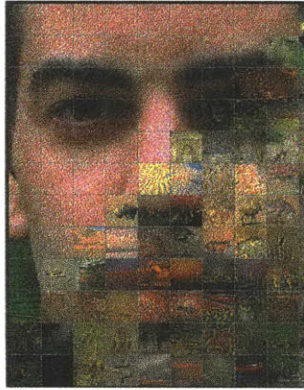# Photomosaics: Putting Pictures in their Place

by

Robert S. Silvers

B.S., Computer Science
University of Massachusetts, Lowell, 1991

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

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1996

Author ...

Program in Media Arts and Sciences
May 10, 1996

Certified by ....

Michael Hawley
Assistant Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by

Stephen A. Benton
Chairman, Departmental Committee on Graduate Students
Program in Media Arts and Sciences
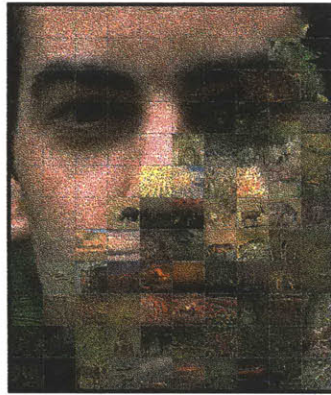
Photomosaics: Putting Pictures in their Place

by

Robert S. Silvers

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, on May 10, 1996
in Partial Fulfillment of the Requirements
for the Degree of Master of Science in
Media Arts and Sciences

Abstract

Photomosaics are mosaics composed of photographic images instead of solid colored tiles. A Photomosaic is a multi-scale image, conveying different information when viewed up close as when seen from a distance. From a distance, only the global subject of the mosaic should be visible, appearing without discrete components, thus the photographic images chosen must be *visually similar* to the local region of the over-arching image they represent. This seamlessness is produced by introducing *sub-picture resolution* — the consideration of regions within images. In addition to producing seamlessness, this resolution increases the potency of each tile, giving it greater global image forming ability. Up close, the Photomosaic should be seen as a matrix of individual photographs, taking the place of the larger image they collectively form. These photographs may be semantically related, directly, indirectly, or metaphorically, to the over-arching image in a global or localized fashion. This thesis presents the concepts, relevant research, essential technology, and creative applications of photomosaics.

Thesis Supervisor: Michael Hawley
Title: Assistant Professor of Media Arts and Sciences

[This page intentionally left blank]

Photomosaics: Putting Pictures in their Place

by

Robert S. Silvers

The following people acted as readers for this thesis:

Reader: ....
<div align="right">Hiroshi Ishii
Associate Professor of Media Arts and Sciences</div>

Reader: ....
<div align="right">Dr. Ken Knowlton
Computer Graphics Pioneer</div>

Table of Contents

List of Figures

Author

Silvers, 27, started taking photographs and playing with computers at the same time – about 18 years ago. He always thought of computer programming as a way to express himself much like he does with his photography. "Combining them was an obvious step," he says. Silvers has melded his interests to produce the images in this thesis.

Silvers received a degree in Computer Science from the University of Massachusetts at Lowell, and went on to work for NYNEX Science & Technology, the research group where he designed and implemented multimedia communications services for video conferencing and interactive television. In 1994, he entered the Media Laboratory at the Massachusetts Institute of Technology in Cambridge. Silvers may be reached by email at rsilvers@media.mit.edu, and has a photo gallery on the Internet accessible from the Media Lab's Web page (http://www.media.mit.edu).

Acknowledgements

A number of people helped make this work possible.

Thanks to Professor Negroponte and others for their role in creating the Lab – it has helped many people pursue their goals.

Thanks to Professor Muriel Cooper. Professor Cooper brought me to the Media Lab, and gave me the chance I needed. Both Muriel and Ron MacNeil helped build a foundation that included art as well as science, and taught me that no work is exempt from the importance of design and aesthetics.

Thanks to Professor Michael Hawley. Professor Hawley gave me the freedom to explore many wildly diverse ideas, and stirred up the pot with a continuous stream of creative input. He contributed to my projects, and provided the support I needed to carry them through.

Thanks to Professor Mike Bove, Professor Hiroshi Ishii, and Dr. Ken Knowlton for their direct contributions to this thesis.

A special thanks to my previous employers Sergio Canetti, Nathan Felde, and Tom Super at Nynex for providing a great environment, and for personally helping me in too many ways to list.

Lastly, thanks to Motorola and Toppan, who have supported my work at the Media Lab and have also been very good friends.
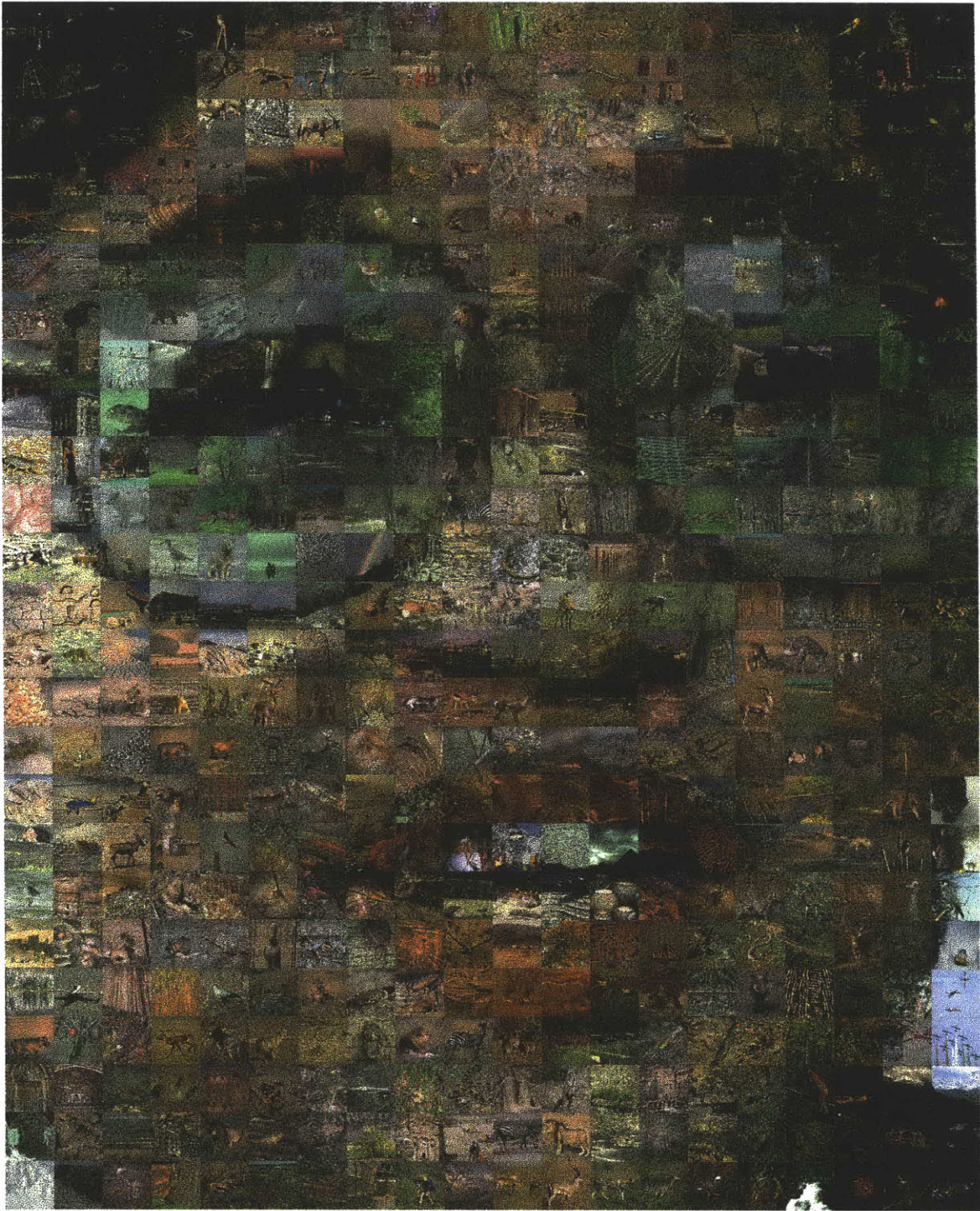
Kay Nishi, April 1996. 500 images were selected from 25,000.

Chapter 1

Introduction

1.1 The Goal

MOSAIC IS THE ART of decoration: variously colored pieces of material are closely set to form an image. Traditionally, mosaic is unlike intarsia or a jigsaw puzzle, for when disassembled, it cannot be simply reassembled by considering each component's shape. Mosaics have existed for thousands of years, and have been assembled from materials ranging from stone, wood, and stained glass, to dabs of paint, pixels of phosphor, and ink.

I have developed a process to create *Photomosaics*, new, more complex mosaics composed of digital images instead of solid bits of color. To make these mosaics, *visually similar* photographs, photographs that look alike when viewed from afar, must be located and organized so that one image may substitute for a region of another. This is the main technical subject of my thesis. My goal is to create mosaics from carefully selected and arranged photographs to allow a realistic over-arching image to emerge when viewed from a distance.

Photomosaics are *multi-scale* images, which when properly constructed, allow only discrete images to be perceived when viewed up close and only one seamless over-arching image to be discernible from a distance. This containment of multiple sources of information gives Photomosaics the power to play with the contents's relations. These relations may be ironic, metaphoric, surprising, or even inject a message.

To make Photomosaics, thousands of images must first be classified and stored in a large database. A *master image*, the global image represented by the mosaic, is then divided into a grid and analyzed section by section. Appropriate images from the database are selected during this process to become tiles in a particular grid space. These images will have attributes and features, such as similar colors, textures, or semantics, to the region being matched. The subject of each tile may also be relevant in some way to the master image or to the specific localized region of the master image.

1.2 Why this Project?

Mosaics have a time-tested charm. The digital techniques I have created attempt to preserve this charm. Photomosaics combine photography and computer science to create a new visual art form. They invite new possibilities such as applications in advertising, entertainment, and art work.

1.3 Possible Applications

Image banks may be sorted and shuffled into a pattern that communicates additional information. Important points may be conveyed by creating a Photomosaic with a relationship between the two scales of images – whether the connection is a direct, inverse, or a metaphoric one.

1.3.1 Advertising

Photomosaics are suitable for advertisements because they are novel, attention getting, and can relate content to the product or message in a powerful gestalt. Figure 1.1 shows a social statement from a clothing company, using a similar method by depicting the acronym AIDS made from human faces.

Figure 1.1: United Colors of Benetton Poster

Figure 1.2: *Where's Waldo* Example

1.3.2 Entertainment

There is a market for picture puzzles. *Where's Waldo* [1], shown in Figure 1.2, is a popular series of books by Martin Handford. Over five million of these books have been sold in the first two years. One searches for Waldo in a confusing array of similar looking objects. For example, the red stripes in Waldo's shirt may match a table cloth or apron. A Photomosaic could also contain hidden images. For instance, a skull could be made with tiles from Alfred Hitchcock's movie *Psycho,* and the challenge would be to find the tile containing an image of Hitchcock himself.

Figure 1.3: Painted Inlay by Kit Williams

Another visual puzzle is Kit Williams's 1984 book [2], which has no title. The title was left for the reader to discover, and express without words, as part of a contest. The person with the most original entry won a fabulous prize. Each page in the book is an image of painted and inlaid wood, and offers a clue (see figure 1.3). Similar mysteries and challenges could be encoded into Photomosaic art work.

1.3.3 Art Work

The history of art shows an ongoing interplay between different media. Mosaic emulated painting by using subtle and realistic hues before taking on its own unique characteristics, such as the use of bright colors and absence of middle tones. At this point paintings began to emulate mosaics. Photography provides a similar example. When photography was first introduced, it was not considered to be a serious art form like painting, as the photographer merely captured reality. Only later did photography become its own art form, with room for individual style and expression. Computer graphics borrows from painting (automatically generating Mirós, Mondrians, or painterly effects), and painting occasionally borrows from computer graphics. Figure 2.12, a painting by Salvador Dalí, shows an image of President Abraham Lincoln, based on the blocky computer graphics image of Lincoln made by researcher Leon D. Harmon. Photomosaics, in turn, can draw from painting, photography, mosaic, and computer graphics to develop its own unique artistic identity.

Chapter 2

Mosaics, Other Composite Images, and Computerized Selection Techniques

PHOTOMOSAICS ARE GENERALIZED MOSAICS. If the techniques used in earlier mosaic forms are sorted in chronological order, their evolution can be seen as a function of the latest technology enabling new possibilities.

2.1 Conventional Mosaics

2.1.1 Development

Mosaic is the art of using closely set colored materials, such as tile, glass, or stone, for decorating a surface. Mosaic tiles differ from intarsia (fitted wooden inlays) in that the shapes do not relate to a large portion of the image. Through combining many small pieces of various colors, a large image may be formed. This larger image cannot be assembled like a puzzle simply by seeing the individual pieces.

The earliest discovered mosaic art work is from the 8th century BC, and mosaic became the leading pictorial art in Byzantium from the 4th to 14th century. Early mosaics were constructed of pebbles which limited the range of colors. Larger pebbles often were used as filler in areas of lower detail, while smaller ones resolved more critical sections. This technique was well developed by the 5th century BC. By the 4th century BC, pebbles were painted to increase the range of available colors. It was at this point that mosaic started to compete with painting for its ability to convey a sense of space and detail. At the beginning of the 3rd century BC, cut pieces replaced the pebble technique and new materials such as glass enriched the medium.

Figure 2.1: Early Roman Floor Mosaic

Mosaic moved to Roman soil by a process of diffusion, and developed into a rich art form in the 3rd century AD. Figure 2.1 shows a Roman floor mosaic from 240 AD with color and shading as elaborate as a painting. Mosaic style, in fact, paralleled that of paintings during the growing use of mosaic through the 4th century. Mosaic advanced during the Roman imperial times of the 4th through 6th centuries when it took on unique characteristics. While the first mosaics had gradual painting-like changes in color made possible by both painted and unpainted pebbles, the newest designs focused on higher contrast and greater brilliance. Many of the middle tones were eliminated altogether. Once this new style was developed and mosaic became popular in church decoration, painting itself started to be influenced. Byzantine and western European paintings of the Middle Ages had greater contrast and brilliance, which may have been influenced by mosaic art.

Mosaic did not sit still. One important advance in mosaic technology is called *tessera*. Tesserae are pieces cut into triangular, square, or other shapes with a size that varied with the level of detail needed in a given region of the art work. Figure 2.2 is a close-up of the empress Theodora created in 548 AD for the church of San Vitale. The bright decoration and colors are in the Hellenistic Roman tradition yet the tesserae form the face with a gentle shading more common in an earlier period. Mosaics in the 10th century, the height of development, took advantage of the unique characteristics of the medium, and materials were chosen based on their level of appropriateness for the task. Glass makers were now producing glass in thousands of tones. Mosaic artists, in an effort to compete with painters of the Renaissance, often duplicated their work and some

Figure 2.2: Empress Theodora

times went a few steps further. Images were made that produced true dimensional effects that, when combined with the rich colors and materials available, went beyond what was possible in paintings.

2.1.2 Materials

Materials used in mosaic art varied through the phases of development. Stone was one of the first materials because of its availability. Uncut pebbles of uniform size became floor decorations in the earliest Greek and Roman works. The range of tint, of course, was limited to natural hues. Glass, first used in the 3rd century BC, could be made rich in color and brought a wealth of new possibilities to the art. By the 5th century, mosaics were often constructed entirely of glass. Since glass is brittle, mosaic using these rich color schemes often became wall art, for durability. Materials were also used to signify semantic information. Gold leaf, for example, was chosen not only to enhance the range of colors, but also to signify light emanating from God. Silver signified light from Christ, and often was restricted to this use. In the 6th century, these metallic pieces were being set at an angle to enhance the light reflecting effects. It was ceramic, however, that became the most popular material in modern mosaic art due to its versatile nature.

2.1.3 Technique

There are many considerations when planning a mosaic, including the overall design, materials, lighting, tile size, and shape of cut. This requirement of a broad range of skills has limited the art to those with a firm grasp of technology, not unlike computer art today. Mosaics were made by first drafting a sketch and detailed outline, which often included the colors desired, on to the surface. When the tiles were set, their texture and hence their appearance was often defined by the location of the mosaic. A floor mosaic, for example, must have more tightly set pieces for increased durability and perhaps for the comfort of bare feet. Wall mosaics, therefore, had more room for artistic control because they could be closely or loosely set, and made from a wider range of materials. The tiles in any mosaic are placed in one of two ways – either one by one directly onto the surface, or by an indirect method of building sections at a time and then placing them into a larger composition. Most mosaics made today are by the indirect method. It should also be noted that earlier mosaics were often designed and constructed by the same artist. Later ones, including most made today, were not set by the artist or designer – perhaps lowering the overall quality of the finished art work.

2.2 Image Mosaics

The process for Photomosaics is unique, but this is not the first time images have been tiled into a mosaic. Following are some examples of previous work.

