Title
VideoMob: building interactive digital crowds:

Permalink
https://escholarship.org/uc/item/7vx1g582

Author
Grenader, Emily Ruth

Publication Date
2014

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

VideoMob: Building Interactive Digital Crowds

A Thesis submitted in partial satisfaction of the requirements for the Master of Fine Arts

in

Visual Arts

by

Emily Ruth Grenader

Committee in charge:

Amy Adler, Chair
Amy Alexander
Teddy Cruz
Nadir Weibel

2014
The Thesis of Emily Ruth Grenader is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________
_____________________________________________________________________

Chair

University of California, San Diego

2014
TABLE OF CONTENTS

Signature Page ........................................................................................................ iii

Table of Contents .................................................................................................. iv

List of Figures ........................................................................................................ v

List of Tables .......................................................................................................... vii

Acknowledgements ............................................................................................... viii

Abstract ................................................................................................................ x

Introduction ........................................................................................................... 1

Chapter 1: Motivation and Related Work .............................................................. 3
  1.1: Group Portraiture ......................................................................................... 3
  1.2: Interactive Art Installations ....................................................................... 5
  1.3: Browsing Crowds Online ......................................................................... 12

Chapter 2: VideoMob ............................................................................................ 13
  2.1: Implementation ......................................................................................... 14
  2.2: Installation and Deployment .................................................................. 18
  2.3: Releases .................................................................................................... 19
     2.3.1: Green Screen Video Booth ............................................................... 21
     2.3.2: Houston Release .............................................................................. 22
     2.3.3: Bear Release ..................................................................................... 23
     2.3.4: San Francisco Release ................................................................. 23
     2.3.5: Athenaeum Release ....................................................................... 24
     2.3.6: Founders Day .................................................................................. 25
     2.3.7: TeleCrowd ....................................................................................... 26
     2.3.8: Filmatic ............................................................................................ 28

Chapter 3: Analysis ............................................................................................... 30
  3.1: Behavioral Analysis ................................................................................. 30
     3.1.1: Tagging Results ............................................................................... 31
  3.2: Visitor Insights ......................................................................................... 35

Chapter 4: Discussion ........................................................................................... 37
  4.1: Future Work ............................................................................................. 38
  4.2: Limitations ............................................................................................... 40

Chapter 5: Conclusion .......................................................................................... 42

References ............................................................................................................ 43
LIST OF FIGURES

Figure 1: *Civic Guard Group Portrait of 1529* ................................................................. 4
Figure 2: *Crowd Painting, Patrons* ..................................................................................... 5
Figure 3: *VIDEOPLACE* ...................................................................................................... 6
Figure 4: *Hole in Space* ........................................................................................................ 7
Figure 5: *Mi Casa Tu Casa* .................................................................................................. 8
Figure 6: *Sense of Space* ....................................................................................................... 9
Figure 7: *The Shadow Dance* ................................................................................................ 10
Figure 8: *Body Movies* ......................................................................................................... 10
Figure 9: *Gathering (spatial location)* .................................................................................. 11
Figure 10: *Gathering (hue and saturation)* .......................................................................... 11
Figure 11: *MegaFaces* ........................................................................................................ 11
Figure 12: *VideoMob Main Setup* ....................................................................................... 14
Figure 13: *Kinect Processing Pipeline* ............................................................................... 15
Figure 14: *VideoMob Architecture* .................................................................................... 16
Figure 15: *VideoMob State Machine* .................................................................................. 17
Figure 16: *Green Screen Booth* .......................................................................................... 21
Figure 17: *VideoMob, Houston Release* .............................................................................. 22
Figure 18: *VideoMob, San Francisco Release* ...................................................................... 24
Figure 19: *Athenaeum Release* ........................................................................................... 25
Figure 20: *Founder’s Day at UCSD* .................................................................................... 26
Figure 21: *TeleCrowd Installation Diagram* ....................................................................... 27
Figure 16: *TeleCrowd. Left: inside recordings; right: outside recordings.* ....................... 28
Figure 23: Filmatic Festival Installation .............................................................. 29
Figure 24: TeleCrowd, tagging. ......................................................................... 31
Figure 21: TeleCrowd sub-groups videos as defined by a variety of the tags. .... 35
LIST OF TABLES

Table 1: Release History ................................................................. 20

Table 2: Results of tagging ............................................................ 32

Table 3: What kind of behavior did you enact while recording? .............. 36

Table 4: Were you influenced by others recording before you? ............... 36
ACKNOWLEDGEMENTS

VideoMob was created by a small crowd of people who shared with me their technical expertise, valuable feedback, and insights about the future of the project. Amy Adler, the chair of my committee, has been with me since the beginning generously sharing her time, support, and wisdom.

This paper, in a similar form, has been submitted for publication in ACM Transactions on Interactive Intelligent Systems on Behavior Understanding for Arts and Entertainment as a paper titled The VideoMob Interactive Art Installation: Connecting Strangers through Inclusive Digital Crowds, 2014, Emily Grenader, Danilo Gasques Rodrigues, Fernando Nos, and Nadir Weibel. The thesis author was the primary investigator and author of this paper.

