UNIVERSITY OF CALIFORNIA
Santa Barbara

WorldMediaViewer: 3D Interactive Display of Multimedia Collections Using Spatial, Temporal and Orientation Metadata

A Project submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Multimedia Engineering
Media Arts & Technology

by

David E. Gordon

Committee in Charge:

Professor Curtis Roads, Chair

Professor Clarence Barlow

Professor JoAnn Kuchera-Morin

December 2016

The Project of
David E. Gordon is approved:

Professor Clarence Barlow

Professor JoAnn Kuchera-Morin

Professor Curtis Roads, Committee Chair

December 2016

WorldMediaViewer: 3D Interactive Display of Multimedia Collections Using

Spatial, Temporal and Orientation Metadata

by

David E. Gordon

To John Cage.

Acknowledgements

I would like to thank the members of my committee for their help in preparing my project and this document: Curtis Roads, Clarence Barlow, and JoAnn Kuchera-Morin.

Abstract

WorldMediaViewer: 3D Interactive Display of Multimedia Collections Using Spatial, Temporal and Orientation Metadata

David E. Gordon

WorldMediaViewer is software for interactively displaying multimedia collections as 3D virtual environments using time, spatial and orientation metadata. Users can create, display and navigate scenes made from rectangular images, panoramas, sounds, and videos, in which media appear over time in virtual space at the approximate locations they were captured in the real world.

A hybrid system linking data visualization, worldmaking, and interactive cinema, *WorldMediaViewer* uses a *metadata-based* spatial and temporal model to create virtual environments from multimedia collections. This metadata-based model allows data and metadata to reinforce each other, providing a flexible alternative to similar systems using image-based geometric models, while bringing browsing large media collections closer to the experience of taking photos, capturing video, or recording sounds in the field.

In addition to offering an innovative method for browsing media collections, *WorldMediaViewer* has a wide range of creative applications. Through displaying media captured in close spatial and temporal proximity, the software lets users

intuitively perceive metadata in a way that not only facilitates detection of incorrect GPS locations, orientations, dates and times, but also opens a window onto the creative process.

With *WorldMediaViewer*, a photographer can visualize the progression of seasons at a particular location, while a sound artist can create an audio installation that moves the viewer smoothly between field recordings in the order they were captured. Through the same framework, the two also can collaborate, simply by capturing and importing images and sounds taken in the same geographical area.

By harnessing the power of metadata, *WorldMediaViewer* offers a flexible framework for artists with little to no technical background and without costly equipment to create and collaborate on 3D virtual environments using small or large multimedia collections.

Professor Curtis Roads
Master's Degree Chair

Contents

List of Figures

Chapter 1

Introduction

The ubiquity of digital technology has made capturing, storing and sharing images and videos some of the most familiar contemporary activities. While debates about the pros and cons of the endless proliferation of recorded media have become common, relatively less attention has been paid until recently to the fact that the majority of today's digital cameras and camcorders not only store image pixels or video frames, but also record *metadata*.

Often referred to as "data about data," metadata is defined by the National Information Standards Organization (NISO) as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource"[11]. Software on most smartphones and digital cameras is now capable of recording metadata for geographic location, altitude, focal length, exposure time, file type, capture device and many other aspects of media files.

Of these metadata fields, geographic location receives by far the most attention, perhaps due to the wide popularity of digital mapping services, the frequent appearance of geolocation in news media and films, or because it requires the least technical expertise to understand.

Another important contributor to this growing awareness of location metadata is that many social media platforms have begun to incorporate it in their image and video sharing services. Map displays with small overlaid icons of the photos, known as "thumbnails," allow users to view images on a satellite map or browse photos taken by friends who traveled to the same areas. Besides social media platforms, photo library management programs, such as Apple Photos or Adobe Lightroom, also offer tools such as map browsing, sorting and filtering based on location metadata.

Map image overlays attempt to provide a simpler, more intuitive method of browsing and retrieving media at particular locations than the list or ordered grid. However, if one considers the typical map image overlay as a data visualization, the majority of data seems to be missing. A map image overlay usually incorporates only latitude and longitude, omitting features such as compass direction, altitude, focal length, aperture and focus distance. How might new browsing interfaces unlock the potential of this unused metadata? What aesthetic value could this metadata hold for artists developing interactive multimedia artworks?

WorldMediaViewer is a software framework for the display of large multimedia collections as interactive, navigable, virtual environments based on metadata. By ordering media in virtual space and time using a wide range of spatial, temporal and orientation metadata currently available in most smartphones, it not only attempts to provide a more intuitive interface for media browsing than the map image overlay, but to serve as a platform for a wide range of creative multimedia projects.

By ordering media in virtual space over time according to location, orientation, and time metadata, *WorldMediaViewer* allows viewers to experience the sights and sounds of a place in the approximate virtual locations and temporal sequence that they appeared in the real world. *WorldMediaViewer* enables multimedia artists, photographers and composers, without an extensive technical background or costly equipment, to create and collaborate on fixed, interactive or hybrid audiovisual installations that simultaneously capture the subjective experience of a particular place and time and reveal aspects of attention, memory and the creative process.

WorldMediaViewer attempts to harness both the ubiquity of digital photography and the increasing prevalence of metadata to reveal the full depth of information contained in multimedia collections. Its *metadata-based* model for spatially and temporally ordering media might be described as a method for displaying metadata *through* media, while also presenting media *through* metadata.

Due to the flexibility and low computational cost of this metadata-based model, unlike the computationally intensive image-based geometric models, *WorldMediaViewer* allows users to create virtual environments spanning a wide range of sizes, spatial arrangements, time sequences, and media types. It also allows the modification of parameters of the scene such as altitude scaling or default object distance, while viewing the results in real-time.

The incorporation of sounds, videos and 360-degree panoramas, in addition to rectangular images, distinguishes *WorldMediaViewer* from previous systems as a tool for the interactive display of truly *multimedia* collections. The ability to combine time-based media with images and panoramas that fade over time allows multimedia artists to design a wide array of potential viewer experiences, while offering an innovative solution to the problem of efficiently browsing media collections of ever-increasing size.

WorldMediaViewer creates a virtual space where media data and metadata reinforce each other, freeing pixels from the confines of the rectangular frame, while unlocking the potential of metadata for more intuitive, meaningful methods of user interaction. Although its metadata-based technique often produces a less ”realistic” result than feature-matching methods, its visualizations can be surprisingly accurate, while being much less computationally intensive.

This metadata method not only gives *WorldMediaViewer* the flexibility to incorporate time-based media, such as videos and sounds, and create scenes covering large geographical areas; it enables, for instance, importing and following real world GPS tracks as virtual paths, navigating to locations where media were recorded both in the morning and evening, or selecting and filtering media based on parameters such as orientation, pixel brightness, focal length, or keyword.

As the cost of digital recording devices and media storage decreases, media collections will only continue to expand in size and complexity. As a result, metadata will likely continue to gain relevance as a method for organizing, browsing and retrieving media files. By harnessing the power of metadata, *WorldMediaViewer* offers one possible solution to the problem of navigating media collections of increasing size, while offering an innovative multimedia platform for artistic creation and collaboration.

The following chapters give a brief history of related multimedia systems and artworks, explain the software's design and methodology, describe the author's use of *WorldMediaViewer* in creating virtual environments for digital installations, and outline possibilities for future work.

Chapter 2

Previous Work

2.0.1 Cubism

Cubism has been called the earliest and most influential of Modern art movements. Cubist paintings, in particular, works from what is commonly called the "Analytical Cubist" period of Pablo Picasso and Georges Braque, between c. 1908-1912, are one of the major inspirations for *WorldMediaViewer*. These paintings depict space using a combination of shifting viewpoints and the depiction of volume using flat planes, or "planar faceting"[2].

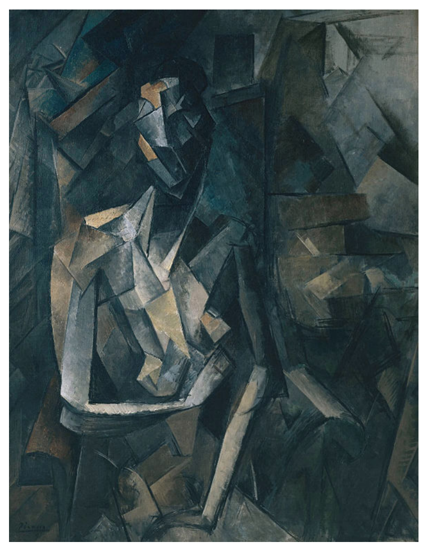

Figure 2.1: *Figure dans un Fauteuil (Seated Nude, Femme nue assise)*, Pablo Picasso (1909-10)

In their book *Du cubisme*, painters Albert Gleizes and Jean Metzinger describe one of the central concerns of the Cubists in this groundbreaking period as the expression of the subjective notion of time proposed by the philosopher Henri Bergson in his notion of "duration." They explicitly relate "this sense of time to multiple perspective and to Bergson's insistence on the elasticity of our consciousness of both time and space"[2].

The techniques of compositing multiple viewpoints and planar faceting, as well as the interest in the subjective perception of time, are among the many influences of Cubism on the *WorldMediaViewer* software. Similar to Cubist painting, in building up 3D scenes out of 2D images and videos, the program essentially treats media as individual samples or windows on perception, through which both the larger picture and a document of the artist's particular attention to a space emerge.

2.0.2 Hockney's Photocollages

Besides Cubism, perhaps the most direct artistic precursor to *WorldMediaViewer* is found in David Hockney's photocollages (Fig. 2.2). In the early 1980s, Hockney, already a famous painter, began making large composite landscapes and portraits by stitching photographs and Polaroids together by hand.

A major breakthrough, Hockney's photocollages are often cited as being as influential on the history of photography as Cubism was for the history of painting.

Just as the Cubists created paintings that composite multiple viewpoints, Hockney uses up to 120 single photographs in each of his collages, to attempt what he sees as overcoming the limited perspective of a stationary camera. In these photocollages, "as in Cubistic painting to a certain degree, the eye searches for and finds many points of focus and many moments of time."[3].

Figure 2.2: *Merced River, Yosemite Valley*, David Hockney (1982)

Through bringing out both large and subtle juxtapositions between different scales, viewpoints and exposure settings, Hockney's photocollages achieve the remarkable feat of simultaneously portraying a subject and revealing many aspects of the photographic process. In this way, they paradoxically give the impression of time in a medium usually considered inherently static.

Along the same lines, *WorldMediaViewer* seeks to further pursue the incorporation of the temporal dimension into photography, by incorporating multiple viewpoints, allowing photos to fade into each other, displaying still images in the context of time-based media, and enabling the viewer to navigate freely around each scene.

2.0.3 Aspen Moviemap

Another significant artistic precursor, anticipating many of the ideas investigated in *WorldMediaViewer*, is the Aspen Moviemap project, produced at MIT in the late 1970s. Video footage was recorded of the streets of Aspen, Colorado, captured between 10 a.m. and 2 p.m. each day. This footage was then turned into an interactive system, which was displayed at the MIT Architecture Machine Group and Media Laboratory demo in 1979.

This interactive display used several laserdisc players, a computer and a touch screen display to allow viewers to choose an arbitrary path through the city in real-time. Users could even touch any building in the current view and jump to a virtual tour of that building, including interior shots, historical images and other related media [9] [16].

One of the earliest examples of a *surrogate travel* system, the Aspen Moviemap not only inspired subsequent Moviemaps in other cities, but was also a major influence on today's Google Street View and similar projects. After working on the Moviemap project while at MIT, media artist Michael Naimark has continued to create powerful installations investigating landscapes and public spaces. His *Be Now Here* installation was produced between 1995-97 using a custom-built recording system consisting of two 35mm motion picture cameras position to capture input for each eye, for capturing 3D footage. In the installation, 3D glasses, four-channel sound, and a 16-foot rotating floor, immerse viewers in a virtual environment composed of 3D panoramic videos of various public plazas across the world [10].

2.0.4 Norman Klein, *Layers of Los Angeles*

A more recent work, which helped inspire the concept of *WorldMediaViewer*, is Norman Klein's *Layers of Los Angeles*, an interactive DVD created in 2003, in which the viewer encounters modern photographs of Los Angeles, layered with early 20th century photographs taken at the same locations.