Figure 2.3: Kai Krause, Mona Lisa made from Mona Lisas

2.2.1 Kai Krause - PixnPix

Krause (http://www.metatools.com) has created images made from smaller, modulated images, and calls the process PixnPix. Figure 2.3 shows one of Krause's first attempts – a Mona Lisa made from tiny stylized Mona Lisas. It is important to note that each of the smaller images are identical except that the brightness has been altered.
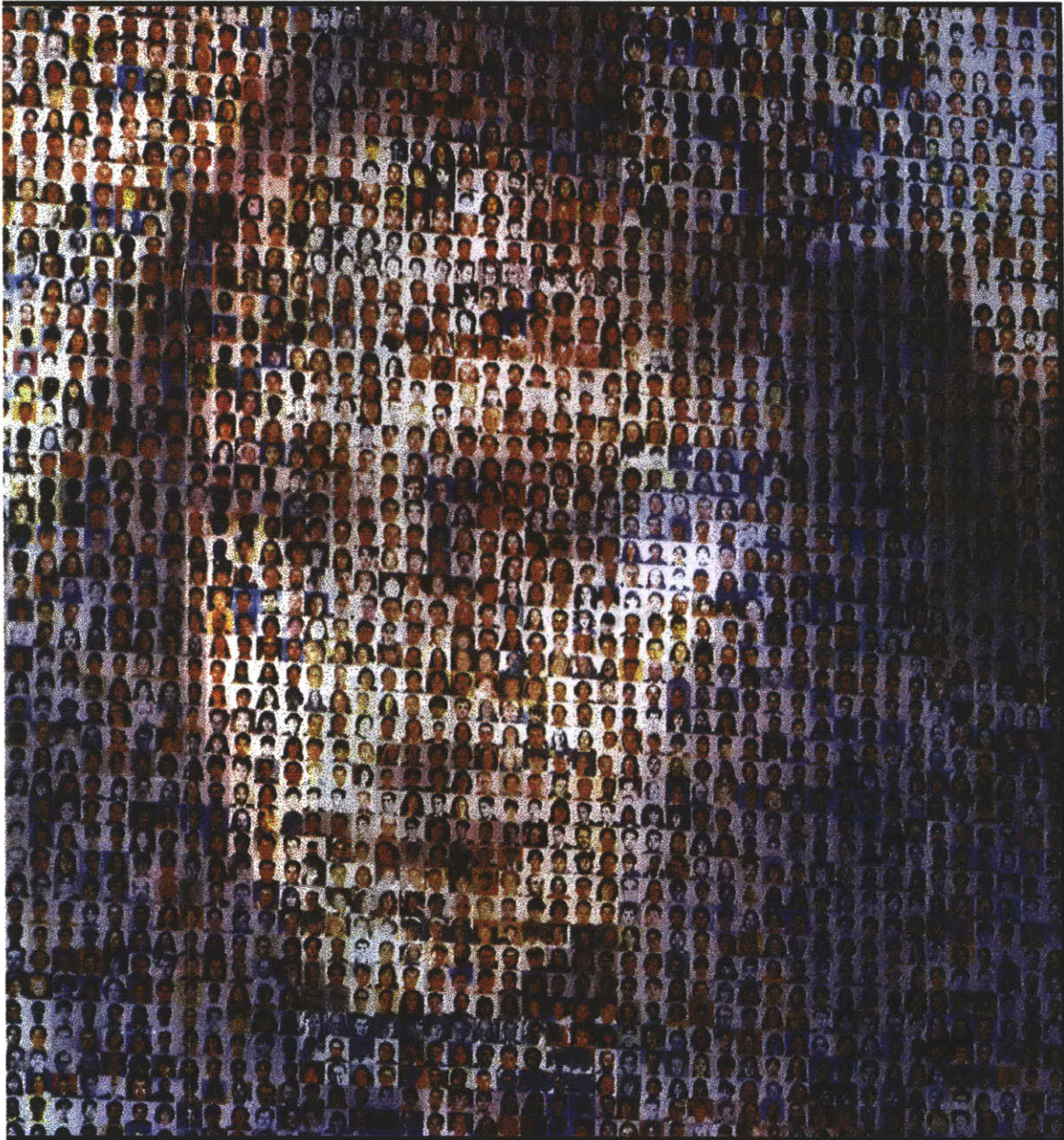
Figure 2.4: Mona Lisa made from 10,000 Passport Photos

2.2.2 Mona Lisa made from Passport Photos

Ramón Núñez Centella had the idea to use ten thousand passport photos and a computer to render an image of the Mona Lisa (partially seen in figure 2.4). This image, called *Gioconda Sapiens,* can be seen at Domus, The House of Man – an interactive science museum in La Coruña, Spain. Notice how the color space is very limited because of the choice of materials (human faces) – a similar problem faced by ancient mosaics made from pebbles.
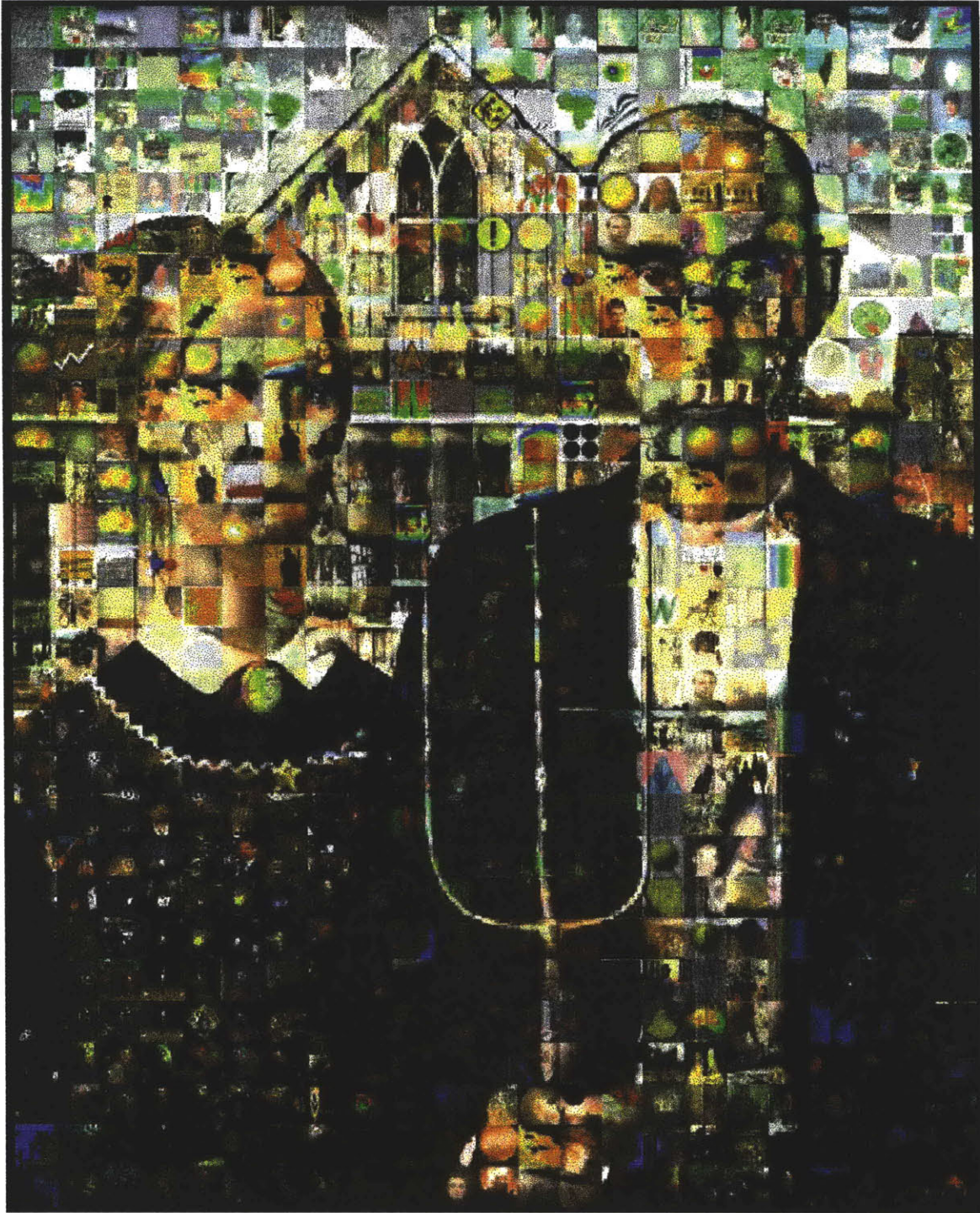
Figure 2.5: Adam Finkelstein, *American Gothic*, 1995

2.2.3 Adam Finkelstein

Finkelstein (http://www.cs.washington.edu/homes/adam/) created image mosaics as a graduate student at the University of Washington. Figure 2.5, one of the several images he created, is an adaptation of the

oil painting *America Gothic* by Grant Wood. Finkelstein made it in early 1995 by collecting images from the Web and arranging them to suggest the underlying image. The grayscale values of each sub-image were altered to be more compatible with the main image. The result appears to be a 50 percent blend of the painting and tiles.

2.2.4 Alain Omer Durançeau

In early November 1995, Durançeau of Canada saw the Photomosaic cover that I did for WIRED magazine and sent me an earlier mosaic that was also composed of images. The difference was that he produced his *by hand*, with enormous patience and labor. Mr. Durançeau used a scene from Georges Méliès's *A Trip to the Moon* as his target image.

To achieve such a feat, Durançeau painstakingly digitized scenes from movies, then carefully processed and hand placed the frames (see figure 2.6).

Being from movies, the mosaic tiles are related to the theme, and the contents of each sub-image match the contours of the master image. Durançeau considered shape and color of sub-objects *within* tiles, not just the overall color. It was this image that motivated me to create more intelligent software. My goal became to produce results that were on par, while taking advantage of the computer's ability to thoroughly consider many more variables than a human artist.
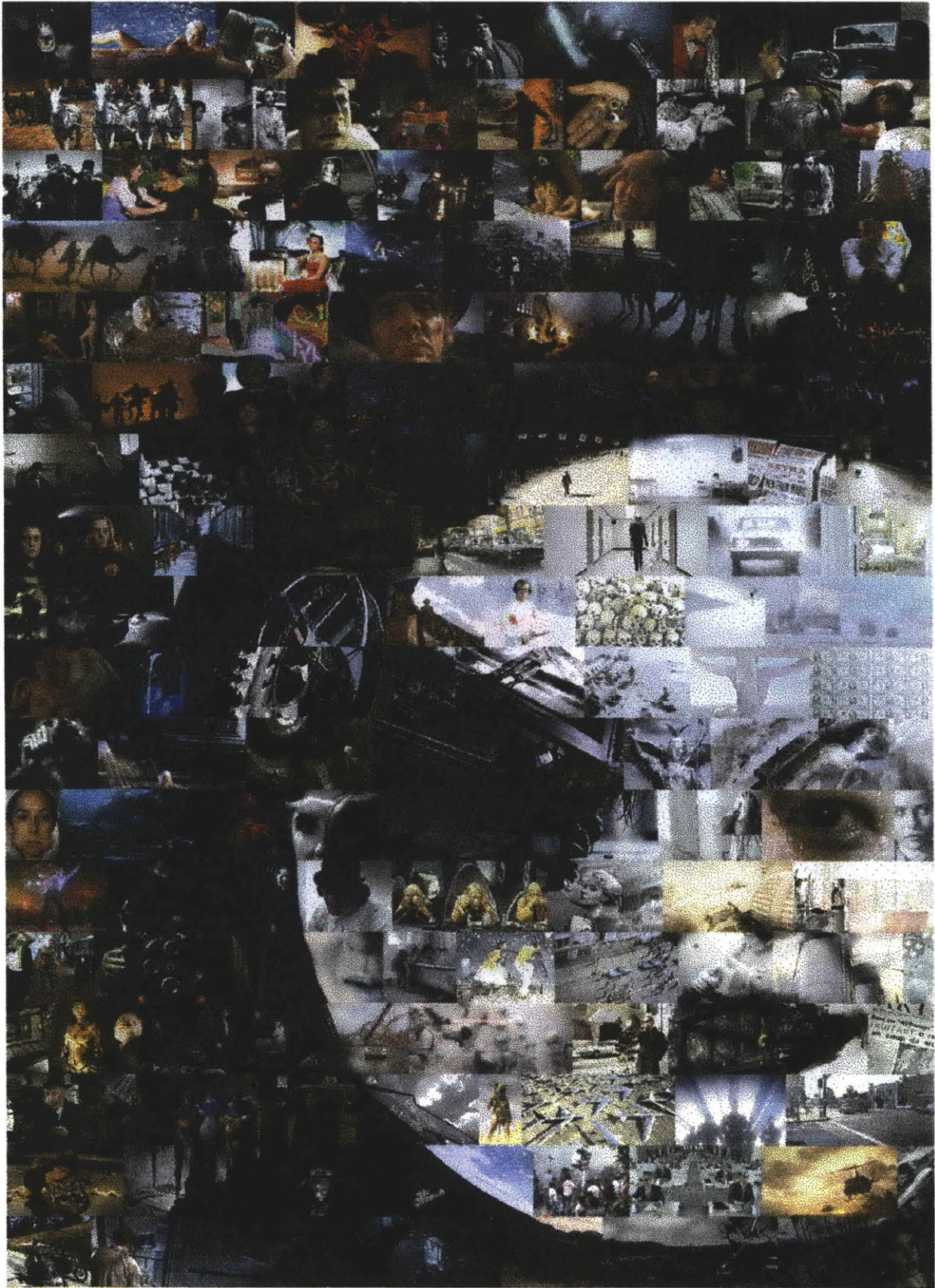
Figure 2.6: Alain Omer Durançeau, Hand-Created Image Mosaic, 1995

Figure 2.7: Giuseppe Arcimboldo, *Vertumnus*, 1590 or 1591

2.3 Other kinds of Composite Images

There are many non-mosaic art forms that achieve similar goals to Photomosaics. Perhaps they make an image from other objects, or images of other objects, that have some relationship to the overall theme.

2.3.1 Giuseppe Arcimboldo

Arcimboldo (1527-1593) was an Italian painter who rendered portraits as a carefully arranged collection of other objects. Images of things such as fruit, vegetables, flowers, animals, and fish were used in combination to form something else, perhaps a face, such as in figure 2.7.

Figure 2.8: Arthur S. Mole, Living Photograph

2.3.2 Arthur S. Mole

Mole, born in 1889, composed "living" images by organizing thousands of people into patterns that form another image. These images are interesting for several reasons: the difficult nature of involving so many people, the artistic and creative ideas that they evoke, and because of the social record that they form. "The Cross and Crown," Mole's first image, was organized in 1913 and included 1500 men and women from the congregation of the Christian Catholic Church [3]. Mole's first widely published image was "The Human American Flag," made in 1917 for Memorial Day at the United States Naval Training Center in Great Lakes, Illinois.

These images were created by Mole standing at the top of a tower with a megaphone and an 11 x 14 inch view camera. He would have the shape drawn on the ground glass of the camera, and then direct the men to stay within the lines. This allowed the perspective of the image to remain correct even though the men were not photographed from directly above. Mole's most challenging image was the Statue of Liberty made at Camp Dodge in Des Moines, Iowa (figure 2.8). It contained 18,000 men (the largest image was of a US shield with 30,000 men), and required complex detail and control over perspective to render. It is interesting to note that the final formation used for this image was 1,235 feet long and 12,000 men were required to fill the flame (because of their distance from the camera) while only 17 were in the bottom line. How did the participants feel about having to stand for two or more hours to make an image? They seemed to be excited to become a part of history [4]. In a 1968 interview with Mole, he described how excited he was to first see an image form right before his eyes [5]. The power to move armies is exciting, I would imagine.
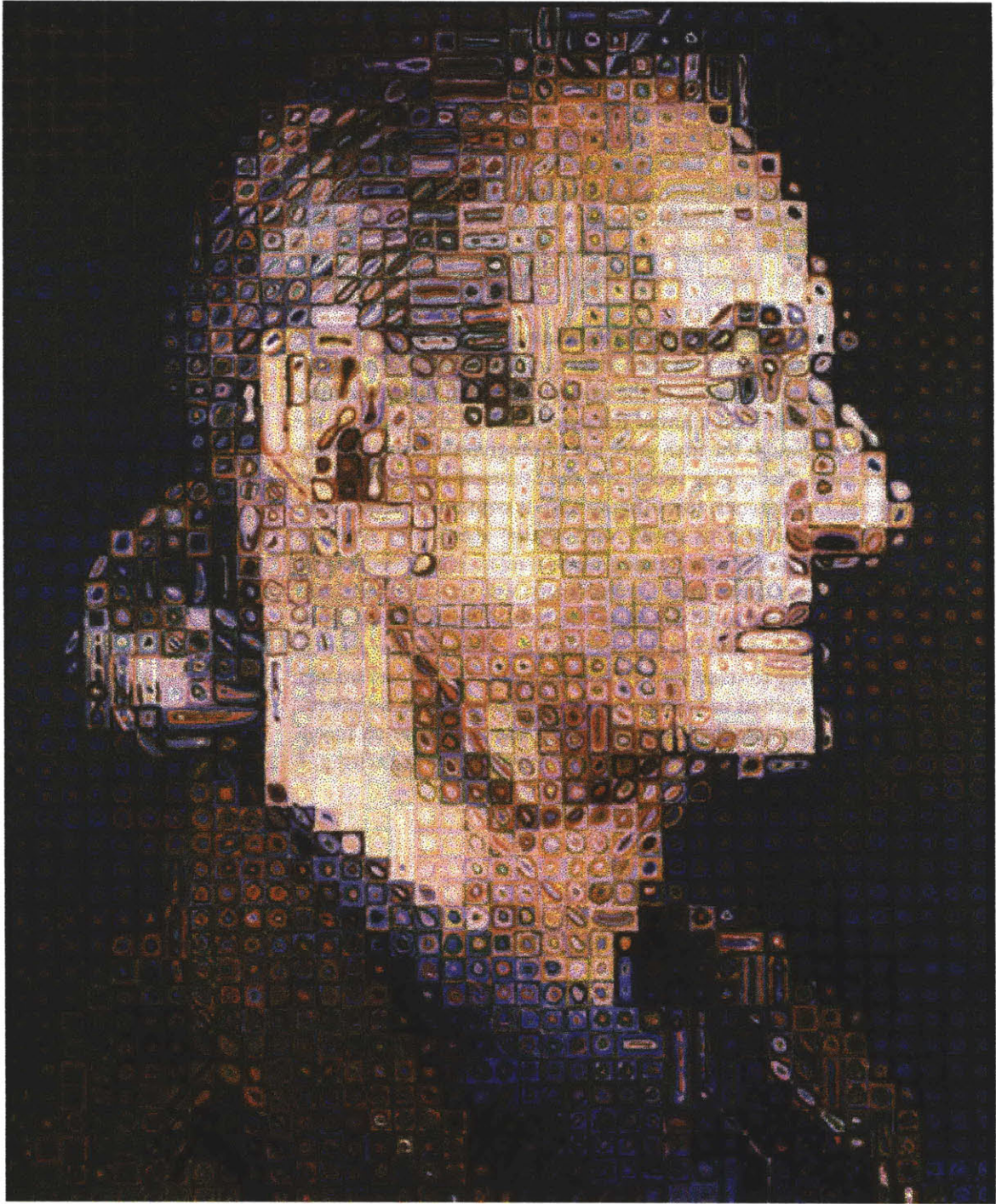
Figure 2.9: Charles Close, *Roy II*, 1994

2.3.3 Charles Close

Close (1940 -) is an American artist who gained attention during the mid to late 1960s as a member of the photorealist movement of painters who based their work on photographs. Mr. Close creates a grid on a real photographic image and then copies the value of that area onto canvas by interpreting it as an abstract shape. Circles, lines, stars, and diamonds are abstractions that become realism in his work. Figure 2.9, created in 1994, is 102 x 84 inches.
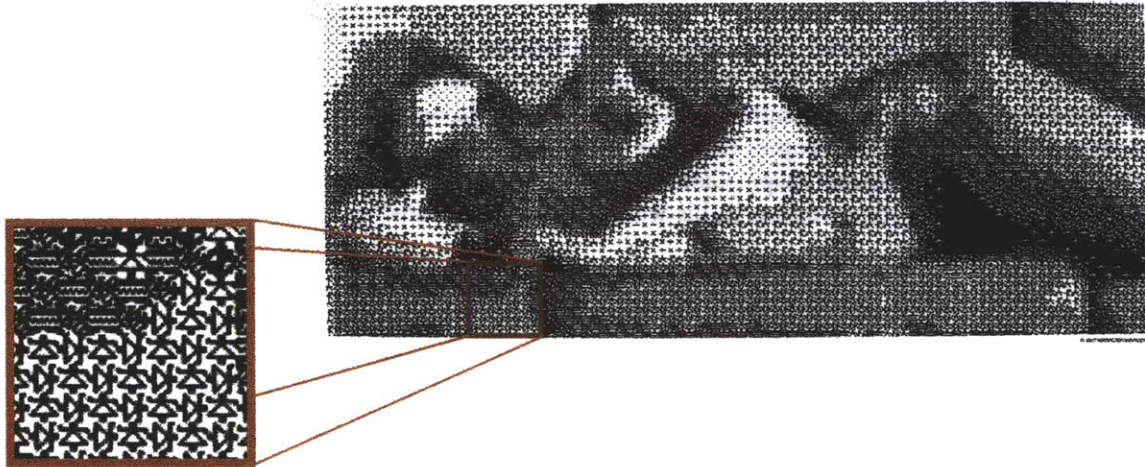
Figure 2.10: Knowlton and Harmon, Nude, 1967

2.3.4 Knowlton and Harmon

In 1966, while at Bell Laboratories, Ken Knowlton and Leon Harmon produced grayscale images from binary (pure black and white) symbols by considering black-dot occupancy on an 11 x 11 matrix [6]. These images were more interesting than ones which simply used dots because they contained symbolic patterns on two levels. From up close, only the pictographic symbols were visible. From a distance, a larger image formed and the smaller pictograms disappeared. Figure 2.10, one of the first they produced, was an attempt to bridge the gap with classical art. Picture processing by computers was so rare at the time, that they used the term "supernumerical" to mean an application for computers not directly related to mathematics.