I met Danilo and Fernando in CSE118, a course taught by Nadir Weibel, PhD. Since that course we have continued to collaborate on a number of projects even though Danilo and Fernando both have returned to Brazil. They were able to translate my crowd video vision into workable code, share important creative decisions, and make the whole process a lot of fun! Nadir mentored us through the entire process sharing valuable insights and feedback, helping with testing and problem solving, and managing the research process. I would like to especially thank Nadir for allowing me to enroll in CSE118 which began a wonderful relationship for which I am very grateful.

I want to thank Amy Alexander, Teddy Cruz, and Brett Staubaum for sharing valuable feedback throughout the progression of VideoMob along with sharing insights toward a number of influential artists. Both Teddy and my brother Sam Grenader were especially great sounding boards for many installation schemes. My companion Bryan
Barry deserves many thanks for helping to build more than his fair share of *VideoMob* installations and for constantly being supportive throughout my adventure through graduate school.

I am also grateful to the CSE department at UCSD for hosting the Bear release, to UCSD Humanities for hosting *VideoMob* as part of Founders Day, to the Athenaeum Museum in La Jolla, CA for exhibiting the project, to ArtPower! for hosting the Filmatic release, and to the UCSD Visual Arts Department and Experimental Media Lab who hosted the *VideoMob* team in a three month artist-in-residence. I would also like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) in Brazil for aiding this multi-country collaboration through the Brazilian Scientific Mobility Program, and to the hundreds of viewers and participants that interacted with *VideoMob*. 
ABSTRACT OF THE THESIS

VideoMob: Building Interactive Digital Crowds

by

Emily Ruth Grenader

Master of Fine Arts in Visual Arts

University of California, San Diego, 2014

Professor Amy Adler, Chair

VideoMob is an interactive video installation platform that enables strangers visiting the artwork to interact across time and space through a computer interface that video-records their presence and co-situates visitors as part of the same digital
environment. By visually combining individual user videos to form a digital crowd, strangers are connected through the graphic display. This artwork is inspired by the way distant people can interact with each other through technology and influenced by artist working in the realm of interactive art. I deployed VideoMob in a variety of settings, locations, and contexts to observe hundreds of visitors’ reactions. By analyzing behavioral data collected from the users through depth cameras and questionnaires, I studied how these participants behave when given the opportunity to record their own video portrait into the artwork. In particular, I report and analyze the specific activity performed in front of the camera and the influences that existing crowds impose on new participants.
INTRODUCTION

Individuals interact every day through technology. When the World Wide Web became globally available in the 1990s, a new way for strangers to interact with one another became possible. Suddenly, individuals could join a chat room or forum that was dedicated to one’s own personal interest and reach a huge population of others with common interests from all over the world. Although technology is often seen as a barrier with respect to face-to-face communication, this same technology enables distant people to connect, and more and more people feel open to communicate with strangers, often creating collective and asynchronous interactions between people living in different places or interacting with technology at different times.

Collective and asynchronous forms of communication can take different forms. In chat rooms and forums this is typically based on text, while on social media such as Facebook, Flickr, or Instagram this is also oriented towards photography. YouTube and other video-based online platforms together with the advent of cheap high-resolution video cameras integrated within smartphones and mobile devices, are changing this panorama, enabling users to create online communities based on video media too. However, when we observe people interacting with online communities through personal technology such as their smartphone, it is becoming clear that the increasing risk of people isolating themselves in the real world while they are connecting in the digital world is worrying.

Interactive technology can be integrated in physical environments creating new opportunities to connect people in the real world through a digital environment. VideoMob attempts to create a portal for strangers who may be visiting the same place or walking on the same path to meet and interact through a digital system. Using
interactive technology and integrating it into the physical environment may increase the chances that strangers who merely pass each other on the street would interact and connect.

VideoMob records a video portrait of a user randomly passing by and interacting with the system, and instantly inserts that portrait into a collective display, forming a dynamic crowd of individuals. The installation aims to investigate the relationships between strangers who have visited the same physical location at different times by storing videos of those strangers and displaying them together even after the strangers have left. The participant becomes both subject and viewer, playing an active role in the process of joining a crowd.

While I wanted to assess the reaction of users to our interactive art installation, my hypothesis was that users behave differently based on viewing the behavior of the crowd displayed in front of them, and on the kind of prompt the user is given to participate, as well as the location and context of the VideoMob installation. To investigate this relationship, VideoMob was deployed in many different locations with hundreds of participants. VideoMob is based on Microsoft Kinect for Windows (http://kinectforwindows.org), which allows the recording of raw data from the visitor’s activity (such as distance from the camera, number of participants, body movements and depth maps) and analyzes and studies behavior in the different contexts and deployment sites.
CHAPTER 1: MOTIVATION AND RELATED WORK

_VideoMob_ was created as an artwork with the goal of connecting individuals who normally would not interact with one another. Within the virtual space enabled by _VideoMob_, visitors from different backgrounds, cultures, languages, etc. and visitors viewing the installation at different times have the opportunity to create a new community. As part of this virtual space _VideoMob_ creates a powerful visualization of self with others which leads to the formation of a novel sense of unity. By enabling the viewer to be instantly part of the virtual crowd, individuals who are not involved in the contemporary art world can feel an instant connection and understanding to the work by visually seeing themselves as a part of it.

