of overall process and desired outcome. A resulting frame from a Video Mosaics is shown in Figure 2.34.

The Video Mosaic process begins with collection of source video footage and a target video which will serve as the basis for the mosaic. Their approach first preprocesses the source footage to extract image information. The target video is then divided into a grid of tiles, and the process searches for source videos video frames to put in the target tile.

Once video frames have been selected which best represent each tile in the target image, the resulting videos are *color corrected* to make each more closely resemble the

Figure 2.34: Video Mosaic

target video. The color correction process modifies the colour and brightness of *each pixel of each source frame* to make the source frame more closely match the target frames. It takes into account the colour distribution and brightness of the target image, and adjusts each of the selected source frames by shifting and scaling their corresponding colour histograms and brightness distributions, to more closely match the colour and brightness of the target image.

The Video Mosaic, its colour correction techniques and its similarity to the TIM approach will be discussed at greater length in following chapters. Suffice it to say at this point that the TIM and the Video Mosaic share a common goal of extending the static, two-dimensional image mosaic into the third dimension of time. Each approach constructs a temporal representation of a base video sequence from a pool of candidate video sequences, by searching through the pool to find the best-fit sequence for each tile of the base video sequence.

This brings the Chapter 2, the background of the mosaic, to a close. The history of the mosaic art form, from its beginnings through to the current digital age has been discussed. A study of Western film art has been covered in an attempt to determine the grammatical implications of the temporal image mosaic in a motion picture setting. Concluding the chapter was a look at related research and art forms.

In the following chapter, the Technical Background, technical concerns and computer science concepts are detailed. These form the basis for the implementation of the temporal image mosaic.

Chapter 3

Technical Background

The following chapter will provide a technical background, discussing the different techniques and methods that have been used in the past to create image mosaics. The chapter will begin by looking at colour. Colour space is one of the most important considerations as it determines the basis of image matching as well as being one of primary qualities the temporal image mosaic wishes to conserve. After colour, the basic brute force algorithm to create an image mosaic will be discussed. Improvements to this method will be covered, both in terms of speed and quality, and alternative methods will be discussed. Chapter 3 will then conclude with a look at the algorithms used by variants of the image mosaic.

3.1 Colour

Colour is an important consideration when attempting to produce a visually pleasing image mosaic. Since the desired result of the image mosaic process is an approximation of an original image, the best solution will be as close to the original image in terms of colour as possible. But how de we determine the "closeness" of one colour to another? Defining a metric on which to base image closeness in terms of colour is a crucial element.

There are several colour models which could be employed during the image matching process. The *RGB scheme* in one of the simplest. It is composed of an integer triplet which describes a colour as containing three parts — red, green and blue — with a range of 0 to 255 assigned to each. The original patented version of Robert Silvers' Photomosaic$^{\mathrm{TM}}$program likely used the RGB colour model. Although Silvers' technique and software are proprietary and patented, the patent literature itself shows the use of the RGB model:

> "31. The article of claim 29 wherein the process includes the further step of computing the average Root-Mean Square error of Red, Green and Blue channels."

The benefit of using the RGB colour model is that it is supported by most graphic applications, and both television and the computer monitor physically represent colour using red, green and blue phosphors. A second benefit of RGB is that it is a three-dimensional orthogonal space [45], which reduces colour matching to a simple Euclidean distance calculation, of the form:

$$\Delta(RGB) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2}$$

The main problem with the RGB colour space is that human beings to not perceive colour in terms of the RGB model. There are other models to consider which do more closely model human perception, notably the CIE XYZ space [45]. This space is a complex colour model, but although the space does represent all the colours humans can distinguish, it is not a perceptually uniform colour space [45]. That is to say, distance calculations between two colours in this space do not accurately reflect the relative closeness of the two. Two related standardized colour spaces which do exhibit perceptual uniformity are CIE L*a*b* and CIE L*u*v*, but they too are regarded as inadequate models for the perception of colour [45].

The colour model used by Finkelstein and Range for their image mosaics is the YIQ colour space [9]. Developed by the video and television industry as a means of perceptually uniform — yet highly compressed — colour representation, the YIQ model is standard for NTSC-television, and represents colour using *luminance*, Y (perceptual brightness), and two *chromaticity* channels, I and Q (perceptual colour). Klein *et al* use this same colour model for their video mosaic approach, weighting each channel in terms of relative perceptual importance with a ratio of 8:3:1, which are the ratios assigned to these channels in NTSC-broadcast television [19]. A related colour model is YUV, the Betamax standard [45]. The conversion from RGB to YIQ is a linear transform, whereas the conversion from RGB to YUV is an affine transform (i.e. a linear transform plus an offset).

Several commercial image mosaic programs use a weighted RGB calculation, which is beneficial because of its simplicity and speed of calculation, making it a favourable alternative to YUV while achieving visually comparable results [45]. The formula is as follows:

$$\Delta(RGB) = \sqrt{(2 + \frac{\bar{R}}{256})(\Delta R)^2 + 4(\Delta G)^2 + (2 + \frac{255 - \bar{R}}{256})(\Delta B)^2}$$

One limitation of the RGB space is that it does not account for the fact that the human perception of colour is non-linear, and different for each of the three colour values, R, G and B [45]. The coefficients in the formula attempt to account for these variations.

This formula is easy to compute, allows the use of the RGB color space for difference calculations, and produces a metric visually similar to the L*u*v* and YUV colour models [45].

3.2 Brute Force Method

The heart of the image mosaic process is the comparison algorithm. The brute force approach taken by Silvers' Photomosaics$^{\text{TM}}$(as documented in the patent), as well as many commercial image mosaic programs, is quite simple.

Given a Base Image (the image that will serve as the basis for the image mosaic) and a pool of Candidate Images, divide the Base Image up into a grid of tiles. Then, for each tile in the Base Image, loop through the pool of Candidate Images to find the image which most closely matches the tile location in the image, based on some match metric.

The process guarantees the *best* solution is found, based on the Candidate Pool and image comparison metrics used, since the algorithm compares every image in the Candidate Pool to every tile position in the base image. This does not imply that the algorithm will produce a visually-pleasing mosaic, as the quality of the end result is related directly to the quality and colour variation of the Candidate Pool. Generally, the larger the Candidate Pool, the better the resulting mosaic [28].

Similarly, the quality of the final mosaic is also dependant on the tiling dimensions of the base image. The smaller the tile dimensions, the better the resulting image will tend to be, up to the point where the tiles are so small, they will becomes pixels and the resulting “image mosaic” will be an exact duplicate of the original image.

3.3 Enhancements

3.3.1 Pixel Comparisons

There are several enhancements which have been made to the brute force method, resulting in both improvements in speed and quality. The first of these improvements deals with pixel comparison, and is an enhancement which improves the speed of the algorithm by reducing the number of comparisons necessary.

Comparing pixels is the heart of the image mosaic process. When determining how close an image is to another, all the pixels of one image are compared to the corresponding pixels in another image and an overall difference value is computed — the sum of the differences between each of the pixels. The speed of this difference calculation is directly dependent on the number of pixels compared, that is to say, the size of the images.

Comparing a 200x200 image from the Candidate Pool to a particular 200x200 tile on the base image would require 40,000 comparison operations. Keep in mind, this is just for one candidate image compared to one tile. Multiplying this by the total number of candidate images results in a serious amount of processing time.

An image mosaic, by definition, is an approximation of the original base image, which allows a degree of freedom in the pixel comparisons. It is highly unlikely (and undesirable) that a candidate image would exactly replace a given tile section of the original image, i.e. the sum of the differences across all pixels is 0. In this case, there would be little point in creating an image mosaic, since the resulting mosaic would be an exact duplicate of the original image.

Image resolution is not a major constraint, because finding an exact duplicate is not the goal. Finding the candidate image which best matches the base image section, based on colour and shape distribution, is the major constraint. This allows us to reduce the resolution of both the images in the Candidate Pool and the tiles in the base image, before computing the colour difference. This will dramatically increase the speed of the difference calculation by decreasing the number of per image comparison calculations.

Consider the example depicted in Figure 3.1:

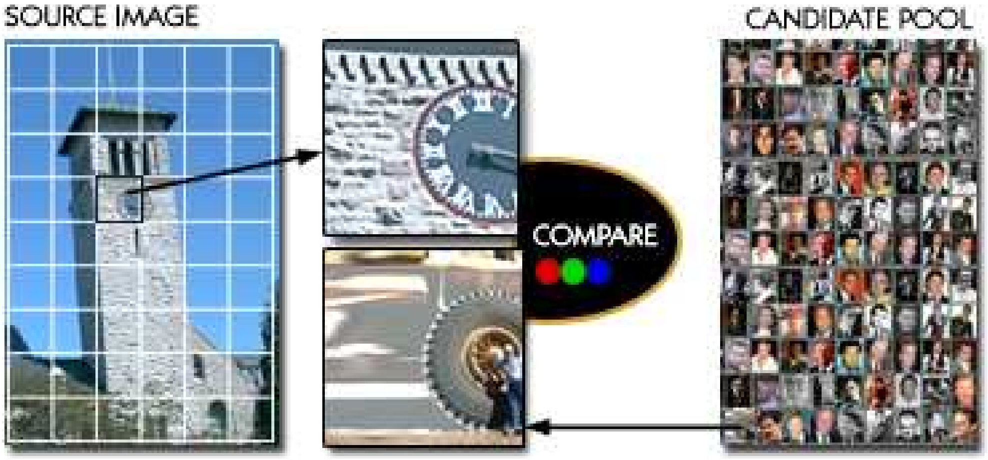


Figure 3.1: Image Comparison Example

In this example, the original base tile resolution is 72x72 pixels, the candidate image size is 72x72 pixels, and the Candidate Pool size is 10,000. Without reducing the images in size before making comparisons, the algorithm would be required to perform 5,184 pixel compare operations for one candidate image comparison. Multiplying this by the number of candidate images, 10,000, results in 5.18x10^7 compare operations — for a single tile in the base image.

Reducing the images in size before comparison will allow for the retention of overall image colour distribution and structure, while drastically reducing the number of compare operations, thus increasing the speed of the overall process. As shown in Figure 3.2, it is possible to reduce the image in size, while still retaining structure and colour characteristics.

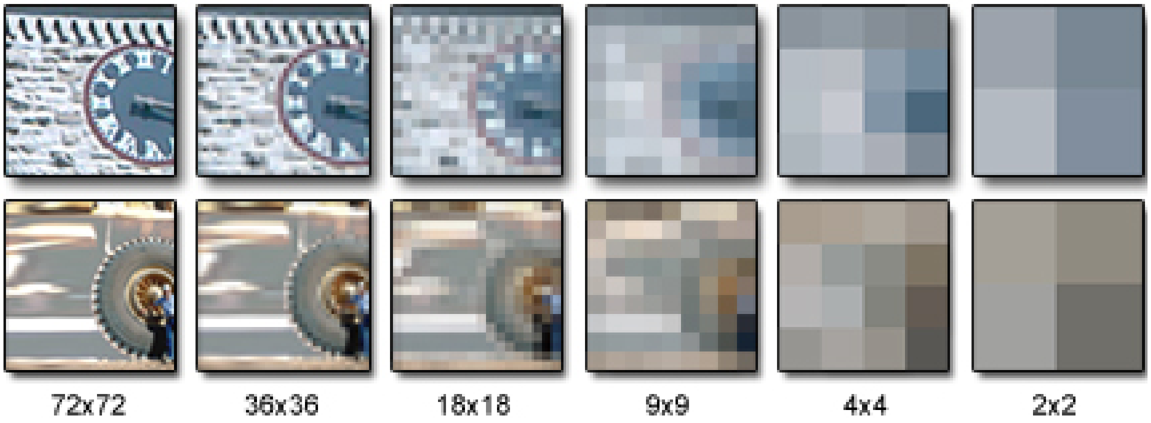


Figure 3.2: Image Reduction Comparisons

The benefit of reducing image size becomes evident. The per-image comparison operations can be reduced from 5,184(72x72), to 1,296(36x36), to 324(18x18), to 81(9x9), to 16(4x4), to as little as 4(2x2). Note however that as the image is reduced, so is the degree of structure and colour apparent in the images. There is a lower limit to reducibility in this example that is dependant on image complexity. In order to produce a visually-pleasing image mosaic, it is desirable to err upwards, at the cost of processing, not downwards, at the cost of quality.

3.3.2 Pre-processing Stage

A second major improvement in performance can be achieved by introducing a pre-processing stage to the overall process. Dealing with images in real time can be costly, as opening the image, scaling and taking pixel readings from each pixel before computing the sum of the differences all take valuable processing time. As every image in the Candidate Pool must be compared to every tile in the base image, the same image files will be opened many times over the course of comparison.

A solution is to introduce a pre-processing stage to the overall process, which opens the images once, scales them accordingly, takes pixel readings and then stores these readings in a data file. Klein *et al* [19] take this approach in their *Video Mosaics* paper, by pre-processing the images to extract information they later use in comparison operations.

3.3.3 Repeated Tile Images

The improvements described above deal predominantly with speed of the algorithm and overall processing time, but there are also improvements which increase the visual quality of the final resulting mosaic. One of the largest qualitative gains comes by solving the problem of repeated tile images.

Repeat images occur when the same candidate image is selected as the best fitting image for more than one tile of the base image. It is problematic when repeated images occur in adjacent tiles and groups of tiles, which is often the case. This occurs frequently in areas of solid colour. Shown in Figure 3.3 is an example of a mosaic with repeated tiles. The result is visually unpleasant, and the groupings of like images becomes distracting. Repeated images, especially in large quantities, can distort the original image by adding visual aspects, such as lines, shapes and colour patterns, which were not originally present in the image.

Figure 3.3: Repeated Images - William Hunt PhotoTile

There are several approaches for dealing with the problem of repeated images. The simplest method is restrict the number of times a candidate image may appear in the final mosaic. By adding a flag for each image in the Candidate Pool, a candidate image can be flagged when it is chosen as a best fit image for a particular tile location in the mosaic. By then checking this flag before making a best fit image selection, images which have already been used can be prevented from being used to fill subsequent tiles.

This method solves the repeated image problem by ensuring that a particular candidate image is used only once per mosaic. However, it is somewhat limiting. In many cases, it is aesthetically pleasing to have repeated images in a mosaic, as it allows for a degree of repetition and familiarity in the end result — it is just not desirable to have large groups of repeated images.

A second solution is to introduce a proximity value, in tiles, which is the minimal allowable distance between like images. When images are being processed by the selection algorithm, this proximity value can be taken into account. If the best fit image as chosen by the algorithm falls outside the proximity range of all other like images, it is deemed to meet the proximity constraint, and the image is placed. However, if the image falls within the proximity range, the image will be discarded and the next best fitting image will be tested.

This approach does quite well at solving the problem, but in some areas with large areas of uniform colour, it introduces a visually undesirable banding pattern due to images placed according to the fixed proximity value. The result of this approach, as produced by Jimage Mosaic [39], can be seen in Figure 3.4 (a).

A third solution, which eliminates this banding pattern, introduces a random variation to the proximity value. When the algorithm selects an image as the best fit image for a particular tile, the algorithm checks to see if the image falls outside the proximity range *plus a randomized fluctuation value.* If so, the image is deemed to meet the proximity constraint, and the image is placed. As before, if the image falls inside the range, the image is discarded and the next best fitting image will be tested.

A fourth solution, which also solves the banding generated by introducing a fixed proximity value, is to process the tiles in random order [39] instead of sequentially. Figure 3.4 demonstrates the results of this approach, (a) before, and (b) after randomization.

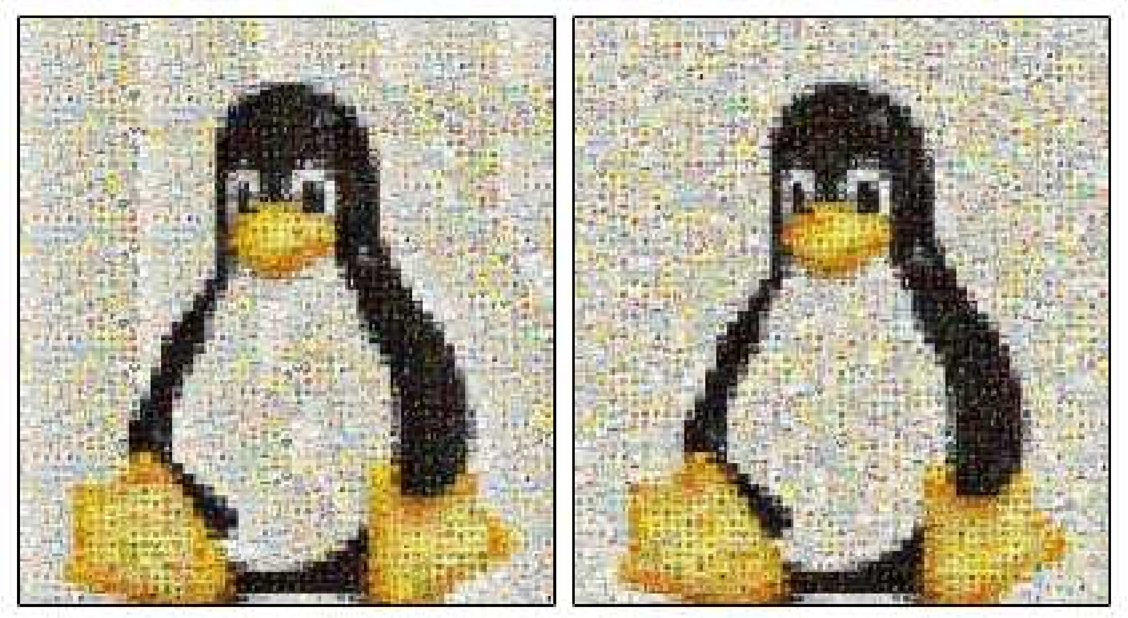


Figure 3.4: Randomized Proximity - (a) Before, (b) After - produced by Jimage [39]

3.3.4 Colour Correction

More an attempt to improve the final result of the mosaic process as opposed to an improvement in the process itself, colour correction modifies the resultant mosaic in order to make it more closely resemble the original base image. Purists will regard this as “cheating.” The motivations behind attempts at colour correction are understandable, and several modifications are possible: changes in colour, changes in brightness, changes in contrast and emphasis of edges. In many cases, subtle changes in colour or brightness can make the end product more visually pleasing. However, some attempts simply go to far, nearing the point where the entire image selection process is negated because the colour correction changes are so drastic.