Both Knowlton and Harmon continued with their research and art. Harmon made the famous blocky image of Lincoln using 14 x 18 pixels and 16 levels of gray in 1970 for research on face recognition [7]. Knowlton experimented with type to convey imagery and started producing images such as that in figure 2.11 in 1971 [8]. The actual Liberty image shown, however, was made in 1986 using an identical technique.

The New Colossus (by E. Lazarus) Not like the brazen giant of Greek fame, / with conquering limbs a-stride from land to land; / Here at our seawashed sunset gates shall stand / A mighty woman with a torch whose flame / Is the imprisoned lightning, and her name / Mother of Exiles. From her beacon-hand / Glows world-wide welcome; her mild eyes command / The air-bridged harbor that twin cities frame. / "Keep ancient lands, your storied pomp!" cries she / with silent lips, "Give me your tired, your poor, / Your huddled masses yearning to breathe free, / The wretched refuse of your teeming shore. / Send these, the homeless, tempest-tost to me, / I lift my lamp beside the golden door!" [This poem was written in 1883 by American poet Emma Lazarus, and in 1903 it was inscribed on a bronze plaque in the pedestal of the statue of Liberty, a gift of the French people to the United States, designed by Frederic Auguste Bartholdi using an iron framework by Gustave Eiffel, dedicated in 1886 by US president Grover Cleveland on Liberty (then Bedloe's) Island, New York Harbor. An extensive restoration of this statue has been coordinated with its 1986 Centennial celebration] ©1986 Ken Knowlton Sunnyvale CA

Figure 2.11: Knowlton, Liberty made from Fonts, 1986
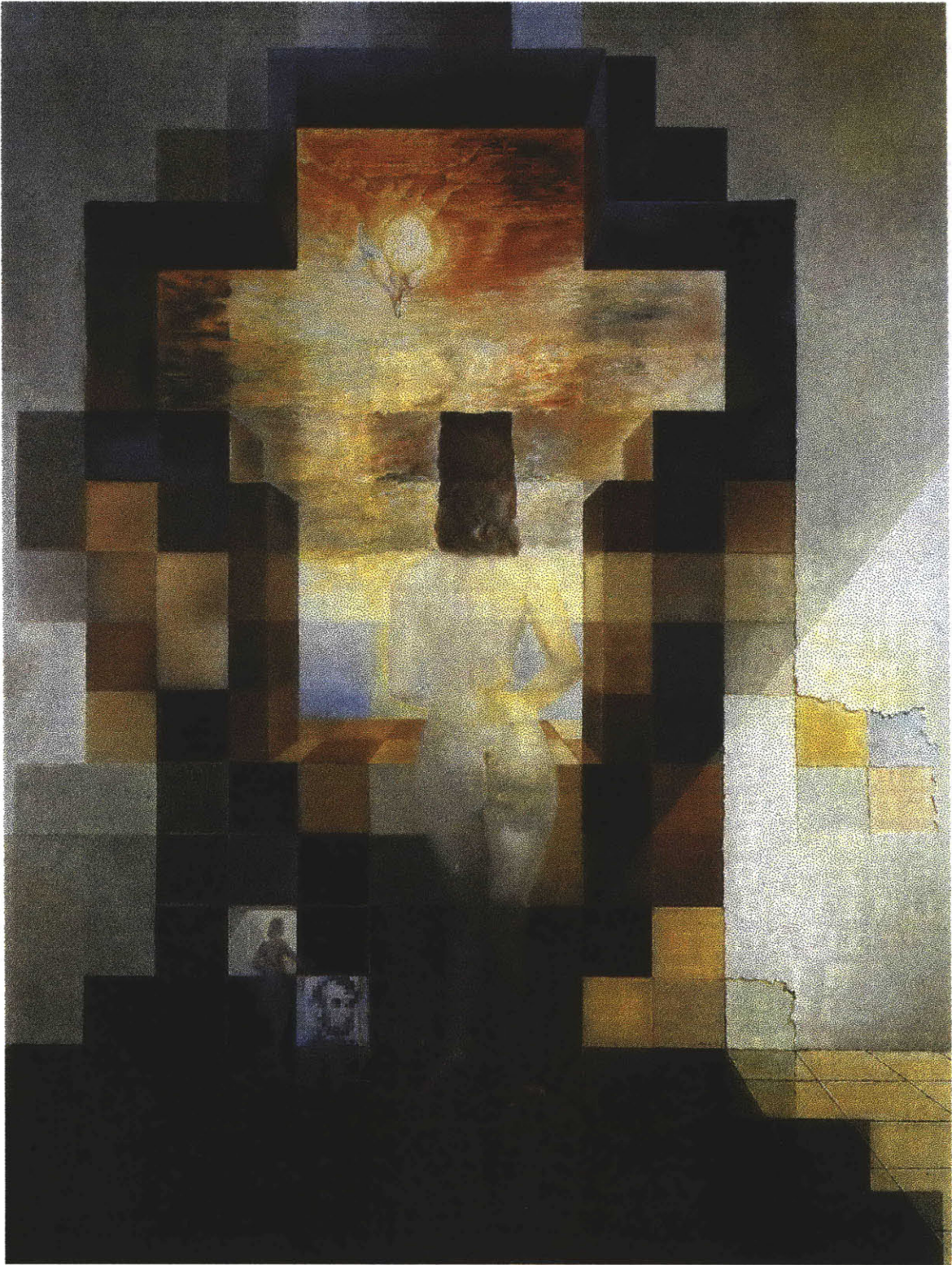
Figure 2.12: Salvador Dalí, *Lincoln in Dalívision*, 1976

2.3.5 Salvador Dalí

Dalí (1904-1989) was a Spanish painter who created classic and surrealistic images. Figure 2.12 shows a particularly relevant 1976 portrait of Lincoln made from other objects, including a woman named Gala. In the lower left quadrant is the 1970 Lincoln image by Leon Harmon that inspired this painting (causing controversy over intellectual property because Harmon believed that Dalí borrowed too much [9]).
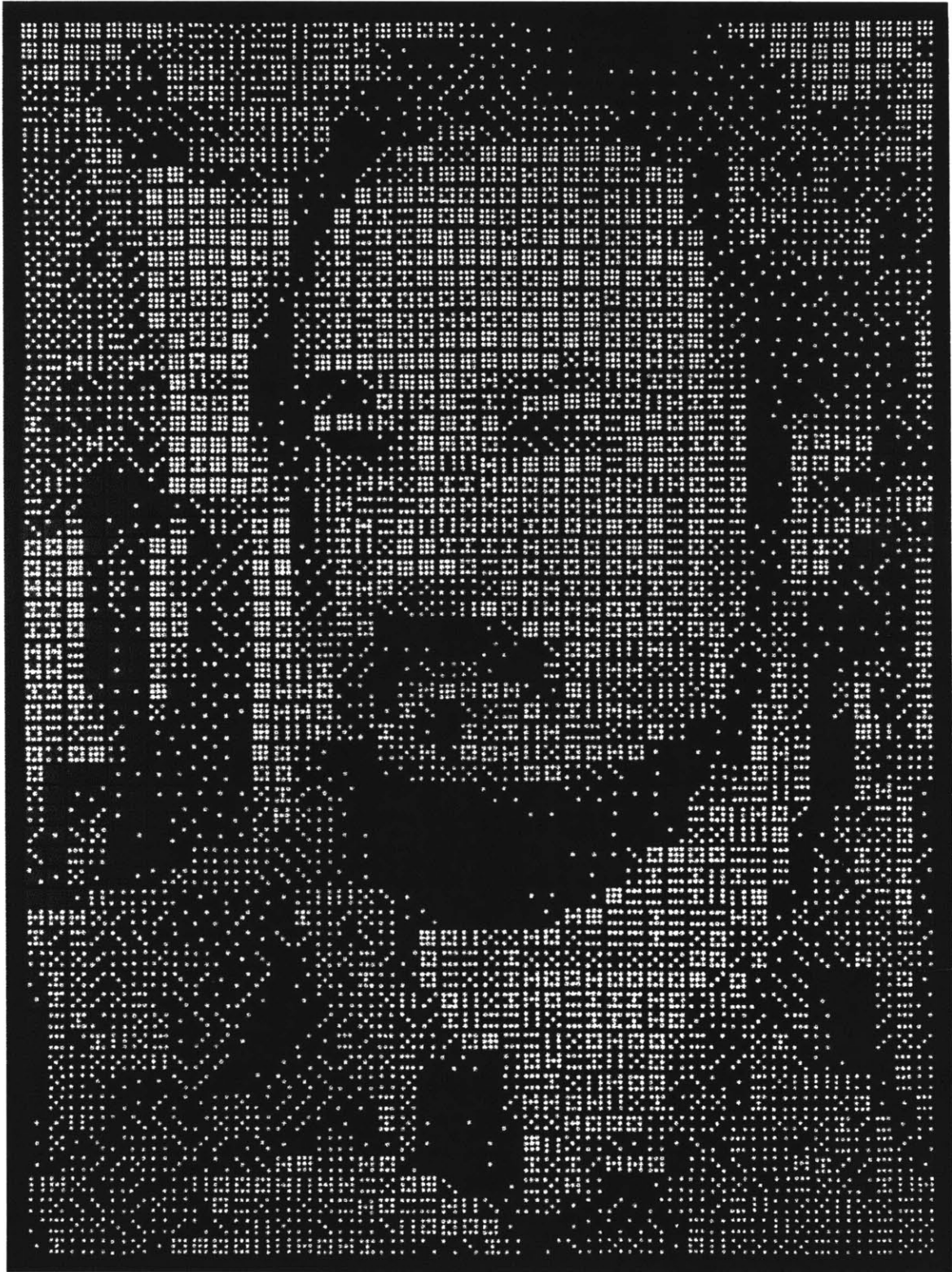
Figure 2.13: Ken Knowlton, *Domino Player*, 1980

2.3.6 DominoPix®

Ken Knowlton (1931-) invented DominoPix,® a process for creating images from complete sets of Dominos. Figure 2.13 is an image of Joseph Scala Jr. made in 1980 from 24 sets of double-nine dominos (1320 tiles). A business was formed to merchandise these designs in which several sets of Dominos came with plans to assemble various images. Custom images – perhaps a portrait, could be prepared on demand. This process is protected under US Patent No. 4,398,890, REPRESENTATIONS OF DESIGNS – a method for making images from complete sets of physical objects. It was especially important to Knowlton that the images contain only complete sets because of the additional challenge this constraint presented.

Figure 2.14: Ken Knowlton, *Cousteau*, 1987

2.3.7 Knowlton Seashell Portraits

Ken Knowlton's seashell images are not renderings but rather physical objects glued to a background. Figure 2.14 (36 x 48 inches) shows a portrait of Jacques Cousteau that was commissioned in 1987 by the San Francisco Exploratorium. Figure 2.15 (32 x 26 inches, 1993) is another example, but with non-grid alignment. It was Knowlton's work that inspired me to create Photomosaics. Around 1990 I visited Wang Laboratories where Dr. Knowlton worked and saw one of his creations. It was a seashell portrait of Dr. Wang – very large and hanging at the end of the hall. Since I viewed it from a distance, I only saw the face and had no idea what it was made from. Upon closer examination, the face disappeared, and was replaced with hundreds of seashells. This led me to think about seeing on multiple scales. Since I was a photographer, I thought about creating such an image from photographs instead of physical objects. It would have been an enormous task to scan enough images at the time, but the idea was forming.
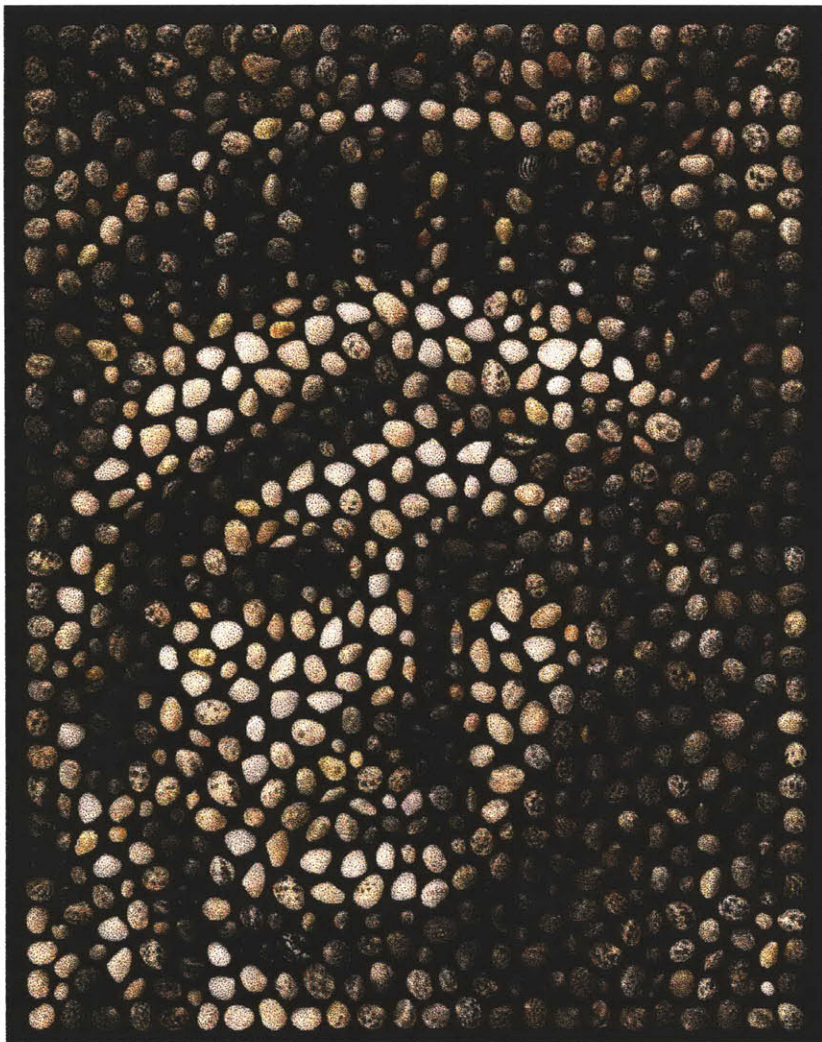
Figure 2.15: Ken Knowlton, *Liberty*, 1993

43

Figure 2.16: Michael Massey, *Marvin Minsky*, 1995

2.3.8 Salient-Stills

Salient Stills, a process created at the MIT Media Lab by Laura Teo-
dosio, Walter Bender, and later improved by Michael Massey, turns
video into a still image [10]. The process, first presented in 1993, creates
a structured representation of a motion video sequence, such as changes
of camera pan, zoom, or subject movement. This representation is then
used to render a still. The resulting image has increased resolution, lower
noise, and added contextual information. The image in figure 2.16, cre-
ated by Michael Massey, is a high-resolution panorama captured with an
ordinary camcorder that panned across the scene and zoomed in to the
main subject. The resulting image has both the wide field of view of the
short focal length frames and the detail captured by the long focal length
frames combined in harmony.

2.4 Survey of Image Selection Software

The automatic selection of images is often done on boiled down representations of the images. This reduction of complexity may either be pre-computed or done during the search itself. It makes sense to pre-compute as much as possible without limiting the range of comparisons because search time will be reduced. Reference [11] proposes *semantics-preserving image compression* which reduces an image to a compact representation that still preserves similarity of content. This representation is so perceptually complete that it is possible to reconstruct the image from its coefficients.

Images may be searched at a high level of abstraction, or on a more raw image processing level. The difference is not important for Photomosaics because of the multi-scale nature, i.e., the match only needs to look similar from a distance without a need for understanding of what the contents are. In other words, a match between a pancake and a brown frisbee of the same size would be better than a match between two frisbees of different size and color. The technique that produces the most visually similar (not necessarily semantically similar) results in the shortest amount of time is best. The matching systems I have considered for the Photomosaic process include ones that search based on color, texture, edge, shape, and RMS difference of color values within sub-regions. I use this last method of RMS error because of its resulting high degree of visual similarity and low relative computational cost.

2.4.1 Color Histogram

Color histogram matching is one metric to consider when comparing two images. References [12][13][14] describe methods of using color for classifying and matching images. A histogram, perhaps with 64 buckets, could be constructed for an image or sub-region of an image. The generated histograms would then be compared and the relative error returned.

Matching histograms is relevant in that from a distance, a red colored image, perhaps of a tongue, would appear visually similar to another red colored object, such as an apple. That semantic content is not considered is more of a shortcoming for other applications than for Photomosaics. This technique is not ideal, however, because it does not consider aligned areas of light and dark to be significant, and hence it will not find matches that are as visually similar as if aligned features were matched.

2.4.2 Texture Matching

Texture is an important characteristic of rapid image understanding [15], and therefore could be used to rate the similarity of images. For applications involving finding *visually* similar images, and not just matches that are similar on a numerical level, it is important that the results match features that humans consider significant [16]. An example of a metric that is not necessarily related to visual similarity is per-pixel comparisons, so it is important to avoid high resolution matched filtering because images can differ at every pixel and still be perceptually similar [17].