The motivation behind _VideoMob_ stems from both contemporary attempts to give new dimensions to the crowd, as well as from a long history of group and crowd portraits that spanned the worlds of painting, photography and visual arts.

1.1: GROUP PORTRAITURE

Dirck Jacobsz b.1496 seems to be the first artist to create non-religious group portraiture (Riegl 1999). At first glance, each painted figure appears to be treated equally. Perspective does not bring members forward and hierarchical structure does not raise leaders high into the composition. The dark uniform wardrobe brings focus to the sea of heads commanding equal attention to each styled face. His painting, _Civic Guard Group Portrait of 1529_, spreads across three panels (see Figure 1). At the time of this artwork, a painted portrait was reserved for the individual, a lord, a king, a religious leader, and was rarely seen in groups. Historian E. deJogh felt he could deduct the religious convictions of the figures through their portrayed faces: “Inflexible, rugged faces, withdrawn into themselves… Let us not forget that times were hard for these
people. Most of them had already embraced the sober and strict Protestant faith” (deJongh 1990). The faces appear to ignore one another, looking forward, out to what lies before the canvas. Their anonymity begins to fade after studying individual’s tools or hand gestures. Physical actions and gestures give us the tools to recognize individual differences. With VideoMob we want to explore this dimension by looking at the signals that each individual creates when joining a virtual group, and how viewers of the work interpret and interact with those signals.

With my Crowd Painting, Patrons from 2010 (see Figure 2), I started to explore physicality in virtual groups. Patrons was painted from submitted photographs of mostly strangers from all over the world who funded the artwork through the website Kickstarter (http://kickstarter.com). This allowed for an egalitarian method of commissioning art in a digital age. Patrons from near and far formed a new crowd on the six-by-nine foot canvas. This work enabled a direct exchange between artist (me) and subject (patron) both in terms of my blog updates and comments on the Kickstarter page of the painting as it was being completed, as well as in terms of feedback through email and online comments from the participants. With VideoMob, the goal was to create an even more

![Civic Guard Group Portrait of 1529.](image)

**Figure 1: Civic Guard Group Portrait of 1529.**
direct interaction between viewer and artwork. Instead of participating in the painting of a crowd from far away, the viewer instantly becomes a part of the video collage in the inhabited space.

Figure 2: Crowd Painting, Patrons

1.2: INTERACTIVE ART INSTALLATIONS

By the mid-1970s Myron Krueger was experimenting with allowing visitors in different locations to appear in the same place through his project VIDEOPLACE (Krueger 1977), illustrated in Figure 3. In one of Krueger’s visions for his VIDEOPLACE technology, a visitor enters a darkened room that contains a large rear projection screen where his life-size image is projected along with the images of other people in other VIDEOPLACE environments. Krueger explains: “The visual effect is of several people in the same room. By moving around in their respective rooms, thus moving their images, the participants can interact within the limitations of the video medium.” VideoMob adds a different dimension to Krueger’s VIDEOPLACE because, as the videos are recorded
and replayed, visitors are able to leave their image even after they have physically left the gallery. Participants are then able to virtually stand next to the image of a person who is gone. "When people are physically together, they can talk, move around the same space, manipulate the same objects and touch each other. All of these actions would appear to be impossible within the VIDEOPLACE. However, the opposite is true. The video medium has the potential of being more rich and variable in some ways, than reality itself" (Krueger 1977). The moment becomes richer because of the opportunity to dance, jump, or slide into a portal where visitors share a space with visitors past.

Figure 3: VIDEOPLACE

_Hole in Space_, a project by Kit Galloway and Sherrie Rabinowitz, was a live event using satellite technology in 1980 where viewers from the Lincoln Center in New York City and the Century City shopping mall in Los Angeles were connected and could communicate (Figure 4). Kit Galloway describes the event as "so exciting that people started calling their friends and family on the other coast to try to see them and talk to them on the screen". When this project was restaged in _The Art of Participation_ show at the SF-MOMA in 2008, both videos were played on parallel screens replaying the
original 1980 interaction as if the groups were meeting again (Frieling 2008). With one installation of VideoMob I built on this idea exploiting facing screens to mimic face-to-face conversation between different participants and visitors concurrently attending parallel sessions on different VideoMob installations.

Sheldon Brown’s installation *Mi Casa Tu Casa* (Figure 5) began with a 1996 three-day residency at the San Diego Children’s Museum where six children from San Diego and six children from Mexico City were invited to share their ideas about home. Brown developed a multi-user virtual environment called *Ersatz*, which integrates a number of gaming and scientific visualization libraries to allow participants from two countries to play together in an interactive environment (Brown 2004). Two playhouses installed in San Diego and Mexico City were mapped onto a shared virtual reality space. Children in Mexico City could see and interact with children in San Diego sharing a
virtual world (Brown 2013). *Mi Casa Tu Casa* created the virtual space that Krueger described in *VIDEOPLACE*. Joining others across the border in a shared reality full of color, flight, and avatars was much richer for visitors than normal visits to the museum. Similarly to *Mi Casa Tu Casa*, I want *VideoMob* to enable the formation of crowds from events in multiple locations, allowing the study of how participants react and engage differently to technology based on their location.