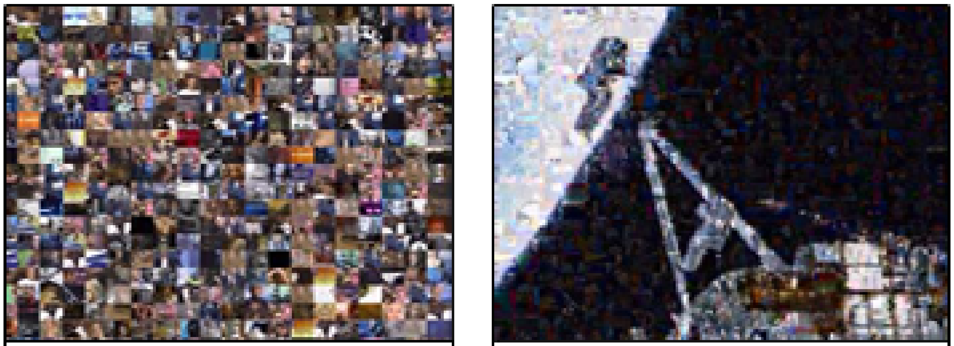
Take, for example, the colour correction as performed by Klein *et al* [19] as part of their *Video Mosaic* process. The correction routine changes each pixel of the resultant mosaic so that it more closely matches the original base image, while still apparently maintaining the integrity of the source frames. By modifying the colour distribution and brightness of the each source frame, a closer approximation of the target frame can be achieved, but retaining the integrity of the source frame itself depends on the amount of color correction used.

The merit of this process is questionable. For instance, as can be seen in Figure 3.5, the result of the colour correction of a mosaic of random noise tiles can produce more visually pleasing resultant mosaic than does the result of colour correcting a mosaic of tiles chosen by their algorithm. Although no indication is given as to the *amount* of colour correction applied in each of these two cases, the fact is that colour correction of random tiles produces a more visually-pleasing resultant image. This calls into question the point of tile selection altogether. Why not just use the color correction process to create video mosaics using whatever videos you prefer as tiles?

Additionally, comparing the results of colour correction on similar tile images within a single corrected frame shows that the correction process is so heavy-handed, that it introduces visual features into the candidate images that were not originally present. Compare the colourful variations of the guitar player shown in Figure 3.6. Each of these images was taken from the frame of video shown on the left. The results of the colour correction are apparent. The particular image of the guitar player has been "corrected" in many instances to resemble the original image in a particular location. The corrections of this image range from red to blue and from light to dark, depending on what the situation calls for. With all these variations, one wonders what the original color of the tile was.

Original Mosaic

Color Corrected Mosaic



Random Image

Color Corrected Random Image

Figure 3.5: Colour Correction of Random Noise

This is not to say that the resulting video mosaic, in its colour corrected state, is visually undesirable — the resulting mosaic is well done and quite striking. However, the heavy-handed correction method does border on cheating, as the correction process introduces characteristics into the original images that were not there before "correction." With so much correction applied to the images, and the correction of random noise patterns producing better visual results, why bother with the initial image comparison and best-fit selection process?

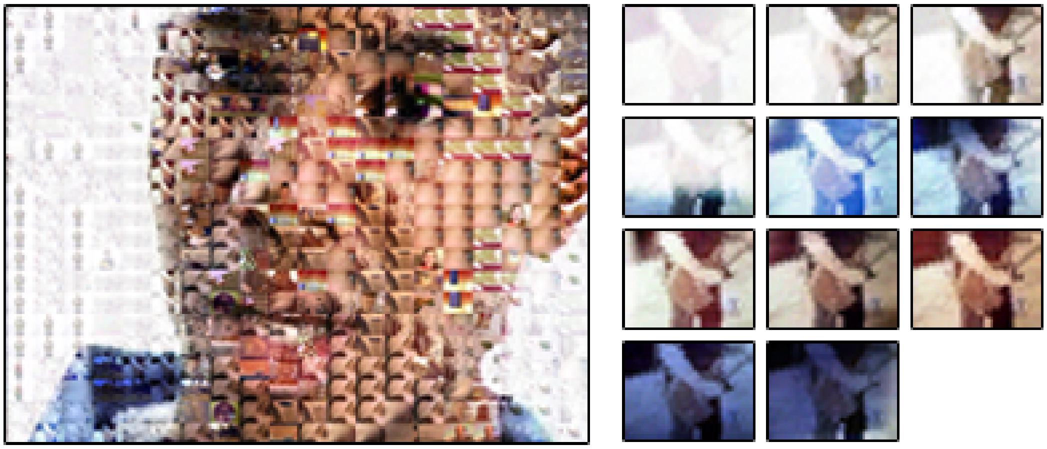


Figure 3.6: Colour Correction on Results on Similar Tiles

The major problem with using color correction in the temporal image mosaic environment, is that although color correction improves the overall result, it does so at the expense of the individual tile sequences. One of the most important characteristics of the image mosaic is the duality between the smaller images and the larger image which they serve to represent. By using color correction to improve the overall representation (the big picture) the individual tiles themselves can suffer to the point of being so "corrected" that they are no longer recognizable, and thus cannot be seen and appreciated by the viewer. This effectively defeats the purpose of the mosaic, and can diminish its overall multi-layered impact.

In some cases, colour correction is quite acceptable. In the Image Mosaics of Finkelstein and Range [9], colour correction is part of their halftone mosaic technique and not simply an altering of the final result. The halftone mosaics from the previous chapter (depicted in Figure 2.31) use identical grey-scale images, and modify them in terms of brightness to achieve a halftone-like effect. From this process came the colour correction shift and scale method. In the case of the halftone mosaics, the colour correction is intrinsic to the process.

3.3.5 Additional Attributes

Another qualitative improvement in the image mosaic process is to account for more image characteristics then simply colour. Depending on what colour model is used during the matching process, images attributes such as brightness, hue and saturation can be taken into account. In fact, it is possible to account for a variety of these attributes during the process.

Some approaches have taken into account only brightness, as the human eye is most sensitive to this attribute [1, p. 348]. This works fine in black and white, since brightness is the measure of graduation from black to white, but in colour using this metric alone produces a unique outcome — visually coherent, but devoid of colour preservation of the original image, as shown in Figure 3.7.

An approach by Yue Zhang [34] takes into account several image attributes by using Content-Based Image Retrieval (CBIR). The method treats the image comparison and tile selection process involved in image mosaic creation as a CBIR problem. By leveraging current CBIR methodologies, such as Query by Example as used by IBM's QBIC system [30] which selects matching images based on similarities to an image presented by the user, a system was constructed which selects images from a CBIR database based on such features as color, shape, color histogram, and texture. Figure 3.8 depicts an image mosaic created using the CBIR method. It should be noted that this approach necessitates pre-processing the candidate images which form the database to extract the image attributes on which queries will be made.

Figure 3.7: Image Mosaic using Brightness

3.4 Wavelet Approach

A relatively new approach to generating image mosaics is the technique of wavelet decomposition. Used by Finkelstein and Range in their image mosaic paper, and as a modified variation in Klein's Video Mosaics, the wavelet method breaks an image down into a very small amount of data called a signature [9]. This signature, depending on its encoding, can then be used to represent broad forms and colours of the image — both statically and through time.

The wavelet technique for image mosaics came out of the work by Jacobs *et al* [14] in content-based image retrieval. Using this technique, it was possible to match images on such features as colour histograms, texture analysis, shape analysis, edge matching or even a combination of these methods [9]. In the case of the Finkelstein mosaics, the

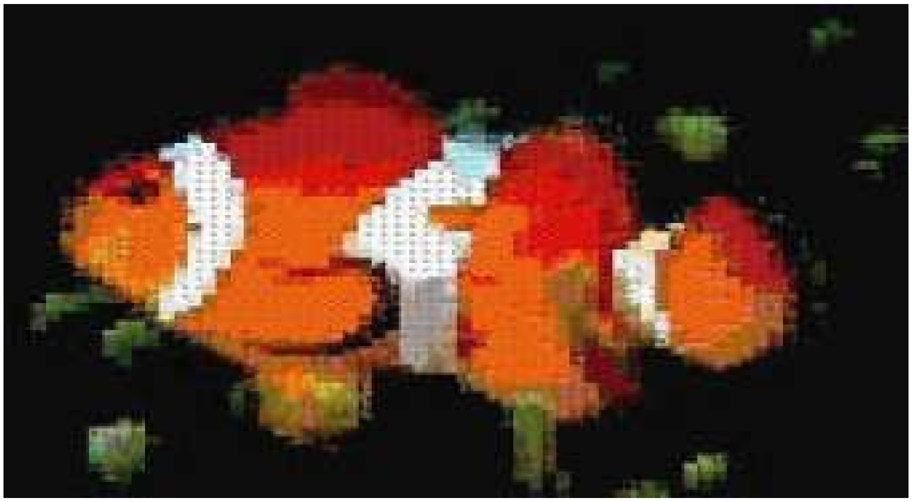


Figure 3.8: CBIR Image Mosaic

best matching candidate image from a particular portion of the base image can quickly be found by comparing the signature of the base image to those in the Candidate Pool, by selecting the closest matching signature. According to Finkelstein and Range, the best matching image for a given base tile can be determined from a Candidate Pool of 20,000 images in less than a second, using a conventional desktop computer (circa 1998).

The resolution reduction approach described in Section 3.3.1, which reduces a 72x72 pixel tile to, say 4x4, is equivalent to the first level of a wavelet decomposition using the Haar basis. In the Video Mosaic approach by Klein *et al*, they eschew the brute force method using pixel-wise comparison as being too costly, and the brute force method using image reduction as having a drawback of not preserving small but important image details [19].

Their chosen approach works by performing a standard two-dimensional Harr wavelet decomposition on each colour channel for every frame of video in their Candidate Pool. Then for each channel, they store the 30 coefficients with largest magnitude in a search-optimized data structure. Pre-processing the Candidate Pool allows this information to be stored on disk. Searching for the best-matching image sequence then becomes simply a matter of comparing the number of significant wavelet coefficients that the base tile shares with a potential candidate image.

3.5 Video Mosaics

As discussed in the previous chapter, the Video Mosaic paper by Klein *et al* appeared during the implementation of the TIM. Although the desired end result is the same, and the implementations similar, there are several major differences in both the end result and overall approach between the TIM and the Video Mosaic.

In terms of implementation similarity, consider how the Video Mosaic is created. The Video Mosaic approach uses wavelet decomposition to break down the video sequences and represents them in the form of coefficients in a search-optimized data structure. Given this structure, a Video Mosaic is built by using a specialization of dynamic programming to find the best-fit of the source to the target video, for each tile. A matrix of size s x t — where s is the number of frames in the source video, and t is the number of frames in the target video — is employed to facilitate what amounts to a breadth-first search of the data. This search is functionally equivalent to the shortest path selection process used to create the TIM, as described later in Section 4.2.3.

There are three major differences between the TIM process and the Video Mosaic process. The first major difference is in repeated tile images. The TIM approach addresses this problem and its visually-unpleasant side effects. The Video Mosaic does not concern itself with repeated tile images. The likely reason why they do not account for repeated images is that their color correction process can change like images to the point where they are no longer simply copies of one another, but effectively different from the point of view of the mosaic. Again, consider Figure 3.6.

The second difference between the TIM and Video Mosaic deals with one of the most important image mosaic considerations, the candidate pool. The use of color correction allows the Video Mosaic approach to effectively side-step the requirement of a large, well-distributed candidate pool, since (as depicted in Figure 3.5) the color correction process is able to correct even a block of random tiles to the point where it becomes visually pleasing.

As discussed earlier in this chapter, color correction is a technique used to make the selected source sequences more closely match the original base sequence in terms of color and brightness. The Video Mosaic approach relies heavily upon color correction to achieve visually-pleasing results, whereas the TIM approach does not use any kind of color correction. In fact, by using color correction, it is likely that the Video Mosaic approach is able to side-step the problem of repeated tile images. Consider again Figure 3.6 and see how the color correction approach has changed various instances of the same image — the guitar player — to the point that they can no longer be considered “similar images”, since some are now light, some are dark, blue, red, and so on.

The TIM approach does not side-step any of the issues with the use of color correction. Solutions were implemented to deal with repeated tile images in a temporal environment, as well as efficiently dealing with a large candidate pool. As a result, the TIM process retains the characteristics of a true image mosaic which can be appreciated from the both levels — the individual tiles and the overall presentation — untouched and un-"corrected."

Chapter 4

The Temporal Image Mosaic

The Temporal Image Mosaic (TIM) is an extension of the static two-dimensional image mosaic, with the added dimension of time. The TIM is a temporal image – a moving picture – composed of a grid of tiles. Each of these tiles is itself a moving image. Seen from a distance, these individual tiles compose the overall subject of the TIM.

4.1 Goals

The three major goals of this thesis are as follows:

1. To produce an intensive survey of the mosaic and the mosaic arts (done in Chapter 2).
2. To create a temporal image mosaic by extending the static image mosaic (described in Section 4.2 and 4.3).

3. To use the temporal image mosaic in an artistic setting, by demonstrating one possible use of the effect in a short film (described in Section 4.4).

The technical goals and constraints of the TIM are:

- The colour sampling and resolution used must produce coherent, visually pleasing results. Because of the subjective nature of this goal, the success of a "visually pleasing" result will be measured on the degree to which rapid cuts can be eliminated, and on how closely images can be matched.
- Colour matching must be perceptually uniform.
- The image matching process must find the best match for every tile in the mosaic.
- Matching tile sequences must be maximized in length, while at the same time, minimized in colour difference.
- Processing optimizations must be used to decrease the enormous amount of processing required to generate the TIM.
- A large Candidate Pool must be established from which to draw the matches, as image mosaic papers indicate the pool is a critical component.

The artistic sub-goals and constraints are as follows:

- The effect must be utilized in an artistic setting in a cinematic manner.
- The effect must support the narrative of the film, respective of its own grammatical implications.

- The technical process must preserve the original colour palette, native resolution and frame rate as closely as possible.
- The images should remain uncompressed throughout the process as not to degrade the final output.

With these goals and constraints in mind, the next section will give a general overview of the process of creating the temporal image mosaic, and will detail the three main components of the TIM: the Pre-Processing Stage, the Sliding Window Comparison Stage and the Composition Stage.

The final sections of this chapter will describe optimizations and improvements to the technique as a whole, will detail the film production element of the thesis, and will cover the usage of the TIM in an artistic context.

4.2 Process Overview

The Temporal Image Mosaic process takes a *Base Sequence*, the sequence of frames which will be approximated by the TIM, and divides this sequence up into a specified number of *Base Tile Sequences*, each corresponding to one tile on the screen. From this point, the process loops through each Base Tile Sequence in the Base Sequence, comparing each one to every *Candidate Sequence* in the *Candidate Pool*. The comparison process looks for Candidate Sequences that best match the Base Tile Sequence. The relationship is depicted in Figure 4.1, and for the remainder of this chapter, we will refer to this single five-frame Base Tile Sequence for all examples.

There are three components to the implementation process: the *Pre-Processing*

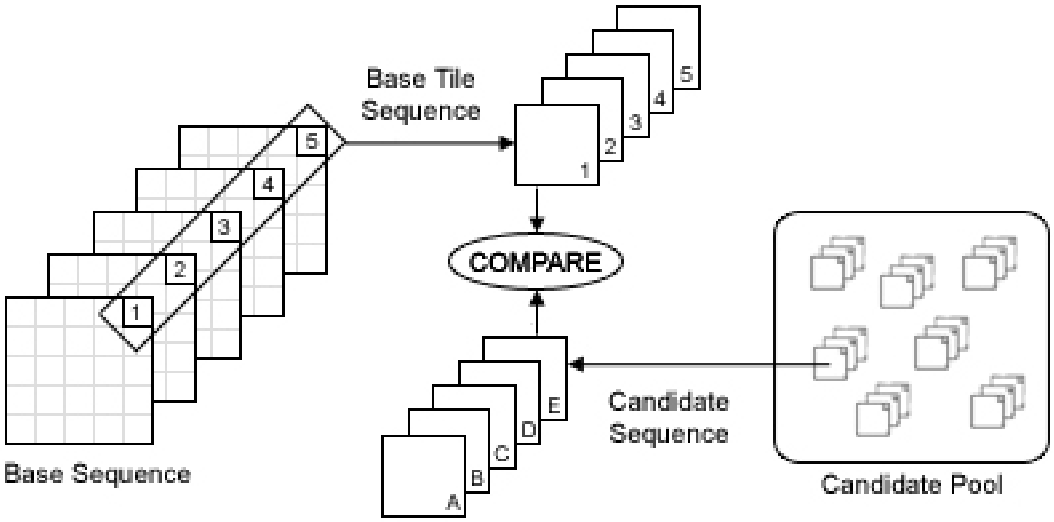


Figure 4.1: Process Relationship

Stage, the *Image Matching Stage*, and the *Composition Stage*. During the Pre-Processing Stage, color information is extracted from the Base Sequence and the Candidate Pool on a frame-by-frame basis. During the Image Matching Stage, every Base Tile Sequence is compared to the Candidate Pool, recording the best-matching Candidate Sequences for each Base Tile Sequence. In the Composition Stage, the best-matching Candidate Sequences are selected from the resultant matches to compose the TIM.

4.2.1 The Pre-Processing Stage

In order to reduce costly file operation and to make the image comparison routine less intensive, a pre-processing stage (as described in Chapter 3) is introduced which samples the colour information for every frame of every Candidate Sequence in the Candidate Pool, and stores the colour information in a file. The same pre-processing is done for each Base Tile Sequence.