Klein "uses digital images in which vintage photos of city streets dissolve, or bleed, into shots of the same scenes as they look today"[17]. On the motivation for the project, he explains to an interviewer: "Los Angeles is a place where memory is elusive. That means we have to find a new way to see. This new way has to be based on layers. Suppose we think of the city as a collage... [in] about 15 layers, all stacked and translucent and opaque" [17].

This concept of merging and blending media from the past and present to evoke ideas of memory and the subjective experience of a place over time is a driving motivation behind the integral roles of time and layering in *WorldMediaViewer* environments. Designed to allow creation of multimedia works with this particular aesthetic in mind, the program, like *Layers of Los Angeles*, uses technology to enhance the experience of photos, while focusing viewers' attention on their local environments.

2.0.5 Audio Walks

The connection between ambulatory movement, time, place and recorded sound has also been explored by Canadian artist Janet Cardiff in a series of audio walks starting in 1991 with *Forest Walk*. For each walk, the viewer is given a CD player or iPod and headphones and told to sit in a particular spot and press play. At this point, the artist's voice gives instructions on where to walk and which direction to turn. Non-verbal sounds such as footsteps, birds or traffic indicate that these were recorded on the same site where the viewer is standing.

In the manner of Richard Long's seminal performance piece, *A Line Made by Walking*, or the approach of British artist Hamish Fulton, Cardiff's audio walks invite the viewer to consider the acts of walking and listening as an artform in themselves. As one participant notes, "Cardiff's cool, breathy voice was my nearly constant companion, pointing out features of the buildings history, or sharing a dream, or issuing simple instructions"[19]. The viewer described walking and listening to sounds and narration as a powerful, even somewhat surreal, experience, identifying the central problem investigated in the work as "the impossibility of being in two times and places at once" [19].

While examples of prior work previously discussed have all included a visual component, Cardiff's audio walks depend on sound for the primary element; though they may also be considered sound pieces with visuals "curated" by the artist, through the act of determining each of the locations and points of view to walk. More recently, Cardiff's work has begun incorporating printed or digital images and iPhone videos.

Similar to the other works discussed, issues of place, time and memory play a major role in Cardiff's audio walks, though, in their case, primarily in the context of sound. The process of layering is another key component in her work, as Cardiff explains: "The virtual recorded soundscape has to mimic the real physical one in order to create a new world as a seamless combination of the two"[1].

Through these sound collage elements and her narrations, Cardiff's audio walks, in effect, *implant* virtual elements within the real world. Conversely, *WorldMediaViewer* transports sounds, videos and images from reality into the virtual world. In spite of these different approaches, both Cardiff's work and the software rely on processes of layering and collage to propose that walking can be an indispensable source of art.

2.0.6 Narrative Maps

Narrative maps describe real or fictional accounts in which spatial relations, geographic locations or movement play an integral role. One researcher explains: ”A narrative map tells a story plotted through space... to provide explicit visual counterpart to the implicit spatial underpinnings of a narrative or argument” [8]. Research centers investigating narrative maps include the SLAB at the University of Southern California Sol Price School of Public Policy, which is interested in ”spatial ethnography and critical cartography... [to seek] new insights by incorporating empirical, field-based research with visual experimentation” [12].

In their recent *Sidewalk City* project, a team led by director Annette M. Kim investigated the experiences of people in the historic Saigon and Cholon neighborhoods of Ho Chi Minh City, Vietnam, from both geographical and ethnographical perspectives. The team recorded both ”the spatial patterns and the social relations of over 6,490 people out on the sidewalk engaged in activities beyond pedestrianism... [including] field surveys and participant observations coded into GIS, interviews with over 270 vendors as well as police and other residents, photography, and videography”[13]. This project culminated in a 2014 exhibition in Ho Chi Minh City and a book about the project, ”Sidewalk City: Remapping Public Space in Ho Chi Minh City” published in 2015.

In locating multimedia collections in virtual space using metadata, while offering several different navigation methods, *WorldMediaViewer* may be seen as a kind of narrative mapping tool for illustrating the spatial movements of photographers, videographers and sound artists over time.

2.0.7 Narrative Collage

A time-based analogue to narrative maps using only images, the form of "narrative collage" has also recently been investigated by researchers including Lei Zhang and Hua Huang at the Beijing Institute of Technology. Their *AutoCollage* software aims to digitally create "a collage form that enables apparent chronological presentation [of] the attracting events recorded by... photos" by adopting "the structure and organization of narrative to construct the photo collage"[20].

Through computer vision analysis techniques, *AutoCollage* detects regions of interest in each image, builds a narrative structure using hierarchical clustering to reconstruct the three main narrative elements: character, setting and plot; and blends the results together to create each collage. Both the spatial aspects studied by narrative maps and temporal order captured in narrative collages inform the metadata-based spatial and temporal model used in *WorldMediaViewer*.

2.0.8 Photosynth

Perhaps the best known tool for creating 3D navigable environments from image collections is Microsoft Photosynth. The research behind this groundbreaking system took place in 2006-7 as the *Photo Tourism* project by researchers at the University of Washington and Microsoft Live Labs. Shortly after it was completed, the technology was implemented as the online sharing service *Photosynth* by Microsoft.

To create highly accurate, seamless virtual scenes out of image collections, Photosynth uses feature matching and 3D modeling techniques to create image-based geometric models of where the images were taken. In developing the technology behind *Photosynth*, the authors describe their main goals as ”to geometrically register large photo collections from the Internet and other sources, and to use the resulting 3D camera and scene information to facilitate a number of applications in visualization, localization, image browsing, and other areas”[14]. They further note that their system does not ”rely on the camera or any other piece of equipment to provide us with location, orientation, or geometry. Instead, we compute this information from the images themselves using computer vision techniques”[14].

Under the right conditions, this approach can produce remarkably convincing environments from images. Many impressive renderings of buildings, monuments and well-known public places are available on the Photosynth website for viewing and sharing. At the same time, the method has a tendency to produce certain characteristic distortions, which can become distracting when large enough to be noticeable.

Perhaps a more significant problem, at least in terms of the subjective experience of place, is that, in creating the image-based models, Photosynth effectively ignores the time element as an aspect of the scenes. In describing their method, the authors note their interest in solving computer vision research problems, and clearly define their goal as assimilating all the images into a geometric model. Their paper does not mention time, except to define their primary challenge as finding a method to "robustly match and reconstruct 3D information from hundreds or thousands of images that exhibit large variations in viewpoint, illumination, weather conditions, resolution, etc., and may contain significant clutter and outliers"[14].

If one considers the problem of turning images into a 3D environment from a strictly aesthetic viewpoint, the merits of eliminating variations in weather, illumination and resolution, or of subsuming all the individual images into a single geometric model, are less clear. In considering blurry images, glare spots, and under- or overexposed images to be mistakes, current image-based geometric models leave behind photographers who have used each of these "errors" for aesthetic purposes for over a century. The further assumption that gaps, distortions, and juxtapositions are inherently undesirable also leaves out elements that are central to many contemporary artists.

Precisely this image-based geometric model, which allows Photosynth to create highly realistic, seamless renderings, also prevents the user from generating incomplete or ambiguous spaces, incorporating heavily digitally manipulated images, showing progression over time, or adding time-based media to a scene.

To create 3D environments that change over time using media collections, artists currently must build their own custom frameworks or work around the implicit assumptions of existing systems. With the ability to fade images, reveal both subtle transitions and contrasts, and incorporate time-based media, *WorldMediaViewer* provides a flexible alternative to image-based methods: a system that embraces the aesthetic and perceptual qualities of motion, change, incompleteness, and discontinuity.

2.0.9 Photo Library Management Software

Photo library management programs, such as Apple's Photos, Adobe Lightroom or ACD System's ACDSee, offer various tools for photographers to import, browse, organize, edit and export image data and, in some cases, metadata. In media library management programs, the ordered grid remains the most common browsing interface[6].

Compared to a simple file list, the ordered grid makes excellent use of limited screen space and serves as a logical method for sorting by one parameter at a time. However, like the file list, it seems to offer little hope of improving file browsing and retrieval for large media libraries. As one researcher observes, "current applications for handling image collections still lag behind the varied and fluent ways of interacting that traditional informal collections offer"[6].

Unfortunately, in spite of the growing popularity of aerial map views, both the ordered grid and file list still dominate the market in commercial media library management software tools. However, the increasing presence of metadata in digital media offers one possible foundation for new interfaces that augment or even someday replace the current ones.

While some programs, such as ACDSee and Apple Aperture, have supported viewing, searching and even editing of metadata, others choose to hide metadata from users. Apple's removal of custom metadata filters in its transition from Aperture to the current Photos app eliminated one of the few mainstream tools for photographers to sort and organize by a wide range of metadata fields. A new feature in iOS 10, "Memories," seeks to make browsing more intuitive; however this service currently allows little customization, performing all the organizing tasks behind the scenes and leaving users unaware of its criteria or methods for selecting images for albums.

In addressing the problem of efficiently and intuitively organizing and browsing multimedia libraries of ever-increasing size, photo and media library management tools require new interfaces that break with "the grid aesthetics and [offer] more loose and expressive visualization" to support "a more informal and goal-oriented interaction instead of the common task-oriented approach"[6].

2.0.10 My Own Work

The *WorldMediaViewer* software grew out of the author's own practice as a photographer, composer, and media artist. As a part of this practice, many hours are spent organizing, labeling, editing, and exporting image, sound and video files. While this process may seem tedious, there is also a certain satisfaction in going through media captured during past travels. In breaking with the ordered grid, file list, and map image overlay interfaces, one of *WorldMediaViewer*'s main goals is to perceptually link browsing large multimedia collections with the experience of traveling itself.

Two major themes in my previous work that have informed *WorldMediaViewer* are an interest in the natural world and in the processes of layering and collage. As a composer, my interest in field recordings began as early as *Birdscape* (2011) for stereo sound and *Transformation* (2012) for eight channel sound. These fixed media works build up layers of time-stretched and processed field recordings to create complex sound collages, in which the origins of sounds often become obscured.

This layering approach and interest in the natural world are further evident in three video pieces, *Patterns* (2014), *Emergence* (2015), and *Parallels* (2015), investigating the abstraction of visual motion from video footage of Santa Barbara streams, ponds and the Pacific Ocean. Although recently my technique has shifted away from the use of digital processing towards the use of unprocessed footage, this interest in images and sounds captured from natural areas has persisted in my work.

An important breakthrough in my work, which took place in the early stages of developing *WorldMediaViewer*, was the video installation, *Transitions* (2015). The piece was premiered in the UCSB Allosphere at the 2015 Media Arts and Technology End of Year Show. The material for *Transitions* consists of hundreds of 360-degree panoramic images taken in each of three Santa Barbara parks.

Figure 2.3: *Transitions*, Screenshot 1, David Gordon (2015)

These panoramas, taken over several months using Occipital's 360 Panorama app for iPhone, are ordered according to geographical proximity, and blended together to create a continuous visual transition that reflects a gradual movement through the different environments. The layering and blending method results in the viewer's always seeing three of the videos simultaneously: one at full brightness, one fading in, and another fading out. Sound recordings taken in the corresponding locations were time-stretched and blended to create gradual sound transitions analogous to this visual process.

Figure 2.4: *Transitions*, Screenshot 2, David Gordon (2015)

In composing *Transitions*, metadata was an indispensable tool. While the Aperture map overlay view saved much time in finding the correct order of panoramas, the still tedious work of manually ordering the videos using Final Cut Pro motivated the addition of the Orientation Mode feature in *WorldMediaViewer*, which enables the creation of these spatial transitions algorithmically.

Chapter 3

Process

This chapter analyzes the process used to create virtual environments out of large multimedia collections based on metadata in *WorldMediaViewer*. The first section provides an overview of basic metadata concepts and explains their uses in digital images, sounds and video files. The second section introduces the main *WorldMediaViewer* class structure. The third section describes the core of the system: a metadata-based model for ordering media in space and time. The fourth section discusses the main display views used in the software: World View, Model View and 2D Map View. Finally, the last section covers the important topic of user interaction: navigation, metadata viewing, panoramic stitching, and user modifications of the model, graphics and navigation settings.