![Figure 5: Mi Casa Tu Casa](image)

In 1998 Jennifer and Kevin McCoy used interactive video projection in *Sense of Space* (Figure 6). In this project, viewers found themselves displayed alongside previous visitors to the gallery using a pre-programmed algorithm to layer images. “A soundtrack places the viewer in fictional transit zones where instructions, descriptions and warnings are issued which alternately assist and confuse the viewer as they navigate the exhibition space” (McCoy 2012). The McCoys use the “instantaneousness of the electronic image” as a way to move quickly through a conceptualized space and allow the user to be a part of it (Bosma 1999). For *VideoMob* the concept of “instantaneous image” is engaged in a similar way to engage users in a different kind of art. Instead of viewing a painting or a drawing, visitors experience a work that changes as they enter its space.
In *Body Movies* from 2001 (Figure 8), Rafael Lozano-Hemmer displays portraits through large-scale projection that are only revealed when a passer-by creates a shadow. Inspired by Samuel van Hoogsraten’s engraving *The Shadow Dance* (Figure 7), Lozano-Hemmer creates an interaction between the visitor’s silhouette and the previously photographed pedestrian. Thousands of nearby street photographs are projected along with a bright light that cancels the projection. When a viewer obfuscates the light, the projected portraits are revealed. A surveillance tracking system detects when each portrait of an image has been revealed and cycles through to show new street photographs (Lozano-Hemmer 2001). *Body Movies* creates a dialog with the audience without using words. The piece makes visitors more aware of their own physical location, their relationship to others at different locations, and how their own size and scale relates to the architecture around them (Graham 2003). I was inspired by
the way Lozano-Hemmer created interactions to study “the brief temporary relationships that emerge” (Lozano-Hemmer 2006). I want the relationships that are created between individuals through VideoMob to inspire the visitor to think of the installation space as a new kind of community.

Figure 7: The Shadow Dance

Figure 8: Body Movies

In 2004, David Rokeby used surveillance video from outside the exhibition’s building to create Gathering, a large dataset of videos projected onto an immersive circle of eight video projections (Figure 9-10). In his system, a panning camera locates moving people who are separated into moving fragments based on color. These videos are rearranged along the screens based on a series of rules. In Figure 9, the fragments are organized by spatial location, and in Figure 10, they are sorted by hue and saturation (Rokeby 2007). By using surveillance video, Rokeby creates a one-way mirror in the sense that users inside the space can see those walking outside. A visitor entering the exhibition may find his image transformed by Rokeby’s algorithm. Here, the “content is contained in this difference between the gesture and it’s transformed or recontextualized reflection” (Rokeby 1995). In VideoMob, we hope to recontextualize the video of a user
by placing an individual within an environment full of other individuals who are also participants in the same environment.

**Figure 9:** Gathering (spatial location)  **Figure 10:** Gathering (hue and saturation)

Asif Khan’s project *MegaFaces* (Figure 11), designed for the Sochi Winter Olympics in 2014 used a Kinect Sensor to create 3-D portraits of visitors “giving everyone the opportunity to be the face of the Olympics” (http://asif-khan.com). Users changed their perception of a photo booth when experiencing the twist on the traditional photographic medium (Khan 2014). Like *MegaFaces*, *VideoMob* also aims to elevate the viewer to the additional role of being the subject.

**Figure 11:** *MegaFaces*
1.3: BROWSING CROWDS ONLINE

In addition to contemporary art in museums, interesting examples of visual collective interaction interfaces can be found online. For instance, Gigapan.org is an interface that enables the exploration of extremely large and high-resolution images, named Gigapixels. In 2007 Gigapan.org was launched by Carnegie Mellon University with NASA and Google, as a place to view and share these high-resolution images. The rationale behind Gigapan.org was to encourage “intercultural communication worldwide” (Heckbert 2010). While many of these images represent landscapes or cities, a Gigapan by David Bergaman shows President Barack Obama’s Inaugural address in 2009, and, more importantly, it shows the crowd attending the speech at a very high resolution, enabling investigations of the crowd at different levels of granularity. In this Gigapan, visitors to the website have the ability to zoom, click, and pan – creating waves of viewpoints that flow from the crowd to the individuals and back to the crowd again. The dots of faces can be zoomed endlessly to see the details of each individual (Bergman 2009).
CHAPTER 2: VIDEOMOB

The main goal of VideoMob is to expand the idea of my "crowd paintings" to incorporate a live video feed and a continuous interaction and feedback loop. By giving the viewer an immediate feedback image, I hoped to invite the viewer to participate in the artwork in a new way. In developing and deploying VideoMob I wanted to find out how this method of engaging strangers with the artwork and with one another would affect the way people act and behave. To visualize a group of people that had their videos recorded at different times, each video was required to have a transparent background where only the person in the video was visible while everything else was not recorded. This effect, widely used in film and television, can be achieved by filming someone that is in the front of a single colored panel. This single color can be removed digitally with a computer program or with a special filter exploiting the "green screen effect."