45

What metrics should be considered for matching images based on texture? Rao and Lohse [18][19] identify relevant texture features as "repetitiveness," "directionality," and "granularity and complexity," while "periodicity," "directionality," and "randomness" were proposed as features by Liu and Picard [20]. Repetitiveness is the degree to which features are repeated. Directionality is the orientation of features in magnitude and phase. Granularity and complexity are the size of features, and the variation of identifiable properties. Periodicity is related to repetitiveness, and randomness is related to granularity and complexity. By quantifying the relative amount of these features and comparing them to other images in a database, you may be able to find visually similar images. The degree to which the computer finds images similar should agree with human perception, so testing with human subjects is important. For example, texture directionality is an important visual cue [21], and Liu and Picard [20] find the direction of the lowest fundamental frequency the most perceptually dominant – not the direction of the frequency with the greatest computed magnitude. Gorkani [22] notes that contrast compensation may be needed to find the texture orientation in a way that best agrees with human test subjects. Texture matching will work as a Photomosaic classification system, but is unnecessary as a level of abstraction because RMS error over sub-regions produces the required level of visual similarity with more efficiency.

2.4.3 Edge Matching

Edge matching has been used as a way to compare a human made sketch with images [23]. In this example. the collection of images were first reduced to edges, and then thinned to produce lines for meaningful comparison with the sketch. Because there is often a lack of unique feature information within a small area of interest and since the primary advantage of this technique is to provide a high-level interface for humans to describe an image, edge matching is not appropriate for Photomosaics.

2.4.4 Shape Matching

Shape matching is a good way to uncover visual similarities, but is a difficult problem [24][25]. If a number of classifications are made, they can be compared between images. For example, IBM's paper [26] considers area, circularity, eccentricity, and major axis orientation. These parameters can be sorted and stored in a table for a high speed (binary) search algorithm to reference. Pairs of images with features of similar dimensions and position would return a higher score. Shape matching at this high a level does not seem to offer a large advantage for Photomosaics (where content of small images does not matter because it would be obscured by viewing distance) over simply comparing color values within sub-regions.

2.4.5 RGB/HSV in Sub-Regions

RGB/HSV matching finds the RMS error between two images within N by N sub-regions. This efficient technique can be used as the universal matching routine for producing visually similar images. If, for example, the number of sub-regions are set to something like eight or more, object shape and color within images will naturally be considered. An image that fades from red to blue, for example, would best match another that fades from red to blue. Reference [11] agrees that RMS error provides a useful metric for perceptual similarity based on shape or appearance. While this method does not work well as a way for humans to provide a high-level description for an image they desire (such as with a sketch) [23][27], it does result in a close perceptual match when viewed from a distance, and hence is what is used.

Chapter 3

The Photomosaic Process

3.1 Introduction

ON APRIL 28, 1995, I digitized 370,000 photographs from National Geographic's collection for a class project with Professor V. Michael Bove. My software cropped each image, scaled it to a small size, and classified it by average color. Then, for each pixel in a master image, my software found the image in the 370,000 picture database with the closest average color to that pixel.
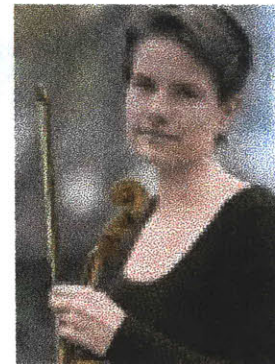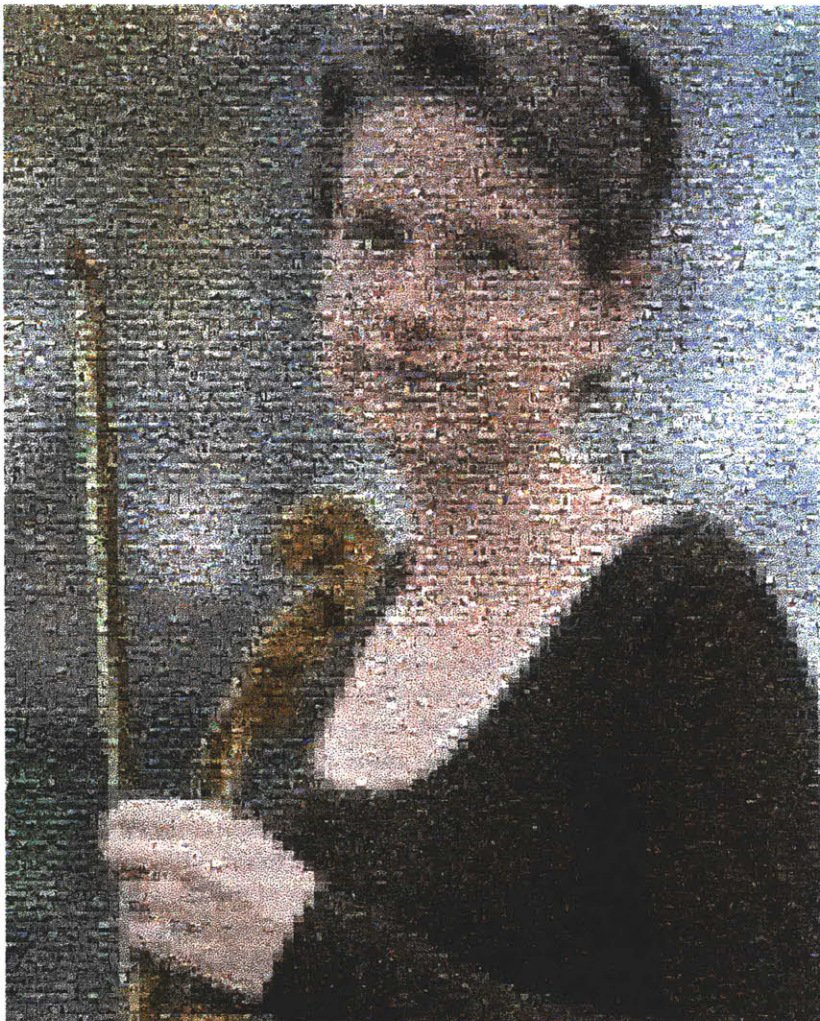
Figure 3.1: Rob Silvers, *Julia*, 1995

The first successful test image was completed on April 30th, 1995. The first real image mosaic, as seen in figure 3.1, was a portrait of Julia Ogrydziak and was created on May 2nd, 1995. This image illustrated the

possibilities of creating mosaics from photographs. After discussions with Professor Bove, I decided to address the problems which had emerged in my initial mosaic process. What should one do when a subimage is not a perfect fit? In the first version, I shifted the actual average color values of each subimage to the desired average value. Altering the images, though acceptable in art, violated my technical inclinations. Another solution was to apply error-diffusion dither to the process. In this process, if an image did not match perfectly, the error was stored so the software could over-compensate with images to the right and below. This enhancement worked as of May 16th 1995, though it often produced a noticeable checkerboard effect because of the spatial resolution being sacrificed for improved color resolution. My first published image (also created on May 16th), a portrait of Jerome Wiesner for the Media Lab publication *Frames*, used this dithering process with no ill effects. See figure 3.2.

Figure 3.2: Rob Silvers, *Jerome Wiesner*, 1995

My initial results were promising, but I knew it was possible to do better by assigning more parameters to an image. I reclassified all of my database images to include the standard deviation of their pixel's brightness values (texture contrast). I then modified my matching program to use a master image with its height and width in pixels equal to the number of vertical and horizontal mosaic image tiles multiplied by the height and width in pixels of the database images used for matching. In other words, the master image now had more horizontal and vertical pixels than mosaic tiles needed for the final rendering, thus allowing the software the opportunity to compute an identical texture contrast coefficient at each grid-spot for comparison with the database. This enhancement was first used on May 17th, 1995. For the cover of the November 1995 issue of WIRED magazine, a mosaic was created using "smoother" images (less deviation of pixels) for smoother areas, such as the face, while "rougher" ones (greater deviation) fell into noisier places, such as the hair. The image was finished Aug. 29th. Figure 3.3 shows the master image next to the final 1,548 tile mosaic.
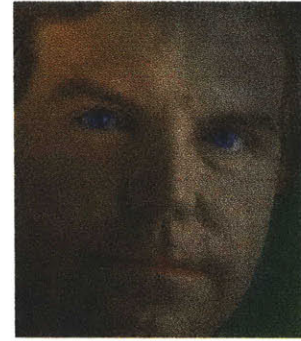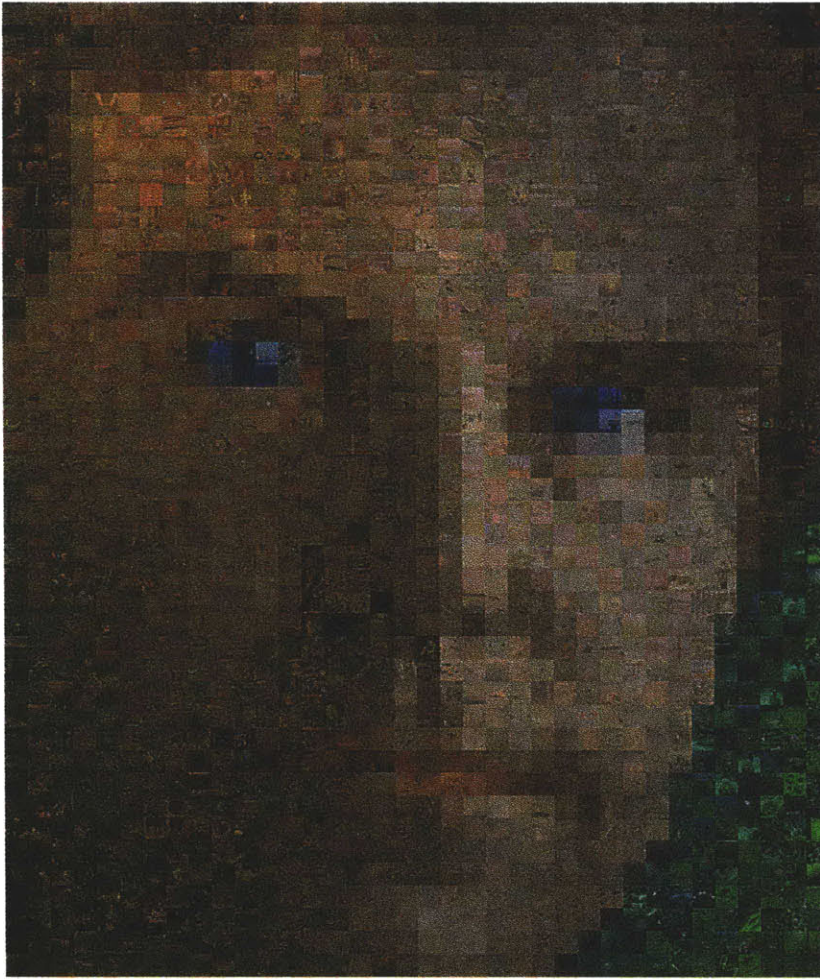
Figure 3.3: Rob Silvers, *Nicholas Negroponte*, 1995

For the Media Laboratory's 10th anniversary, I created a poster of hands (Julia Ogrydziak's) counting to ten (figure 3.4). This was the largest image yet, containing 5,400 photographs. Texture was considered, and the images were slightly tinted instead of using dither, avoiding the checkerboarding visible in the WIRED mosaic.[i] A green to blue wash was selected for the background, as green and blue images were featured prominently in the database. This increased the quality of matching for the large area. Toppan, a Media Lab sponsor, generously printed 4,000 copies of the poster, but the poster was never released due to NATIONAL GEOGRAPHIC's wishes (see "NATIONAL GEOGRAPHIC," page 74).

[i] Decision to tint or to look for another solution was made under the pressure of a production schedule.
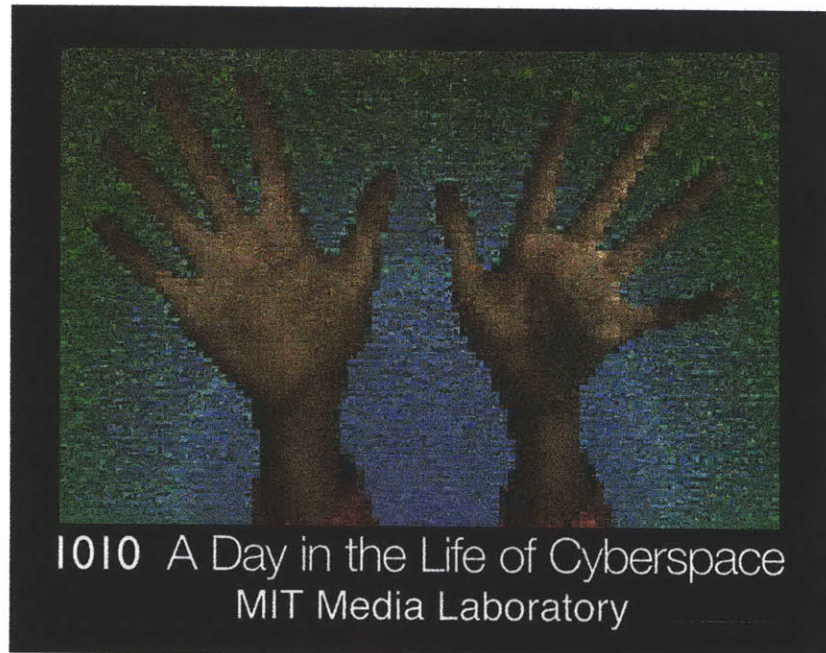
Figure 3.4: Rob Silvers, *1010 Poster*, 1995

3.2 Basic Process

The Photomosaic process involves three steps: creating the database, analyzing the master image to find appropriate photographs, and generating a bitmap suitable for display.

3.2.1 Useful Resources

The following is a list of hardware, software, and other materials useful for creating Photomosaics.

Hardware:

- One workstation with at least 96 MB RAM, 4GB of disk space, and video input. This is needed to capture images and run the Photomosaic software. 128 MB of RAM is preferable for making complex images (i.e. considering 16 x 16 sub-regions of 30,000 images).

- One Macintosh with at least 80 MB RAM. While not strictly required, I found this useful for editing and printing images after generation. A PC or UNIX system could be substituted if desired.

- One VHS video tape player. Source material is captured from tapes either by controlling the player to single-step through the program or by capturing in real-time during ordinary playback.

- One computer controllable laser disc player. This is very useful for gathering source material from movies – it is better than using video tape if the desired content is available from both sources due to the higher quality and easy random access to stills.

- One high-quality output device to print the images.

Software:

- Photomosaic software (created for this thesis).

- Adobe Photoshop or comparable image editing software.

Other:

- Database of images. 1000 to 10,000 or more images are needed. If the images are to be published, copyright clearance must be obtained.

3.2.2 Database Design

The database design is simple. A UNIX file system holds the images in directories by category and image size. Before creating a database, thousands of images must be obtained. Possible image sources are described in Chapter 4. Once the images are located, the program *make_square*, described in the appendix, crops a square out of each original rectangular image, resamples that square to a standard size, then moves it to the proper database location. Image categorization by subject must be done in advance of this stage.
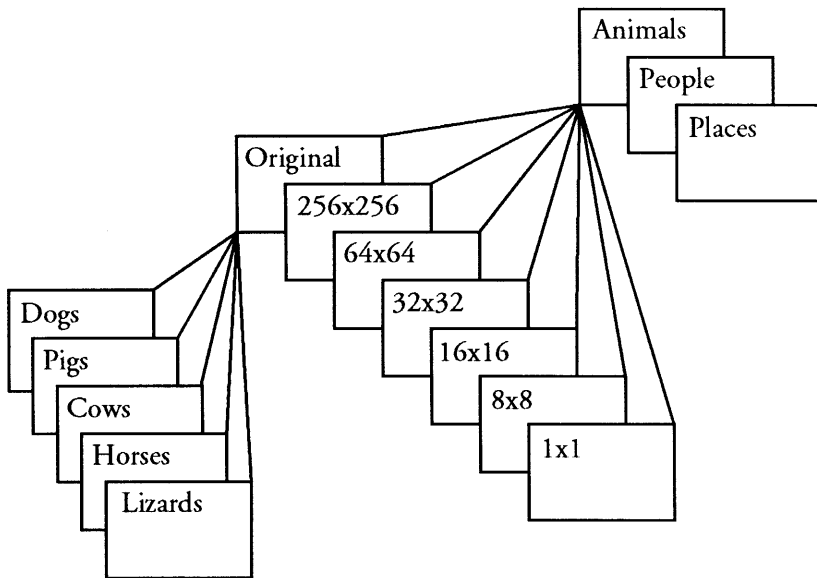
Figure 3.5: Organization of Database

Figure 3.5 shows the organization of database images. If one wanted to make a Photomosaic of animals from around the world, one would start the matching program set for the root node "Animals." The software would then use any images contained within the subdirectories under that specific node. If semantic content, as described later in this document, is specified, then image sub-categories will be relegated to specific regions of the mosaic.

Directly under the root node are sub-directories containing identical image files at different resolutions. One sub-directory is labeled "original," and contains the uncropped version of each image file at full size. This is needed because a tile may be recropping during mosaic creation,

if the artist does not like the original (automatic) crop. The next directories, labeled 256 x 256 and 64 x 64 are large versions of the square tiles used primarily for outputting the final bitmap. The 32 x 32 directory is used for viewing the mosaic on the screen during the construction process. The smallest sub-directories, 16 x 16, 8 x 8, and 1 x 1, are preloaded when the mosaic building program starts up and are used to match images during the selection process.

3.2.3 Generating the Image

The first steps in creating a Photomosaic are selecting a master image, a database to represent it, and the number of tiles to use. Section 3.3.1 describes in detail some of the finer issues of subject selection and the corresponding number of tiles. I will describe the details of the process beginning after these factors are determined. See figure 3.6 for a schematic overview of the system.

When the *glmosaic* program (see appendix) is run with a database path specified on the command line, a sub-sampled version of each image is loaded into a linked list of structures.