3.1 Metadata

3.1.1 Overview

In a general sense, *metadata* refers to data describing other data. Before digital databases entered widespread use in the 1980s, metadata was primarily used in library card catalogs. In subsequent years, several metadata standards have emerged for establishing common classifications to facilitate organizing and preserving electronic resources and archives. These standards allow users to easily identify digital files, especially on a distributed network such as the internet.

For digital media files, such as images, sounds, and videos, "metadata" essentially describes any information in the file besides the media data themselves, whether image pixels, sound samples or video frames. Digital recording devices, such as cameras, camcorders, smartphones, and audio recorders, commonly include metadata information, or "tags," describing: time and location captured, recording device and software used, metadata standards used, file size, and various device-specific settings. Other common descriptors include fields for authorship, keywords and licensing information.

Many tags have meaning only for specific media types, for example video frame rate or audio bit depth. Since the core of *WorldMediaViewer* is a metadata-based model for displaying multimedia files, the following overview of the uses of metadata in digital images, sounds and videos provides a necessary introduction to understanding the project's methodology.

3.1.2 Image Metadata

Metadata for images, compared to other media, has received the most widespread public attention, due to the recent popularity of image sharing on social media websites and applications. The popular term "geotagging" refers to the process of adding location metadata, whether by naming a large geographical area such as "Los Angeles" or "Balboa Park," or by including more precise GPS coordinates.

The most frequently used metadata standard by digital cameras and smartphones to record technical information about an image is the Exchangeable Image File Format (EXIF) standard, initially released in 1998 [5]. EXIF metadata contains fields such as the GPS location information, altitude, direction, orientation, camera make and model, focal length, shutter speed, and field of view. In addition to EXIF metadata, images often include metadata under the International Press Telecommunications Council (IPTC) standard. The IPTC website describes it as "the most widely used standard to describe photos, because of its universal acceptance among news agencies, photographers, photo agencies, libraries, museums, and other related industries" [4].

```
File Name                 : 2016-08-12_13-42-03.jpg
File Size                 : 3.5 MB
File Type                 : JPEG
Make                      : Apple
Camera Model Name         : iPhone
Orientation               : Horizontal (normal)
Software                  : Aperture 3.6
Exif Image Width          : 4032
Exif Image Height         : 3024
Date/Time Original        : 2016:08:12 13:42:03
Create Date               : 2016:08:12 13:42:03
Subject Area              : 2015 1511 2217 1330
Color Space               : sRGB
GPS Latitude              : 35 deg 9' 16.10" N
GPS Latitude Ref          : North
GPS Longitude             : 119 deg 51' 56.35" W
GPS Longitude Ref         : West
GPS Altitude              : 649.4 m Above Sea Level
GPS Altitude Ref          : Above Sea Level
GPS Img Direction         : 346.2011834
GPS Img Direction Ref     : True North
Hyperfocal Distance       : 1.82 m
Field Of View             : 63.7 deg
Exposure Time             : 1/2564
F Number                  : 2.2
Flash                     : Off, Did not fire
ISO                       : 25
Shutter Speed Value       : 1/2564
Description               : vert_angle_deg=-4.1 / horiz_angle_deg=-2.3
Aperture Value            : 2.2
White Balance             : Auto
Exposure Mode             : Auto
Exposure Compensation     : 0
Metering Mode             : Multi-segment
Brightness Value          : 11.49591002
Light Value               : 15.6
Lens Make                 : Apple
Lens Model                : iPhone 6s Plus back camera 4.15mm f/2.2
Focal Length              : 4.2 mm (35 mm equivalent: 29.0 mm)
Digital Zoom Ratio        : 1
GPS Map Datum             : WGS-84
Exif Version              : 0221
```

Figure 3.1: Selected ExifTool image metadata output

In general, while EXIF describes technical details, IPTC conveys semantic information about the content and purpose of an image. Image Description and

Keywords are two examples of commonly used IPTC metadata fields. To illustrate the full range of metadata fields for a JPEG image, the list in Fig. 3.1 was generated using Phil Harvey's ExifTool software. As the program used within *WorldMediaViewer* to extract metadata from images, ExifTool was indispensable to this project.

Most smartphones already record GPS location, compass orientation and altitude. Elevation and rotation angles, not currently a part of standard EXIF metadata, yet also important for recreation of a scene, can be imported into *WorldMediaViewer* from photos taken using the popular iPhone app Theodolite.

3.1.3 Panorama Metadata

Not only traditional, rectangular images, but 360-degree panoramic images can be imported by *WorldMediaViewer* as well. While dependent on the device used, virtual reality cameras and smartphone apps often record much of the same metadata as for rectangular images. Currently, the program can import panoramas taken with the Ricoh Theta and Theta S cameras or with the Occipital 360 iPhone app. These systems store EXIF fields for GPS Latitude and Longitude, Altitude, and Image Direction.

Since panoramic stitching software automatically adjusts for elevation and rotation angles, no Theodolite image placeholder is needed, as is for videos. However, some additional considerations regarding elevation and rotation are necessary. In Occipital 360, one must begin a panorama with the camera oriented horizontally; if one starts by holding the camera vertically, as is common, the Image Direction field will be incorrect by about 120 degrees. Similarly, the rotation angle when taking panoramas with the Ricoh Theta cameras must be zero; if the camera is held upside down or sideways, the Image Direction field is not recorded, likely to avoid similar errors to the one previously mentioned with the Occipital 360 app.

3.1.4 Video Metadata

Video files, like images, typically use the EXIF metadata standard. These include GPS coordinates, aperture, shutter speed and other camera settings, although the relevant metadata for reconstructing a scene is relatively more limited. To view common metadata fields for video, see Fig. 3.2 for a list again compiled from ExifTool output.

Since video metadata taken with a smartphone may contain GPS information, but lack compass direction, elevation angle, rotation angle, and other fields relevant to the spatial layout of a scene, the preferred method for capturing video to use in *WorldMediaViewer* is to take a Theodolite image immediately before recording a video at the same location and orientation.

```
File Name                       : IMG_5894.MOV
File Size                       : 159 MB
File Type                       : MOV
Make                            : Apple
Model                           : iPhone 6s Plus
Duration                        : 26.35 s
Image Width                     : 3840
Image Height                    : 2160
Media Create Date               : 2016:05:10 02:44:59
Media Modify Date               : 2016:05:10 02:45:25
GPS Latitude                    : 34 deg 25' 14.88" N
GPS Longitude                   : 119 deg 53' 44.52" W
GPS Altitude                    : 25.855 m
GPS Altitude Ref                : Above Sea Level
Compressor Name                 : H.264
Bit Depth                       : 24
Video Frame Rate                : 29.977
Audio Format                    : mp4a
Audio Channels                  : 1
Audio Bits Per Sample           : 16
Audio Sample Rate               : 44100
Media Time Scale                : 600
Media Duration                  : 26.35 s
Media Language Code             : und
```

Figure 3.2: Selected ExifTool video metadata output

After loading the available video metadata, the program compares each video's capture dates and capture times to those of images taken at nearby GPS locations. If an image was captured before the video within a specified amount of time, by default a few seconds, this image becomes identified as a "placeholder" from which the missing metadata fields are obtained.

3.1.5 Sound Metadata

Compared to images and videos, audio files generally contain much more limited metadata. Since the original MP3 standard did not include a method for storing file metadata, a de facto standard called ID3 was developed in 1996 [15]. ID3 tags include descriptive categories often used by digital media players such as iTunes, including: Title, Artist, Album, Year, Track Number and Genre. While designed with MP3 in mind, ID3 tags have also been applied to other file types, including AIFF and WAV files.

Digital audio recorders do not typically record ID3 tags, and the author is aware of no readily available method for adding GPS attachments to digital audio recorders. Due to the lack of a geotagging method for audio files using the ID3 standard, *WorldMediaViewer* determines the location information for sound files from GPS track files via a two step method of (1) loading the Creation Date / Time from the file itself, rather than its metadata, and (2) refer to the GPS track for that Date / Time to retrieve the GPS location.

3.2 Software

WorldMediaViewer is written in Java using the Eclipse IDE. This section describes external libraries and software used in the project, while the next section explains the basic class structure.

3.2.1 Metadata Extraction

Metadata extraction in *WorldMediaViewer* is performed using Phil Harvey's excellent ExifTool software. ExifTool offers not only metadata reading, but also writing and editing capabilities, for both images and video. Without it, this project would have been impossible or taken much longer to complete.

3.2.2 Graphics

For 2D and 3D graphics display, *WorldMediaViewer* uses the Processing and Processing Video libraries. Built using the Open Graphics Library (OpenGL), Processing was developed by Ben Fry and Casey Reas at MIT Media Lab, starting in 2001, as a visual arts programming tool for educational contexts. Recently, it has also come into widespread use as a tool for rapid prototyping and data visualization.

3.2.3 Other

Hierarchical clustering in *WorldMediaViewer* is achieved using the Hierarchical Clustering Java implementation by Lars Behnke. Panoramic stitching is done using the Stitcher class from the JavaCV implementation of OpenCV, an open source library aimed at real-time computer vision functions. Several interpolation methods are used from the open source ToxicLibs library, containing many functions useful for computational design tasks.

3.3 Class Structure

3.3.1 WorldMediaViewer

The WorldMediaViewer class is the main application class, which contains fields for the highest-level status modes, and methods for handling the initial selection of media library folder and program setup, including instantiation of the WMV_World, WMV_Stitcher and WMV_Utilities classes. Since WMV_World contains the primary classes and methods for creating a virtual scene from media, the WorldMediaViewer class itself contains only the setup methods, restart(), and methods required by the Processing library to be top-level, such as onKeyPressed() and onMousePressed() for handling input, and movieEvent() for retrieving video frames from a buffer.

3.3.2 WMV_World

The WMV_World class represents the virtual world, consisting of one or more media environments contained in WMV_Field objects, the WMV_Viewer class containing the virtual camera, and a WMV_Display class for 2D Map View, on-screen messages and other Heads-Up Display functions.

Only the main application class and two helper classes, WMV_Stitcher and WMV_Utilities are not contained in the WMV_World object. The fields of WMV_World are limited to the system modes, graphics modes, time cycle variables, and other global settings.

3.3.3 WMV_Field

The WMV_Field class represents a single virtual environment, consisting of at least one media file. The most significant components of WMV_Field are the WMV_Model class, representing the spatio-temporal model of the field, and arrays of media objects.

For a detailed description of WMV_Model, see **Section 3.4: Model**. Media objects in the virtual scene belong to either WMV_Image, WMV_Panorama, WMV_Sound or WMV_Video classes. Each of these specific media type objects inherits from WMV_Viewable, an abstract class representing all media types viewable in the 3D scene.

3.3.4 WMV_Viewable

The WMV_Viewable object represents a single media file for display in virtual space over time. Vertex calculation, object drawing, graphics settings and other media type-specific methods are left to the inheriting WMV_Image, WMV_Panorama, WMV_Sound, and WMV_Video classes. As the abstract class for all viewable objects, WMV_Viewable contains only methods and fields related to capture location, spatial clustering, and fading over time.

The WMV_Image class and WMV_Video class both have rectangular geometries, while WMV_Panorama has a spherical geometry and WMV_Sound consists of only a single point in space: the capture location. Each class has methods for retrieving that particular media object's individual metadata field values, distance from the camera, and angle facing the camera, and media data, among other values, many of which depend on media type.

While the visibility of panoramas and the audibility of sounds depend solely on the viewer's distance from the media capture location, the visibility of images and videos depends on the media *viewing location*, determined from capture location by adding a displacement vector calculated from a combination of sensor size, focus distance, estimated subject size, and orientation.

To prevent media from suddenly appearing in a single frame, objects fade out as the viewer approaches either the far or near clipping planes. To avoid letting them appear as a thin rectangle from the side, images and videos also fade as the camera moves around them, so that once they no longer face the camera, they have disappeared. In high photo density areas where overlapping images would become a distraction, these fading methods, along with fading over time, allow the user to view the scene as clearly as possible.