To allow VideoMob to be as mobile and natural as it could be, a Microsoft Kinect for Windows sensor was the best device to use in lieu of a large green screen. The lightweight Kinect could be easily carried around and required no modification to the environment in which it was being installed. Moreover, the "green screen effect" could be implemented using only the Kinect device. Although the edges of the filmed person presented some glitches and the resolution of video was not extremely high, the benefits of using the Kinect outweighed the visual imperfections. In the future advanced filtering in the graphic pipeline could be introduced to improve image quality. Also, the Kinect technology will advance and Kinect v2 will help in enhancing the quality of the video recordings.
Figure 12: VideoMob Main Setup

Figure 12 shows the main interface of VideoMob as setup in one of the releases (Bear release, see below). Users stand in front of the VideoMob and align their hands with the sketched hand-figure (left panel, and bottom-left on the right panel) to start recording. In specific releases visitors could record using audio detection as a trigger to start. Once recording starts, users can freely move and VideoMob records their video, automatically removing the background. While recording, visitors can behave as they wish, making speedy motions, waving, dancing, turning from the camera or just sitting still only slightly adjusting their position. After a pre-defined (and configurable) time the recording stops and interaction with VideoMob is ended. Upon exiting the recording mode, the user is suddenly gratified, as his recently recorded video appears immediately among all the video portraits of previous users in a large, ever changing crowd (Figure 12, right panel, top).

2.1: IMPLEMENTATION

The Microsoft Kinect SDK is implemented to interface with the Kinect sensor and activate recording of the scene. Both the RGB images and the depth images are recorded as exposed by Kinect, and the presence of new users is monitored exploiting the body-tracking feature of Kinect. Once a participant signaled he wanted to interact
with the system (see above) recording starts from both cameras (RGB + depth). The Kinect device developer interface allows the depth camera picture to be matched with the regular RGB picture. By doing so – with some loss of precision – the pixels that are further away from the camera (the background) can be removed. Also, the Kinect provides user information (through the body-tracking feature) for each pixel, giving information about whether that pixel is recognized as being part of a person. Hence, not only can the background be easily removed, but also anything not related to the users currently interacting with VideoMob. This allows the system to distinguish different people from each other. Figure 13 illustrates this process and workflow in detail.

**Figure 13:** *Kinect Processing Pipeline detailing processing of depth information (in yellow), RGB data (in green), and body (user presence) information (in blue). The merging of these three detection algorithms enables the creation of the visitors’ videos and their insertion into the virtual VideoMob crowd.*

*VideoMob* is built using the open source C++ toolkit OpenFrameworks (http://www.openframeworks.cc) as the underlying software framework and Kinect for
Windows as its main sensing device. OpenFrameworks, a portable framework aimed at leveraging creative coding in C++, works as a wrapper for a big set of technologies (e.g.: OpenGL, sound libraries, image libraries, event system, etc.). The Kinect for Windows SDK enables working with a C++ Application Programming Interface (API) for Microsoft Windows. A depth camera abstraction layer allows implementation of VideoMob’s functionalities independently of the Kinect sensor. This enables the use of cameras that are suitable for environments where a Kinect won't work (e.g. under direct sun light as it overpowers the infrared light emitted from Kinect, disabling Kinect from obtaining the depth information), without requiring changes to the architecture. The overall architecture of VideoMob outlining the interplay of Kinect and OpenFrameworks, and outlining storage of the recorded videos is illustrated in Figure 14.

**Figure 14**: VideoMob Architecture: all data is received through Kinect’s depth and RGB cameras and processed by the “Depth Camera Abstraction Layer”. OpenFrameworks provides the software base for working with OpenGL, the sound system for recording /playing sound streams, and a simple XML parser used to load VideoMob’s configuration. All videos and sounds recorded are saved to a permanent storage.
**VideoMob** is a created as single window software application that usually extends to fill up to four monitors in a multi-screen computer environment. As shown in Figure 12, this single window displays a real-time feedback from the camera that is used as the main interaction platform with the user, as well as the resulting animated crowd—usually spanning multiple screens. In order to ensure proper communication and interaction with users, VideoMob is based on a multi-stage state machine that is responsible for handling all user interactions, as described in Figure 15.

**Figure 15:** VideoMob State Machine: once visitors enter the camera range, the system waits for a first user interaction (hand waving or sound). After a predefined threshold is reached the system changes into recording mode and notifies users with a special “Ready-Set-Go” screen. While video is recorded users do not further interact with the user interface, but the system records any movement or activity, which is eventually saved and displayed within the crowd. When the cycle is over, VideoMob is ready for the next visitor.
2.2: INSTALLATION AND DEPLOYMENT

*VideoMob* has been installed on a variety of computer systems running Windows 7 or 8. A *VideoMob* computer with the standard Kinect drivers must be connected to a Kinect for Windows sensor and one or more displays. Although *VideoMob* targets low cost, mobile hardware (e.g. notebooks, HP all-in-one) and its code is optimized to use as little processing time as possible due to the data-intensive real-time image processing a fast machine is recommended with a minimum of 4GB of memory. All videos are saved in raw format during an interaction, the crowd animation is totally allocated in the memory, and sound files are saved as waveform files.