For the Candidate Sequences, the pre-processing routine reduces each frame down to 8x6 pixels in resolution. Three values, R, G, and B are then sampled for each of the 48 pixels in the reduced image. These values are written to a file, along with the file information about the particular frame. The same is done for the Base Tile Sequences, except that each Base Tile Sequence is sampled at 8x6 (instead of the entire image). This results in a 1-to-1 relationship between the Candidate Sequences and the Base Tile Sequences in terms of resolution, and will allow for direct pixel-to-pixel comparison without the need for any further scaling at a later time.

Once the pre-processing stage is complete, not another image need be opened until the final composition stage of the process, since all necessary colour information is contained in the files which resulted from the pre-processing. These files can be extremely large depending on the length of the Base Sequence and Candidate Pool. In the case of a Candidate Pool file containing 20 hours of Candidate Sequences (2.16 million frames, 56 GB in raw AVI format), the resulting file size was approaching the Windows file size limit of 4 GB. These files now serve as the basis for the Image Matching Stage, which is modelled after a sliding window approach.

4.2.2 The Image Matching Stage

With all the colour information from the reduced images now contained in files, direct comparison of each frame can be done by summing the differences between each of these 48 RGB values. Using the weighted RGB color difference metric as described in Chapter 3, the RGB differences are used to give a single value which represents the colour difference between two frames (that is, the difference between a Candidate

Sequence frame and a Base Tile Sequence frame).

Thus:

$$\Delta(Frame) = \sum_{i=1}^{48} \sqrt{(2 + \frac{\bar{R}_i}{256})(\Delta R_i)^2 + 4(\Delta G_i)^2 + (2 + \frac{255 - \bar{R}_i}{256})(\Delta B_i)^2}$$

Keep in mind that this calculation provides the difference between two static frames, and since a primary goal of this thesis is to extend the image mosaic into the dimension of time, we must compare not just frames of sequences, but whole sequences themselves. This is what the sliding window process accomplishes.

Sliding window matching is a technique used to solve such problems as genetic sequence and sub-string character matching. Consider a sub-string character matching example. Given a candidate character string of ‘abcdefg’ and a base string of ‘cd’, find all occurrences of the base string in the candidate.

The sliding window solution to this problem involves taking the base string ‘cd’ and sliding it along the candidate string ‘abcdefg’ from beginning to end, one character position at a time. At each position, the two strings are compared, and matches are noted as they occur. The process ends when the base string has been completely slid from the first position of the candidate, to the last. Figure 4.2 provides an example.

The end result of the TIM sliding window process is a list of all the matching sequences between the Base Tile Sequence and the Candidate Pool. Matches can occur on single frames, multiple frames, or across the entire Base Tile Sequence. Matches can start and stop at any frame in the Base Tile Sequence and Candidate Pool. This can be seen in the example in Figure 4.3, where the Base Sequence is five

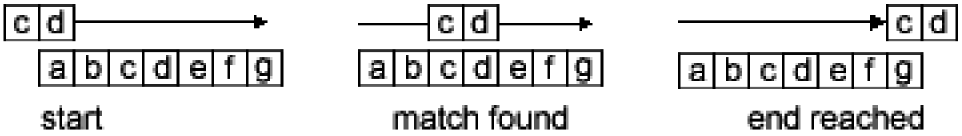


Figure 4.2: Sliding Window String Matching Example

frames long.

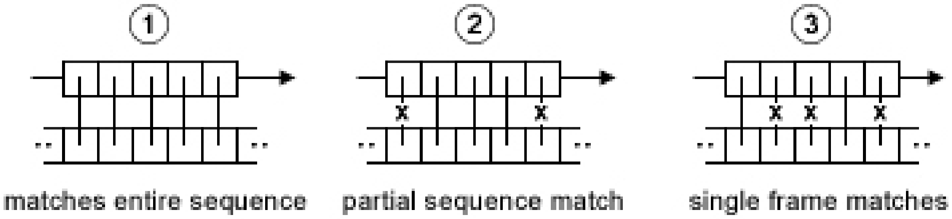


Figure 4.3: Frame Matching example

Example (1) shows a complete match between the Base Tile Sequence and a section of the Candidate Pool. Example (2) shows a three-frame partial match, which starts at the second frame of the Base Sequence and ends at the fourth frame. In example (3), there are two single-frame matches — one at the first frame and one at the fourth.

In each of these match cases, the information written to disk is the name of the Base Tile Frame where the matching sequence beings, the corresponding name of the Candidate Pool frame, the length of the matching sequence and the colour difference between the frames of the segments. This information is all that is required to later rank the matches in terms of length and best match.

For the TIM, we wish to find Candidate Sequences (and subsequences) which most closely approximate portions or all of the Base Tile Sequence. To accomplish this using the sliding window process, the Candidate Pool is considered as one large, contiguous image stream — each Candidate Sequence placed back-to-back. The Base Tile Sequence is *slid* along this Candidate Pool window one frame-position at a time. At each frame increment, the frame-by-frame colour difference between the two sequences are calculated.

At a particular position of the sliding window, match sequences are formed by accumulation. A match sequence begins at a position which the color difference between the two frames is below the matching threshold. The sequence is accumulated contiguously, one frame at a time, until a subsequent color difference exceeds the match threshold, at which point the sequence is terminated, and the details are written to file. The value written to file which describes the color difference for a particular sequence is the average difference per frame.

The process then continues, starting a new match sequence at the next position that satisfies the matching threshold, and accumulating frames one at a time until the sequence must be terminated. In this manner, multiple match sequence are possible for a single sliding window position. Figure 4.4 depicts the sliding window process.

In this example, the Base Tile Sequence is five frames long and the Candidate Pool is twelve frames long. At position (A), the sliding window process begins by comparing the last frame of the Base Tile Sequence to the first frame of the Candidate Pool Sequence. If a match is found, the match information (frame number, match length, Base Tile number, and difference value) is written to a file. Position (B), shows the sliding window matching process after seven iterations. Position (C) shows the process with only two iterations remaining. The entire process ends when the

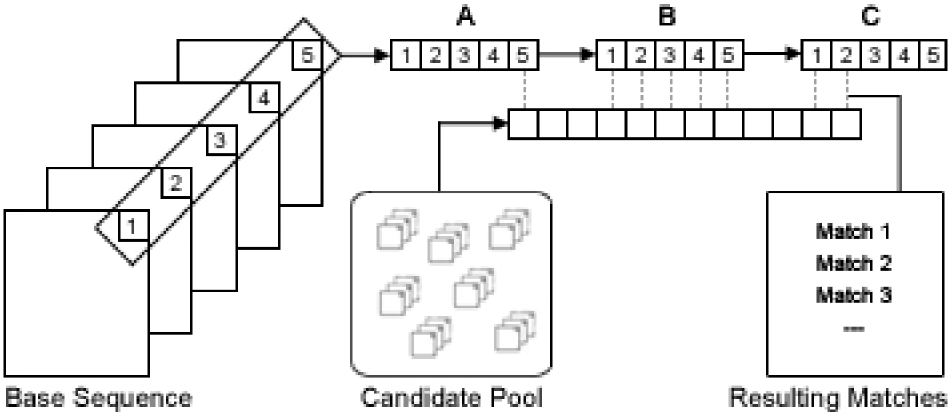


Figure 4.4: Sliding Window Process

first frame of the Base Tile Sequence is compared to the last frame of the Candidate Pool Sequence.

In general, the sliding window process is very time-consuming. A Base Sequence tiled at 34x34 requires the process to be done 1156 times — once for each Base Tile Sequence. Every frame of every Base Tile Sequence is compared to every frame of every Candidate Sequence in the Candidate Pool. On a relatively fast processor, the processing of a modest Candidate Pool containing 30.4 hours (2.16 million frames) of Candidate Sequences against a 6-second Base Sequence divided up into 1156 tiles, takes 18 days (440.7 hours) to complete.

In terms of running time, the algorithm is O(c b t), where c is the number of Candidate Frames, b is the number of Base Sequence Frames and t is the number of tiles in the Base Sequence.

4.2.3 The Composition Stage

At this stage of the overall TIM process, we have a list of all Candidate Sequences which match portions of the Base Tile Sequence. For each match, we have a match length in frames and an associated difference value, Δ between the two matching segments. What remains to be done is to select the "best" Candidate Sequences from this match list to compose the TIM. Candidate Sequences can be selected based on length, lowest difference value or a combination of the two.

To compose the TIM, it is necessary to find a sequence of Candidate frames that span from the initial Base frame to the final Base Frame. Figure 4.5 depicts an example of this situation. In this simple example, there are two Candidate Sequences, with matching values of W_1 and W_2. Starting at Frame 1, we wish to select Candidate Sequences so that we reach Frame 5:

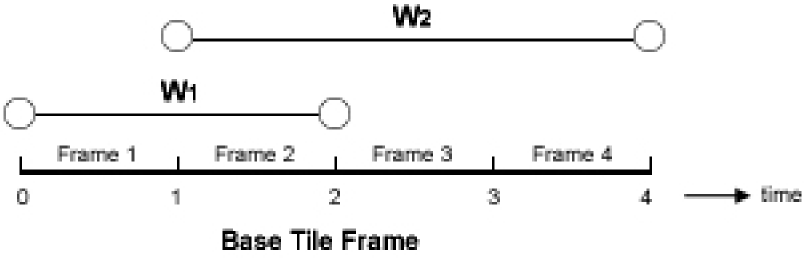


Figure 4.5: Match Sequences

This situation can be best represented as a Directed Acyclic Graph (DAG) by treating each Candidate Sequence as an edges of the graph and their start and end frames as vertices. The graph is built using the match list, which can be thought of as a simple list of graph edges. The resulting DAG contains all available Candidate Sequences, represented as edges, with a distinct start and end point. The corresponding difference value becomes a weight on the edge.

Figure 4.6 depicts a DAG constructed from the two Candidate Sequences in the previous figure:

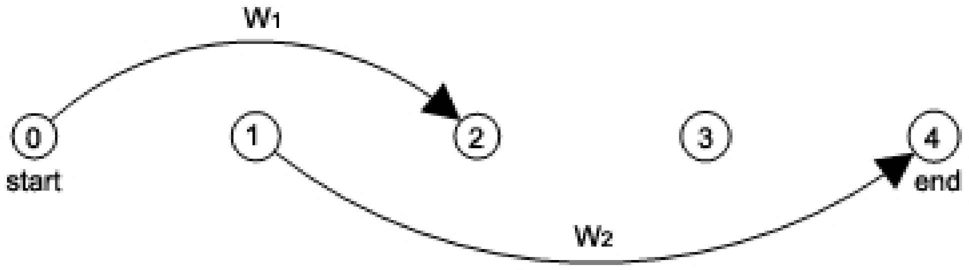


Figure 4.6: Candidate Sequence DAG

With the DAG constructed, the shortest path algorithm for directed acyclic graphs can be used to find the best path through the graph which spans from time 0 to time N of the Base Tile Sequence. Before this process is run, however, there are two additional considerations which must be taken into account.

The first consideration is *edge stranding.* Situations arise during the shortest path process when an edge ends at a particular vertex, where there are no outgoing edges. In this circumstance, although many edges may pass over the vertex, the process would be stranded at this edge, and have to end prematurely without producing a complete path from start to end. To solve this problem, we allow the algorithm to chose from any other edge which happens to pass over this stranded edge. This is achieved by breaking each edge up into *prefixes* and *suffixes* of itself, allowing an edge that starts, for example, at vertex one and ends at vertex five, to also stop and start at every vertex between one and five. A prefix of an edge is an edge which starts at the same start vertex as the original edge and ends before, and a suffix is an edge which ends at the same end edge as the original edge and starts after.

Figure 4.7 depicts the original edges (W_1 and W_2), the prefixes (W_{1a},W_{2a},W_{2b}) and suffixes (W_{1b},W_{2c},W_{2d}) of each edge.

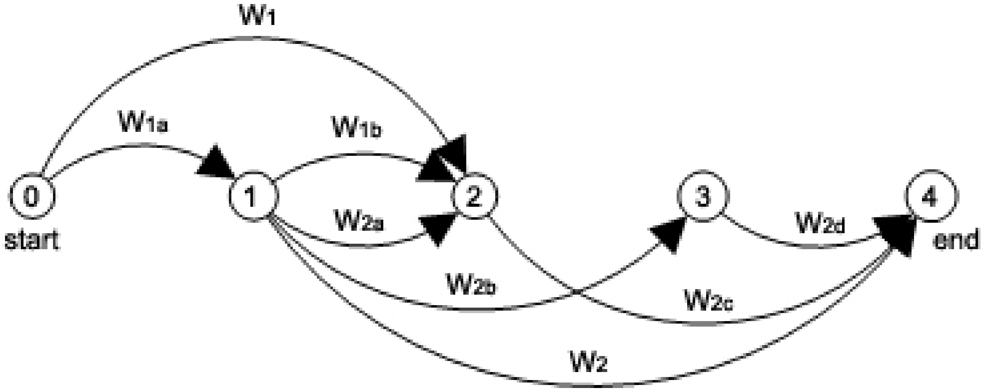
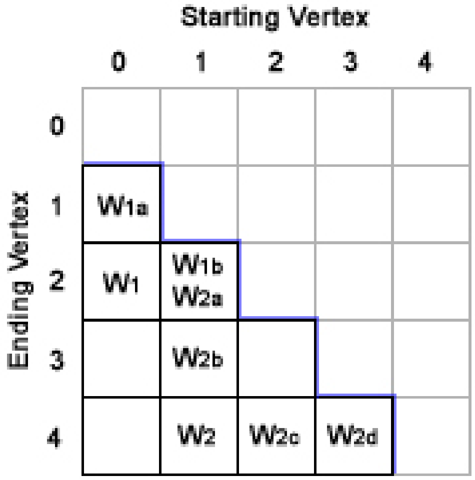


Figure 4.7: Prefixes and Suffixes

The second consideration before the shortest path routine can be used on the DAG is the issue of graph connectivity. In order to ensure that the graph is connected, thus guaranteeing that a shortest path solution can be found, we must add single frame *safety* connections between each pair of adjacent vertices in the graph. Even with a very large Candidate Pool, cases can occur where — due to complex tiles, unique colour variations or strange patterning — no matching Candidate Sequences can be found. In these cases, the single frame safety edges ensure that a shortest path solution can be found in an otherwise unconnected situation. These edges are given exorbitantly large weights as to ensure that they are only chosen as a path of last resort by the shortest path algorithm.

The actual implementation of the DAG was accomplished using an array, where any non-empty array element represents the weight of an edge between vertices in the graph denoted by the column and row indices of the element. The array depicted in Figure 4.8 demonstrates the representation of the graph in the previous figure.



Ending Vertex \ Starting Vertex	0	1	2	3	4
0					
1	W_{1a}				
2	W_1	W_{1b} W_{2a}			
3		W_{2b}			
4		W_2	W_{2c}	W_{2d}	

Figure 4.8: Array implementation

The value of W_2 in [row 4, column 1] represents a directed edge which travels from vertex 1 to vertex 4, with a corresponding weight of W_2 (i.e. a sequence of frames from time 1 to time 4). The shortest path algorithm takes this array as its input, and returns the lowest-cost path from vertex 0 to vertex n, minimizing the sum of the cost (array elements) on the overall path.

With regard to the weighting of the edges during the shortest path selection process, there will be a distinct aesthetic difference in the resultant TIM between edges selected by longest length versus edges selected by lowest colour difference. It is expected that the longest length segments will have a higher overall cost, while the shortest segments, down to single-frame matches, will have lower cost. Aesthetically, the shorter the segments, the more discontinuous and frenetic the resulting TIM is expected to be. At the other end of the spectrum, larger segments, up to the entire length of a Base Tile Sequence, are expected to be less representative of underlying colour, and possibly less desirable overall.

Because of these opposing trade-offs, calculating a parameterized weight for each edge will allow us to account for the contribution of each, length and cost, while at the same time giving a single value with which to weight edges for the shortest path algorithm. It allows for the maximization of length at the same time as minimization of cost. Given the colour difference, ΔC, the length of a particular edge, L, and a range for each with which to normalize their values — CMax and LMax — an overall parameterized cost function can be calculated for a given parameter, p:

$$cost = p \times (\frac{\Delta C}{CMax} \times L) + (1-p) \times ((\frac{LMax - L}{LMax})^N \times L)$$

When producing the TIM, the value of p (0. . .1) can be adjusted, allowing the user to experiment with the aesthetic variations produced by favouring either colour difference or length.

Note another facet of the parametrization, the exponent, N, on the normalized length calculation. Raising the normalized length to a power allows for non-uniform weighting of candidate edge lengths. This non-uniform weighting allows for greater distinction between short and long Candidate Sequences. The effects of different values of N can be seen in Figure 4.9, for a given normalized length L. Increasing the value of N allows longer edges to be favoured (given a lower weight) non-uniformly, and similarly, shorter edges are given a higher normalized weight — but not in a strictly linear fashion, as in the case of N=1.

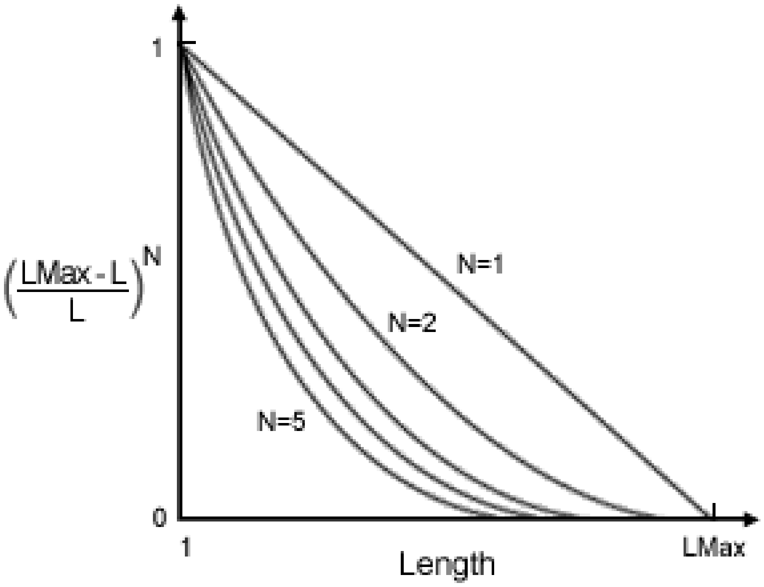


Figure 4.9: Normalized Length Weighting

With this completed, the edge cost calculation, the formation of the edge prefixes and suffixes, and the safety links added to ensure a strong connectivity, the shortest path process begins. The process is repeated for each Base Tile Sequence and the resulting shortest path routes are written to a file.