Source Image

Software

Image Database

Photomosaic Output
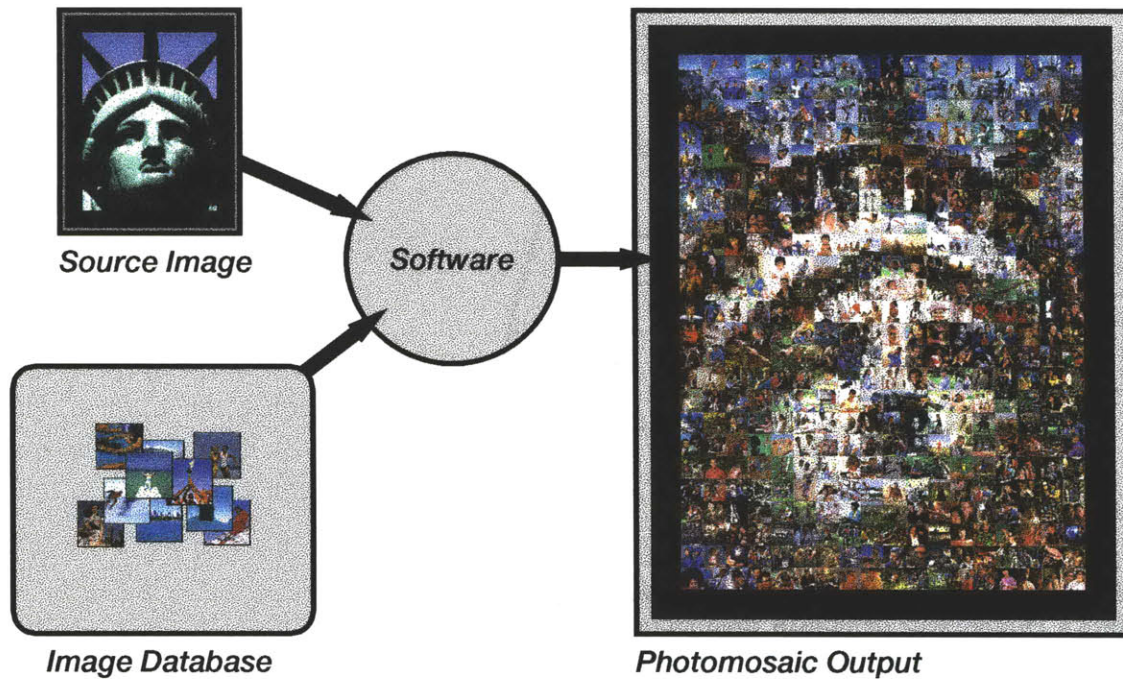
Figure 3.6: Diagram of Matching System

```
struct an_image {
        char *path;                     /* pathname of file in the database */
        char used;                      /* whether or not image has been used in a tile */
        unsigned short *r;              /* r, g, and b image data used for rms matching */
        unsigned short *g;
        unsigned short *b;
        struct an_image *next;          /* pointer to next structure */
        struct an_image *previous;      /* pointer to previous structure */
} an_image;
```

A master image is then loaded under operator control. This master image must be the proper size in pixels, determined by the number of output tiles desired multiplied by the number of sub-sampled regions one wishes to consider during the matching process. The number of sub-sampled regions, in turn, must equal the size of the sub-sampled database images loaded upon startup.

When the matching process begins, the master image is divided into x by Y tiles where (x, y) is:

$$(master\_image\_width / width\_subsamples,$$
$$master\_image\_height / height\_subsamples)$$
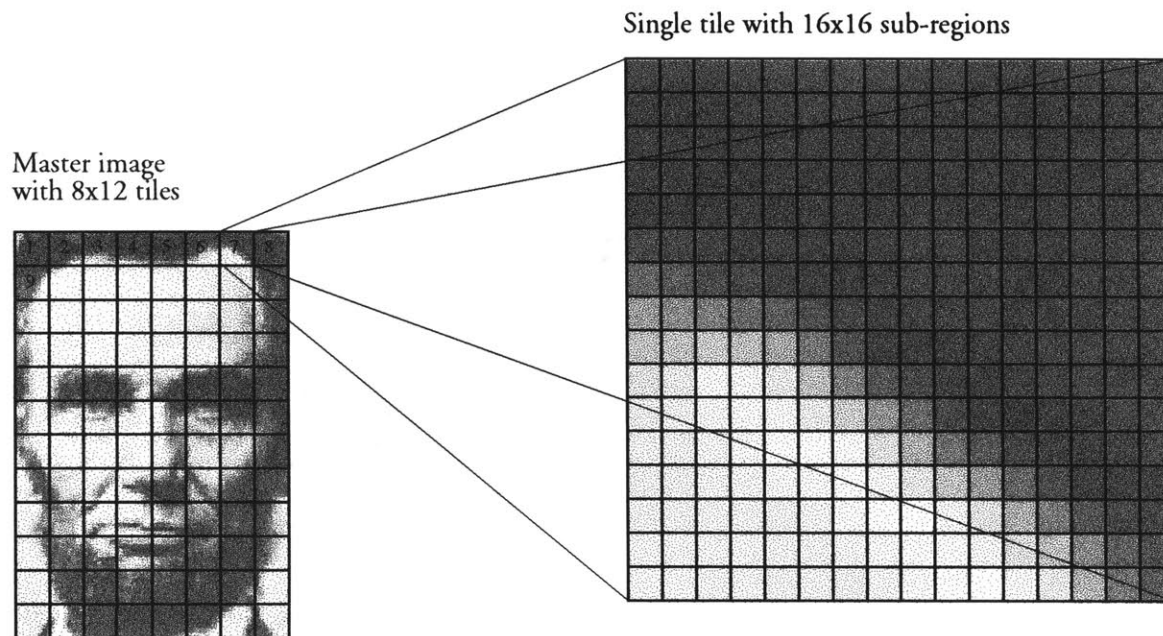
Figure 3.7: Tiles and Sub-Regions

Figure 3.7 shows a master image with 8 x 12 tiles in which each tile contains 16 x 16 sub-regions. The tiles are selected one at a time and the average RMS error of the red, green, and blue channels of each sub-region is compared to each image in the database. If the computed error (degree of visual similarity) between the current images being considered is the lowest thus far, an index to that photograph is stored along with the error. When all of the database images have been tested, the image with the least error that has not already been selected for another location is as-

signed this particular grid-space in the mosaic and marked used. This process is repeated for all the mosaic tiles. Since this selection works from top to bottom, the best images are often used up first, which is more noticeable when using smaller databases (see "Error Diffusion," page 82). Upon completion, a list of image names is written to a text file which is used by the final rendering program to construct a bitmap from the full resolution versions of the images.

Visual similarity, as mentioned, is determined by finding the images with the least average image value RMS error on comparison. Sub-sampled RMS was selected as the matching algorithm because it is a simple and efficient algorithm which results in visually similar matches. The results are so successful, that Photomosaic tiles in most cases produce more resolution than solid pixels. This stems from the matching system comparing many areas within each photograph during the selection process. This effect, called *sub-picture resolution,* enhances the ability of a viewer to see an image formed from a small number of tiles. Fewer photographic tiles are needed than would be possible with solid colored pixels. Thus, tiles are more than pixels.

Figure 3.8 is an image designed to test the search algorithm's degree of finding visual similarity. Giving this image to the software will return 16 images that can be compared to the results from other algorithms.
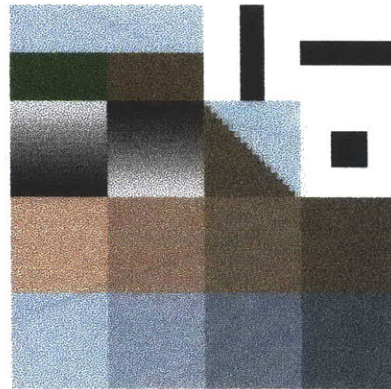
Figure 3.8: Test Image

If we run this test on the first version of my software, which only finds an image whose average color (average of all pixels) is closest to the average color of the input region, this test pattern would first be reduced to the one in figure 3.9. Since all the computer can "see" are these solid blocks, it is not surprising that the output, as seen in figure 3.10, is approximate in color but not in shape. Figure 3.11 shows what happens when we consider sub-regions, in this case four by four. Because the computer can now see some amount of light and dark regions within each tile, it is able to consider them when searching the database. The first output tile in figure 3.11, a jet fighter, is blue at the top and dark green at the bottom, just like the input test pattern for the same region. Similar appropriate choices can be seen in other tiles.
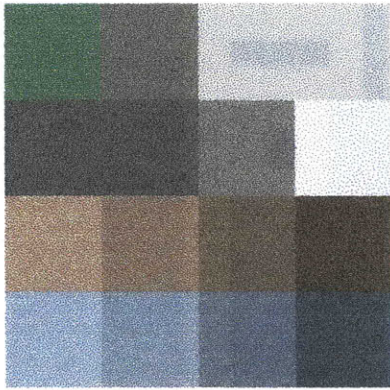
Figure 3.9: What Computer "Sees" When it Averages Values
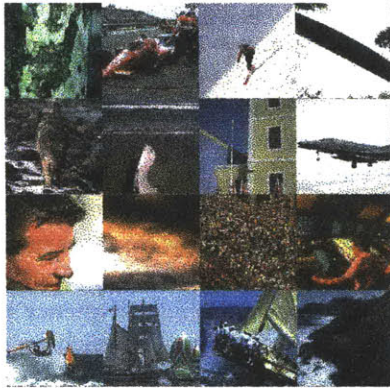
Figure 3.10: Output for Averaging Algorithm
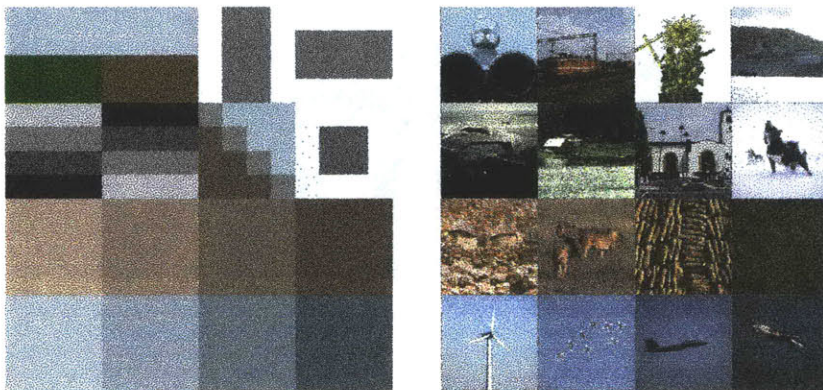
Figure 3.11: Test Image with 4x4 Sub-Regions and Output
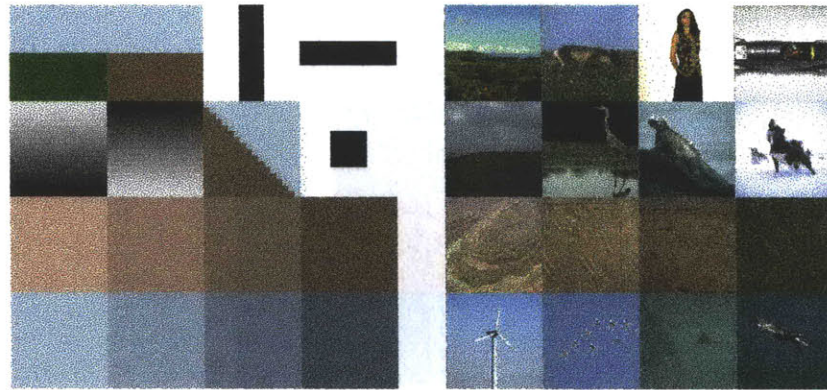
Figure 3.12: Test Image with 16 x 16 Sub-Regions and Output

Now, consider that the computer can see the original pattern of figure 3.8, at full resolution. Figure 3.12 shows that if this amount of detail is considered during the search, highly appropriate matches are found. The woman in the upper right is the same shape as the vertical black bar in the same region of the master image. The lizard in another tile matches the brown and blue diagonal that it was compared to. This high-degree of shape matching has a powerful effect on the image-forming ability of a Photomosaic. Information about the contours and shading in a master image may transcend the boundaries of each mosaic tile.

Not only do more sub-regions allow for more sub-picture resolution, they also result in a more uniform distribution of color by selecting lower contrast images for regions of little high-frequency detail. This can be seen in the lower eight images of figure 3.12 which are more uniform than those selected in 3.10. Considering more sub-regions improves visual quality, but rendering time suffers. This could be fixed by changing the current N^2 algorithm to an $N/2$ algorithm (see "Singular Value Decomposition," page 82).
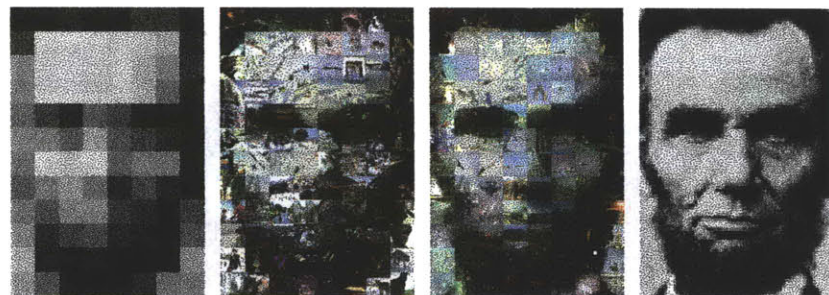
Figure 3.13: Four Abes

Figure 3.13 demonstrates the effects of sub-picture resolution in a real-world example. The first Abe is composed of 8 x 12 *pixels* while the second is 8 x 12 *images*, with 2 x 2 sub-regions within each tile considered

during the selection process. Notice how the second mosaic is a more recognizable Lincoln than the first (the 1970 pixelized image by Leon Harmon had 14 x 18 pixels but was based on a different photograph of Lincoln, so they cannot be directly compared). The third Lincoln, where 16 x 16 sub-regions were considered, is more recognizable still and has less noise. The fourth image is the master used to form the mosaics, shown for comparison.

3.2.4 Rendering a Finished Image

To render a final image, one needs access to the original database and the Photomosaic description file output from the matching program. The rendering program reads in a list of the selected tiles, locates the full sized version of these files in the database, and binds them together to create a bitmap. The tiles may be separated by a line of any color, helping to discretize them when viewed up close. From a distance, the grid-lines are thin enough to disappear completely, not interfering with the seamlessness of the mosaic. The bitmap is saved in a standard format to be displayed on a monitor or output in printed form.

3.2.5 Hardcopy Output

There are several ways to print a digital image, dependent on quality, price, and size constraints.

- Film recording and photographic printing is often the best option. An image can be written to photographic film using a *film recorder*. Once the image is on a chrome or negative, it can be printed onto normal photographic paper. This option is best for a moderate number of small copies as writing the image onto the film is a one-time cost.

- Direct digital printing potentially produces the highest quality, but each print is expensive, making this the best option for one or two copies but a poor choice for hundreds of copies. Digital printers are either continuous-tone or use half-toning. Continuous-tone printers can place an exact color for each pixel in the image. Half-toning printers can only put drops of solid color, and form shades of color by using dots of different sizes or different spacing. Hence the print will look less photographic. It is important to notice that a 300 dpi printer using half-tones actually has less than 300 dpi resolution because it takes several of these dots to represent a color. Such a printer is good for less than 100 true *pixels* per inch (spatial resolution is traded off for improved color resolution).

- Process color printing is the technique used to reproduce images in magazines and books, and the best choice for hundreds or thousands of near photographic copies.

3.3 Refinements

Optimizations are important for allowing one to push the limits of the medium. The Photomosaic process is not automatic; one does not push a button for guaranteed optimal quality. Like a paintbrush or camera, the software is only a tool and must be used with skill. There is room for increased quality of the results using talent, knowledge, and experience.

3.3.1 Optimizing Choice of Subject and its Representation

In this section I will discuss some of the issues an artist must consider when representing a subject. Considerations include the choice of subject, the number of tiles to represent this subject, and the final size of the display. There are three main issues when choosing a subject:

- Complex detail is difficult to resolve in a Photomosaic.

- The colors of the subject must be common in the database.

- Familiar objects, especially faces, are the easiest to recognize.

Consider the level of detail needed to resolve the essence of the subject. If the detail of a subject is too complex, consider minimizing detail. For example, to render a successful Photomosaic of a person, zoom in on their face as tightly as possible, rather than showing their entire body. If the colors of a subject are not common in the database (red is often hard to match if the database is composed of natural images), either change the colors in the subject, or increase the number of similarly colored images in the database (see "Tuning Color," page 62). Aside from these technical considerations, if the goal is an easily recognizable image, a familiar subject such as a face is the best choice. Our perceptual system can identify the shape of a face, even when none exists. People have, for example, seen images of faces on the moon, on mars, and in the bark of trees. Thus, seeing one in a Photomosaic should prove less challenging than seeing something obscure like an automobile carburetor.

Another important decision is the total number of tiles used in a Photomosaic. More tiles will help resolve additional detail, but the average matching error will often increase as the "best" tiles are used up. This is a delicate balance. The aspects to consider in selecting the number of tiles are:

- Desired viewing distance.

- Desired image size.

- Number of images in database.

- Color space of database when compared to master image.

- Detail in master image – a single face needs fewer tiles than two faces.

60

The viewing distance, image size, and number of tiles are inter-related, as a larger image size requires either more tiles or larger ones. If a database has a lot of images in it, one can often use more tiles before the matching becomes poor. The maximum number of tiles one can use should be $1/N$th of the number of tiles in the database. A practical value for N may be as low as 5, but 10 is more desirable. The better the match of the color space between master image and database, the smaller N can be for a given level of perceptible quality. Finally, more detail requires resolving with more tiles. Minimize detail in the target image to use fewer tiles. Using fewer tiles allows one to print the final image at a small size while still seeing the individual images.

The size of the tiles can be tuned by the quantity of tiles. To select the number of the tiles to use, decide if the goal is to maximize the distance between the two levels of images or to minimize it. If one would like to only see the sub-images when the mosaic is viewed up close, and only the over-arching image when viewed from a distance, it helps to make the sub-images large enough so that they are too coarse to be understood from the viewing distance. Having fewer sub-image tiles facilitates this goal. A large amount of tiny sub-images may resolve the larger image too easily, resulting in less surprise from viewers. In this case, it looks as though the tiles are simply squashed into pixels. A Photomosaic displayed on a post card must still have a certain number of tiles in order to resolve detail, but not too many as the viewer will not be able to see the smaller scale sub-images clearly at this reproduction size.

3.3.2 Tuning Color

The colors of a master input image may not fit well into the color space of the photo database. Poor mosaics then result because there are not enough images to choose from to provide visually similar matches. One solution is to analyze the color space of the database and quantize this to the N most popular colors. The source image is then altered to fit this space.

A finite color table can be created from a set of raw colors. One way to accomplish this is to find the average color of each image, then determine N colors where each color is then the average of a set of the original colors in the database, and the size of each set is equal. An excellent survey of color quantization techniques may be found in [28]. Figure 3.14 shows a three dimensional representation of the colors in one of my databases. 26,000 images are each represented as a single point in space, with the location of the point on the X, Y, and Z axis equal to the average R, G, and B color of that image. The cube may be rotated to visualize the clusters of colors and determine the richness of the color space. Black and white images will form a line progressing through all three axes. Figure 3.15 shows the 32 most popular colors in this database from most popular to least popular.