Current graphic cards usually support up to four screens and usually the use of more than two monitors requires an extra graphics board and special adaptors. An implementation allowing for more than four monitors could be achieved with the use of SAGE and OpenGL capture if needed (Renambot 2013).

In order to enable frequent installations of *VideoMob* on several diverse and unique environments, an easy way to customize *VideoMob* was created to adjust the way it would be visualized. Window dimensions, frame rate, features to be used, recorded videos length, and more can be specified as part of a special XML configuration file. XML is used because its syntax is simple and self-descriptive, and OpenFrameworks provides a straightforward interface to this technology. The definition of different XML configurations for each environment where *VideoMob* was deployed (see below) allowed specific detail adjustments such as the number of monitors to use, the monitor resolution, the amount of memory available for graphics, the size of the
window, the positioning of the user interface as well as its dimensions, the recording frame rate and different properties of the crowd drawing algorithm.

### 2.3: RELEASES

Over the past year, starting in July 2013, VideoMob has been released in a variety of venues. After each event user feedback was incorporated to improve the system and integrate new features with each release. The iterative releases also allow light to be shed on different user characteristics and different contexts of use. VideoMob releases were set-off by a pre-release of an earlier “Green Screen”, regular-camera prototype that initiated the development of the Kinect-based VideoMob software.

Table 1 lists VideoMob’s release history indicating the environment, the location and the features that have been integrated in that release. In the remainder of this section we describe each release in detail, highlighting data collection and initial findings.
### Table 1: Release History

<table>
<thead>
<tr>
<th>RELEASE</th>
<th>DISPLAYS</th>
<th>DATE</th>
<th>LOCATION</th>
<th>NEW FEATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-release “Green Screen”</td>
<td>Interface on monitor with crowd projected.</td>
<td>May 22, 2012</td>
<td>Art show, University of California, San Diego, La Jolla, CA.</td>
<td>Max MSP projected crowd</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houston</td>
<td>Interface on monitor with crowd on large screen.</td>
<td>July 29, 2013</td>
<td>After party in Houston, TX.</td>
<td>Uses Kinect sensor to record WebM video files.</td>
</tr>
<tr>
<td>Bear</td>
<td>Single application shown on multi-screen display.</td>
<td>August 2013 – March 2014</td>
<td>CSE Lobby, University of California, San Diego, La Jolla, CA.</td>
<td>10sec uncompressed video files to display videos much more quickly.</td>
</tr>
<tr>
<td>San Francisco</td>
<td>Interface on monitor with crowd on large screen.</td>
<td>August 17, 2013</td>
<td>Wedding reception, Berkeley, CA.</td>
<td></td>
</tr>
<tr>
<td>Athenaeum</td>
<td>Interface on monitor with crowd projected.</td>
<td>November 8, 2013 - January 5, 2014</td>
<td>The Athenaeum Music and Arts Library, La Jolla, CA.</td>
<td>Changed record message to “Ready, Get Set, Go!”</td>
</tr>
<tr>
<td>Founder’s Day</td>
<td>Interface on monitor with crowd on multi-screen display.</td>
<td>November 14, 2013</td>
<td>Outdoor celebration, University of California, San Diego, La Jolla, CA.</td>
<td></td>
</tr>
<tr>
<td>TeleCrowd</td>
<td>Two interface monitors with crowds projected onto rear-projection screens.</td>
<td>March 8 – 14, 2014</td>
<td>Art show, University of California, San Diego, LA Jolla, CA.</td>
<td>Added 10 second sound recording, and record of skeleton data.</td>
</tr>
<tr>
<td>Science Discovery Day</td>
<td>Interface on monitor with crowd on projected multimedia screen</td>
<td>March 26, 2014</td>
<td>Torrey Pines Elementary School, La Jolla, CA.</td>
<td></td>
</tr>
<tr>
<td>Filmatic</td>
<td>Interface on monitor with crowd projected onto campus architecture.</td>
<td>April 26, 2014</td>
<td>Filmatic Festival, University of California, San Diego, La Jolla, CA.</td>
<td></td>
</tr>
</tbody>
</table>
2.3.1: Green Screen Video Booth. The pre-release of VideoMob consisted of a software toolkit created using the MAX MSP interactive media framework (http://cycling74.com/products/max). This pre-release, shown in Figure 16, consisted of a simple photo booth that allowed users to participate in a large projected crowd video. Visitors to the exhibition entered the booth where they were recorded for five seconds by a webcam while in front of a standard green screen. Users had no direction other than a request to stay within the green screen area. The large booth was cumbersome, but users were able to enter into some state of privacy. Unlike later versions, there was a stool for visitors to sit in the booth. Collected videos were then automatically processed by MAX chroma-key algorithms and added to the projected crowd. The Green Screen Video Booth was released at the University of California San Diego, Visual Arts Department, as part of a solo-exhibit, as well as in Houston, TX and in San Diego, CA, as part of local art festivals. No formal data was collected, but users seemed to make small performances for the camera like making funny faces or spinning around.