This resultant file is all that is then required to composite the final frames of the TIM. The file is read and each required frame from the Candidate Pool is composited into the required position in the TIM frames. These frames are then compiled into a video file, and the temporal image mosaic process is complete.

4.3 Mosaic Enhancements

At this stage, all of the initial technical goals, sub-goals and constraints have been met and accomplished. Every Base Tile Sequence frame has been compared to the Candidate Pool frames, and the best matching Candidate Sequences were chosen, based on some combination of colour difference and length, to produce the TIM. The colour metric gives results closely matched to human colour perception.

Colour was preserved and remained unchanged throughout the process, as to provide a resultant TIM as closely matched as possible to the original Base Sequence. Because of the immensity of the Candidate Pool, a method that takes time linear in the size of the Candidate Pool was used to perform the matching in the form of a sliding-window process. And, finally, the parametrization of the colour difference and length of the matching sequences allows us to trade off maximizing the length of the Candidate Sequence versus minimizing the colour difference.

Several improvements were made to the TIM process to improve and enhance the aesthetic result. A large aesthetic improvement was accomplished by increasing the number of Candidate Sequences in the Candidate Pool. As discussed in earlier chapters, many of the papers dealing with image mosaic creating indicate that the single most important factor required to produce a good image mosaic is a large, well-distributed Candidate Pool. The initial size of the pool was 86 minutes, which was increased to 1838 minutes for subsequent runs.

This increase in the size of the Candidate Pool resulted in a dramatic increase in required processing time and disk space necessary to deal with the image files, the intermediate data files, and the results themselves. Changes were made to the overall process to optimize the end-to-end processing time required, from the preprocessing stage all the way to the shortest path selection process. One change included converting intermediate data files to binary format, into which the data was then packed as tightly as possible.

A second change was done to the sliding-window comparison process. With the Candidate Pool increased 22x in size, the sliding-window process became extremely time-consuming, taking five weeks to complete processing. By removing the square root operation from the colour-difference calculation and keeping as much data in memory as possible to reduce file read operations, the overall processing time required for the sliding-window process was cut down by almost three weeks, to just under two weeks to run the process.

4.3.1 Repeated Candidate Sequences

The most effective aesthetic improvement to the TIM process deals with repeated Candidate Sequences. As covered in Chapter 3, repeated image situations can occur when identical images are chosen as best-fit Candidates for more than one tile in the mosaic. It becomes aesthetically undesirable when these identical tiles appear in close proximity to one another, or in large groupings of the same image as may occur in areas of uniform colour. This issue of repeated images in close proximity to one another was evident in the first TIM result.

The problem is solved in generally the same manner as in the static image mosaic, but had to be tailored to function in the temporal environment. The shortest path selection process was modified to allow it to keep track of up to ten alternative Candidate Sequence choices for a particular graph edge. When the shortest path process is selecting a particular Candidate Sequence, the algorithm also takes into account whether or not this image sequence has been used at this time in any other tiles within a certain *search radius*. Since tiles are processed top-to-bottom and left-to-right, tiles are only compared which are above and to the left of the current tile, as shown in Figure 4.10.

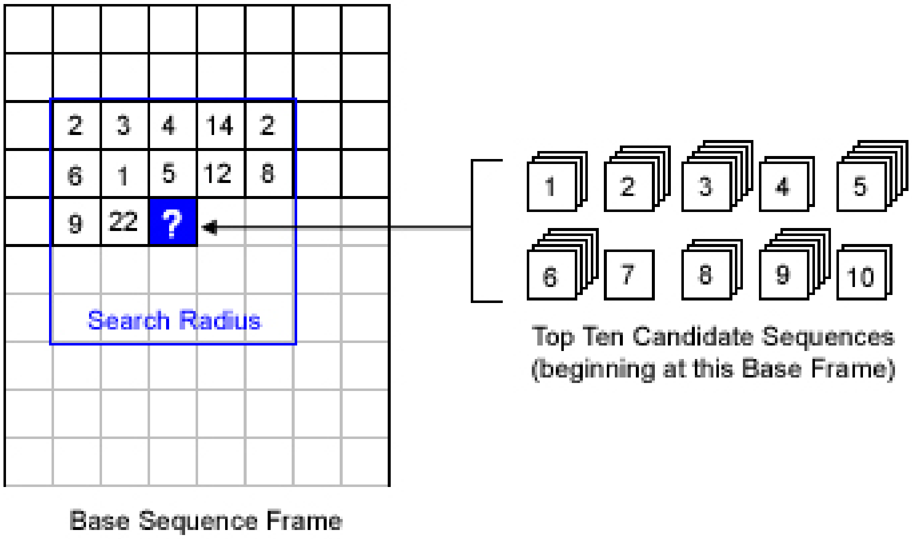


Figure 4.10: Repeated Candidate Sequences

In this case, the only viable Candidate Sequence to choose is Sequence 10, since all other Candidate Sequences have been used within the search radius. In the event that all Candidate Sequences have been used and no selection is possible, the selection process defaults to the best matching choice.

The search radius is a randomized value, similar to that used in the static image mosaic selections process as described in Chapter 3. In the case of the Temporal Image Mosaic, modifications to the static approach search radius allow the process to account for temporal proximity as well. The search radius as used to create the

TIM is calculated using a base radius of 4, with a randomized deviation of ±3.

The process accounts for temporal proximity by requiring that Candidate Sequences differ by more than a specified number of frames. This prevents Candidate Sequences from being selected as acceptable, when in fact only differing from a similar Candidate Sequence by a few frames. The algorithm uses a temporal proximity value of 60 frames (corresponding to two seconds of video).

4.4 Film Production

In addition to the technical goals of this thesis are a set of artistic goals relating to the use of the temporal image mosaic effect in an artistic setting. Under the guidance of Clarke MacKey, co-supervisor from the Department of Film, the process of writing and producing a short narrative film was undertaken. With the grammatical implications of the TIM in mind, several film ideas were developed which would incorporate the TIM, while at the same time, satisfying the initial sub-goals of using the TIM to support the narrative of the film. After several treatments had been written, one idea was chosen and work proceeded to bring the film to life.

The idea, entitled *Sheep*, is set in a metaphorical society comprised of identical drone workers. The daily lives of these drones are controlled by the clock, and each goes about his or her daily life irrespective of others. Some workers in this society have socially productive jobs, while others have socially counter-productive jobs which harm society as a whole. All the while, these workers go about their respective jobs, unaware and unconcerned with what is happening around them.

The story itself centers around one particular worker in this society who occupies the socially counter-productive job of a Clock Tower Sniper. On the day the film depicts, he comes to understand his predicament for the first time, and is forever changed by his insight.

A script for the film was written, proceeding through five revisions. At this point, location scouting began around the general Kingston and Queen's campus area. Once locations were chosen and secured for use, a shot-by-shot storyboard of the entire script was drawn, revised and finalized, resulting in the storyboard which was used during production. The final *Sheep* script and storyboards can be viewed in the Appendix A and B of this thesis.

Once actors were found and auditioned, a film crew was assembled. A producer was hired to secure the necessary props, equipment, filming permits, food and transportation. Principal filming began on July 20th, 2002 and lasted for six days. Reshoots were done at various later dates.

The entire production was shot using a Canon XL-1 digital video camera at 720i (720x480 interlaced), with various effects shots captured using a Toshiba PDR-M4 digital still camera at 1600x1200. The film was edited over a three-month period using Adobe Premiere. In addition to the TIM section, various special effect portions were accomplished in Adobe Photoshop and Adobe After Effects.

The portion of the film which uses the TIM has a very specific intent, in that it is used in a way which achieves the main sub-goal — the desire to use the effect to support the narrative. As discussed earlier in the background chapter, the grid-based structure of the TIM alludes to uniformity and order, while at the same time the multiple on-screen images allow for the depiction of multiple points of view and points in time. The TIM is used in the short to depict the dreams of the main character, the Clock Tower Sniper. The uniformity and multiple points of view and time serve to reflect both the characters' homogeneous existence, while also representing his states of dreaming and imagination.

The impact of the TIM itself, in similar fashion to the image mosaic (in that a conceptual whole is created from the sum of the elemental pieces) allows the effect to go even deeper from a thematic standpoint. The TIM usage occurs at the climax of the film. The effect serves as a representation of the exact moment when the main character becomes aware of his individuality. At this moment of insight, he is day-dreaming, wondering what it might be like to be an individual. The sum of all his day-dreams, all the tiles in the TIM, construct the depiction of the character who facilitates his catharsis, who is also an individual existing in the uniform world of the film.

The Candidate Pool from which the TIM was constructed was created from footage shot during the production period. The pool was shot with a specific intent in mind as to how it would be used in the TIM, and what intent the footage would visually serve. With the themes of imagination and individuality to be depicted by the TIM, the contents of the Candidate Pool had to reflect these themes, since they would be selected by the algorithm to compose it. Thus, care had to be taken during the filming to ensure that captured sequences supported the thematic intent of the piece.

The use of the TIM in this manner, using its grammatical nature to the narrative advantage of the film, accomplishes the artistic goals set out at the beginning of this chapter. The next chapter, the Results, will depict the visual results of the TIM process and the quantitative measurements regarding its generation.

Chapter 5

Mosaic Results

This chapter details the results of each stage of the TIM creation process. The first section discusses the results of the Pre-Processing Stage in which the Base Sequence and Candidate Pool are pre-processed to extract color information. The second section discusses the results of the Sliding Window Comparison Stage. The third section discusses the results of the Composition Stage and the last section displays the visual results for three matching parameters — one favouring length, one favouring quality and one which combines the two.

All results were obtained using a Pentium 4 1.8GHz Intel computer running Windows 2000, using the Magick++ image processing library.

5.1 Pre-Processing Results

The Pre-Processing Stage consists of two parts. For the first part, the AVI video files are exported into a series of sequential frames. In the second part, these frame sequences are sampled on a pixel-by-pixel basis for RGB color information, which is then written to a binary file.

The frame-by-frame export was accomplished using Virtual Dub. Frames were exported at 30 frames per second (the frame rate of the original material) in TARGA format. The export of the candidate footage was batch processed, and the entire 1826 minutes of footage resulted in 2,160,000 frames, each of dimension 64x48. The process took 5 days (116 hours) to complete. The export of the 6-second Base Sequence resulted in 180 frames, each of dimension 640x480, and took 40 seconds to complete. The average export time using Virtual Dub is 0.2 second per frame.

Extracting the RGB color information from the 2.16 million Candidate Sequence frames was accomplished using the Magick++ image processing library. During the process, each frame is reduced in size to 8x6 and RGB information is sampled from each of the 48 pixels, and written to a binary file. The time taken to process all 2.16 million frames was 10 days (259.3 hours), resulting in an average processing time of 0.43 seconds per frame.

Extracting the RGB information from the 180 Base Sequence frames is a slightly different process in that the Base Sequence frames are reduced in size to 272x204 and processed in terms of 1156 8x6 regions — corresponding to each tile in the TIM. The time taken to process all 180 frames was 74 seconds, resulting in an average pre frame processing time of 0.41 seconds per frame.

The results of the pre-processing stage can be seen in Table 5.1.

5.2 Sliding Window Comparison Results

The Sliding Window Comparison stage is the most time-consuming stage of TIM creation. During the process, each frame of the Candidate Pool is compared to every

Stage	Type	Frames	Time	Time/Frame
Frame Export	Base Sequence	180	40 sec	0.2 sec
	Candidate Sequence	2.16 million	116 hours	0.2 sec
Color Extract	Base Sequence	180	74 sec	0.41 sec
	Candidate Sequence	2.16 million	259.3 hours	0.43 sec

Table 5.1: Pre-Processing Results

frame of every Base Tile Sequence. That is, the entire sliding window process is performed once for each tile in the Base Sequence. All matches between frames during this process which meet the predefined matching tolerance value are written to a binary file.

Two variations of the Candidate Pool were run through the sliding window process — a small pool of size 186 min, and a relatively large pool of size 1826 minutes. The raw time taken to process the small pool, through all 1156 tiles of a 6-second Base Sequence, was 40.2 hours, resulting in an average per tile processing time of 2.1 minutes. The raw time taken to process the large candidate pool was 440.7 hours (18 days), resulting in an average per tile processing time of 22.9 minutes. Note that the ten-fold increase in the per tile processing time is a direct result of increasing the size of the candidate pool (in this case, also by a factor of ten). Generally, the larger the pool, the more matching candidate sequences will be found. As a result, the increased number matching sequences per tile take longer to process.

The results of the Sliding Window Stage can be seen in Table 5.2.

It should also be noted that a tile which finds many matches to the Candidate Pool takes longer to process than a tile which finds few or no matches. This is due to the overhead required to process that match, track its length and write the match

Pool Size	Base Seq Length	Pool Length	Raw Time	Time/Tile
Small - 186 min	6 sec	410,000	40.2 hrs	2.1 min
Large - 1826 min	6 sec	2,160,000	440.7 hrs	22.9 min

Table 5.2: Sliding Window Results

to file. In the case of the large library, a tile which generates no matching sequences takes 17.2 minutes to process, while a tile which does generate matches requires a time closer to that or above the average time of 22.9 minutes.

5.3 Composition Results

There are two parts to the Composition Stage — the Shortest Path Selection and TIM Output. The heart of the Composition Stage is the Shortest Path Selection process, which selects the best Candidate Sequences from the list of matching candidates, based on the parameterized combination of match length and match quality. As in the Sliding Window Stage, Base Tile Sequences are processed one at a time. The final part of composition is the TIM Output, where the TIM is composited frame by frame, and encoded into video form.

Given a list of matches, the shortest path process builds the array which represents the graph, then finds the shortest path solution through the array using the parameterized weights. The raw time taken to find the shortest path solution for all 1156 6-second Base Tile Sequences is 41:42 minutes, resulting in an average per tile processing time of 2.2 seconds.

As in the case of sliding window matching, the shortest path selection stage takes longer to produce results for a Base Tile Sequence which has many corresponding matches versus one with a relatively few or no matches. The time taken to run the shortest path process on a Base Tile containing 0 matches is 0.25 seconds, while the time taken to process one containing 7276 matches is 5.9 seconds.

The second part of the Composition Stage is the frame by frame generation of the TIM using the shortest path results for each Base Tile Sequence. Once the individual frames of the TIM are created, they are compacted together to form a video file — the end-resulting TIM.

The raw time taken to generate 180 frames of the TIM, corresponding to the a 6-second Base Sequence, is 45 minutes, resulting in an average per frame generation time of 15 seconds per frame. The raw time taken to compress all 180 frames into an MPEG file is 8 minutes. This result will vary depending on the length of the Base Sequence, the amount of compression applied, the end resolution and file type selected — MPEG, AVI, MOV, etc.

The results of the composition stage can be seen in Table 5.3.

Process	Base Seq Length	Raw Time	Avg Time
Shortest Path	180f/6 sec.	41:42 min.	2.2 sec/tile
TIM Output	180f/6 sec.	95:32 min.	31.8 sec/frame
TIM Encoding	180f/6 sec.	7:20 min.	—

Table 5.3: Composition Results

5.4 Visual Results

This section depicts the resulting TIMs for three different parameter values, and discusses the distinct visual features of each. Because of the temporal nature of the TIM, the reader may find it helpful to view the results on the CD-ROM which accompanies this thesis.

All examples in this chapter were created using a 6-second Base Sequence, divided into 1156 (32x32) Base Tile Sequences, and the 1826 minute Candidate Pool. For the purposes of comparison, Figure 5.1 shows the first frame of the Base Sequence used in the following TIM examples (BaseSequence.mpeg on CD).

Figure 5.1: First Frame of Base Sequence

It will also be helpful to recall the matching function:

$$cost = p \times (\frac{\Delta C}{CMax} \times L) + (1 - p) \times ((\frac{LMax - L}{LMax})^N \times L)$$

5.4.1 Quality (the P=1 TIM)

The first result presented favours match quality. Generated using a matching parameter of P=1, the result is a TIM created from the best-matching Candidate Sequences in terms of color difference, regardless of temporal length. Figure 5.2 depicts the first frame of the P=1 TIM (TIM1.mpeg on CD).

Figure 5.2: First Frame of P=1 TIM

As expected, the result is a TIM of the highest visual quality, most closely resembling the Base Sequence. The resulting frames are soft in appearance, resembling the work of an Impressionist. Distinct lines and shapes in the Base Sequence are clearly respected in the TIM, as are original areas of colour and shading. Fine detail is preserved in many complex areas. Note the picked fence, the line of pedestrians, the windows of the background building and the car on the left. The result is visually pleasing in static form. However, when viewed in temporal form, there are several aspects that make the P=1 TIM undesirable.

Paramount in the undesirable visual qualities of the P=1 TIM is *flicker*. Favouring quality irrespective of length results in short Candidate Sequence choices, in many cases as short as a single frame. Although movement in the original Base Sequence is preserved, these short segments result in a large number of discontinuities between successive frames. Over the course of the TIM, these rapid “cuts” infuse the TIM with a frenetic, high-energy feeling. This freneticism can be distracting, introducing movement and points of attention where none previously existed. From a film perspective, the introduction of energy and flicker may work against the thematic intent of a TIM.

The second undesirable quality of the P=1 TIM, another side effect of the quick cuts, is tile continuity and the inability of the viewer to focus on the content of a particular tile because of the cutting pace. These rapid cuts, at a maximum of one cut per frame, or 30 cuts per second, render the content of a particular frame undiscernible to the viewer. From the perspective of the artist, this is counter-productive to the intent of a mosaic. One of the most important visual aspects of the image mosaic is suggestion of a whole from a variety of conceptually related sub-pieces. The artistic intent is that the viewer will appreciate the mosaic from both perspectives — the whole and the component pieces. Viewer inspection and appreciation of the P=1 TIM on a tile by tile basis is impeded by the rapid cutting, undermining an important component of the mosaic.