3.3.5 WMV_Viewer

As the program's namesake class, WMV_Viewer is one of the most significant and extensive classes in *WorldMediaViewer*, representing the virtual viewer, with the ability to navigate and interact with the current field and its media objects. The most important of the many navigation and interaction functions in the WMV_Viewer class are discussed in **Section 3.6: Interaction** below.

3.3.6 WMV_Display

WMV_Display handles the Heads-Up Display (HUD) Modes, including displaying 2D Map View, Field View, Cluster View and Control View. This subsection describes each of these HUD Modes and how they contribute to the software's functionality.

The most important of these Modes, 2D Map View, shows all media and clusters in the current field as colored points on an aerial map (Fig. 3.3). The

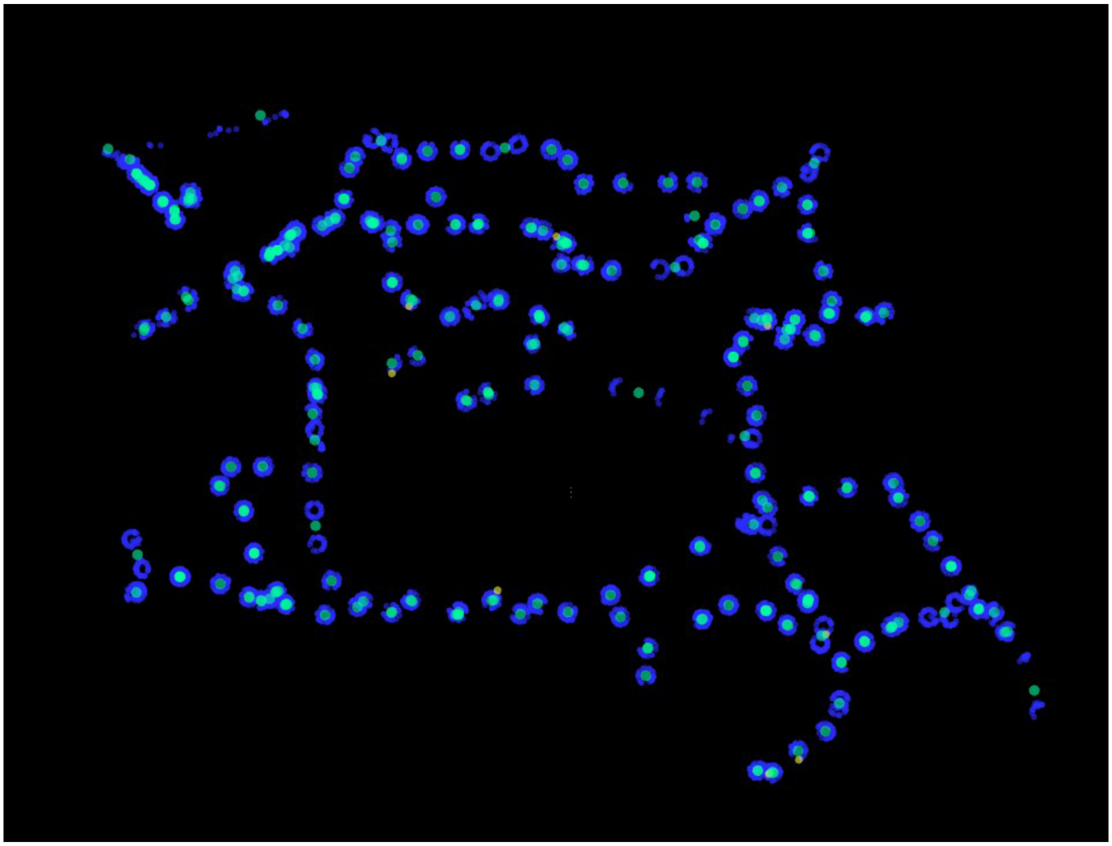

Figure 3.3: *Lake Los Carneros Park*, 2D Map View, Media Locations and Clusters

2D Map View has several Display Modes, which include visualizations of different combinations of three elements: media capture locations, media viewing locations, and clusters. Media can be filtered by type, either image, panorama or video. The size of clusters viewed on the 2D map are scaled based on the number of media points they contain. As the density of media measures the amount of time the artist spent in a particular place, this visualization might be called a "map of attention."

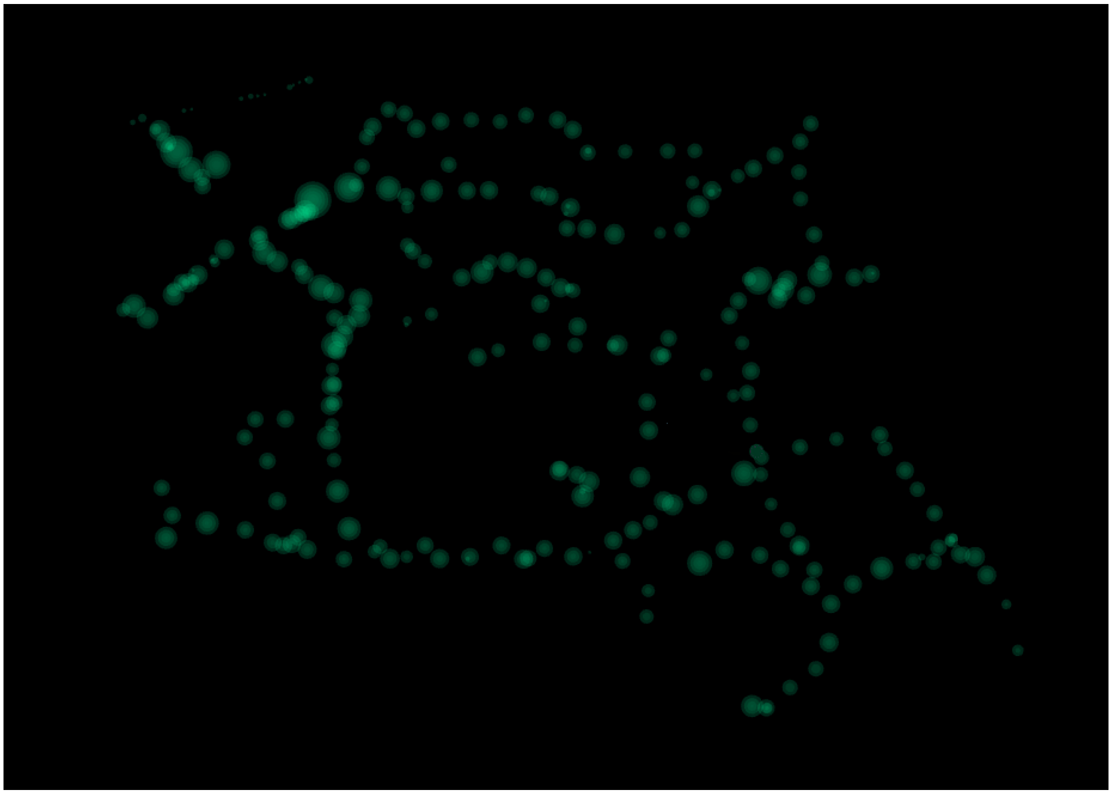

Figure 3.4: *Lake Los Carneros Park*, 2D Map View, Clusters Only

Field View displays statistics about the field and the current program settings. Field statistics include the number of media by type (image, panorama, sound, or video), number of clusters, number of currently visible media by type, and the viewer's geographic location and compass direction. Cluster View allows the user to move through each cluster as a list, showing data about the number and type of media each contains. It also contains a traditional grid view of all the photos in the cluster, similar to common photo library management programs.

Control View shows the keyboard commands for operating the software during World View.

3.3.7 WMV_Utilities

WMV_Utilities is a class containing utility methods for functions such as conversion from GPS distance to meters, converting time formats, image resizing and exporting images of the current scene. It also includes the WMV_Stitcher object, essentially a wrapper for the Stitcher OpenCV object, that implements panoramic stitching functions using the *WorldMediaViewer* interface.

3.4 Model

Every framework for turning images, sounds or videos into 3D environments has at its core a particular spatial model for specifying how the media are arranged in virtual space. For the images or media to change over time, they must also follow a temporal model, determining their order in time. This chapter gives a detailed explanation of the metadata-based spatial and temporal model that distinguishes *WorldMediaViewer* from the image-based modeling approaches used many current systems.

3.4.1 Location

The location of images, sounds and videos in the virtual space is determined by GPS latitude, longitude and altitude, with additional displacement in the case of non-panoramic images and videos determined by a combination of focal length, sensor size and focus distance, if available.

While the Pythagorean theorem is sufficient to convert between GPS distances and meters for short distances, for long distances, a better approximation can be calculated using the Haversine Formula for finding the great-circle distance between two points. *WorldMediaViewer* uses this latter method to ensure the greatest accuracy.

3.4.2 Spatial Clustering

Two clustering algorithms are used in determining the spatial clusters of a field: k-means and hierarchical clustering. K-means clustering is an iterative refinement technique to partition n observations into k clusters, where each observation belongs to the cluster with the nearest mean. This strategy begins by assigning random data points to the clusters. From this point, "For step 1, the centroid is computed for each set. In step 2, every point is assigned to the cluster whose centroid is closest to that point" [18].

These two steps are alternated until the termination criterion is met, often a threshold value, epsilon, representing a minimum amount by which clusters must move between iterations in order to continue refinement. Since k-means clustering requires a number of clusters to be set, many data mining applications that use k-means incorporate other algorithms to first estimate the number of clusters.

Unlike the iterative method of k-means, hierarchical clustering analyzes a data set through building a hierarchy of clusters. The hierarchical partition is presented using the dendrogram, a tree diagram, often used for tasks in biology such as illustrating the clustering of genes or the phylogeny of organisms.

This method offers several advantages, namely its versatility and its production of multiple partitions, "which allow different users to choose different partitions, according to the desired similarity level"[7]. The main disadvantage of hierarchical methods is the inability to scale well, due to the time complexity, which increases exponentially with the total number of instances[7].

Each field in a WMV_World object can be separately initialized using either k-means or hierarchical methods. K-means is recommended for fields with very large numbers of media to achieve satisfactory results in a reasonable time, while the hierarchical method may be used for fields with fewer media to achieve a greater level of control.

Once either the k-means or hierarchical method has determined the spatial clusters, clusters under a minimum distance threshold are merged. Merging clusters prevents two clusters from occuring at nearly identical locations.

3.4.3 Orientation

While, in EXIF metadata, ”orientation” refers to the format of a rectangular image, either ”portrait” or ”landscape,” in this paper, unless otherwise specified, the term refers to how a media object is oriented in virtual space. Specifically, it describes the three Euler angles defining the direction, elevation and rotation of a media object in virtual space.

The meaning of the orientation parameter depends on media type. For rectangular images and videos, orientation determines the direction of displacement of the media viewing location from the GPS capture location. For panoramic images, the image texture is applied to a sphere, the orientation parameter simply determines the orientation of the texture on the sphere. In the current version of *WorldMediaViewer*, orientation applies only to images and videos, not to sounds.

3.4.4 Media Segments

Similar to the spatial clusters described above, media objects in *WorldMediaViewer* are grouped within each cluster by orientation into Media Segments.

Media Segments associate images and videos that overlap with each other by a certain minimum angular threshold. Groups are defined by a left and right bounds of rotation around the Y axis, and high and low bounds of rotation around the X axis.

Besides being used by some Automatic Navigation functions, Media Segments allow exporting images of the same part of a scene, which can be useful for panoramic stitching using external applications. Panoramic stitching is also possible within *WorldMediaViewer* using the JavaCV library's default image stitching algorithm. Stitched panoramas can replace or appear alongside the original images, as well as be exported as .JPG images.

3.4.5 Time

The incorporation of change over time is one of the most significant differences between *WorldMediaViewer* and existing systems such as Photosynth. Time fading for each a media object is based on the EXIF parameters of Capture Date and Capture Time. Media objects reach full brightness at their "center time," which is represented in the software as a normalized value between 0. and 1., where 0. is midnight and 1. is midnight the next day. Date is represented using the same range of values, where 0. means January 1st and 1. is the end December 31st.