Figure 16: Green Screen Booth
2.3.2: Houston Release. The first real version of VideoMob was released at a party in Houston, TX in June 2013 (Figure 17). This issue was an experiment using Kinect to replace the need for a green screen. It mixed a basic OpenFrameworks application that captured the videos with a webpage to display them. Two applications were used (one for the crowd, one for the input interface) and a running webserver collected the videos from the input window to form a crowd on the other window. It was slow as videos were saved using the WebM standard (http://www.webmproject.org) which took 30-60 seconds to convert and save to video. Furthermore, the crowd playback was made of many compact videos playing at the same time, not allowing the application to scale smoothly as people were added to the crowd. Finally, no sound was recorded. Participants were very supportive and enjoyed trying it out but reported that videos were too dark and took too long to save for playback. No formal data was collected, but users seemed to mimic the behavior of other users in the video. Most videos included dancing. Users appeared excited and joyful.

Figure 17: VideoMob, Houston Release
2.3.3: Bear Release. The next release of VideoMob was posted in the lobby of the Computer Science and Engineering building at the University of California, San Diego, UCSD (see Figure 12 above). This release was the first one that used a single application, decreasing the time required to update the crowd. Several improvements were introduced too: the crowd was quickly generated by keeping it on memory and only adding the recently filmed frames using a separate thread of processing. Videos were saved in a raw format to reduce the time users would wait after recording it and the crowd was saved as a single video, making playback smoother. The video was left running for seven months to test the extended use of VideoMob. Computer Science students seemed interested and routinely stopped at the VideoMob installation recording mostly group performances. The addition of VideoMob to the Lobby transformed it into a more social environment where people gathered and played with the technology.

2.3.4: San Francisco Release. The next release of VideoMob was shown at a wedding reception (Figure 18). A directed light was used to make sure users were well lit. Many bugs that made the application crash were solved and a new algorithm for drawing the crowd was introduced. In the Bear Release, the video recording of each person started at the same time in a way that any viewer could easily realize that the videos were being repeated over and over again. In this release, each person's video had a random start time so that no two videos would start together to make it harder for the viewer to realize that the whole crowd was playing in a loop. As the wedding reception was a celebratory event, users appeared happy. Most users chose to dance. The choice to dance was probably aided by music that was also playing in the space. Contemporary
photo booths that capture still photography have become popular for parties, receptions and events. They are easily available as full systems like Smile Booth (http://smilebooth.com) or software like dslrBooth (http://www.dslrbooth.com). Many users of the San Francisco release commented on the creativity and excitement of having VideoMob as opposed to a conventional photo booth.

![Figure 18: VideoMob, San Francisco Release](image)

**2.3.5: Athenaeum Release.** In the Athenaeum release (Figure 19), VideoMob was displayed as part of a public art collection within the Athenaum Music and Arts Library for five weeks. Most of the feedback received during this release was from the museum staff that was able to watch people interact with the installation every day. For the library, which began curating art exhibits in 1990, it was the first digital interactive installation. Visitors participated differently depending on whether they were there for an evening event (music concerts, poetry readings, and even the annual holiday party), or a normal daytime visit to the library. The staff described the final collections of videos as a strong representation of the community of library members, visitors, and other staff members.
Because of this description, the collection of images became a portrait of the Athenaeum Music and Arts Library without needing any images of books, music or the interior/exterior spaces. Instead, members of the community were able to recognize the place based solely on the people who participated in VideoMob.

![VideoMob, Athenaeum](image)

**Figure 17:** VideoMob, Athenaeum

**2.3.6: Founders Day.** During Founder’s Day, an outdoor celebration at UCSD’s campus (see Figure 20), a booth was made resembling an old timey photo booth. Instead of acting as one might in older times, when portrait sitters needed to be still for the slow light sensitive photographic material, visitors mostly danced and laughed for their videos.
2.3.7: TeleCrowd. The next installation of VideoMob was deployed in the art gallery at UCSD. The show, entitled TeleCrowd, displayed two large parallel projection screens that filled the entire gallery. As shown in Figure 21, two separate computers ran two instances of the VideoMob software. One was outside the gallery and one was inside the gallery. Inspired by the Hole in Space installation at SF-MoMA (see Figure 4), two projectors were used to display the crowds from the two recording inputs. The projected crowd created from individuals recording inside the gallery faced the projected crowd created from the individuals recording outside the gallery. This was made possible by using two large rear-projection screens which faced each other within the gallery. In order to be more realistic, this version added a new dimension to the artwork: audio feedback. While installed, VideoMob recorded the individual voices of the visitors and played them back to create the effect of the murmuring noise of a crowd within the gallery.