Once the most popular colors are determined, the quality of matches can be improved by altering the master image to these colors, creating a larger selection of images for the matching algorithm. In figure 3.16, I rendered a George Washington from both popular and unpopular colors. The Washington made from unpopular colors is indeed less recognizable.
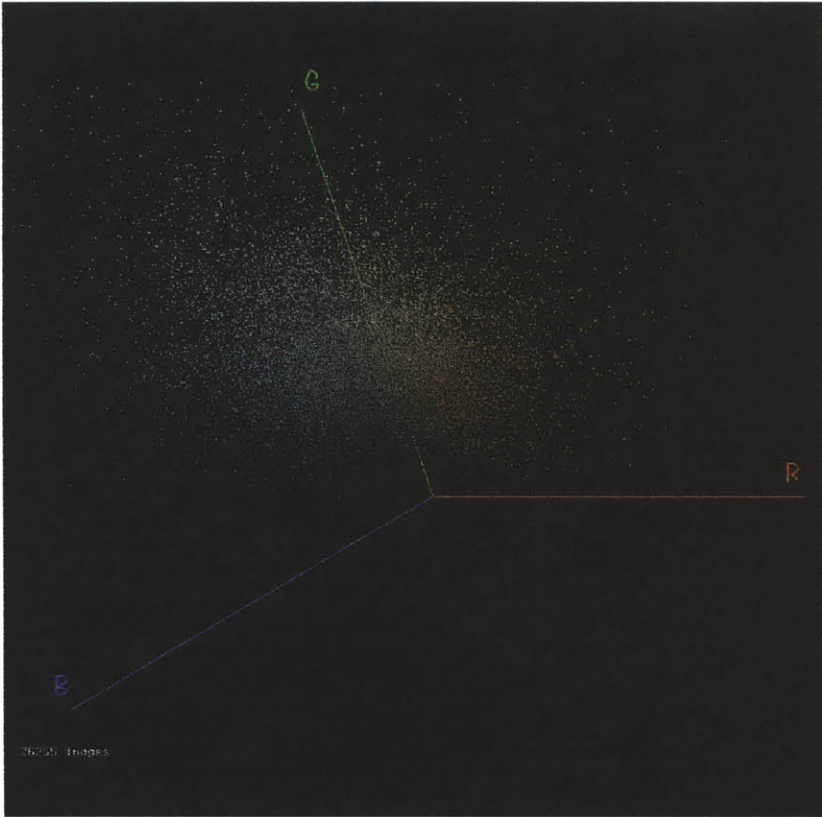
Figure 3.14: Color Cube with 26,000 Images
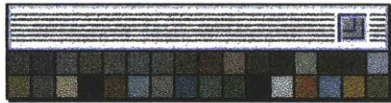
Figure 3.15: The 32 Most Popular Colors
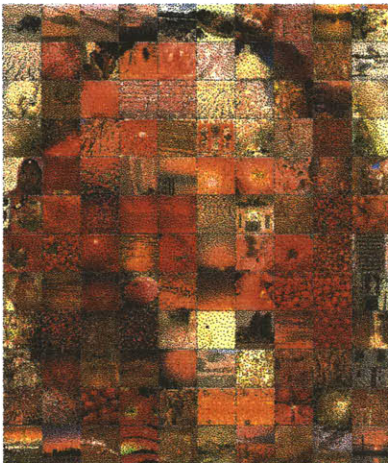
Figure 3.16: Images Made from Popular and Unpopular Colors

3.3.3 Modification of Selected Images

I prefer to not modify the sub-images because constraints can provide a greater challenge. Ken Knowlton, when making his images from Dominos, decided to only use complete sets, thus heightening the interest of the resulting images. I have tried modifications in the past. For example, my first attempts at making mosaics from photographs used colorization when the images did not match perfectly, but viewers were generally less impressed when they knew that modifications were done. Nevertheless, some modifications will create better looking mosaics. These include altering colors, "noise" filtering, and controlling sub-image cropping.

Altering colors can easily be done because the software knows what it is looking for during the matching process. It finds the closest match, then tints the colors of the chosen image to be more similar to the desired values. Constraints can be imposed for subtlety, such as only allowing modifications of up to 10 percent, or only modifying single parameters of hue, value, or saturation.

Another possible modification is filtering the noise which interferes with perception of the master image from the sub-images (see "Issues Relating to Coarse Pictures," page 70). Noise comprises anything in a sub-image that does not match the corresponding region of the master image. Not all noise is harmful, but the greater the noise amplitude and the closer it is to the signal, the more harmful it is. Noise can be filtered out in the frequency domain when it is within the same band as the signal. The filter can be a curve with a weaker effect on noise further from the signal. The result in the spatial domain is colorization and blurring, but only within limited bands, thus it is less detrimental to the sub-images than pure colorization or low-pass filtering (high-frequency detail would remain in sub-images).

Finally, images can be cropped, nudged, and zoomed to achieve a better match. This is a modification that everyone can understand, because it is commonly done with photographs. One must, however, constrain the level of zoom or adjustments to the cropping to keep a very small portion of a photograph from becoming a tile. If this happened, it might not contain an identifiable subject. Both cropping and zooming should be under human control, perhaps with a special program to annotate constraint information or to specify a percentage of allowable automatic change (see "Tweaking to Fit," page 86).

3.3.4 Artistic Choices During Mosaic Construction

Artistic considerations regarding the selection of the master image have been discussed. This section describes the refinement of a computed mosaic to produce a more perfect form. The steps include viewing the image during the process from various distances, re-cropping image tiles, replacing image tiles with alternates, and deciding whether grid-lines should be drawn between tiles.

After the software creates a draft of a Photomosaic, the artist should study the image from a distance to see if the master image is well represented. Is the transition from tile to tile smooth? Are there any tiles that jump out at you because they have the wrong color? If so, enter the editing mode and have the computer offer alternate images for that grid space. The system will display many choices, sorted from best to worst match, with the RMS error for each one. Select the most visually pleasing tile, and repeat this process until all the tiles are well integrated from a distance.

Do the tiles look good up close? Do they convey the right subject matter? If a tile is objectionable because of content, replace it with an alternate. Is a tile appropriate but poorly cropped? All tiles start out as rectangles, and a square is automatically cropped from the center. Quite often, this cropping is not visually pleasing and must be redone by hand. The original version of a selected tile may be opened for recropping. Keyboard controls pan and zoom a cropping square until the user is satisfied, at which time the tile is re-integrated into the database. After these processes, all tiles should look good from a distance and up close.

Finally, the artist may decide to render the image with or without grid lines. The grid-lines are so thin that from a distance they do not interfere with the continuous appearance of a well executed Photomosaic. When viewed up close, it is often desirable for the virtual master image to vanish and discrete tiles to fill the space. Grid-lines help break the continuous nature of tile-to-tile transitions, helping to increase the effect of multi-scale imaging.
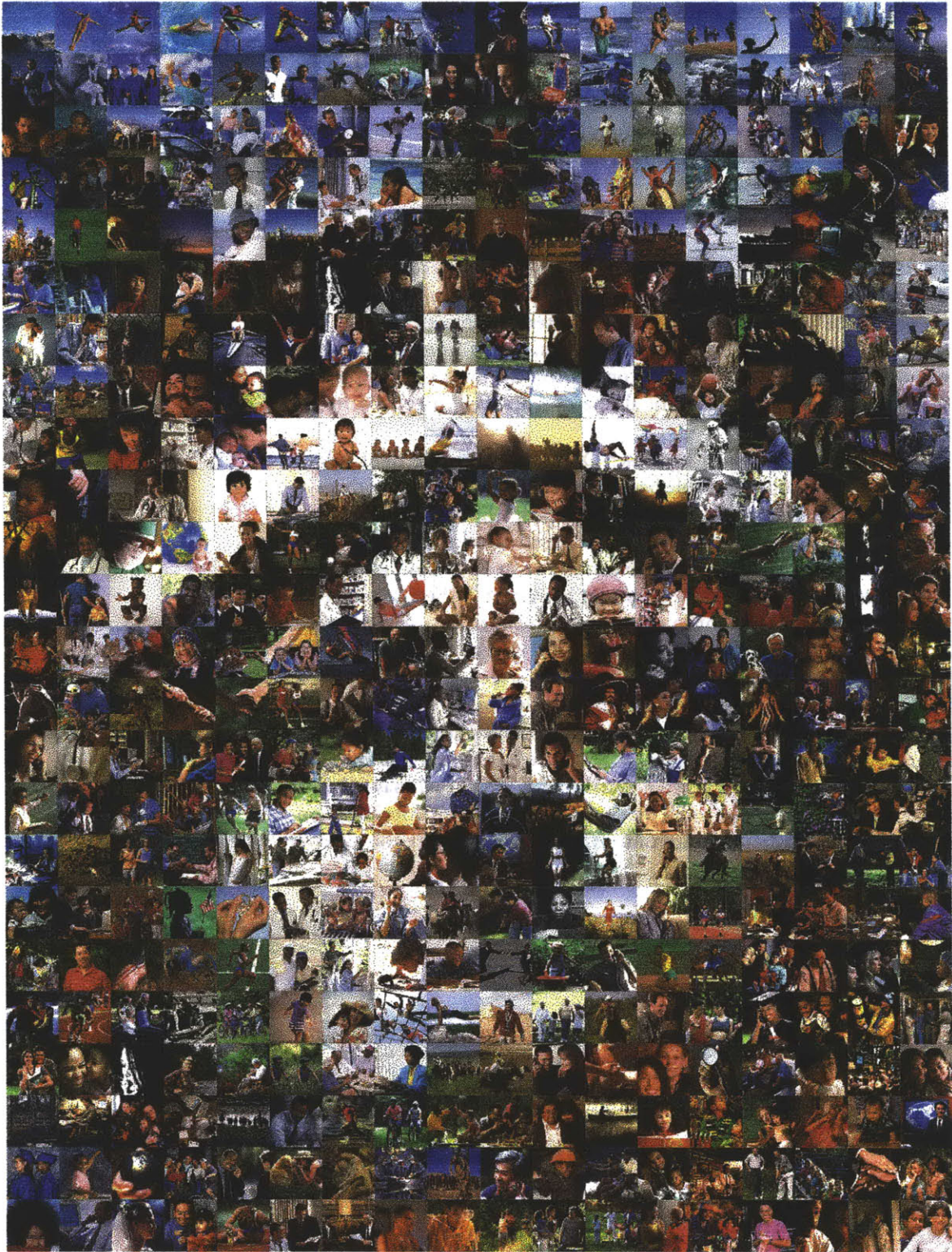
Figure 3.17: Rob Silvers, *Liberty made from Americans*, 1996

3.3.5 Semantic Map - Another Dimension

Photomosaics may be used to communicate ideas by relating the master image to the mosaic tiles. This may be seen in its simplest form in figure 3.17. Another dimension to this axis may be added by relating the content of the tiles to a specific *region* of the master image. Related semantic information is not a new concept for artistic expression. In 3rd century Roman mosaics, gold and silver were used to show light emanating from God and Christ. The capability to place related content has been implemented in the form of a *semantic map*.
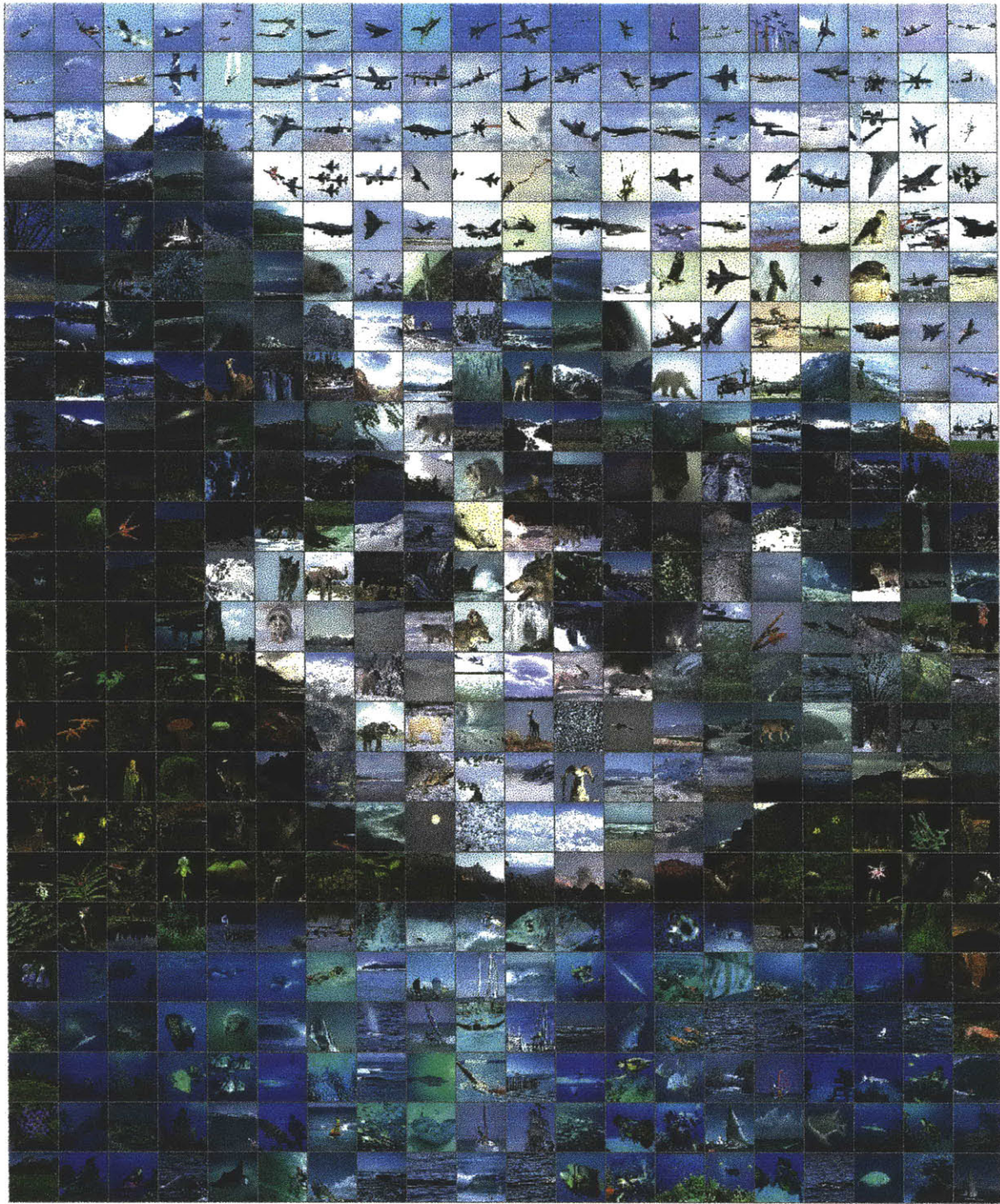
Figure 3.18: Rob Silvers, *Example Semantic Image*, 1996
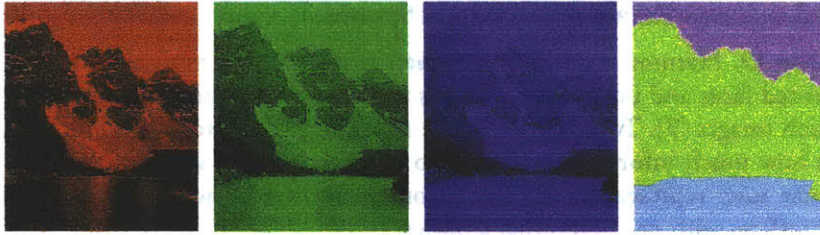
Figure 3.19: Red, Green, Blue, and Semantic Channels

A *semantic map* is a channel associated with a source image expressing desired content in a given region. Content is "painted" on a layer under the image pixels. Figure 3.18 shows a scene of a lake, mountain, and sky. Notice how the image of water, land, and sky is tiled with fish, animals, and birds/airplanes. Normally, the Photomosaic software would only consider the sky and water as blue, fitting blue images to those regions. In this example, there is a semantic layer, with only fish in the water, and only birds and airplanes in the sky. The images in figure 3.19 show the red, green, blue, and semantic channels for the master image. The semantic channel has three areas where content was painted using an ordinary paint program. To do this, colors are first defined in the paint program and given names such as "fish" and "birds" instead of "purple" and "orange." The bit patterns for these colors are put into a configuration file which also contains the image directories to be considered for each "color." When a Photomosaic is generated, this text file is loaded with the master image and used to limit the considered images for each region. A future enhancement would be artificial intelligence and high-level pattern recognition software which automatically "sees" what is in each image and classifies them without human categorization. Finally, the promise of semantic maps is not fully realized without powerful database integration (see "Larger and more Sophisticated Databases," page 81).

3.3.6 Issues Relating to Coarse Pictures

Leon D. Harmon studied the human perception of minimally represented faces and uncovered methods to enhance our ability to recognize such images [7]. Two discoveries are relevant to Photomosaics. The first is how much information it takes to represent a face, and how to optimize such representations. The second is related to noise, and how it interferes with our perceptual abilities.

To determine the minimal information needed, Harmon reduced a number of faces to 16 x 16 pixels and 8 levels of gray (768 bits of information), and noted that recognition accuracy (based on testing with a group of human subjects) was on average 48 percent. Harmon also noted that the *range* of recognition accuracy for all of the images was from 10 to 96 percent. In other words, some were more recognizable than others. Was this because certain facial features were stronger in some images, or something else? Harmon hypothesized that important features in some of the images may have fallen on pixel boundaries, reducing recognizability. To test this theory, he created four images of each face. The first was the image mapped at random, the second was with the source image shifted one half pixel width to the right, the third with half a pixel width down, and the fourth to the right and down. These new images were compared and tested for best recognizability. Harmon noted that if only the number one choice of the four was chosen and considered again for the original experiment, the overall rate of recognition rose from 48 percent to 95 percent. These four possibilities should therefore be tried when making a Photomosaic.