With this in mind, the next result favours temporal length with no regard to color difference.

5.4.2 Length (the P=0 TIM)

At the opposite end of the matching parametrization is the P=0 TIM. It favours length irrespective of color difference. Additionally, given the choice between Candidate Sequences of equal length, the process will select the Candidate of lowest color difference. Figure 5.3 depicts the first frame of the P=0 TIM (TIM0.mpeg on CD).

Figure 5.3: First Frame of P=0 TIM

Immediately evident is the lowered visual quality, as the result is visually coarse compared to the softness of the P=1 TIM. Overall, large areas of color are preserved, as are large linear features, but smaller details are less distinct. Direct comparison of the picked fence and the line of pedestrians to the previous result illustrates the difference in level of detail depiction.

The coarseness and low match quality is visually undesirable when viewed in static form, but temporally, the P=0 TIM is a striking contrast to the high-energy of the previous result. As in the P=1 TIM, movement is preserved, undistorted and continuous, but favouring length allows the longest Candidate Sequences available to be selected for the TIM, resulting in a very continuous result. Flicker is drastically reduced, visible only in areas of rapid change in the original Base Sequence. The overall TIM is calm, reflecting the pace of the Base Sequence, as no unnecessary energy has been added by mass discontinuity.

From an artistic standpoint, the most important improvement is in tile continuity. In contrast to the discontinuous nature of the P=1 TIM, tile continuity in this case allows the viewer to follow the TIM on a tile by tile basis. Many tiles contain a single Candidate Sequence, which runs the entire temporal length of the TIM, producing no discontinuity. In this regard, the P=0 TIM is a *true* image mosaic, in that it allows viewer inspection and appreciation of both the individual pieces and the conceptual whole.

There is a trade-off between the two parameter values. A value which favours only quality produces the best visual results in terms of color matching, but introduces undesirable discontinuity. A value which favours only length of match produces the best results in terms of continuity, but at the expense of color and fine detail representation. The final result accounts for match quality and match length.

5.4.3 Quality and Length (the P=0.5 TIM)

Using a parameter value of P=0.5 allows the Candidate Sequences to be selected in a manner which accounts for both color difference and length of match. The matching function combines the two by normalizing each of the values and adding half of their contributions together to form the final matching value. Figure 5.4 depicts the first frame of the P=0.5 TIM (TIM1.mpeg on CD).

Figure 5.4: P=0.5 TIM

As expected, the visual characteristics of the P=0.5 TIM are a combination of the two previous results. In terms of visual quality, the result is not as soft as the P=1 TIM, yet is certainly an improvement over the coarseness of the P=0 TIM. Large areas of color are preserved, as are large linear features. Representation of small levels of detail, such as the picket fence and the line of pedestrians, is also improved. Overall, in static form, the P=0.5 TIM improves upon the drawbacks of detail representation and color match as seen in the P=0 TIM.

When viewed temporally, much of the tile continuity seen in the P=0 TIM is retained. The appearance of more mid-length Candidate Sequence choices, on average, increase the number of discontinuities per tile sequence, through not nearing the discontinuous flicker as seen in the P=1 TIM. As a result, the TIM faithfully reflects the energy level of the original Base Sequence. Additionally, as in both previous results, movement is preserved and continuous.

From an artistic standpoint, the increase in the tile sequence discontinuity does not detract from the artistic potential of the P=0.5 TIM. In most instances, Candidate Sequences are of sufficient length to allow individual tile appreciation by the viewer. The increase in color matching and depiction of detail as compared to the P=0 TIM is also an artistic asset of this result.

5.5 Summary

Cinematically, each of the three resulting TIMs have potential uses. The P=1 TIM with its high visual quality and low Candidate Sequence lengths results in a high-energy discontinuous presentation, which could be used to cinematic advantage in certain situations.

The P=0 TIM sacrifices image quality and minute detail representation for Candidate Sequence length choices. The result is a TIM which more faithfully represents the original Base Sequence in terms of energy. It also comes the closest to the original intent of the thesis. It is a true mosaic in that it allows for viewer appreciation of individual tile content, in addition to suggesting a whole from the individual pieces.

The P=0.5 TIM provides a balance of image quality and match length. The result remains faithful to the original Base Sequence in terms of energy. The selection of shorter Candidate Sequences results in tile sequence discontinuity, but remains acceptable in terms of the ability of the viewer to appreciate individual tile sequences. Image quality and level of detail representation are improved as compared to the P=0 TIM. As a result of this, the P=0.5 TIM is the result most suited to achieve the original cinematic intent of the this thesis.

Chapter 6

Conclusion and Future Work

This thesis has presented a detailed history of the mosaic and the mosaic arts, which traces the evolution of the mosaic as an art form from its beginnings as a decorative element, up to the present day image mosaics and related art forms in the digital realm. This thesis has also presented the Temporal Image Mosaic (TIM), which extends the static image mosaic by adding the dimension of time.

Comparisons have been made to the technique of the Video Mosaic, and the results show the TIM achieves its initial goals, solving the problems of repeated images and candidate pool size which were avoided by the use of substantial color correction in the Video Mosaic. In terms of parametrization, results have shown that favouring length over colour difference when selecting candidate sequences produces a true image mosaic. However, balancing length and color difference in the form of a parametrization produces the most useful result. This parametrization also addresses the goal of producing a "visually pleasing" result, as it allows the artist to control the trade-off of length versus quality.

As a result of the interdisciplinary partnership of this thesis, the TIM has been considered both as a technical and artistic creation. From the artistic standpoint, the TIM has been evaluated as to its possible contributions to the art of filmmaking, and in the context of film grammar. Though film grammar is largely context-sensitive, the TIM, like the static image mosaic before it, speaks visually from it composition of a larger presentation from conceptually related components, and from its distinct grid-based structure offering multiple points of view and time. Respecting these grammatical implications, the thesis has gone a step beyond the analytical and into the purely creative by producing a short narrative film which exhibits one possible use of the TIM in an artistic setting.

6.1 Future Work

Over the course of TIM development, several avenues for improvement became apparent, and may be addressed in future work:

Candidate Pool

In an effort to improve the visual quality of the results, consider the candidate pool. The importance of the candidate pool to the final results cannot be understated. Many papers on the subject of image mosaic creation stress that the single-most important factor in mosaic creation is a large, well-distributed (in terms of color) candidate pool. The results of the TIM support this — the candidate pool is crucial. The larger the candidate pool, the more options there are during matching, increasing the chances that a good match will be found. The largest candidate pool used in TIM creation was 1826 minutes (2.16 million frames) in length. Although time consuming to create and occupying over 60GB of storage space, increasing the pool by a factor of

10 or 100 with well-distributed candidate sequences would produce noticeably better visual results. Depending on artistic intent, it should also be noted that the type of candidate footage used is also an important consideration. Television and movie footage with rapid cutting patterns may provide good color distribution for the pool, but the pace and rapid discontinuity in the material is likely to reduce the length of resultant matches. Slower paced or continuous footage is preferred, and will provide a good basis for longer matching sequences.

Color Comparison Resolution

As another effort to improve visual quality of the results, consider color comparison resolution. In order to reduce processing time, images in this thesis were reduced in size and compared at a resolution of 8x6 pixels. As shown in Chapter 3 (Figure 3.2), reduction of images before comparison does preserve large overall features and shape within the image, but comparing at higher resolutions will give a more accurate result in terms of fine detail. Higher comparison resolution of 40x30 or 80x60 would provide more accurate matching and allow for greater distinction among candidates. As mentioned, however, this is at the expense of processing time in the pre-processing and sliding window comparison stages.

Image Database

In an effort to provide a higher level of organization to the candidate pool and data files required in TIM generation, consider an image database. With increases in candidate pool size and color comparison resolution comes the complexity of managing the images and the data. A database used to organize the candidate pool and data resulting from the pre-processing and sliding window stages would be a welcome improvement. In fact, it may be advantageous as the candidate pool increases in size to structure the sliding window process around database queries.

Image Attributes

In an effort to increase match quality, consider the use of more image attributes than just color. Several image mosaic approaches consider more than just color during the matching process. Static image properties such as shape and brightness, and temporal properties such as shape and brightness through time could be considered. These attributes could then be accounted for in the match cost calculation. The preprocessing stage could be used to extract these visual properties from the frames and sequences as a whole, and store this information in the form of a database mentioned previously.

Base Sequence Energy

In an effort to adhere to the energy level of the original Base Sequence as close as possible, consider the Base Sequence energy during matching. For each Base Tile Sequence, an energy level could be calculated which represents the energy level (rate of change) of the tile over time. When the parameterized cost calculations are being done to determine the best-fit candidate sequences, the parameter p could be varied over time corresponding to the energy level of each tile. Energy could be individualized for each tile instead of remaining constant for the entire candidate selection process. In this manner, low-energy candidate sequences would be selected to represent low-energy sections of the Base Sequence, and similarly for high-energy sequences.

Color Correction

In an effort to respect the visual and cerebral impact of the image mosaic, *do not* use color correction methods to improve results. Several image mosaic approaches have developed color correction methods which are used to enhance the chosen candidate images and sequences so they more closely resemble the original Base Sequence. Changing the candidate images from their original visual state is counter-productive to the visual and cerebral impact of the image mosaic, which is that natural constituent pieces function in harmony to suggest a large overall presentation. “Correcting” the natural images to get a better visual fit robs the image mosaic of its power — the revelation in the mind of the viewer that what they are seeing is not an illusion, but a natural harmony of images.

Relying on a large and well-distributed candidate pool and adequate image comparison techniques, evocative image mosaics can be created without the need for color correction. Color correction, though used creatively by Finkelstein and Range to producing their own distinctive variation of the image mosaic, in the opinion of this author, remains an unnecessary compensation for an inadequate candidate pool.

Bibliography

[1] Ascher, Steven and Pincus, Edward. *The Filmmaker's Handbook*. Penguin Group, 1999.

[2] Bankston, Douglas. *A Hostile Lodging*. American Cinematographer. Vol. 84, No. 4, April, 2002.

[3] Bankston, Douglas. *Quadruple Vision*. American Cinematographer. Vol. 82, No. 4, April, 2000.

[4] Bordwell, David and Thompson, Kristin . *Film Art : An Introduction*. 6th Edition. McGraw Hill, 2001.

[5] Cormen, Thomas et al. *Introduction to Algorithms*. Second Edition. McGraw-Hill Book Company, 2001.

[6] Cunningham, Michael. “Slave to the Machine: ‘What's the point/pixel?’” *CIRCA Art Magazine*. CIRCA 89, Autumn 1999.

[7] “Cubism.” *Dictionary of the Arts*. London. Brockhampton Press. 1997.

[8] Curand, Jannic. *Byzantine Art*. Paris: Terrail, 1999.

[9] Finkelstein, Adam and Range, Marisa. "Image Mosaics." *Proceedings of Raster Imaging & Digital Typography*, pp. 11-22, 1998.

[10] Haeberli, Paul. "Paint By Numbers: Abstract Image Representations." *Computer Graphics (SIGGRAPH)*, pp. 207-214, 1990.

[11] Harmon, L. D. *The Recognition of Faces.* Scientific American, Vol. 229, No. 5, pp. 71-82, 1973.

[12] Hausner, Alejo. "Simulating Decorative Mosaics." *Computer Graphics* (SIGGRAPH), pp. 573-580, 2001.

[13] Fels, S. S., and Mase, K. "Technique for Interactive Video Cubism." In *Proceedings of ACM Multimedia*, pp. 368-370, October 2000.

[14] Jacobs, C.E., Finkelstein, A., and Salesin, D.H. "Fast Multiresolution Image Querying". *Computer Graphics (SIGGRAPH)*, pp. 277-286, 1995.

[15] Katz, Stephen D. *Film Directing Shot By Shot.* Studio City: Michael Wiese Productions, 1991.

[16] Kim, Junhwan and Pellacini, Fabio. "Jigsaw Image Mosaics." *ACM Transactions on Graphics*, Vol.21, Iss. 3, pp. 657-664, 2002.

[17] Klein, Allison W., Sloan, P.P., Colburn, A., Finkelstein, A., and Cohen, M. F. "Video Cubism." Microsoft Research Technical Report MSR-TR-2001-45, 2001.

[18] Klein, Allison W., Sloan, P.P., Colburn, A., Finkelstein, A., and Cohen, M. F. "Stylized Video Cubes." *ACM SIGGRAPH Symposium on Computer Animation*. pp. 15-22, July 2002.

[19] Klein, Allison W., Grant, T., Finkelstein, A. and Cohen, M. F. "Video Mosaics." Non-Photorealistic Animation and Rendering 2002, June 2002.

[20] Litwinowicz, P. "Processing Images and Video for an Impressionist Effect." *Computer Graphics (SIGGRAPH)* pp. 407-414, 1997.

[21] Mango, Cyril A. *Art of the Byzantine Empire*. Prentice Hall, 1972.

[22] Mascelli, Joseph V. *The Five C's of Cinematography*. Silman-James Press, page 165, 1965.

[23] Meier, Barbara. "Painterly Rendering for Animation." *Computer Graphics (SIGGRAPH)*, pp. 477-484, 1996.

[24] Monaco, James. *How to Read a Film: The World of Movies, Media and Multimedia*. Oxford University Press, 2000.

[25] Salisbury, Michael, Anderson, Sean, Barzel, Ronen and Salesin, David. "Interactive Pen-and-Ink Illustration." *Computer Graphics (SIGGRAPH)*, pp. 101-108, 1994.

[26] Salisbury, Mike, Anderson, Corin, Lischinski, Dani and Salesin, David. "Scale-Dependent Reproduction of Pen-and-Ink Illustrations." *Computer Graphics (SIGGRAPH)*, pp. 461-468, 1996.

[27] Silvers, R. and Hawley, M. *Photomosaics*. Henry Holt, 1997.

[28] Stokstad, Marilyn. *Art - A Brief History.* Prentice Hall and Harry Abrams Publishers, 2000.

[29] Strauss, Michael. *Impressionism & Modern Art.* Sotheby Parke Bernet Publications, 1974.

[30] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glassman, E., Petkovic, D. and Yanker, P. "The QBIC project: Querying Image by Content Using Colour, Texture and Shape". *International Society for Optical Engineering*, pp. 115-132, February 1993.

[31] Sawhney, H. S., Hsu, S., and Kumar, R. "Robust Video Mosaicing Through Topology Inference and Local to Global Alignment." In *Proc. European Conference on Computer Vision*, vol.1407, p.103, 1998.

[32] Szeliski, R. "Video Mosaics for Virtual Environments." *IEEE Computer Graphics and Applications*, Vol. 16, No. 2, pp. 2230, March 1996.

[33] Tran, N. "Generating Photomosaics: An Empirical Study." In *Proc. ACM Symposium on Applied Computing*, pp. 105-109, 1999.

[34] Zhang, Yue. *On the use of CBIR in Image Mosaic Generation.* Technical Report TR 02-17. Department of Computing Science, University of Alberta, 2002.

(All web references below available as of November 7th, 2003.)

[35] "ANRO Commercial Printing - 1996 Newsletter."
http://www.anro.com/mailing-news/ink3.html

[36] "Impressionism." *Artelino.com.*
http://www.artelino.com/articles/impresionism.asp

[37] "Chuck Close - About the Art." *Arts Connected - ArtsNet Minnesota.*
http://www.artsconnected.org/artsnetmn/identity/close.html

[38] Blake, Scott. "Bar Code Art History"
http://www.barcodeart.com/history/history.html

[39] Drewes, Jim. "JImage Mosaic"
http://jimage-mosaic.sourceforge.net/samples.php

[40] "Greek Mosaics - History." *Greekmosaics.com.*
http://www.greekmosaics.com/History.htm

[41] Hunt, William Leigh. *The PhotoTiled Pictures Home Page.*
http://home.earthlink.net/~wlhunt/

[42] "Giuseppe Arcimboldo." *Illum.co.uk.*
http://www.illumin.co.uk/svank/biog/arcim/arcbio.html

[43] Kaminker, Laura. "Chuck Close and the Salvation of Work"
http://www.disabled.gr/gb-arts/paint01.html

[44] *Merriam-Webster Online Dictionary.* http://www.merriam-webster.com

[45] Riemersma, Thiadmer. "Colour Metric." *Compuphase.com.* ITB CompuPhase.
http://www.compuphase.com/cmetric.htm

[46] "Runaway Technology." *Photomosaic.com.* http://www.photomosaic.com/

[47] “The Origins of Mosaic Art.” *Sidarte.it.*
http://www.sidarte.it/doc/storia-del-mosaico-ing.htm

[48] Stark, Joan G. “Development of (ASCII) Text Art”
http://www.geocities.com/joan_stark/textasciihistory.txt

[49] Storr, Robert. “Chuck Close” *The Museum of Modern Art.*
http://moma.org/exhibitions/1998/close/