Figure 3.5: *Venice Canal Historic District*, Morning (top) and Evening (bottom)

When Time Fading Mode is turned on, media fade in and out over the program's time cycle. Fig. 3.5 shows one virtual location in the *Venice Canal Historic*

District virtual environment at two different times in the time cycle: Morning (top) and Evening (bottom). The length of this time cycle and the media display length may be set by the user, providing the flexibility to display media over different time scales. With a short time cycle and long media length, the day passes quickly, but media are displayed for a relatively long time, so that many media are visible at once. If the time cycle is long, while media length is short, the day passes slowly, while media are visible for a relatively short time. If media length is short enough, only one media object is visible at a time, illustrating the precise timing and order in which the images were taken.

3.4.6 Time Segmentation

Time segments in *WorldMediaViewer* are groups of media within a spatial cluster that have temporal proximity. Segment centers are calculated through k-means clustering, similar to the spatial clustering method described above, but in the time domain. After segment centers are found, upper and lower bounds are calculated by finding the highest and lowest media times in the spatial cluster. Time segments are used in fading media over time, as well as Automatic Navigation Mode, to allow navigation to clusters according to their positions on the timeline or dateline.

3.5 Display Views

3.5.1 World View

The World View is the main display view in *WorldMediaViewer*, showing the virtual environment according to the current viewer position and orientation (Fig. 3.6). While most commands in World View are also available in 2D Map View, the software is designed for users to spend most of their time in World View.

While several Program Modes allow the user to customize various program functions, one of the permanent features of World View is that media opacity or brightness is a factor of distance: when the viewer moves too close or too far from a media object, it fades to black. As the viewer moves back into visible range of an image, it fades in. Videos are displayed in the same way as images, except that sound also fades in and out with the video opacity or brightness. This method of fading media with distance not only reduces visual clutter, but improves performance, since media out of visible range can be discarded from memory.

Figure 3.6: *Painted Rock*, World View

3.5.2 Model View

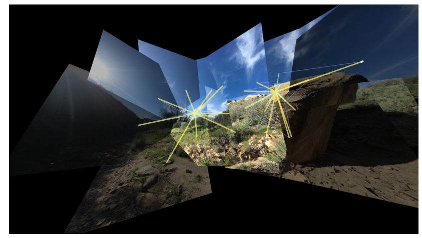

Figure 3.7: *Lost Valley*, Model View

Model View is identical to World View, except that a yellow line extends from to the cluster center to each associated media object (Fig.3.7). This view is useful for visualizing the current cluster's spatial model, either for aesthetic purposes, such as giving a sense of the geometry, or helping with practical tasks such as distinguishing two nearby clusters of images, or finding metadata errors.

3.5.3 2D Map View

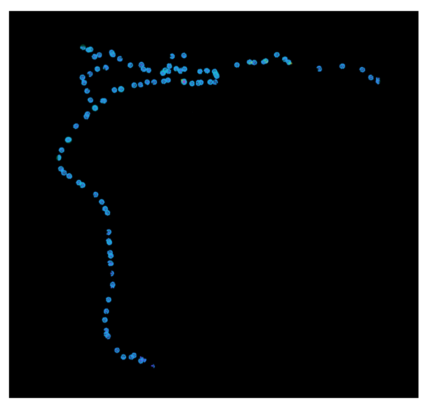

Figure 3.8: *El Escorpion Park*, 2D Map View

The 2D Map View presents an "aerial" view of the current field. Unlike map image overlays in most existing photo browsing software, the 2D Map View in *WorldMediaViewer* does not include geographical data except for the media themselves (Fig.3.8). The decision not to include a more traditional map with street or location names reflects a different aesthetic from existing programs, in which media locations themselves make the spatial layout of an environment visible. Media objects are color coded by media type, while several different display settings enable the user to better understand each environment's spatial layout. The 2D Map View is used in Interactive Clustering and as an alternate view, helpful for navigation tasks, such as moving across large gaps in which no media are visible.

3.5.4 Program View

Program View displays global statistics about the current field (Fig. 3.9). The screen divides variables into six sections: Program Modes, Graphics, Field, Model, Viewer, and Time. The Program Modes section indicates which Program Modes are currently active, including Orientation Mode, Alpha Mode, Time Fading, and Date Fading. See **Subsection 3.6.3: Program Modes** for more information on the different Program Modes.

The Graphics section of Program View displays the current media length, image size factor, default focus distance and current transparency level, or alpha, used for displaying media. The Field section contains statistics about the current field, including name, visible objects by media type, total objects by media type, number of videos playing, and many other parameters. The Model section gives the number of spatial clusters, number of merged clusters, clustering method used, and minimum and maximum distances between clusters.

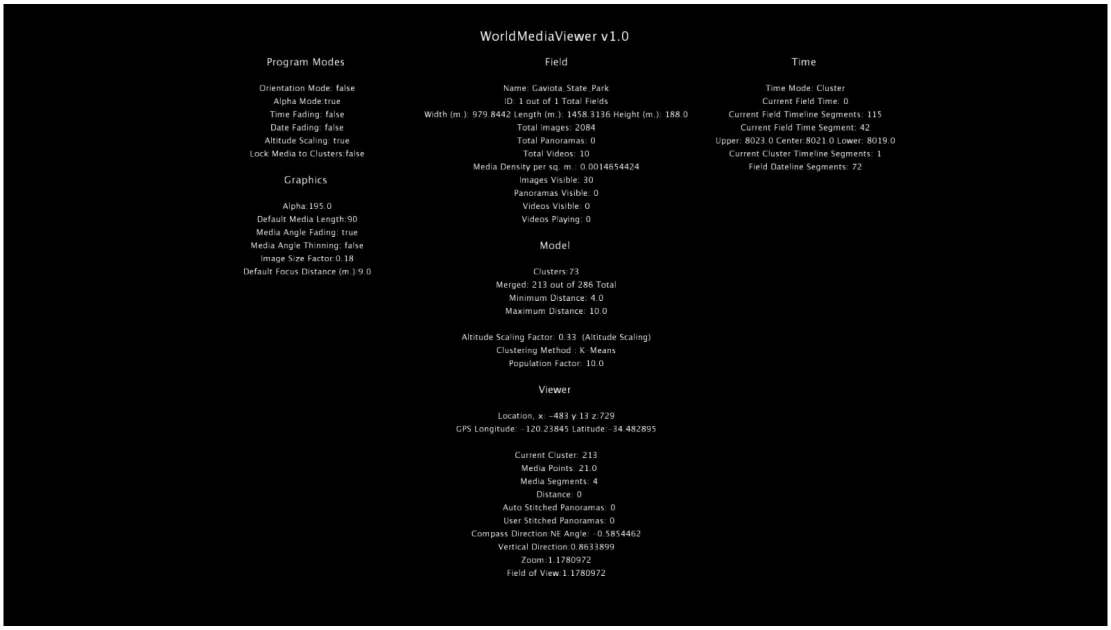

Figure 3.9: *Gaviota State Park*, Program View

Program View's Viewer section gives the current virtual viewer position in two formats: GPS coordinates and world coordinates. World coordinates consist of X, Y and Z values representing meters from the origin. Finally, the Time section displays information about the time cycle, including the current time, the number of time and date segments in the field, and number of time and date segments in the current cluster.

3.5.5 Cluster View

Cluster View displays statistics and thumbnail images for spatial clusters in the current field (Fig. 3.10). By default, Cluster View shows statistics for the

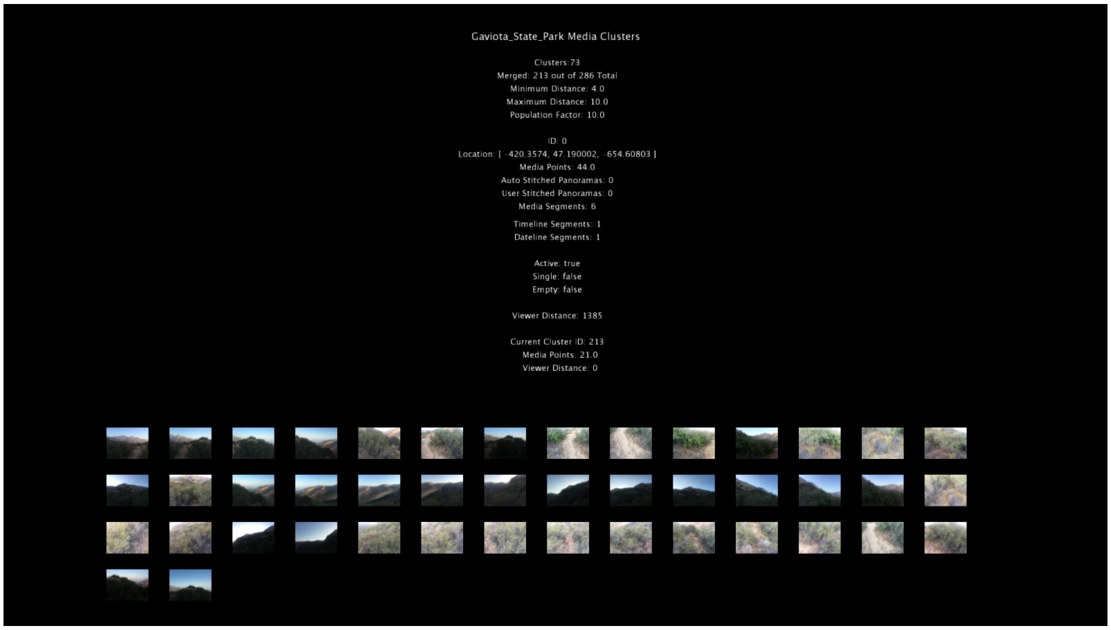

Figure 3.10: *Culver City*, Cluster View

closest cluster to the viewer's virtual position. The user can also browse through the complete list of clusters using the left and right arrow keys. Statistics about the cluster are displayed at the top of the screen. These values include number of associated media items, number of orientation segments, and number of timeline and dateline segments. At the bottom of the screen, an ordered grid lets the user preview the cluster's media in a format familiar to users of existing photo browsing systems.

3.5.6 Control View

The last user Display View is Control View, which shows the keyboard and mouse controls for viewing, navigating and interacting with the virtual environment. The many commands are divided into eleven sections: Display, Time, Time Navigation, Model, Graphics, Navigation, Auto Navigation, Interaction, GPS Tracks, Memory and Output.

3.6 Interaction

3.6.1 Manual Navigation

Navigation in *WorldMediaViewer* has two main types: Manual and Automatic. Both methods are implemented using a force-based model to create smooth transitions between moving and standing still. Manual navigation is based on a walking metaphor, in which the user can move forward, backward, left or right using the keyboard or trackpad: similar to the original controls of the Aspen Moviemap project.

In addition to walking, the viewer can turn to look up or down, as well as zoom in, which changes the virtual camera's field of view. To prevent the user from moving into a void without any media, once the viewer stops walking, if no media are within visible range, the virtual camera is gradually pulled toward the nearest spatial cluster until reaching its center.

3.6.2 Automatic Navigation

Both spatial clusters and time segments play a fundamental role in the *World-MediaViewer* Automatic Navigation system. In Automatic Navigation Mode, many different commands are available for the viewer to explore a field via both metadata and data parameters. Users can move to the nearest spatial cluster, to the nearest cluster with a specified media type, or to the nearest time segment, either in the same spatial cluster or in the entire field.

Automatic Navigation Mode also features three advanced functions: Timeline Navigation, Memory Navigation and GPS Track Navigation. Timeline Navigation lets the user easily move between clusters in chronological order, based on the order of time segments in each cluster. The user can move one time segment at a time, or let the program automatically follow the complete timeline of the field.

Memory Navigation allows the user to specify locations of interest, including the direction the viewer is facing, and follow these locations in order as a virtual path. GPS Track Navigation lets the user import and follow a GPS track file as a virtual path. GPS tracks record a person's movement through the real environment, and can be created with a number of freely available smartphone apps.

The only currently supported format is .GPX, however many software tools exist for converting other GPS track formats, such as .KML, to .GPX.