Harmon also questioned why it was easier to recognize a pixelated image when viewed from a distance, while squinting, or while shaking it. He believed the reason was noise which obscures the relevant content of the image. Since the content is at a low spatial frequency of, in this example, 8 cycles per picture height (half of the height in pixels), any frequency higher than this is noise and interferes with the signal. Do we need to remove all noise? Harmon found that the noise within about two octaves of the sampling frequency was the most important because it was in the same band as the content. This problem is called *critical-band masking*. Figure 3.20 illustrates this point. Let us say that we have two signals, A and B, of differing spatial frequencies. There is also a noise N in the domain that visually interferes with our ability to see the signal in direct proportion to its amplitude and proximity. In our example, the signals are the same strength so the noise N only interferes more with A than B because it is closer in the frequency domain. In fact, A is within the *critical-band* and overwhelmed, while B, being further away, is still perceived. Harmon's specified fixed range of two octaves was an over-simplification because the further the noise is from the signal, the less effect it will have. The range and shape of the critical band, therefore, is a curve and should be based on actual experiments with human subjects.
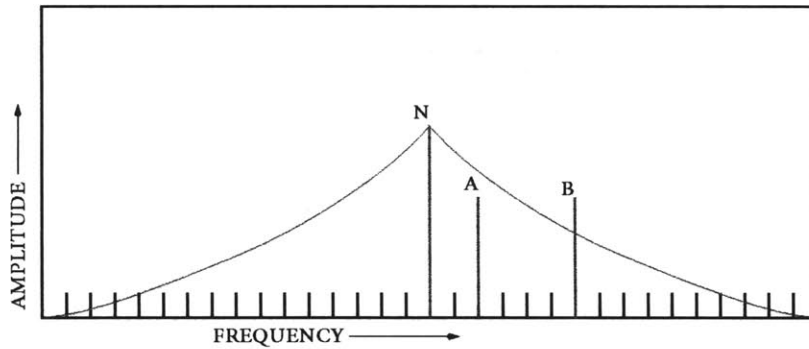
Figure 3.20: Critical-Band Masking

To prove this point, Harmon added random noise to an image at two spatial frequencies. In one image, the noise was added within the range of the signal frequencies (left image of figure 3.21, from Harmon's paper [7]). In the other, it was two octaves away from the signal (right image). Notice that the image on the right, with the same amount of noise but with it further from the content, is more recognizable.
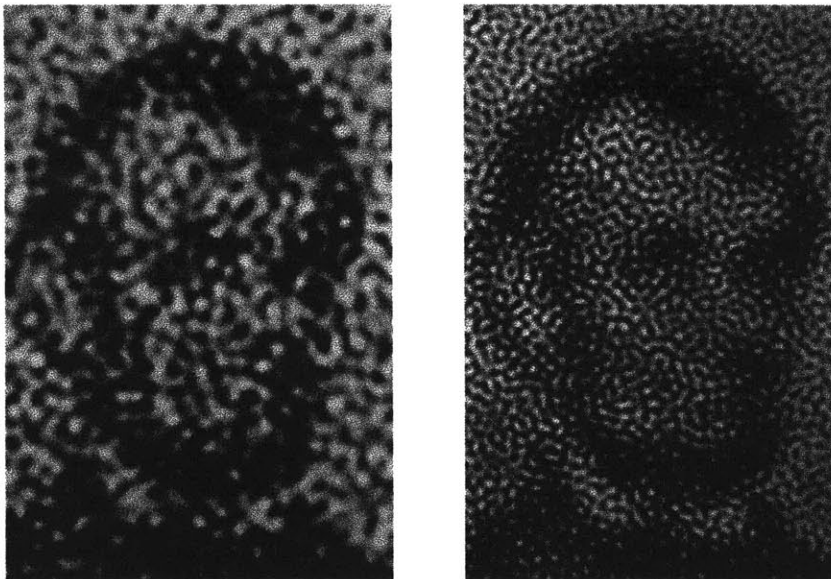
Figure 3.21: Random Distributed Noise

These findings are important to Photomosaics because any error between the photographic tiles and the corresponding region of the master image is considered to be noise interfering with our ability to perceive the intended subject. To remedy this, two solutions are proposed: the first involving modifying the sub-images, the second does not need to. After the matching process, analyze the spatial frequencies of the master image within a grid-spot and reduce the "noise" in the sub-images within the

same band (which would have the effect of selective blurring and colorization). If this modification of sub-images is not acceptable, perhaps because it may destroy the quality of the image or the artistic integrity of the completed art work, then consider the amount of in-band noise during the matching process as another weighted parameter.[ii]

The expected result from either enhancement is a closer range for discerning the global subject, which reduces the distance between the two-scales of images. However, these results may be tuned to create the greatest or least difference between the two-scales. There will be cases where enhancing a viewer's ability to perceive the main subject of a Photomosaic is desirable, especially where the Photomosaics have a smaller number of tiles, difficult subjects, or a very small image database.

[ii] Suggested by Mike Hawley during personal communication.

Chapter 4

Copyright and Image Acquisition

4.1 Copyright Issues

ACQUIRING A LARGE COLLECTION of images brings up the issue of permission for use. Who owns the images, and can they just be used or do you need a license?

4.1.1 Fair Use

Even images that are not in the public domain may sometimes be used and distributed without permission from the holder of the copyright under what is called *fair use*. Fair use is based on case law, meaning that there are no distinct rules that if you follow you are guaranteed to be safe. Rather, there are guidelines based on past rulings, and if you stay within them you increase your chances or winning if you are sued. Following are the four factors that may be used to judge is fair use applies (from 17 U.S.C. § 107):

- the purpose and character of the use, including whether the use is of a commercial nature or is for nonprofit educational purposes – that is, commercial uses are less likely to be fair use; nonprofit educational uses more likely quality as fair use;

- the nature of the copyrighted work – that is, copying from factual works is tolerated more than copying from more creative fictional works;

- the amount and substantiality of the portion used in relation to the copyrighted work as a whole – that is, the more that is copied, or the more significant the portion that is copied (regardless of the quantity), the less likely that fair use will apply;

- the effect of the use upon the potential market for or value of the copyrighted works – that is, if the use has an adverse impact on the market for the original work, it will not constitute fair use.

Each of these factors must be considered, although one specific case, *Harper & Row Publishers, Inc. v. Nation Enters*, may be used as a precedent to demonstrate that the fourth is the most important.

In addition to educational use, fair use may apply when the material is used for criticism or reporting (editorial use). Fair use probably would not apply if the images are used directly for profit. Even though only small pieces of photographs are needed for a Photomosaic, they might still be considered valuable and protectible parts – perhaps the "heart" of each image. It ultimately comes down to good faith and reasonableness.

4.1.2 NATIONAL GEOGRAPHIC

This thesis would not be complete without telling more of the story about how Photomosaics came about. In early 1995, NATIONAL GEOGRAPHIC had a research agreement with the Lab to work with large image collections. They supplied us with laser discs containing almost 400,000 still images stored as NTSC frames. A proposed project was to convert them into digital form and write software that could index and search them based on high-level content. I decided instead to make mosaics out of photographs. After succeeding at this task, I sought permission to publish one in the Media Lab's newsletter Frames, and permission was granted (see figure 3.2 "Rob Silvers, Jerome Weisner, 1995," page 50).

For the Lab's 10th anniversary, I created a poster of hands counting to ten. Once again, we sought permission to distribute this work. This time we were refused because they believed the images were too clear and recognizable. (see figure 3.4 "Rob Silvers, 1010 Poster, 1995," page 52). They were only 64 x 64 pixels each – or almost 1/1500 the bits of a standard Kodak digital image (fair use clause 3). NATIONAL GEOGRAPHIC was concerned that the relationships with their photographers would be jeopardized. Would these cases fall under fair use? I would think yes because they were for research (fair use clause 1) and would not reduce the salability of the individual images (clause 4). We did, however, have an agreement with them preventing us from distributing the images outside of the laboratory.

4.1.3 Possible Copyright Solutions

One possible solution to the problem of payment to copyright holders is to use the radio model.[iii] In the early days of radio, there was need for a method to play songs and compensate the owner in a manageable way. Since it was impossible at the time to track how many times your song was played, its airtime was estimated and royalties were payed based on this estimation. Perhaps all photographers that had work shown in a specific image collection could be payed each time the collection was used, even if their image did not appear in the Photomosaic? This makes some sense because images are not chosen based on any type of established merit system. Still, legislation on some level would need to be agreed upon. Perhaps when you allow a stock photo agency to carry your work you agree to the possibility of it showing up in a Photomosaic in exchange for increased compensation.

4.1.4 Public Domain Images

Images that are in the public domain are of course fair game. The Library of Congress, for example, has a large collection of public domain images but few are in digital form and accessible. This becomes the challenge. Until public domain images are packaged and distributed, they will be hard to come by in quantity. Still, just as public domain software is packaged in bulk and sold on CD-ROMs, this should happen with photographs.

[iii] Vice President Al Gore suggested this idea during personal communication.

4.2 Image Acquisition

Images are increasingly digital. In the future, all of our memories will live in a server ready for access. For now we need to search for Photomosaic content, or expend effort to create it. The most practical locations to search for images are the Internet, stock photo agencies, CD-ROM collections, and television.

4.2.1 Internet

There are millions of images on the Internet and it is growing by the day. Harvesting these images for Photomosaics would keep your collection growing. One could either acquire a list of image locations from the database of a Web searching system, or write a special software agent to move through links and directly fetch image files. The technical challenge is to filter desirable images from the rest.

Images on the Web are used for photographs, buttons, backgrounds, textures, and advertising. If you are willing to accept all such images for an art work, the task is easy. A filter is needed, however, if you only want photographs. Such a program could be written because photographs do not have the same large regions of a solid color that buttons do. Still, there are many images that are photographs but also have solid colored boarders around them. A segmenting filter is needed to identify such images and crop the interesting center from the unneeded outer regions.

4.2.2 Stock Photo Agencies

Stock agencies maintain a large collection of images to sell to clients in need of something specific. There are several issues relevant to stock photo agencies. They charge for each image based on use, they often do not own the copyright to the images they sell, and their images may not be in digital form.

Stock agencies charge money on a per-image basis so being in need of hundreds or thousands of images would be prohibitively expensive unless a special arrangement can be made. Such an arrangement may include having them agree to send you thousands of images in advance, but only charge for the ones that end up in a Photomosaic. An even more desirable solution would be to be charged for the composition that you create as if you purchased a *single* image from them. They could then perhaps divide the payment between many photographers.

This brings up the second, and perhaps most important issue: most stock agencies do not own their images, but rather resell them for independent photographers. Most agencies, therefore, are not free to let you use images for free even if they want to and must make arrangements that are agreeable to their photographers. Simply contacting hundreds of photographers on your behalf is prohibitive.

The ideal agency would own their images outright and would have a large number of them already in digital form. Unfortunately, a typical agency may only have part of their collection digitized. The general solution for them is to digitize the most recent catalog collections and the newest material as it arrives, but leave most of the older images in file cabinets. This problem is fading fast.

4.2.3 CD-ROM collections

CD-ROM image collection are often provided by stock photo agencies and have similar restrictions. Typically, you are free to search their collection of thumbnails and use them in mock-ups of your work, but must purchase the actual image for any real use. There are a few collections that you may purchase outright, but they are often of lower quality than those available from a stock agency, and therefore provide weaker components from which to construct a composite.

4.2.4 Image Acquisition from Television

One way to collect thousands or millions of images is to collect them from broadcast television, video disc, or tape. Images are being flung through the airwaves and published on videos at a blinding rate, why not capture them? Assuming the copyright issues are satisfied (which may fall under fair use), only technology can get in the way.

A video frame-grabber can monitor a signal and grab all frames, frames at regular intervals, or a frame when it significantly differs from the last one captured. This third method seems preferable because it produces similar looking frames which may be considered duplicate images by viewers of the finished art work. Video capture from a stored medium may be handled by single stepping to ensure that each frame is considered.

Motion video will have interlace artifacts caused by the way television is displayed (first the odd video lines, then the even ones). The individual frames will need to be de-interlaced. This can be done in a number of ways. The simplest is to simply discard one set of lines, called fields, and replace them with an interpolated version of the lines around it. The disadvantage is a loss of resolution. A more sophisticated approach is to interpolate only where there is significant motion (perhaps a person running), and keep both sets of fields where there is none (the stationary background).

There is another issue to consider when capturing video sources, whether from video tape, disc, or broadcast television. This is the problem of *3:2 pulldown*. Whenever film that was shot at 24 frames per second is transferred to video running at 60 fields per second, there is a conversion process called 3:2 pulldown that can have strange effects in still images. A single frame of the original film becomes three fields of video. The next frame in the sequence becomes two fields. This alternates back and forth and has the following consequence: a still frame of video may have one frame of film that is the odd field and another non-related film frame in the even field. If this happens on a transition in the movie, the results are especially bad. The simple solution is to use only one field all the time and suffer the loss of resolution. The complex solution is to track the 3:2 pulldown and recover each original frame of film. This does matter even for Photomosaics because the images are not always printed very small.

The handling of captured images can be treated in several ways. All images may be grabbed and stored in a database for future use, or they can be added "live" to an image being worked on. When a frame is grabbed, it moves to the grid-spot that it most closely matches – displacing any previous image that may have been there if and only if it is a closer

match. Over time, the final mosaic image will become more and more complete. One interesting implementation would be for the software to work on many images simultaneously. It could tune in different programs when they are broadcast and work on a relevant image at that time. For example, when *David Letterman* is on, a portrait made from his show could be constructed. Every day at the same time, the program will tune in and work on this image. Over time, it will get better and better, and will not be done until the artist is happy with it and all tiles are full. If the artist does not like a tile that the program selected, he could see all of the tiles previously selected for that grid-spot and use another instead. When Letterman ends, the software could switch to a different channel and work on another image for that time period. As months go by, many of these batch jobs will develop into completed works.
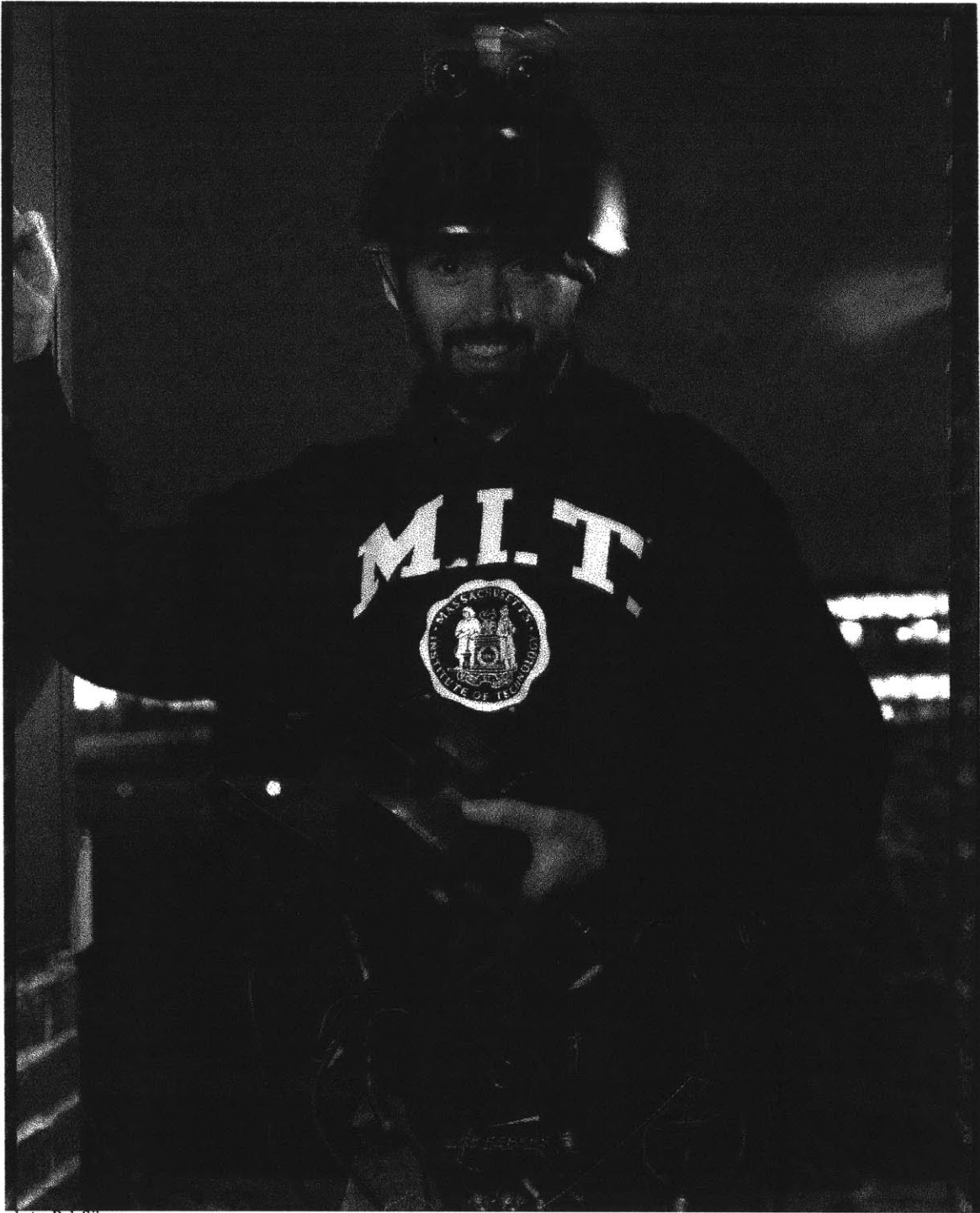
Figure 4.1: Wireless Netcam

4.2.5 Your Life as a Mosaic – Filled in Over Time

Steve Mann, at the time of this writing, is a Graduate Student at the MIT Media Lab, who wears a head-mounted video camera and display (http://www-white.media.mit.edu/~steve/netcam.html) that provides him with both augmented and mediated reality. Augmented reality means that he can see more than is normally available (perhaps name-tags could be superimposed over people's heads). Mediated reality means that the visor provides a *visual filter* [29] that governs what Steve can see. Suppose this stream of images – what you see during your lifetime – were pumped into a Photomosaic? Everything that you see since birth has the potential to become a tile in the mosaic of your life.

Chapter 5

Future Possibilities and Conclusion

5.1 Speed and Quality Enhancements

I have made a recognizable Lincoln using 8 x 12 photographs, which is better than possible with solid pixels. This has led me to set a new goal of making a recognizable Lincoln from 4 x 6 photographs. There are two ways to go about this: either consider millions of images during the selection process, or improve the criteria used for selection. I propose to use both options.

5.1.1 Larger and more Sophisticated Databases

The Photomosaic system should be integrated with a relational database that provides for advanced queries on numerous parameters. Images could then be stored in the database along with these other parameters as binary large objects (blobs), or simply as pointers to the real image file. An example query would be "find me an image that contains any animal or bird and is visually similar to this blob."

5.1.2 Search Optimizations

Because we are searching a static database, it is possible to sort or categorize the data and reduce the compute time by limiting the search to a smaller region. If we consider the color cube in figure 3.14, we can see that database images inhabit a region of space. This space can be segregated into smaller regions. For example, if we are looking for an image that is mostly green, do we need to search in the red portion of the 3D representation of the color space?

One possible method is to find the average color of the grid-space of the master image that we are trying to represent, and only search images in the database that are within a certain distance of this. If a match is not found, we can expand the search to a larger region. The problem is that this only approximates the ideal situation because it is possible that a better match was outside of the initial search area. Still, an N-fold increase in speed may be worth a small reduction in quality.

To implement this technique, it would be worth creating a Photomosaic the traditional way and making a histogram of the distance each selected tile was away from the estimated starting search region. This will help decide how much distance much be searched, what parameters should be used to determine the starting position, and how to best sort the database.

5.1.3 Singular Value Decomposition

My current algorithm is N,[2] making it memory inefficient, slow, and poorly scalable. One possible solution is to break down the image matrix into its salient components using a singular value decomposition, or SVD. The singular values may be sorted by significance and used for matching instead of the matrix values themselves. A large array of values could be pre-computed for all images in the database and sorted in advance, or they could be computed from images as they are being read in. The quality and speed trade-off could be adjusted during run-time by only matching the first N elements of this array.