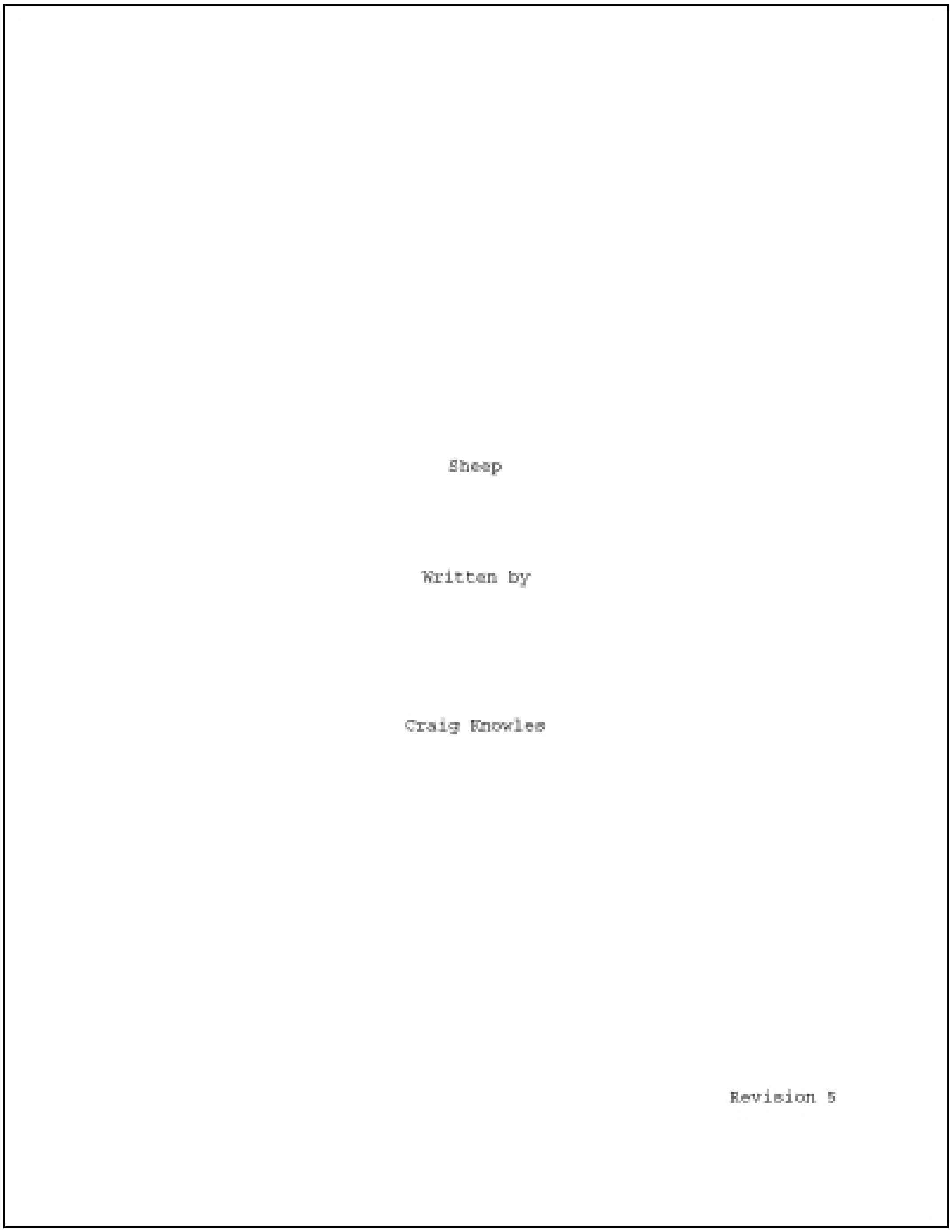
[50] Troebs, Micheal. *Algorithms to Create Image Montages.*
http://www.stud.uni-hannover.de/∼michaelt/juggle/

[51] Zucman, Glenn. “Strange Pixels - Timeline - Leon Harmon.” *Artboy 2.6.*
http://www.ioc.net/∼artboy/P/harmon.html

Appendix A

Script

What follows is the fifth and final revision of “Sheep”, the script written under the guidance of film co-supervisor Clarke MacKey as part of artistic contribution to this thesis:

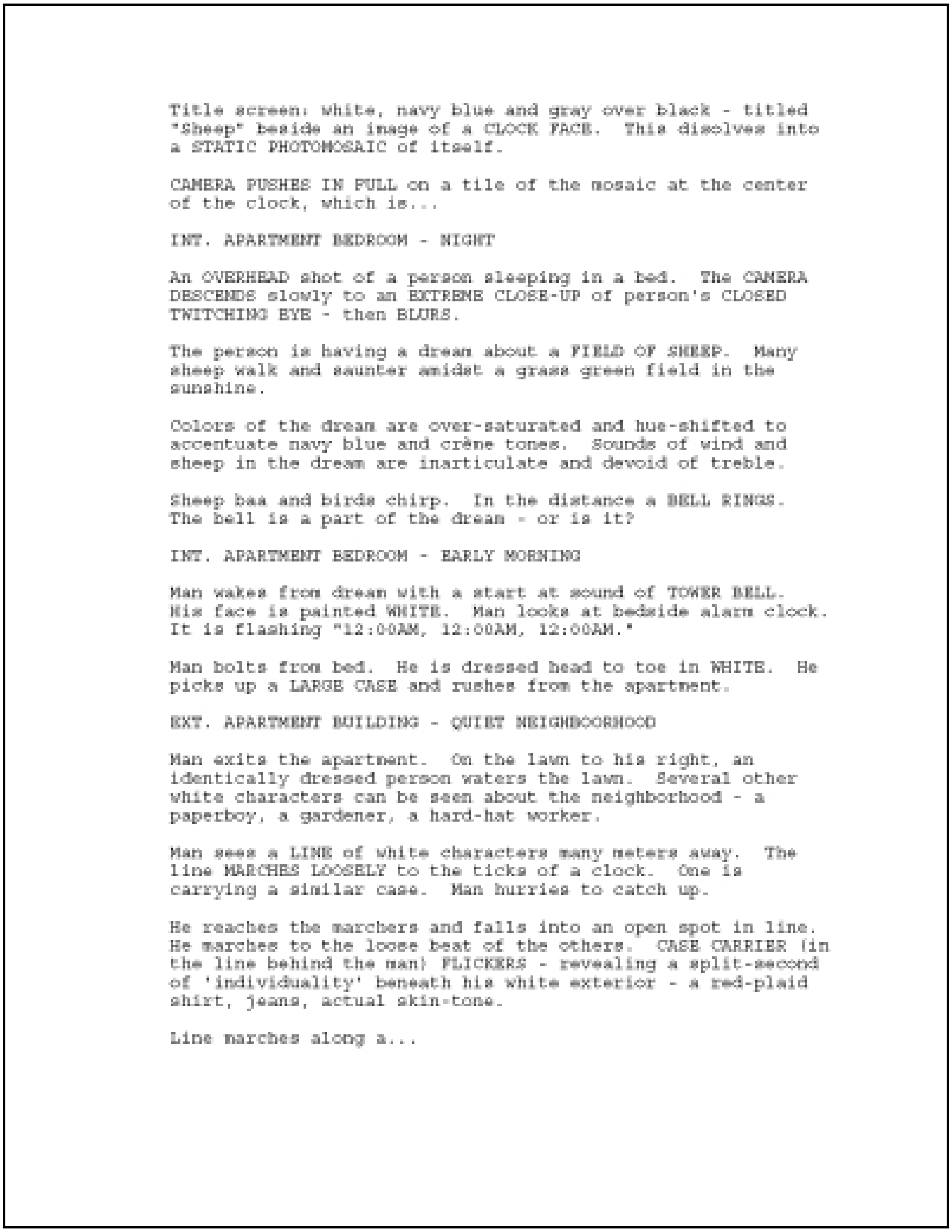


Sheep

Written by

Craig Knowles

Revision 5



Title screen: white, navy blue and gray over black - titled "Sheep" beside an image of a CLOCK FACE. This disolves into a STATIC PHOTOMOSAIC of itself.

CAMERA PUSHES IN FULL on a tile of the mosaic at the center of the clock, which is...

INT. APARTMENT BEDROOM - NIGHT

An OVERHEAD shot of a person sleeping in a bed. The CAMERA DESCENDS slowly to an EXTREME CLOSE-UP of person's CLOSED TWITCHING EYE - then BLURS.

The person is having a dream about a FIELD OF SHEEP. Many sheep walk and saunter amidst a grass green field in the sunshine.

Colors of the dream are over-saturated and hue-shifted to accentuate navy blue and crêne tones. Sounds of wind and sheep in the dream are inarticulate and devoid of treble.

Sheep baa and birds chirp. In the distance a BELL RINGS. The bell is a part of the dream - or is it?

INT. APARTMENT BEDROOM - EARLY MORNING

Man wakes from dream with a start at sound of TOWER BELL. His face is painted WHITE. Man looks at bedside alarm clock. It is flashing "12:00AM, 12:00AM, 12:00AM."

Man bolts from bed. He is dressed head to toe in WHITE. He picks up a LARGE CASE and rushes from the apartment.

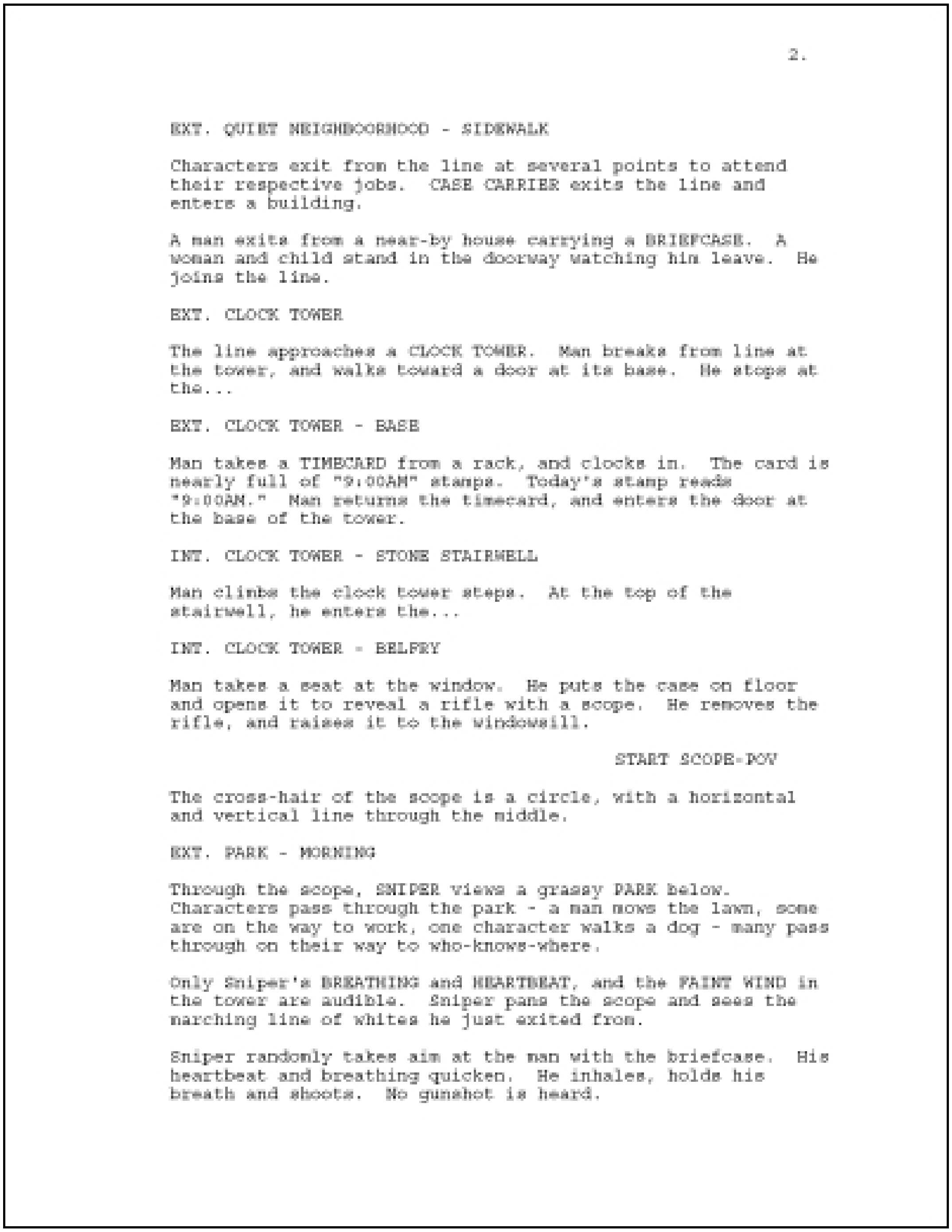
EXT. APARTMENT BUILDING - QUIET NEIGHBOORHOOD

Man exits the apartment. On the lawn to his right, an identically dressed person waters the lawn. Several other white characters can be seen about the neighborhood - a paperboy, a gardener, a hard-hat worker.

Man sees a LINE of white characters many meters away. The line MARCHES LOOSELY to the ticks of a clock. One is carrying a similar case. Man hurries to catch up.

He reaches the marchers and falls into an open spot in line. He marches to the loose beat of the others. CASE CARRIER (in the line behind the man) FLICKERS - revealing a split-second of 'individuality' beneath his white exterior - a red-plaid shirt, jeans, actual skin-tone.

Line marches along a...



2.

EXT. QUIET NEIGHBOORHOOD - SIDEWALK

Characters exit from the line at several points to attend their respective jobs. CASE CARRIER exits the line and enters a building.

A man exits from a near-by house carrying a BRIEFCASE. A woman and child stand in the doorway watching him leave. He joins the line.

EXT. CLOCK TOWER

The line approaches a CLOCK TOWER. Man breaks from line at the tower, and walks toward a door at its base. He stops at the...

EXT. CLOCK TOWER - BASE

Man takes a TIMECARD from a rack, and clocks in. The card is nearly full of "9:00AM" stamps. Today's stamp reads "9:00AM." Man returns the timecard, and enters the door at the base of the tower.

INT. CLOCK TOWER - STONE STAIRWELL

Man climbs the clock tower steps. At the top of the stairwell, he enters the...

INT. CLOCK TOWER - BELFRY

Man takes a seat at the window. He puts the case on floor and opens it to reveal a rifle with a scope. He removes the rifle, and raises it to the windowsill.

START SCOPE-POV

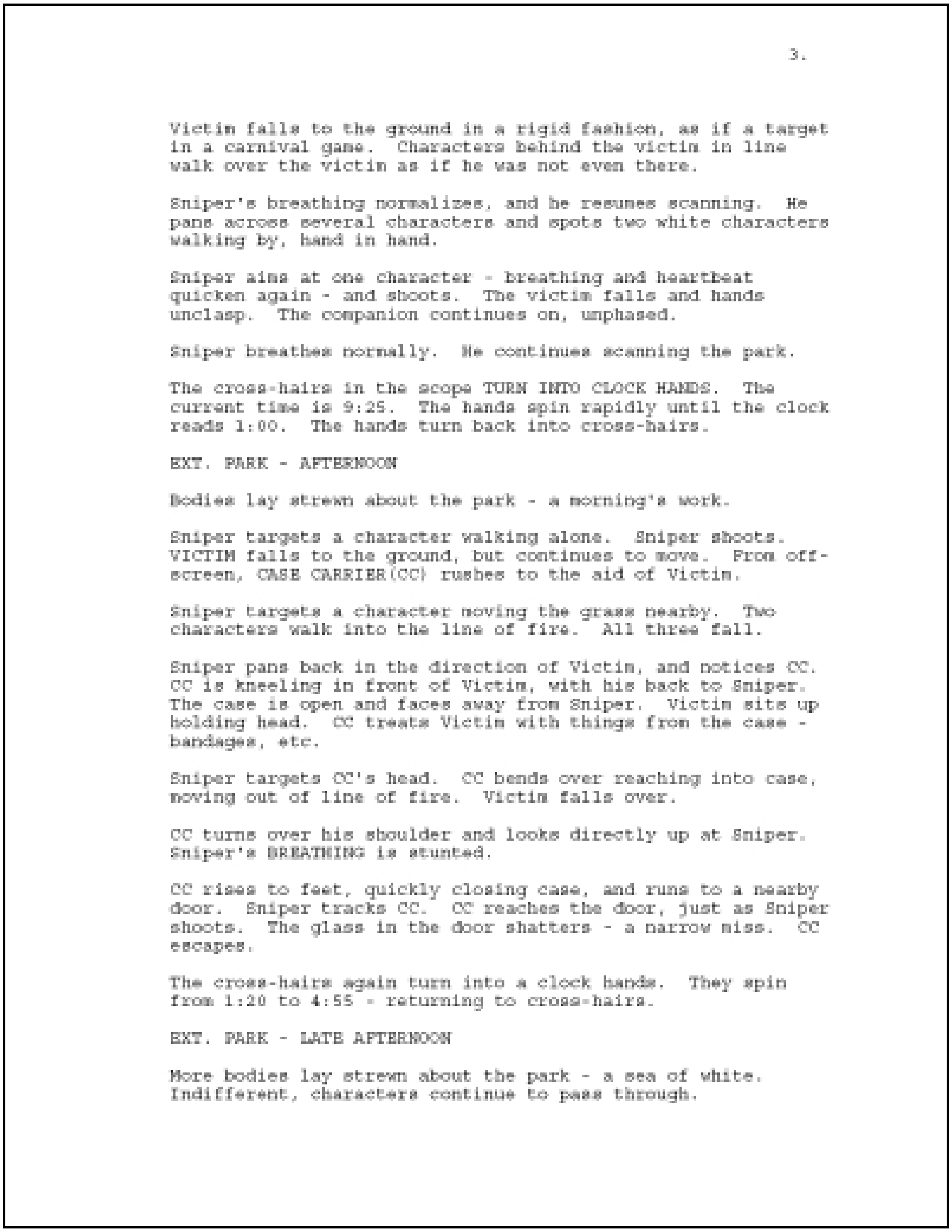
The cross-hair of the scope is a circle, with a horizontal and vertical line through the middle.

EXT. PARK - MORNING

Through the scope, SNIPER views a grassy PARK below. Characters pass through the park - a man mows the lawn, some are on the way to work, one character walks a dog - many pass through on their way to who-knows-where.

Only Sniper's BREATHING and HEARTBEAT, and the FAINT WIND in the tower are audible. Sniper pans the scope and sees the marching line of whites he just exited from.

Sniper randomly takes aim at the man with the briefcase. His heartbeat and breathing quicken. He inhales, holds his breath and shoots. No gunshot is heard.



Victim falls to the ground in a rigid fashion, as if a target in a carnival game. Characters behind the victim in line walk over the victim as if he was not even there.

Sniper's breathing normalizes, and he resumes scanning. He pans across several characters and spots two white characters walking by, hand in hand.

Sniper aims at one character - breathing and heartbeat quicken again - and shoots. The victim falls and hands unclasp. The companion continues on, unphased.

Sniper breathes normally. He continues scanning the park.

The cross-hairs in the scope TURN INTO CLOCK HANDS. The current time is 9:25. The hands spin rapidly until the clock reads 1:00. The hands turn back into cross-hairs.

EXT. PARK - AFTERNOON

Bodies lay strewn about the park - a morning's work.

Sniper targets a character walking alone. Sniper shoots. VICTIM falls to the ground, but continues to move. From off-screen, CASE CARRIER(CC) rushes to the aid of Victim.