3.6.3 Program Modes

This subsection describes the various Program Modes available to the user for modifying how the program displays media in World View. Alpha Mode specifies whether image fade in and fade out transitions should determine media opacity or brightness.

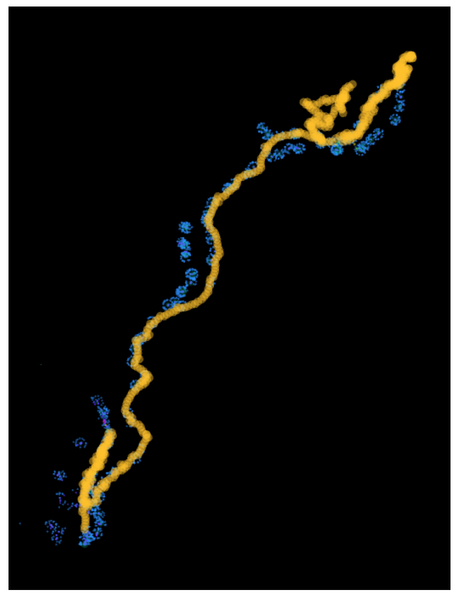

Figure 3.11: *Rattlesnake Canyon Park*, GPS Track

In Angle Fading Mode, media fade in as the viewer turns towards them, reaching full opacity or brightness in front of the viewer. Media Thinning Mode sets a minimum angular threshold between images or videos, where if the angle between two media is below the threshold, one of them is hidden. Angle Fading Mode and Media Thinning Mode serve to improve both aesthetics and visibility, especially when many overlapping media items are visible at once.

Time Fading Mode determines whether media fade in and out based on capture time. When it is turned off, all media within visible range are displayed. Date Fading Mode indicates whether media fade in and out based on capture date. Date Fading Mode and Time Fading Mode can operate independently of each other, or at the same time. When both are turned on, media fade through the time cycle for each date on the dateline of the current cluster.

Orientation Mode is an alternate display mode from the Standard Mode, where the virtual camera "moves by standing still." Rather than moving about a virtual scene, the virtual viewer instead sees a single panorama at all times, where the closest spatial clusters within a distance threshold are all visible *simultaneously.* Movement causes the viewer simply to see different clusters fade in and out, rather than being able to move around the media. Orientation Mode was inspired by the author's panoramic video installation, *Transitions* (See **Subsection 2.0.10: My Own Work**).

3.6.4 Interactive Clustering Mode

Interactive Clustering Mode allows the user to choose the clustering method and determine clustering settings, including minimum cluster difference used in merging, as well as and settings specific to the current clustering algorithm. For k-means clustering, these settings include number of initial clusters, population factor and termination threshold distance. For hierarchical clustering, since the cluster hierarchy is only calculated once, the primary parameter available to the user besides minimum cluster distance is dendrogram depth.

In Interactive Clustering Mode, the user sees the 2D Map View and the values of the current clustering settings. Each time a parameter is changed, such as minimum cluster distance, dendrogram depth or k-means population factor, the selected clustering algorithm runs, letting the user observe any cluster movement or resizing that occurs. At this point, the user may choose to revert to the previous setting, make additional modifications, or exit Interactive Clustering Mode and return to the World View.

3.6.5 Selection Modes

The three Selection Modes allow users to view metadata for a file or to choose images or groups of images for panoramic stitching. In Single Selection Mode, individual media are selected by the user looking at an object and pressing the 'x' key or double clicking with the trackpad (Fig. 3.12). When a new media item is selected, the last selected item is automatically deselected.

In Multiple Selection Mode, more than one media item may be selected. The last Selection Mode, Segment Selection Mode, selects whole Media Segments at a time (Fig. 3.13). Both these Selection Modes allow panoramic stitching of

Figure 3.12: *Fryman Canyon*, Single Selection Mode with Metadata Display

rectangular images using the JavaCV library. Once a multiple image selection has been made, the user presses the "—" key to attempt to stitch the images into a panorama.

If the panoramic stitching is successful, by default, the new panorama appears alongside the original images. Additional commands allow the user to control whether the panorama, original images, or both, are visible.

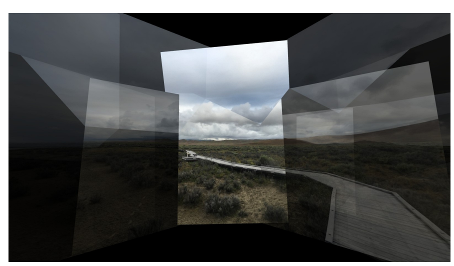

Figure 3.13: *Soda Lake*, Segment Selection Mode

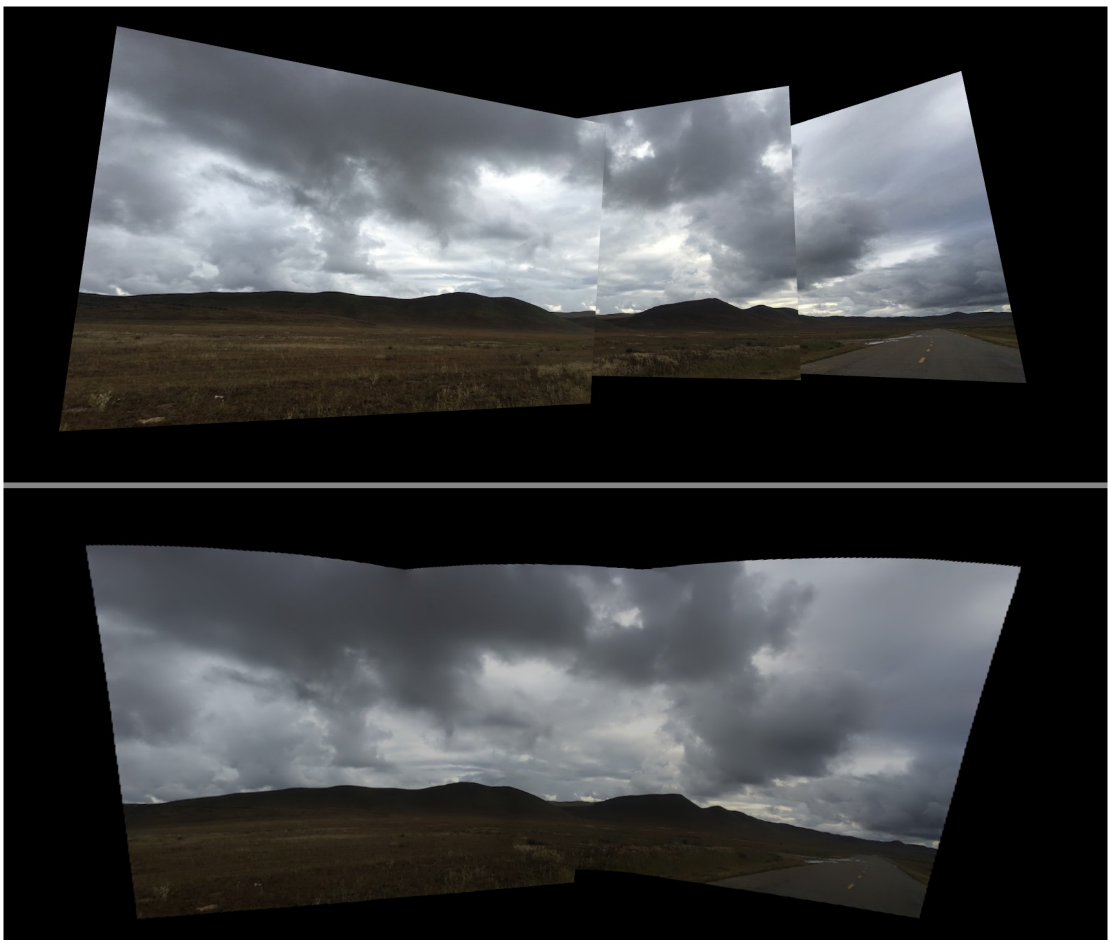

Figure 3.14: *Soda Lake*, Panoramic Stitching, Before (top) and After (bottom)

Chapter 4

Results

WorldMediaViewer was conceived and developed as a result of the author's own work as a composer, artist and photographer, in order to contribute both aesthetically and practically to the activities of capturing, selecting, and ordering media files. During its development, the software has been used to create a series of 3D virtual environments based on the author's own images, panoramas, sounds and videos.

These environments cover dozens of geographical areas of varying sizes and dimensions. More than a testing ground for solving technical problems, these environments constitute an ongoing artwork called the *Perceptual Geography* project, which aims to address a number of significant questions involving the role of multimedia in art and society:

- How do media reflect the places where they were captured?
- What does metadata reveal about the creative process?

- Are personal media libraries artworks?
- How can the proliferation of digital media lead to reinvestment in the local, rather than a growing sense of dislocation or distraction?

This chapter describes these various environments and their presentation in three exhibitions, including both interactive and non-interactive digital installations.

4.1 Environments

One benefit of the flexibility of a metadata-based system is the ability to progressively add media to a scene over time without worrying about minimum or maximum size or media density. All of the environments created for the *Perceptual Geography* project are open-ended artworks. Over time, as media are added, each will continue to grow in size and complexity. Some may eventually overlap. These changes record both the places themselves and the artist's attention progress over time.

The environments are classified in two main categories: natural areas and urban areas. Initially conceived with only natural areas in mind, *Perceptual Geography* was originally titled "Open Spaces." This preference for natural areas partly reflected the fact that GPS data is best captured outdoors, without large buildings or other structures that can interfere with a satellite signal. The project was renamed *Perceptual Geography* after several urban environments were included. In spite of the risk of greater GPS inaccuracies, the different shapes, light patterns, types of motion, and pace of life in cities creates an interesting aesthetic contrast between the urban environments and the natural areas.

4.1.1 Santa Barbara

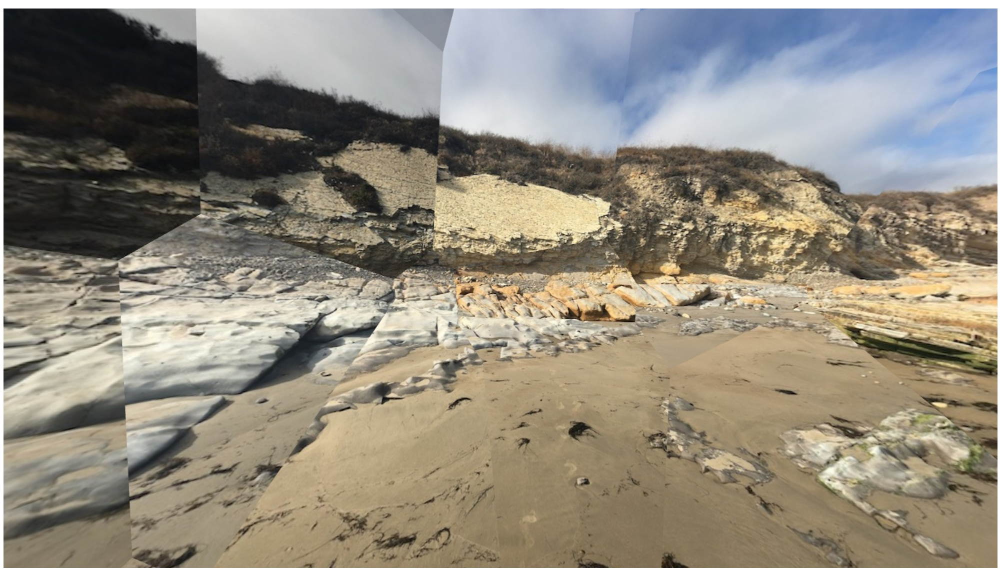

Figure 4.1: *Refugio Beach* David Gordon (2016)

WorldMediaViewer and the *Perceptual Geography* project were conceived in Santa Barbara, where the author lived for six years until June 2016. Situated along California's only stretch of south-facing coast, Santa Barbara is bordered to the north by the vast wilderness of the Los Padres National Forest. Since 2015, the author created over 20 virtual environments of varying sizes and densities in natural areas in and near Santa Barbara. The most significant, based on size and density, are: Ellwood Bluffs, Lake Los Carneros Park, More Mesa Bluffs, Cold Spring Trail, Rattlesnake Canyon Trail, Refugio Beach (Fig 4.1) and the Santa Barbara Botanic Garden. The urban environments in Santa Barbara are the extensive UC Santa Barbara campus, including the adjacent beaches and lagoon, and the Downtown area.