Because the values are sorted by importance, each would be more important than the original values in the image matrix. This smaller list of values will consume less memory, may be matched more quickly, and will only double in size as the image matrix goes from say 8 x 8 to 16 x 16 instead of being squared.

5.1.4 Error Diffusion

Currently mosaics are made from the top down. The negative consequences of this are that the quality of matches is usually worse at the bottom of the mosaic than at the top because the best images are used up first. I have four solutions for this problem. The first is to simply randomize the order in which the tiles are worked on until they are all done. The second is to start in the center and spiral out.[iv] The third is to sort the tiles in their order of importance – perhaps starting with the eyes, nose, and mouth.[v] While these methods will help prevent viewers from seeing the progression of error in the final piece, they do not solve the greater problem. What if an image is selected to represent a tile early on, and is not free to be used later in the mosaic even if it would be a much more perfect fit in another location? To solve this, I propose the fourth solution: the matching process could first try each image on every tile location, while keep a sorted list of errors for these possible locations. While this is memory intensive, it is not much more work than what is currently being done (with the exception that images are not eliminated from the matching test as they get "used up"). Now with everything precomputed, optimally place the images until the least *overall* error is satisfied. This technique, unfortunately an N^2 algorithm, would be especially valuable when the database is small in relation to the number of mosaic tiles.

[iv] Suggested by Ken Knowlton during personal communication.
[v] Suggested by Mike Hawley during personal communication.

5.1.5 Adjacent Tile Communication

Tiles could "hold-hands" and communicate with each other. Perhaps a tile would "know" that the left side of itself helped define important detail in an image and that the tile attached to its left should have an aligned feature? Maybe the error of a tile could be made up for with the tiles around it? This comes down to the question of whether or not it is good to reduce the quality of a match because another tile was a poor match. I currently do not believe in this for large-tiled Photomosaics because this "dithering" effect only works well when there are a lot of tiles to dither. The other issue is that if tiles are dependent on each other, and you change a tile by hand, the ones around it may change on their own. If so, it makes hand-tuning more difficult. You could always not have the others change by themselves after you change one by hand, but then you are compromising the dithering process.
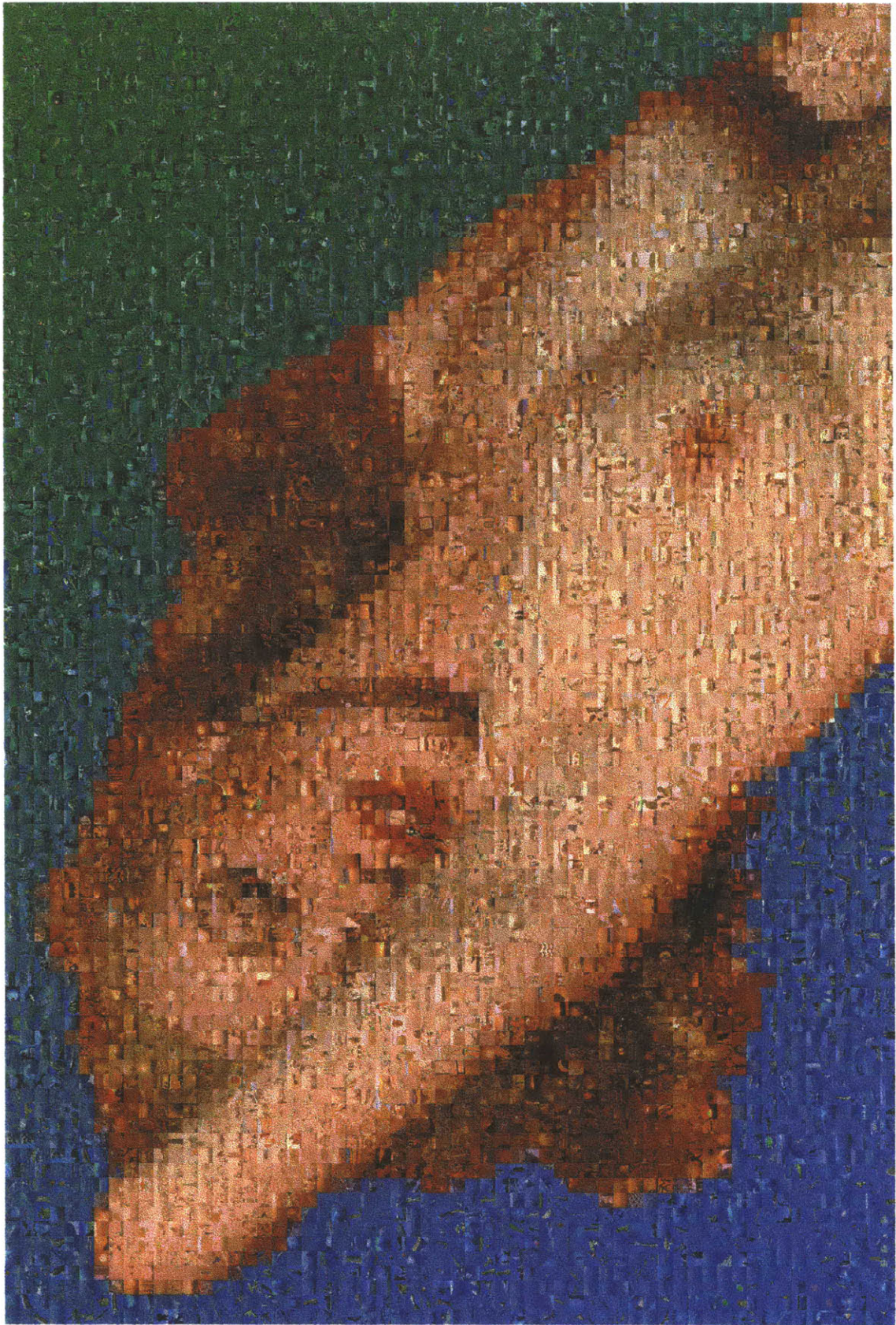
Figure 5.1: Rob Silvers, *Marilyn*, May 14th, 1995

The mosaic in figure 5.1, based on an image that first appeared in Playboy magazine, represents an iconic figure from Americana. Because of the attention-commanding nature of such provocative images, Marilyn has traveled from centerfold to postage stamp, becoming a symbol of our collective American culture. Such imagery evokes emotion in tune with our personal experiences upon viewing. Reactions to this image have ranged from finding it beautiful to offensive, depending the viewer's perspective.

Marilyn was an ambitious image that challenged my early mosaic constructing technology, and it was a breakthrough to achieve this range of color and smoothness of color graduation. Over 370,000 images were considered during the tile selection process, contributing to these properties. Marilyn is both my favorite and most popular mosaic. Is this because of the cultural baggage she carries, or because the composition is visually pleasing? I believe that both of these factors play a role.

5.1.6 Non-Grid Placement

Tiles could be placed in a non-grid arrangement. Hexagonal (honeycomb) is just another form of a grid, but what about completely arbitrary placement? Take an image in the database – each image – and try placing it at every possible location with different degrees of scaling. If there is a lower degree of error than what was underneath, place it and keep on trying. When all placements of a given image have been tried, move on to the next image. Any given image should only be placed once, so earlier matches should be removed.

5.1.7 Considering Noise Band

We may have two database images which the matching algorithm considers approximately equal quality and worthy of selection. One may contain errors between itself and its relative section of the master image at different spatial frequencies than the other. If this difference, considered noise, is within the band of the master image's salient content, it will hinder overall perception more than if it is out of band. The presence of noise should be considered during the matching process (see "Issues Relating to Coarse Pictures," page 70).

5.1.8 Tweaking to Fit

Images must currently be cropped from rectangle to square in the Photomosaic process. This square is chosen from the center, although a human operator may recrop it by hand later. What if the computer could crop, wiggle, and tweak the image on demand to optimize the similarity with the underlying section of the master image?[vi] This would greatly increase the quality of the matches and add additional diversity between finished mosaics.

   If one wanted to make a Lincoln portrait from images of the Civil War, and only had 1300 such images, dynamic tweaking would help a great deal. In one possible implementation, a human operator would mark the range of cropping and tweaking allowed using an automated tool. The computer would then work within these constraints. Another method would be to allow automatic changes up to a specified percent. Perhaps a 10 percent range of automatic variation would result in a significant percent improvement of match-quality.

5.1.9 Placing Required Images

What if an artist needs to ensure that an image of Hitchcock appeared in a mosaic made from his movie frames? The software should take a list of images that must be placed and put them into their best-fitting location before starting the rest of the mosaic. The currently implemented software only considers placement based on merit and does not affirmatize tiles with this placing-holding.

[vi] Suggested by Mike Hawley during personal communication.

5.1.10 Infinite Zoom

Photomosaics are not limited to producing just an image from images. They may also be used to make an image from images which are in turn made from more images.

5.2 Conclusion

Photomosaics are mosaics made from images that are selected because they are visually similar to the region of a larger, master image that they replace. The method used for finding the visual similarity of two images was comparing the difference of RMS error within sub-regions.

Photomosaics are multi-scale images in that they take on a different character when viewed up close than when seen from a distance. Having fewer image tiles that each convey more information makes for a more seamless and surprising result. The relevant information content of each tile and level of seamlessness was increased by sub-picture resolution, or the consideration of regions within images. Having more sub-regions not only helped for images containing detail, but also for "empty" images because the match did not simply have a similar average color but also a similar color distribution.

Beyond the basics of matching images, several enhancements were uncovered relating to colors, modifications, and additional search parameters.

Images have color and databases provide a color space, or range of color coverage. If these colors fit or are forced to fit into the space, recognizability is improved.

Tweaking images is a "destructive" (to the images) method of increasing matching ability, and should turn many near misses into better matches. Images could be twiddled, zoomed, cropped, or wiggled into place. Handling images in such a way would have the effect of expanding the searchable space.

Considering noise as a parameter, or the difference between a selected sub-image and the region of the master image within the same gridspot, is a non-destructive method of improving the image-forming ability of a Photomosaic. Noise has the ability to mask a signal in proportion to its proximity to the relevant information, so considering noise should unmask more relevant features.

Finally, semantics can play an important role in the application of Photomosaics. Enhanced methods of categorizing images by content are needed.

The results of this thesis provide only a first attempt at creating mosaics from photographs. However, the attempt has produced results that show promise for more optimization and consideration of salient features, whether on the low-level of image attributes, or on a higher, semantic plane.

Appendix

Tools for Creating Photomosaics

The process of creating Photomosaics requires custom software as well as off-the-shelf image processing, conversion, and paint programs. Following is a summary of the most important tools that I created for the Photomosaics project.

glmosaic images_location

This is the main Photomosaic program and is responsible for reading in the database, analyzing a master image, and selecting tiles to construct a mosaic. Glmosaic, when first invoked, reads in a version of the image database that is subsampled to the same degree as the number of sub-regions being considered for matching. For example, if 8 x 8 sub-regions will be considered, then 8 x 8 pixel images are read into memory for use in the matching process. After startup, a menu gives the operator a number of choices. These choices include the ability to read in a master image and start rendering, to read in an output file previously generated so that the operator can continue to refine a previously computed mosaic, to write out a datafile, and to hand-edit a computed image. If a Photomosaic has been computed, the output file (filename.out) saved is needed for the next step of rendering.

The ability to hand edit a computed Photomosaic is an important one because it gives the artist control over the final result. If the edit option is selected, the operator may click on a tile and see many alternatives – scrolling through a list of other images to not only visually see what they look like in their grid-space, but also what their numerical error is. If the artist likes a tile, but not the way it was cropped square from its original rectangular form, then he may opt to re-crop the image. If the re-crop key is pressed, the image opens in its raw, rectangular form. Cropping lines may be moved and zoomed to select a new area of interest, and when done, the new image is reintegrated into the database as the artist prefers. Over time as the artist fixes more and more problems, the database becomes ever more refined.

render images_location datafile

Render must be given two parameters: a location where the database images are stored, and the datafile that is output from glmosaic. This datafile contains a list of images to use for building a mosaic. Each image in the list is pulled from the database and rendered into a final bitmap with the same name as the input datafile, but with a .sgi extension replacing .out. A grid of colored lines is optionally placed between tiles to help set them apart.

makequant `image_directories`

This program is used to analyze the colorspace of a set of images. It outputs an image where each pixel is the average color of an image found in the directory of images. This new file may then be loaded into an image processing program and quantized to a color-index image. The palette of this indexed image may be used to colorize a master image so that the colors more accurately map into the database.

3Dcube `image_directories`

This program draws a cube in three-dimensional space where R, G, and B maps into X, Y, and Z. Images in the database are read in one at a time and the image's average color maps to a related point in space. The usefulness of this utility comes from the ability to show the artist clusters of colors – where the most colors are. If the artist sees a lot of blues, he may choose to use this specific database for mosaics that are mostly to be blue. Black and white images, incidentally, may be seen as a line that moves through all axes uniformly.

make_square `size output_path directories`

This program accepts a list of directories, an image size, and an output path on the command line as input. It then opens each directory and searches for rectangular images in which to crop the largest possible square image from. This square image is then resampled down to the specified size and moved to the output path. If the original image was in landscape format, the square is cropped from the center. If the original image was in portrait format, the square is cropped closer to the top so that if the subject is a person, the face will (hopefully) be included.

resamp `size output_path directories`

This program will take a list of image directories and search them for images to convert to a new size, and places the converted images on the output path. Resamp is useful for converting and storing images at multiple resolutions.

References

[1]   MARTIN HANDFORD. *Where's Waldo*. Little Brown and Company, 1987.

[2]   KIT WILLIAMS. *Untitled*. Alfred A. Knopf. New York, 1984.

[3]   DAVID L. FISK. Arthur S. Mole: The photographer from Zion and the composer of the world's first "Living Photographs." *The History of Photography Monograph Series*, No. 3, Arizona State University.

[4]   "I didn't raise my boy to be an eyebrow," Parting Shots, *Life*, 24 September 1971, p. 82.

[5]   DAVID HERREN. From Peach Picker to Photo Artist, Arthur S. Mole, *The Times (Fort Lauderdale)*, October 9, 1968.

[6]   LEON D. HARMON AND KENNETH C. KNOWLTON. Picture Processing by Computer. *Science* April 1969, Volume 163, pp 19-29.

[7]   LEON D. HARMON. The Recognition of Faces. *Scientific American*, November, 1973, pages 71-82.

[8]   KEN KNOWLTON AND LEON HARMON, Computer-Produced Grey Scales. *Computer Graphics and Image Processing*, Vol. 1, No. 1, April 1972.

[9]   SCOT MORRIS. GAMES. *OMNI Magazine*, May 1981, pages 136-137.

[10]   Laura Teodosio and Walter Bender. Salient Stills from Video. *Proceedings of the ACM Multimedia '93 Conference*. August, 1993.

[11]   A. PENTLAND, R. W. PICARD, S. SCLAROFF. Photobook: Content-Based manipulation of Image Databases. MIT Media Laboratory Perceptual Computing technical Report No. 255, Nov. 1993.

[12]   ELIZABETH BINAGHI, ISABELLA GAGLIARDI, AND RAIMONDO SCHETTINI. Indexing and fuzzy logic-based retrieval of color images, In *Visual Database Systems, II, IFIP Transactions A-7*, pages 79-92. Elsevier Science Publishers, 1992.

[13]   MIKIHIROI IOKA. A Method of defining the similarity of images on the basis of color information. Technical report RT-0030, IBM Tokyo Research Lab, 1989.

[14]   MICHAEL J. SWAIN AND DANA H. BALLARD. Color indexing. *International Journal of Computer Vision*, 8(1):11-32, 1991.

[15]   MONIKA M. GORKANI AND ROSILAND PICARD. Texture Orientation for Sorting Photos "at a Glance." *IEEE Conference on Pattern Recognition*, Jerusalem, 1994.

[16]   R. W. PICARD AND M. GORKANI. Finding perceptually dominant orientations in natural textures. Spatial Vision, Spec. Perceptual Computing TR #229, MIT Media Lab, 1993.

[17]   R. W. PICARD AND T. KABIR. Finding similar patterns in large image databases. In *Proc. ICASSP*, pages V-161-V-164, Minneapolis, MN, 1993.

[18]   A. R. RAO AND LOHSE. Towards a texture naming system: Identifying relevant dimensions of texture. In *IEEE Conference on Visualization*. San Jose, 1993.

[19]   A. R. RAO AND LOHSE. Identifying high level features of texture perception. *Computer Science* RC17629 #77673, IBM, 1992.

[20]   FANG LIU AND ROSALIND W. PICARD. Periodicity, directionality, and randomness: Wold features for perceptual pattern recognition. *Proceedings of the 12th International Conference on Pattern Recognition*, Jerusalem, Israel, 1994.

[21]   H. TAMURA, S. MORI, AND T. YAMAWAKI. Textural features correspond-
        ing to visual perception. *IEEE T. Sys., Man and Cyber.*, SMC-8(6):460-
        473, 1978.

[22]   M. GORKANI. Designing an orientation finding algorithm based on hu-
        man visual data. Master's thesis, MIT, 1993. Perceptual Computing
        TR #222, MIT Media Laboratory, 1993.

[23]   TOSHIKAZU KATO, TAKIO KURITA, NOBUYUKI OTSU, AND KYOJI HIRA-
        TA. A sketch retrieval method for full color image databases. In *Interna-
        tional Conference on pattern Recognition (ICPR)*, pages 530-533, The
        Hague, The Netherlands, September 1992. IAPR.

[24]   DAVID MUMFORD. The problem with robust shape descriptions. In
        First International Conference on Computer Vision, pages 602-606, Lon-
        don, England, June 1987. IEEE.

[25]   DAVID MUMFORD. Mathematical theories of shape: Do they model per-
        ception? In *Geometric Methods in Computer Vision*, volume 1570, pages
        2-10. SPIE, 1991.

[26]   E. GLASMAN, M. FLICKNER, W. NIBLACK, D. PETKOVIC, W. EQUITZ, R.
        BARBER, AND P. YANKER. QBIC project: Querying images by content us-
        ing color, texture, and shape. In *IS&T/SPIE 1993 International Sympo-
        sium on Electronic Imaging: Science & Technology, Conference 1908,
        Storage and Retrieval for Image and Video Databases*, February 1993.

[27]   KYOJI HIRATA AND TOSHIKAZU KATO. Query by visual example. In *Ad-
        vances in Database Technology EDBT '92. Third International Conference
        on Extending Database Technology*, Vienna, Austria, March 1992.
        Springer-Verlag.

[28]   HECKBERT, P. S. Color Image Quantization for Frame Buffer Display,
        ACM Computer Graphics (SIGGRAPH). 16(3), 297-307, 1982.

[29]   MANN, S. Mediated Reality. MIT Media Laboratory Perceptual Com-
        puting Section Technical Report No. 260. 1994.