Sniper targets a character moving the grass nearby. Two characters walk into the line of fire. All three fall.

Sniper pans back in the direction of Victim, and notices CC. CC is kneeling in front of Victim, with his back to Sniper. The case is open and faces away from Sniper. Victim sits up holding head. CC treats Victim with things from the case - bandages, etc.

Sniper targets CC's head. CC bends over reaching into case, moving out of line of fire. Victim falls over.

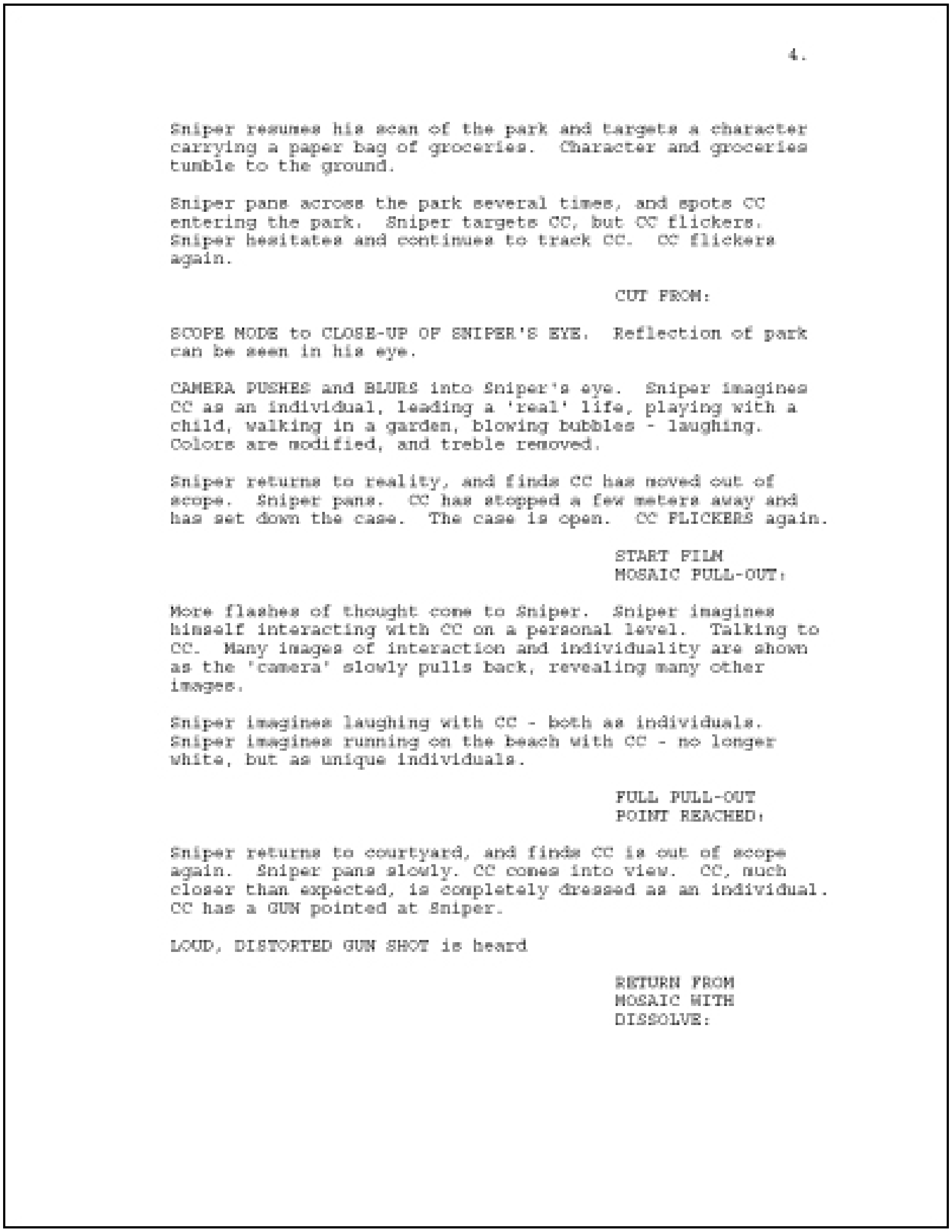
CC turns over his shoulder and looks directly up at Sniper. Sniper's BREATHING is stunted.

CC rises to feet, quickly closing case, and runs to a nearby door. Sniper tracks CC. CC reaches the door, just as Sniper shoots. The glass in the door shatters - a narrow miss. CC escapes.

The cross-hairs again turn into a clock hands. They spin from 1:20 to 4:55 - returning to cross-hairs.

EXT. PARK - LATE AFTERNOON

More bodies lay strewn about the park - a sea of white. Indifferent, characters continue to pass through.



4.

Sniper resumes his scan of the park and targets a character carrying a paper bag of groceries. Character and groceries tumble to the ground.

Sniper pans across the park several times, and spots CC entering the park. Sniper targets CC, but CC flickers. Sniper hesitates and continues to track CC. CC flickers again.

CUT FROM:

SCOPE MODE to CLOSE-UP OF SNIPER'S EYE. Reflection of park can be seen in his eye.

CAMERA PUSHES and BLURS into Sniper's eye. Sniper imagines CC as an individual, leading a 'real' life, playing with a child, walking in a garden, blowing bubbles - laughing. Colors are modified, and treble removed.

Sniper returns to reality, and finds CC has moved out of scope. Sniper pans. CC has stopped a few meters away and has set down the case. The case is open. CC FLICKERS again.

START FILM
MOSAIC PULL-OUT:

More flashes of thought come to Sniper. Sniper imagines himself interacting with CC on a personal level. Talking to CC. Many images of interaction and individuality are shown as the 'camera' slowly pulls back, revealing many other images.

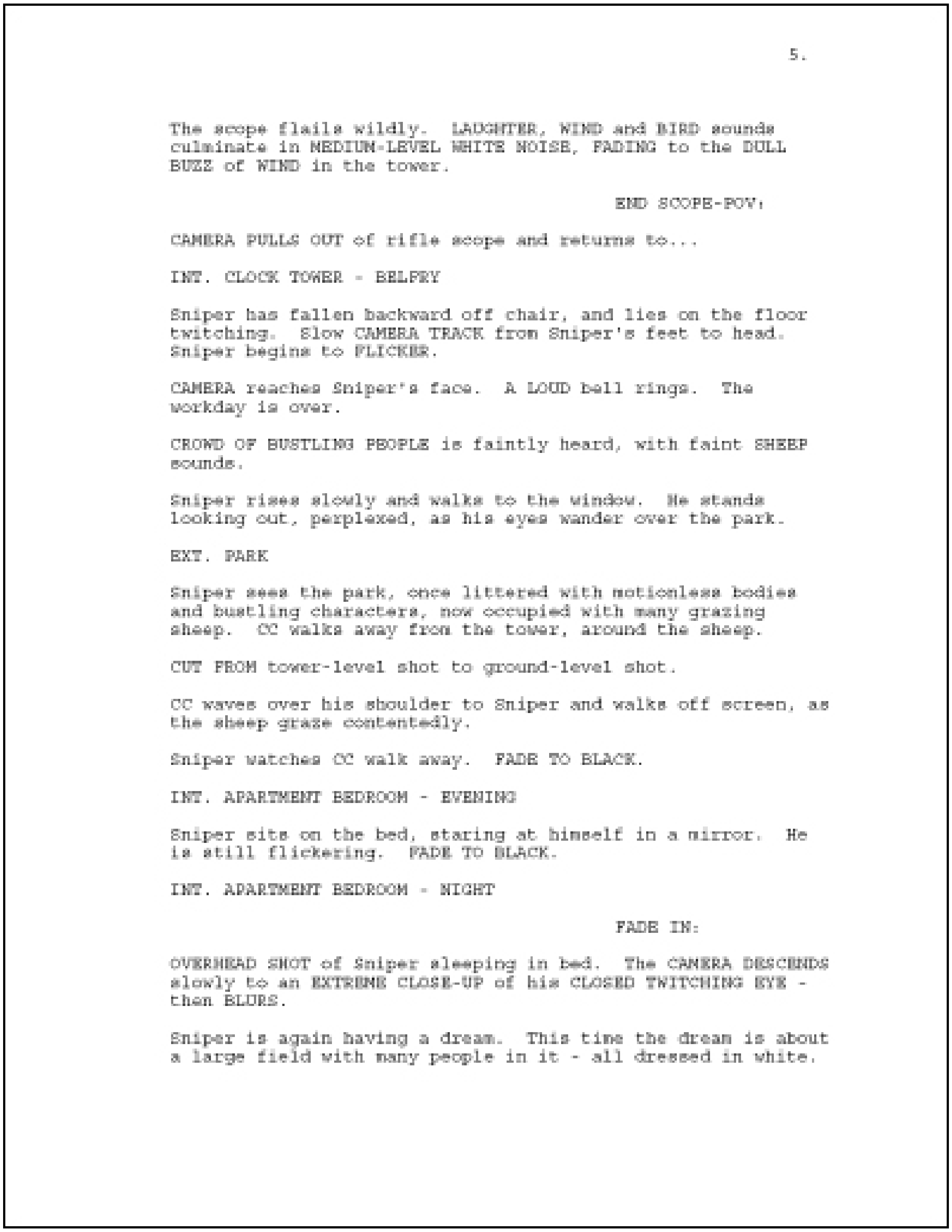
Sniper imagines laughing with CC - both as individuals. Sniper imagines running on the beach with CC - no longer white, but as unique individuals.

FULL PULL-OUT
POINT REACHED:

Sniper returns to courtyard, and finds CC is out of scope again. Sniper pans slowly. CC comes into view. CC, much closer than expected, is completely dressed as an individual. CC has a GUN pointed at Sniper.

LOUD, DISTORTED GUN SHOT is heard

RETURN FROM
MOSAIC WITH
DISSOLVE:



The scope flails wildly. LAUGHTER, WIND and BIRD sounds culminate in MEDIUM-LEVEL WHITE NOISE, FADING to the DULL BUZZ of WIND in the tower.

END SCOPE-POV:

CAMERA PULLS OUT of rifle scope and returns to...

INT. CLOCK TOWER - BELFRY

Sniper has fallen backward off chair, and lies on the floor twitching. Slow CAMERA TRACK from Sniper's feet to head. Sniper begins to FLICKER.

CAMERA reaches Sniper's face. A LOUD bell rings. The workday is over.

CROWD OF BUSTLING PEOPLE is faintly heard, with faint SHEEP sounds.

Sniper rises slowly and walks to the window. He stands looking out, perplexed, as his eyes wander over the park.

EXT. PARK

Sniper sees the park, once littered with motionless bodies and bustling characters, now occupied with many grazing sheep. CC walks away from the tower, around the sheep.

CUT FROM tower-level shot to ground-level shot.

CC waves over his shoulder to Sniper and walks off screen, as the sheep graze contentedly.

Sniper watches CC walk away. FADE TO BLACK.

INT. APARTMENT BEDROOM - EVENING

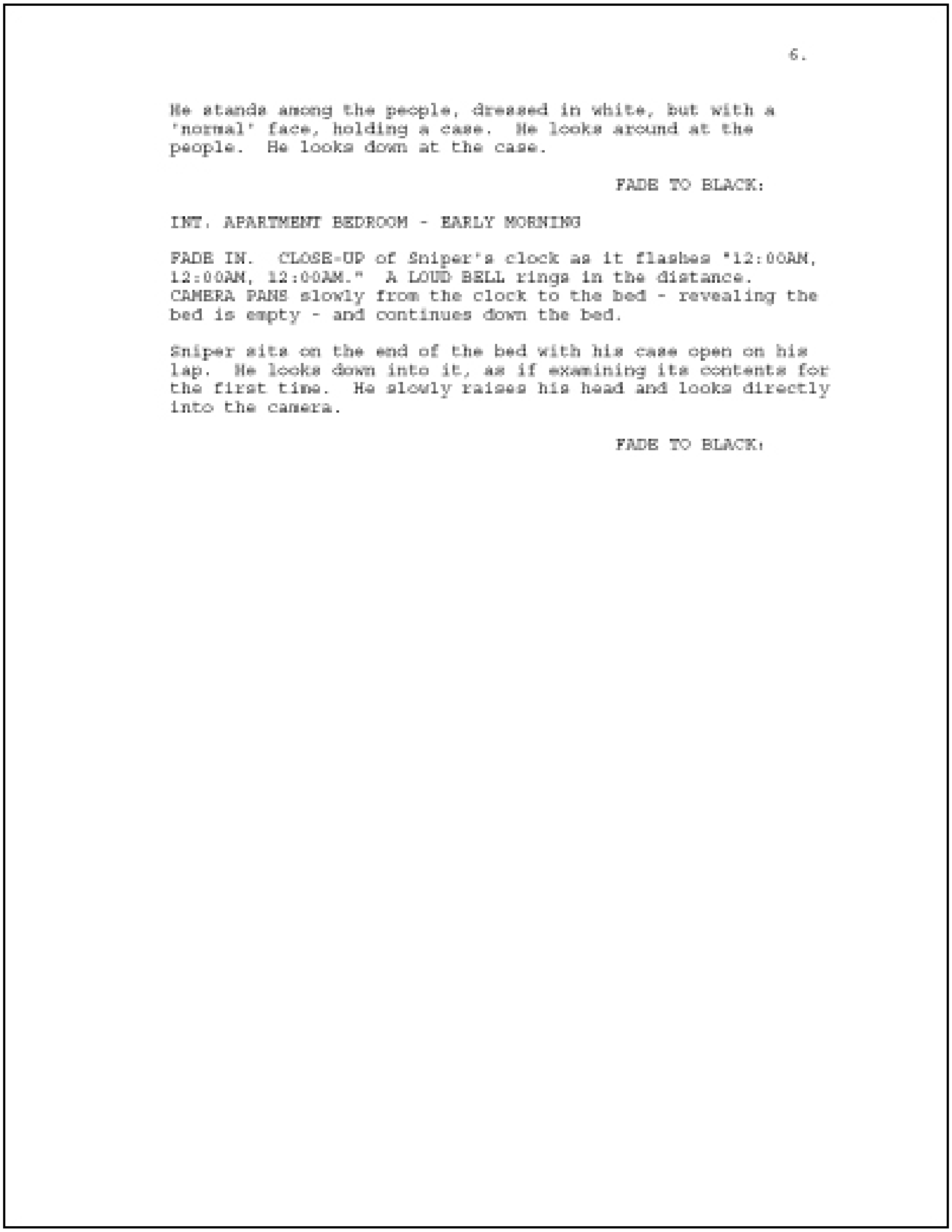
Sniper sits on the bed, staring at himself in a mirror. He is still flickering. FADE TO BLACK.

INT. APARTMENT BEDROOM - NIGHT

FADE IN:

OVERHEAD SHOT of Sniper sleeping in bed. The CAMERA DESCENDS slowly to an EXTREME CLOSE-UP of his CLOSED TWITCHING EYE - then BLURS.

Sniper is again having a dream. This time the dream is about a large field with many people in it - all dressed in white.



6.

He stands among the people, dressed in white, but with a 'normal' face, holding a case. He looks around at the people. He looks down at the case.

FADE TO BLACK:

INT. APARTMENT BEDROOM - EARLY MORNING

FADE IN. CLOSE-UP of Sniper's clock as it flashes "12:00AM, 12:00AM, 12:00AM." A LOUD BELL rings in the distance. CAMERA PANS slowly from the clock to the bed - revealing the bed is empty - and continues down the bed.

Sniper sits on the end of the bed with his case open on his lap. He looks down into it, as if examining its contents for the first time. He slowly raises his head and looks directly into the camera.