4.1.2 Los Angeles

Figure 4.2: *Venice Canals Historic District*, David Gordon (2016)

The author's hometown and current residence, Los Angeles, holds a wide array of natural areas, including the expansive Santa Monica Mountains National

Recreation Area, one of the world's best examples of a Mediterranean climate ecosystem. Griffith Park, another expansive park in the center of Los Angeles, contains both open spaces and public attractions such as Griffith Observatory, the L.A. Zoo and the Autry Museum. Distinct urban areas in Los Angeles are too numerous to name. They cover many different geographical areas, ranging from communities on the edge of the Santa Monica Mountains, to areas such as Venice and Marina del Ray along the Pacific Ocean, to inland areas as diverse as Downtown L.A., Highland Park, and Pasadena.

Perceptual Geography currently includes 10 environments in the Los Angeles area. The nature areas are Woodley Avenue Park, Upper Las Virgenes Canyon, Fryman Canyon, Dixie Canyon and Oak Forest Canyon, while the urban areas consist of Venice Canals Historic District (Fig. 4.2), Bergamot Station (Fig. 4.3), Downtown Los Angeles, and Downtown Culver City.

4.1.3 Other

Besides Santa Barbara and Los Angeles, *Perceptual Geography* includes environments from several wilderness areas in Santa Ynez, the San Luis Obispo area, and remote places such as Carrizo Plain and the Anza-Borrego Desert. Urban environments outside Southern California include Detroit, London and a number

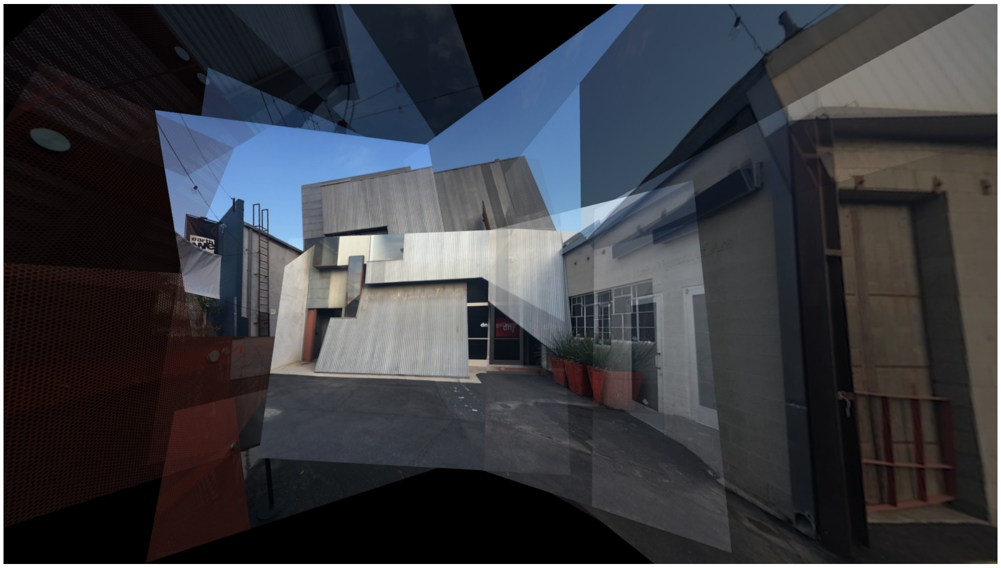

Figure 4.3: *Bergamot Station*, David Gordon (2016)

of cities in Germany visited in the summer of 2016, most notably Cologne and Berlin.

4.2 Exhibitions

The author has exhibited three digital installations so far, all in the Santa Barbara area. In the *Pathways* (2015) and *Mosaic* (2016) installations, the navigation system was interactive, while *Geospatial Collage* (2015) used automatic navigation.

4.2.1 Pathways (2015)

Figure 4.4: *Rattlesnake Canyon*, David Gordon (2015)

Figure 4.5: *Rattlesnake Canyon*, David Gordon (2015)

Pathways is the first digital installation to use *WorldMediaViewer* software, initially named "PhotoScene," which was presented at the Media Arts and Technology Department End of Year Show in May 2015 in UC Santa Barbara Elings Hall. This interactive installation took place in the Translab, directed by Marcos Novak. Two side-by-side projectors created an immersive setting for viewers to navigate virtual scenes, consisting of several thousand layered images and dozens of field recordings recorded along Rattlesnake Canyon Trail near downtown Santa Barbara (Figs. 4.4 and 4.5).

The exhibit was a successful proof of concept, though improvements were needed to bring the idea to full fruition. Several viewers expressed an interest in the process of linking images and sounds by place and time, while others found elements distracting, such as the lack of motion when the viewer is still and the presence of gaps in the media where nothing is displayed.

In updating the code for the current *WorldMediaViewer* version, new classes added and others were rewritten largely from scratch, partly in response to this viewer feedback. For instance, the addition of Automatic Navigation Mode, the incorporation of video and the introduction of time fading respond in part to the issue of a lack of motion and gaps in the media.

4.2.2 Geospatial Collage (2015)

The second installation of work built with the software, *Geospatial Collage*, was exhibited at the UCSB College of Creative Studies Gallery over two weeks in July 2015. The project consisted of a single environment built from thousands of images of the UCSB campus and adjacent lagoon, dozens of field recordings, as well as music.

As the first non-interactive *WorldMediaViewer* installation, the exhibit was also the first successful use of the software in building a collaborative project. The author recorded about two dozen short field recordings across campus. Afterwards, Ori Barel, a composer and PhD student in the UCSB Music Department, composed over 20 short electronic pieces using these recordings. Barel's compositions were assigned to spatial locations within the field, where they could be heard along with the original recorded sounds captured in the same spatial locations.

In addition to introducing the collaborative aspect to the project, design improvements to *WorldMediaViewer* included adding image transparency, changing the background to black, and an early version of Automatic Navigation Mode, which enabled a continuous presentation over 8 hours each weekday, following the format of an ambient installation, where viewers could come and go as they wanted.

In both this and the earlier *Pathways* installation, photos were taken with a Nikon D3200 DSLR camera with GPS attachment. This attachment records GPS location, orientation and altitude, but not elevation and rotation angle. The lack of elevation and rotation metadata required the manual input of these values via the Image Description metadata field: a method of manual approximation that worked some of the time, but often led to distracting inaccuracies and distortions.

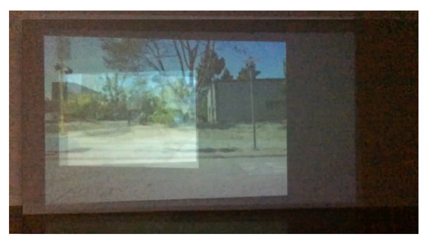

Figure 4.6: *Geospatial Collage* (2015), Installation at College of Creative Studies Gallery

Not only the images' elevation and rotation, but also the GPS locations for sounds needed to be input by hand, a quite tedious process. Since this need for manual input would have inhibited making *WorldMediaViewer* widely accessible to artists, it motivated the incorporation of the Theodolite image importing feature and addition of the GPS track method for extracting the geospatial position of sounds to the software in the current software version.

4.2.3 Mosaic (2016)

Mosaic, the most recent interactive installation, was shown at the 2016 MAT End of Year Show, *White Noise*, in May 2016. The installation was also presented in August 2016 at the Santa Barbara Center for Art, Science and Technology as part of MAT/SBCAST First Thursdays (Fig 4.7). These exhibitions featured images and videos from six environments, including both urban and natural areas. They showcased many of the features of the current *WorldMediaViewer* version, including trackpad navigation, rotation and elevation angle input through the Theodolite app.

Figure 4.7: *Mosaic* (2016), Installation at Santa Barbara Center for Art, Science and Technology

Another new feature in the SBCAST show was the inclusion of five 6 in. x 9 in. and five 12 in. by 18 in. digital prints of images created by the author while exploring the same environments on display in the show. The exhibit was an overall success, with the trackpad navigation allowing an increasing number of viewers to navigate the environments. Major improvements in visual accuracy achieved through the inclusion of rotation and elevation angles seemed to retain viewers' interest longer and encourage them to explore all the possible viewpoints with the mouse.

Chapter 5

Conclusions

Most existing software systems aimed at recreating scenes from images are based on creating an image-based geometric model through the processes of feature matching, alignment and stitching algorithms. While these systems are capable of highly realistic, seamless composites, they are designed primarily for viewing a single, relatively compact geographical area at a time. As such, they are highly suited for giving virtual home tours and for showing popular landmarks, monuments, or city centers in aerial views. However, the computational intensiveness of image-based geometric models currently inhibits the creation of systems capable of rapidly adding or removing images, modifying object distances, including sound and video, or building virtual scenes of any size or dimension.

In contrast to these applications, *WorldMediaViewer*'s more flexible metadata-based spatial and temporal model takes advantage of the fact that reading spatial and orientation metadata from a file is much less computationally intensive than the multistage process of constructing an image-based model. *WorldMediaViewer* applies the computing power saved for tasks such as the incorporation of time, sound and video playback, and the ability to navigate large media collections.

Although the spatial accuracy of the metadata-based model is less reliable, the metadata-based model allows *WorldMediaViewer* the flexibility to display multimedia libraries of any size and spatial arrangement. Preserving the original image borders creates an "artificial" effect due to gaps and overlapping images. However, this artificiality, similar to Cubist paintings or Hockney photocollages, can have aesthetic value of its own. Furthermore, it serves as a record of the photographer's process: a document of his or her attention to each particular place at each particular time.

Besides flexibility, wide accessibility is another primary motivation behind *WorldMediaViewer*. Most smartphones already record GPS location, compass orientation and altitude, while the popular iPhone Theodolite app records the elevation angle and rotation angle fields. In using metadata as its foundation, *WorldMediaViewer* allows artists, photographers and composers with little technical background to create interactive installations, fixed audiovisual compositions, or combinations of predetermined and interactive elements.

Chapter 6

Future Work

6.1 Automatic Field Detection

Currently, *WorldMediaViewer* requires users to organize media into folders by large geographical area prior to importing them. While this task is relatively easy using the map views in several popular commercial software packages, including Apple's Photos, using K-means or other targeted clustering methods, future versions of the software could automatically group images into fields from a single folder. This application of clustering, not just to grouping media in close proximity, but also to finding the larger geographical areas, would be an added convenience for users and make further progress toward the goal of creating a self-contained solution to 3D display of multimedia collections.

6.2 Metadata Editing

Another feature planned for future software versions is a Metadata Editing Mode, in which fields such as GPS location, altitude, orientation, time and keywords can be modified in real-time using Scene View. This functionality would have practical applications, allowing photographers, videographers and sound artists not only to intuitively identify and correct errors in capture location or time, but also to conveniently add media descriptions to keywords in a particular geographic area or at a specific time. Using the ExifTool command, not only reading but editing common metadata fields is possible for single images or image lists captured by many devices.

6.3 Feedback System

An intriguing possibility, once metadata editing capabilities have been added to future versions of *WorldMediaViewer*, would be the creation of a feedback system, in which exported images from a virtual scene are given EXIF spatial, time and orientation metadata based on the position and orientation of the virtual viewer. These images can then be fed back into the system, creating a potentially infinite loop.

6.4 Virtual Reality

Navigation using Virtual Reality (VR) headsets such as the Oculus Rift or HTC Vive offers another interesting direction for potential future research. *WorldMediaViewer* demonstrates that a metadata-based spatial and temporal model allows an intuitive interface for viewing panoramic images in combination with sounds, videos and traditional digital images. The added immersive element of a VR headset and ability to track user movements through room sensors, in particular, seems to hold great potential for investigations of 3D display of multimedia libraries.