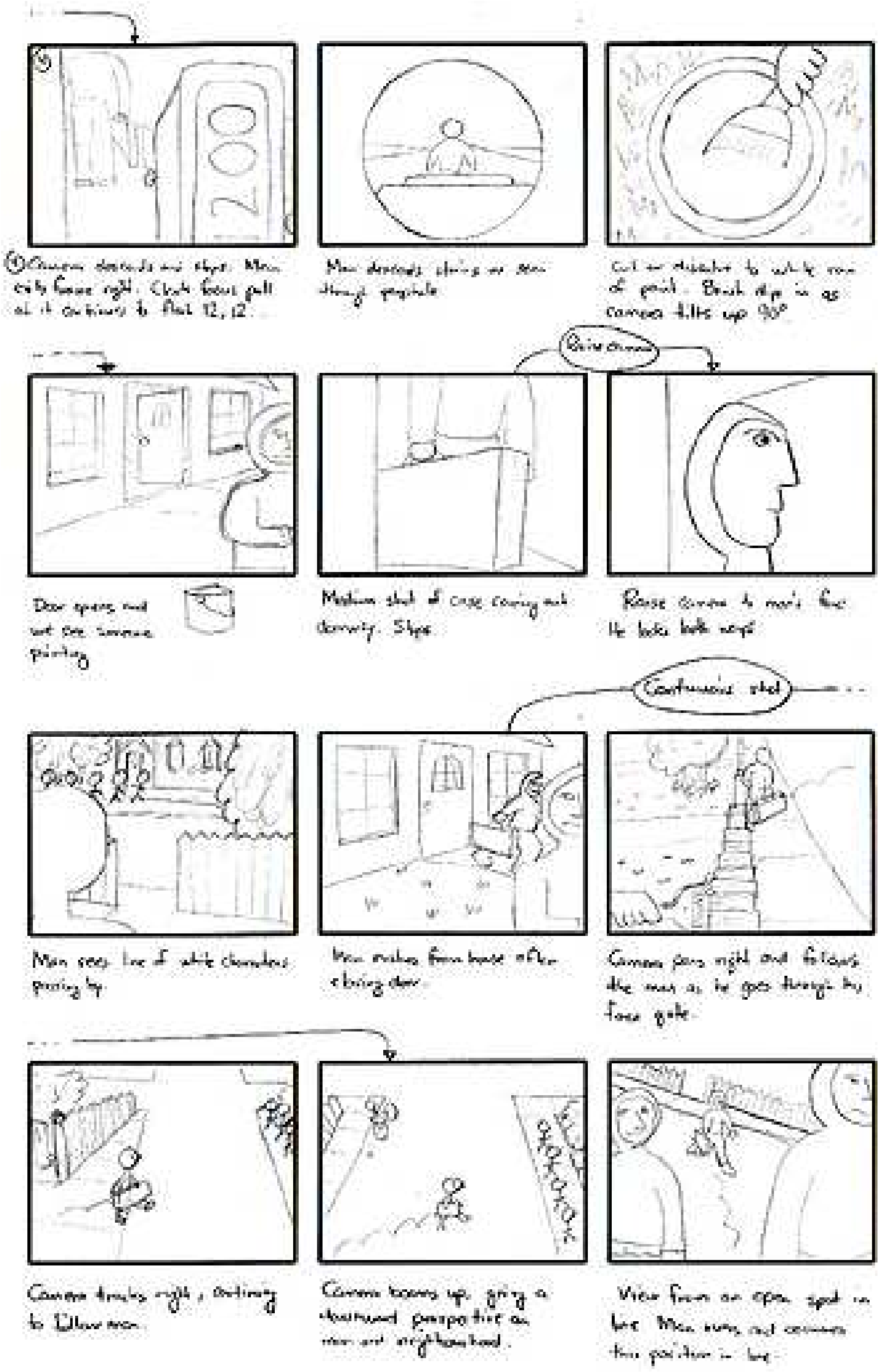
FADE TO BLACK:

Appendix B

Storyboards

As part of the artistic development for the short film, the following storyboards were prepared after the script was written and approved by film co-supervisor Clarke MacKey. The storyboards were used during filming:


Long Push in
SHEEP
Push in
Push in until pixelated
Long push in
Blur and pixelate
Overhead Guy asleep. Shadow of his window cast across left side.
Push in. Eye twitches and blinks. Can see green light from clock flashing
Surreal shot of sheep
Opens eye.
Tilts head & looks at clock
Pulls out a little to see...
Continuing Shot
12:00
Focus pull from clock to guy. Guy walks from bed to screen right. Hold as he walks back and forth several times
① Open on key. Man grabs key in a rush, and knocks clock off desk. ② Camera tracks over desk and down the side of it
③ Camera descends slowly as man puts case out from under bed


④ Camera descends and stops. Man exits frame right. Clock focus pull as it continues to flash 12, 12.
Man descends stairs as seen through peephole.
Cut or dissolve to white can of paint. Brush dips in as camera tilts up 90°
Raise camera
Door opens and we see someone painting
Medium shot of case coming out doorway. Stops.
Raise camera to man's face. He looks both ways
Continuous shot
Man sees line of white characters passing by.
Man rushes from house after closing door.
Camera pans right and follows the man as he goes through his fence gate.
Camera tracks right, continuing to follow man.
Camera booms up, giving a downward perspective on man and neighborhood.
View from an open spot in line. Man turns and assumes this position in line.

Hold on man in line as his eyes wander slightly to [illegible] him.

Camera pans down to see what they carry.

Camera pulls up to a corner to capture the continuous shot

Characters walk around corner and change screen direction R to L. Each passes by camera. Camera follows, but stops in front of [illegible]. CC, back in flashers as he walks by.

Flower door opens, man walks out. Two characters look on [illegible].

Two characters watch as man join the line

Line crosses a street. ① exits from the line, ② join the line, ③ exits from line. All [illegible]

Another character exits the line

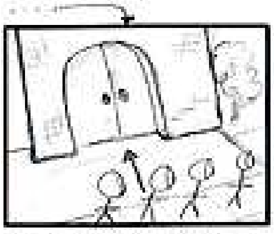
3/4 close shot. CC (seen behind [illegible]) flashers.

Elevated shot of the line walking toward camera. Slowly descends

Camera reaches ground level as characters turn onto sidewalk. All turn screen right. CC exits screen left.

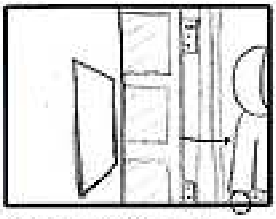
Camera pans and follows characters to screen right

Tilted up shot of clock

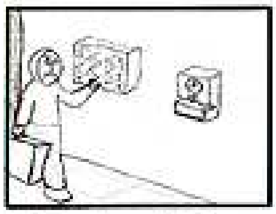


Camera tilts down to reveal Sniper exiting from the line and walking toward the door.

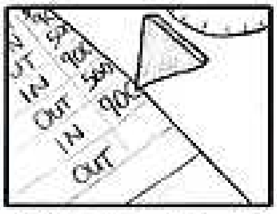
Reverse of Sniper walking toward door.



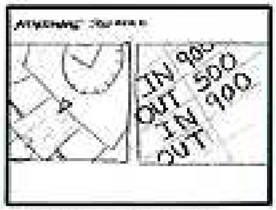
Sniper walks through door.



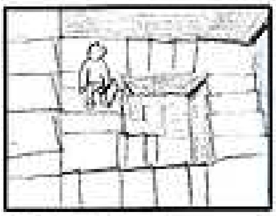
Sniper picks up his timecard.



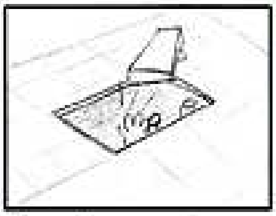
Extreme closeup showing today's punch is 900 am.



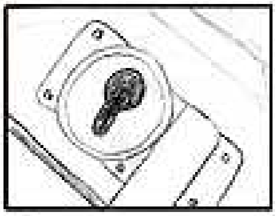
Alternate shot depending on feasibility of punch clock closeup.



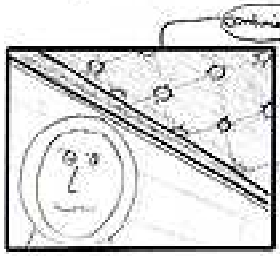
Overhead shot of man ascending stairs.



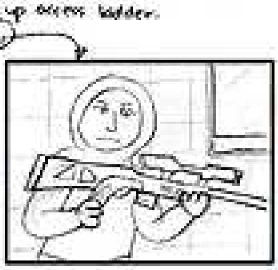
Man pushes case up through a hole in the floor and climbs up access ladder.



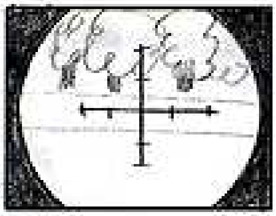
Close up of lock as man inserts key and opens.



Inside case. Lid opens and light pours in.

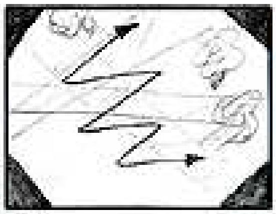


Man takes gun from case and walks to window and looks through scope.

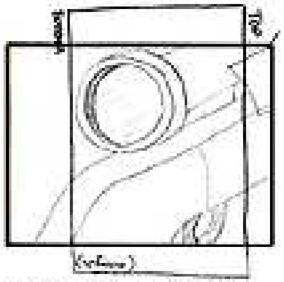


The park below comes into focus as Sniper adjusts the scope.

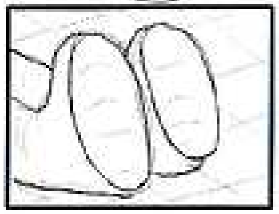
Sniper pops from case, and CC comes into view. Both close their [illegible]. [illegible] completely as [illegible]



CC shouts and steps back wildly then goes black.

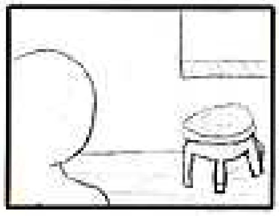


Camera pulls out of black shape, pans across floor . . .

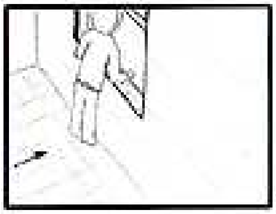


. . pan and see Sniper's feet, twitching slightly. Sniper begins to flicker . .

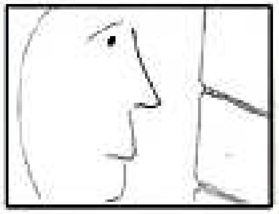
. . Camera pans completely up Sniper's body and smile on his face.



Sniper sits up and looks toward window

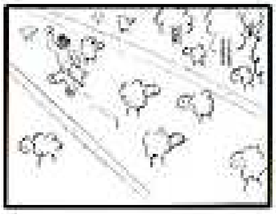


Sniper walks to window

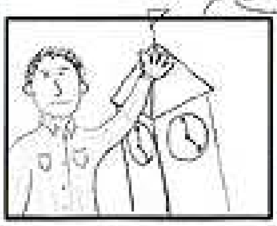


Sniper looks left and right, unsure of what he's seeing

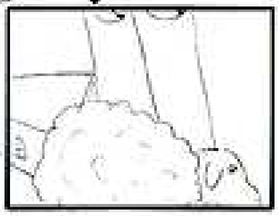
Sniper sees courtyard area, now full of sheep. Sniper looks left and right



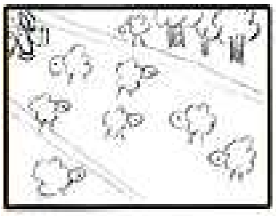
Sniper sees CC walking away. CC waves over his shoulder as he walks away.



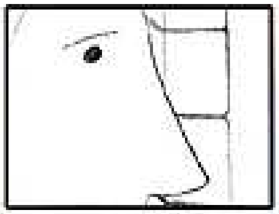
CC walks out of frame and waves over his shoulder



Camera pans down and we see a sheep

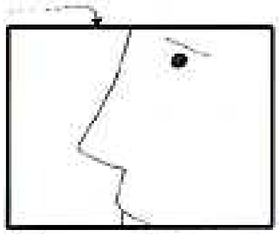


CC exits screen left, continuing to walk down the street

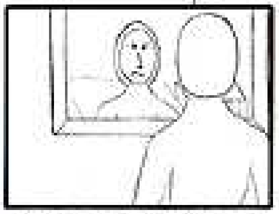


Close up of Sniper looking. Fade to black.

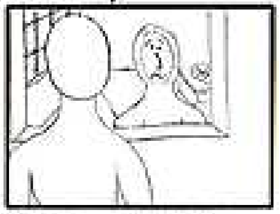
Reflection of clock in the mirror in Sniper's bedroom



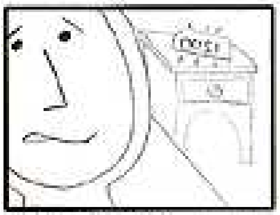
Pan screen right from previous to Sniper's face. He looks in the mirror. He is flickering.



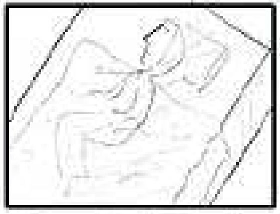
Cut to over the shoulder shot of Sniper staring at himself in the mirror. He continues to flicker.



Cross the axis by cutting from left to right. Sniper continues to flicker

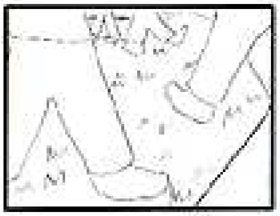


Close up shot inserted
Two options: 1. Hold on this shot and fade to black, as Sniper flickers
2. Zoom past face to clock. Fade to black as Sniper flickers.



Sniper is in bed sleeping. As before, camera descends to his twitching eye.

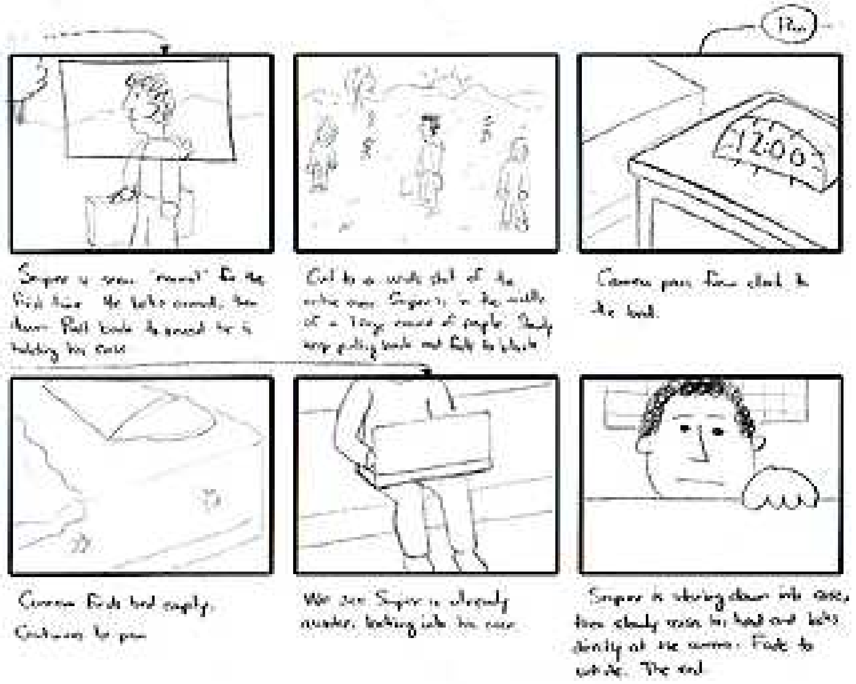
Eye twitches. Camera blurs.



Subjective camera, walking through a crowd of people — all while on a grassy green field

Characters walk aimlessly about the grassy field. Several shots of this.

Sniper walks into screen, and go back to objective. Sniper's face, now "normal". His head is down.



Sniper is seen "[illegible]" for the first time. He walks around, then down. Pull back to reveal he is holding his case.

Cut to a wide shot of the entire area. Sniper is in the middle of a large crowd of people. Slowly keep pulling back and fade to black.

Camera pans from clock to the bed.

Camera finds bed empty. Continues to pan.

We see Sniper is already outside, looking into his case.

Sniper is staring down into case, then slowly raises his head and looks directly at the camera. Fade to white. The end.

Appendix C

Glossary

The following is a glossary of terminology found throughout this thesis, often associated with *italicized* words. These definitions span the variety of backgrounds covered in the preceding pages, including computer science, digital arts, film, painting, and history.

Anti-Aliasing A technique used to smooth the jagged edges in a bitmapped image. When diagonal or curved lines are put in bitmapped form, these shapes must be made with square pixels; any lines that are not vertical or horizontal have a stair-stepped appearance. Anti-aliasing changes the pixels along the edges of the line into varying shades of colour, in order to make the edge appear smoother.

Byzantine Era The era of the Roman Empire, beginning with the conversion of Emperor Constantine to Christianity in 312 A.D. and stretching to the 1453 and the fall of Constantinople to Ottoman Turks [8, p. 8,14]. This period is considered to be the height of the mosaic art form [40].

Cubism A revolutionary movement in early 20th-century art, particularly painting, pioneering abstract art forms. Its founders, Georges Braque and Pablo Picasso, were admirers of Paul Cezanne and were inspired by his attempt to create a new visual language. In analytical Cubism, three-dimensional objects were split into facets and analyzed before being “re-assembled” as complex two-dimensional images. [7]

Impressionism A theory or practice in painting especially among French painters circa 1870, of depicting the natural appearances of objects by means of dabs or strokes of primary colours in order to simulate actual reflected light [44].

Mise-en-Scène The arrangement of actors and scenery in the frame of the filmed image [44]. The relations and associations between the placement of actors and objects within the frame can be used to communicate deeper meaning and narrative connections within a film. Generally, there is nothing in a film frame that is not meant to be there, that is, not planned. Elements of the frame itself have a purpose in the act of constructing the meanings in a film [4].

Montage The production of a rapid succession of images in a motion picture to illustrate an association of ideas [44]. This juxtaposition between successive frames and sequences is one of the primary methods of conveying ideas and thematic associations within a film.

Opacity Opacity (opposite of *transparency*) is the measure of how opaque or *see-through* a graphic or image layer is [44]. The opacity of a graphic can be controlled using a graphics program by layering the image and controlling the transparency of the individual layers. The more opaque an image or layer is, the less it can be seen through.

Persistence of Vision The human eye holds an image for a short period of time after it has disappeared. If a series of still images is projected quickly enough, they merge physiologically and the illusion of motion is maintained. [24, p. 91]

Pixel Short for "picture element." Pixels are the small, discrete elements that together constitute an image, as on a television or computer screen [44].

Photorealism A style of painting in which an image is created in such exact detail that it looks like a photograph.

Pointillism The theory or practice in art of applying small strokes or dots of colour to a surface so that from a distance they blend together [44].

Pompeii An ancient city in southern Italy, southeast of Naples, that was buried under volcanic ash in 79 A.D. as a result of the eruption of the nearby Mount Vesuvius [44]. Many well preserved examples of decorative and artistic mosaics have been unearthed by the archaeological excavation of Pompeii.

Resolution Resolution is the number of *pixels* contained on a display monitor, expressed in terms of the number of pixels on the horizontal axis and the number on the vertical axis. The sharpness of the image on a display depends on the resolution and the size of the monitor.

Temporal Occurring through time [44]. In this case, used to distinguish between the static image, and the motion picture. A temporal image is one that has an associated component of time with its rendering.

Vita

Name	Craig Bradley Knowles
Place and Year of Birth	Bathurst, New Brunswick, Canada, 1975
Education	Queen's University, 2001-2003 Cleveland Institute of Art, 1999-2000 University of New Brunswick, 1993-1999
Experience	Teaching Assistant, Department of Computing and Information Science, Queen's University, 2001-2003 Teaching Assistant, Department of Computer Science, University of New Brunswick, 1997-1999 Teaching Assistant, Department of Geodesy and Geomatics Engineering, University of New Brunswick, 1994-1996
Publications	J. S. Shell, J. S. Bradbury, C. B. Knowles. "eyeCOOK: A Gaze and Speech Enabled Attentive Cookbook". Video Program (UbiComp 2003). Seattle, WA. October 2003. J. S. Bradbury, J. S. Shell, C. B. Knowles. "Hands on Cooking: Towards an Attentive Kitchen". Extended Abstract (CHI 2003). Fort Lauderdale, FL. April 2003.