6.5 3D Spatial Audio

The spatial element for imported sounds in *WorldMediaViewer* is currently limited to triggering and fading audio recordings based on geospatial location and time. Viewer orientation has no effect on the sound playback. For a more immersive, realistic auditory experience, 3D audio techniques such as Vector Base Amplitude Panning (VBAP) or Ambisonics seem a natural extension for future versions of the program. Since the goal of the system is the display and navigation of large media collections, rather than creating the most realistic simulation possible, the computational expensiveness of incorporating 3D audio prevented its exploration in the current version. Another reason for this choice is that, given the lack of spatial orientation information in sound files and GPS tracks, acquiring precise information about sound sources through metadata is currently impossible.

6.6 Sonification

Sonification, or auditory display of data, is still a relatively new field with exciting potential for both artistic and scientific applications. Artists and researchers have explored image and video sonification in recent years, often focusing on pixel data, such as the parameters of brightness, color or saturation. However, the sonification of media spatial, temporal or orientation metadata is yet largely unexplored. Not only the 3D display of images and videos for pixel-based sonification, but numerous avenues for, both aesthetic and practical, are possible for harnessing the advantages of conveying information sonically. For instance, sonification could inform users about media blocked from view by other media, could highlight media with a certain keyword or author, or convey nearly any metadata parameter to the user if custom settings are implemented.

Bibliography

[1] Cardiff, J. [n.d.], 'Introduction to the audio walks'.
URL: *http://cardiffmiller.com/artworks/walks/jena.html*

[2] Green, Christopher and Musgrove, John [2016], 'Cubism', *Grove Art Online* .

[3] Hurlbert, R. [1986], 'A Blend Of Cubism And Photography: Photographs By David Hockney', *Sun Sentinel* .

[4] IPTC [2016], 'Photo metadata'.
URL: *https://iptc.org/standards/photo-metadata/*

[5] JEITA [2016], 'Exchangeable image file format for digital still cameras: Exif version 2.31'.
URL: *http://www.jeita.or.jp*

[6] Keller, I., Stappers, P. J. and Vroegindeweij, S. [2004], Supporting informal collections of digital images, *in* 'the conference', ACM Press, New York, New York, USA, p. 17.

[7] Maimon, O. and Rokach, L., eds [2005], *Data Mining and Knowledge Discovery Handbook*, Springer-Verlag, New York.

[8] Mullen, L. [n.d.], 'Narrative maps'.
URL: *http://lincolnmullen.com/projects/spatial-workshop/narrative-maps.html*

[9] Naimark, M. [2016*a*], 'Aspen moviemap'.
URL: *http://www.naimark.net/projects/aspen.html*

[10] Naimark, M. [2016*b*], 'Be now here'.
URL: *http://www.naimark.net/projects/benowhere.html*

[11] NISO [2004], 'Understanding Metadata'.

[12] SLAB [n.d.*a*], 'Research'.
URL: *http://slab.today/research/*

[13] SLAB [n.d.*b*], 'Sidewalk city'.
URL: *http://slab.today/2014/08/sidewalk-city-ho-chi-minh-city/*

[14] Snavely, N., Seitz, S. M. and Szeliski, R. [2007], 'Modeling the World from Internet Photo Collections', *International Journal of Computer Vision* **80**(2), 189–210.

[15] Sundstrom, M. N. . J. [2016], 'History id3.org'.
URL: *https://web-beta.archive.org/web/20160413030431/http://id3.org/History*

[16] The Architecture Machine, a. t. M. M. S. I. G. [1981], 'The interactive movie map: A surrogate travel system'.
URL: *https://www.youtube.com/watch?v=Hf6LkqgXPMU*

[17] Ulin, David L. [2004], 'City of Ruins: How Norman Klein Has Emerged As L.A.s Most Innovative Social Critic', *Los Angeles Magazine* .

[18] Weisstein, E. W. [n.d.], 'K-means clustering algorithm, from mathworld a wolfram web resource'.
URL: *http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html*

[19] Wray, J. [n.d.].

[20] Zhang, L. and Huang, H. [2013], 'Hierarchical Narrative Collage For Digital Photo Album', *Computer Graphics Forum* **31**(7), 2173–2181.

Appendix A

WorldMediaViewer

A.1 Results

A.1.1 Natural Areas

A.1.2 Urban Areas

Figure A.1: *Carrizo Plain*, David Gordon (2016)

Figure A.2: *Santa Barbara Mountains*, David Gordon (2016)

Figure A.3: *Malibu Creek*, David Gordon (2016)

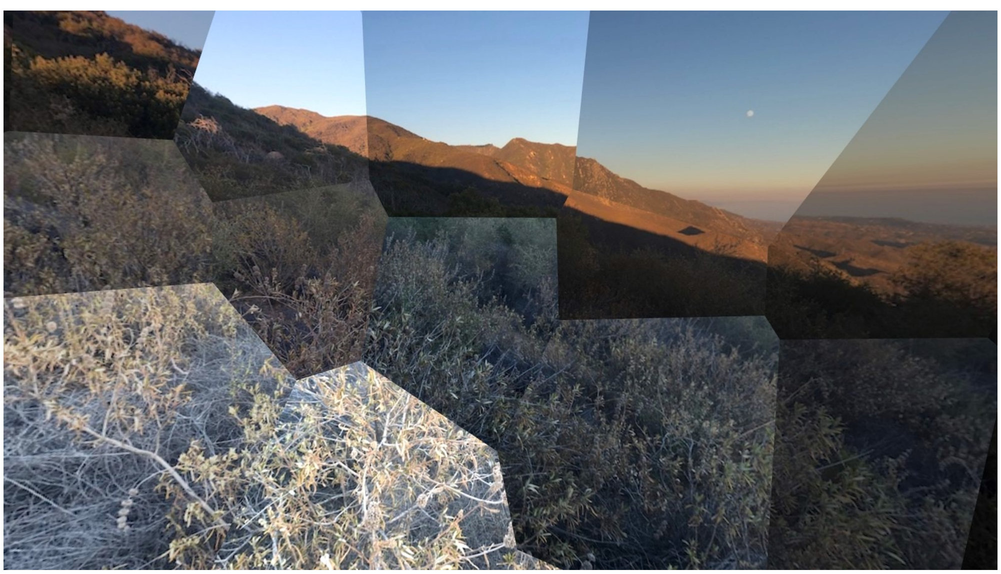

Figure A.4: *Santa Barbara Mountains*, David Gordon (2016)

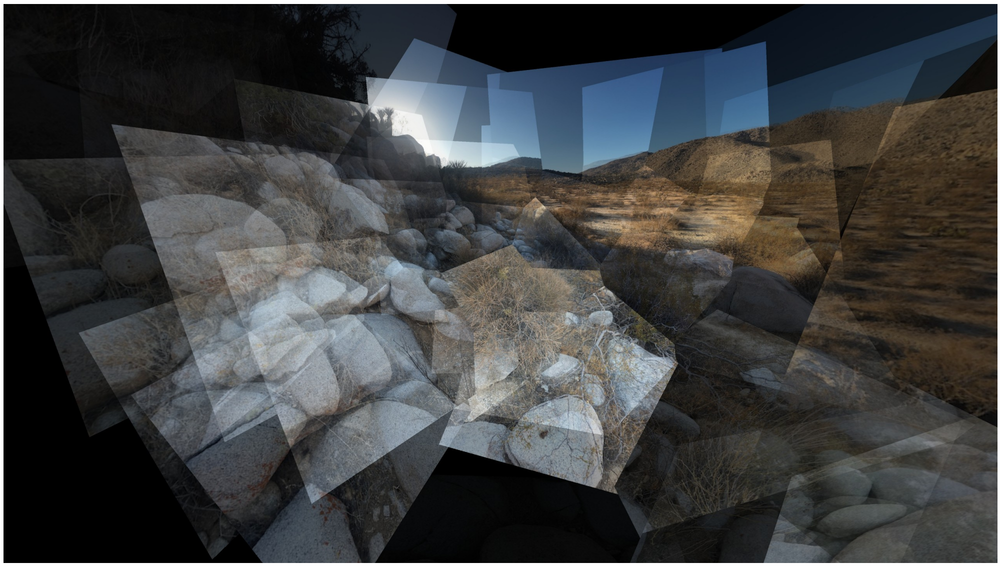

Figure A.5: *Smuggler's Canyon*, David Gordon (2016)

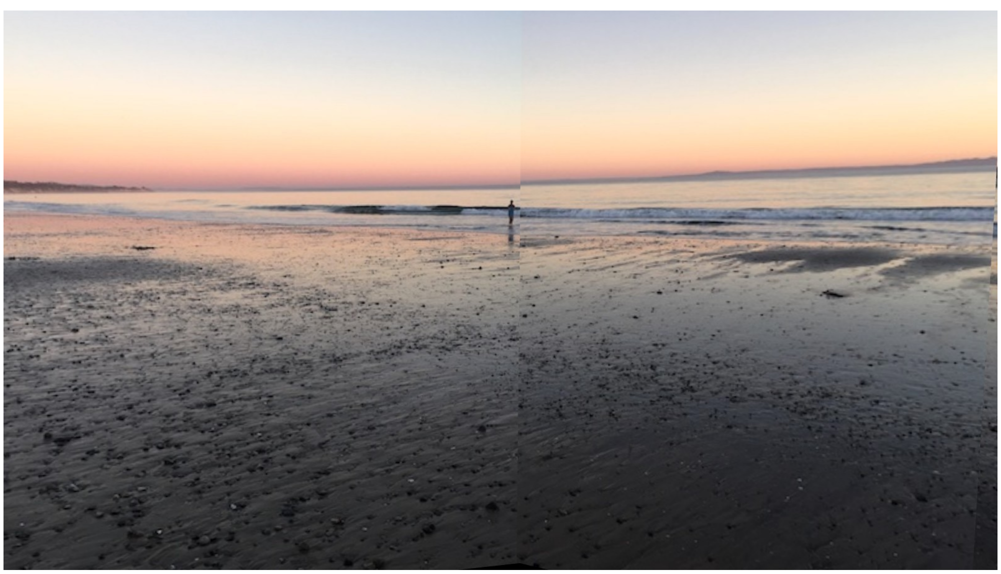

Figure A.6: *Malibu Creek*, David Gordon (2016)

Figure A.7: *Bergamot Station*, David Gordon (2016)

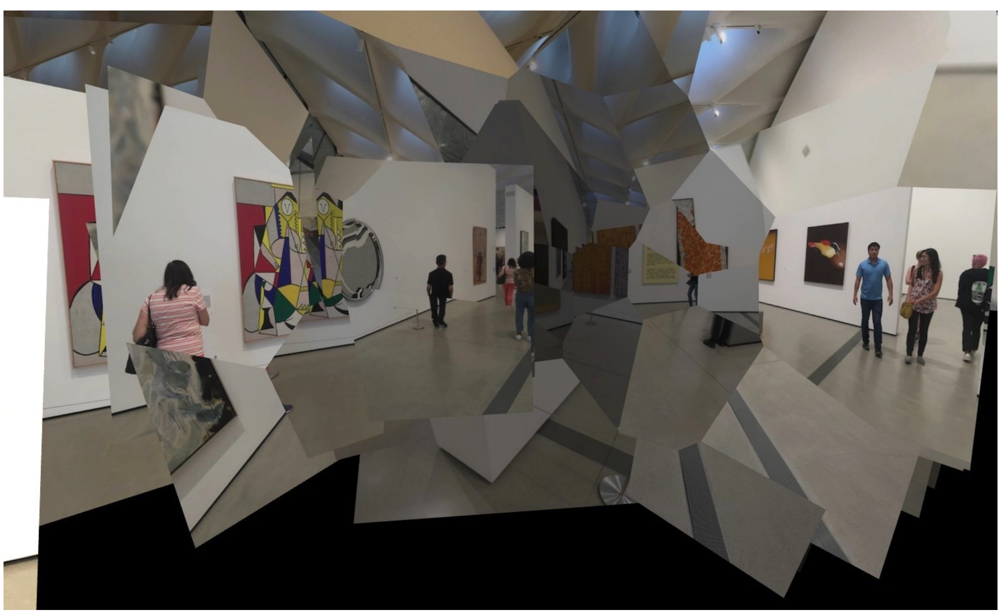

Figure A.8: *Broad Museum*, David Gordon (2016)