Figure 18: Founder’s Day at UCSD
From all the recorded audio clips, only the last twenty are played simultaneously with a fixed but randomly chosen volume. For each of the loop iterations, one of the audio recordings was chosen to have an outstanding volume.

Given that audio was a novel dimension of the artwork, *TeleCrowd* exploited sound also as a new way to interact directly with the installation. Visitors of *TeleCrowd* were in fact able to start interacting with it and initiate video recording by speaking in the direction of the Kinect microphone. The *VideoMob* software detected that someone was close to it and talking at a volume level that was above the average and initiated recording. In the loud space of an art exhibition with constant background sounds, visitors of *TeleCrowd* were forced to come closer to the computer in order to be detected by the Kinect microphone and camera. Figure 22 shows the resulting crowds for both inside and outside settings. Overall, individuals appeared to enjoy participating in this installation of *VideoMob*. Because one of the inputs was outside, many people walking by with no intention of participating were drawn in to try it out.
2.3.8: Filmatic. The next installation of VideoMob was displayed during UCSD’s Filmatic Festival (http://filmaticfestival.com) in April 2014. This version explored relational architecture, as it was projected onto the wall of UCSD’s Price Center & Student Center as shown in Figure 23.

Students, staff, and visitors of UCSD were engaged by physically placing their image on the campus walls. Although most participants were attending the festival, many passersby also paused to be recorded into the crowd.
Figure 23: Filmatic Festival Installation
CHAPTER 3: ANALYSIS

At every release of VideoMob, a variety of behaviors were expressed, and the ways visitors of different backgrounds and ages interacted with the artwork influenced these behaviors. It seemed also that users behaved differently based on viewing the behavior of the crowd displayed in front of them. In order to better understand this behavior and map it to the different dimensions of interaction enabled by VideoMob, the installation version TeleCrowd, the latest and most complex release, was used to investigate interactions by means of a categorical analysis of the recorded videos.

3.1: BEHAVIOR ANALYSIS

The infrastructure of VideoMob automatically saved every single video of every visitor recorded in the releases described above. Collected videos range from 30 to 550 per installation across the different releases. Given time constraints and for practical manners it was not possible to analyze all available videos, I therefore selected a subset of them out of the most complete release, TeleCrowd. Although this represents only a small subset of the whole data, we feel that this is representative of interactions with VideoMob for the following reasons: (a) TeleCrowd was the most recent and complete installation of VideoMob, (b) TeleCrowd had the most users per hour of all releases so visitors who recorded themselves frequently had the opportunity to see other people who also recorded themselves in the same space, (c) by choosing videos that were recorded consecutively, we were able to study visitors who recorded themselves multiple times and (d) by choosing consecutive videos we were able to study how users were influenced by the users recorded before them.

The first 100 videos from this release were studied and tagged depending on the recorded visitor’s action, age, quality of recording, apparent emotion, amount of
movement and people in the video. Figure 24 illustrates the process of tagging one of the videos. Since only one coder subjectively performed the visual video analysis, this might lead to inherent classification errors. However, we feel that classification of this data is based on broad and well-identifiable categories, and it is reasonable that the tagging will be mostly correct.

Figure 24: TeleCrowd, tagging.

3.1.1: Tagging results. To discover whether users behaved in similar ways their recording were classified by action. Actions were categorized as: acting (performing for the camera), dancing (displaying dance moves or moving to a beat), fighting (making punching or kicking motions), goofy (making silly faces), jumping, kissing, modeling (striking model like poses for the camera), standing still, talking, using a prop (like sunglasses or a camera), and waving. In some cases, users displayed multiple actions in each video. In that case, the most prominent action was tagged. Table 2 shows a summary of the tagging and analysis.
### Table 2: Results of tagging

<table>
<thead>
<tr>
<th>Action</th>
<th>Still</th>
<th>Dancing</th>
<th>Acting</th>
<th>Goofy</th>
<th>Modeling</th>
<th>Fighting</th>
<th>Waving</th>
<th>Using Prop</th>
<th>Kissing</th>
<th>Jumping</th>
<th>Talking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
<td>20</td>
<td>9</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Recordings</th>
<th>Full</th>
<th>Left before</th>
<th>Accidental</th>
<th>Multiple Times</th>
<th>Frame Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>51</td>
<td>8</td>
<td>26</td>
<td>17</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Unaffected</th>
<th>Happy</th>
<th>Cool</th>
<th>Confused</th>
<th>Awkward</th>
<th>Engaged</th>
<th>Sad</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39</td>
<td>27</td>
<td>13</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>Mixed</th>
<th>35+</th>
<th>17-35</th>
<th>Under 16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22</td>
<td>24</td>
<td>40</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image Quality</th>
<th>Good</th>
<th>Dark</th>
<th>Too close</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>66</td>
<td>5</td>
<td>4</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of People</th>
<th>Group</th>
<th>Pair</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>39</td>
<td>54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Movement</th>
<th>Big</th>
<th>Medium</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>27</td>
<td>61</td>
</tr>
</tbody>
</table>

**Age**: Users were classified by age to determine broadly the age group that participated. A lot of mixed age groups have been tagged, representing mostly parents with children. The biggest single age group of visitors is the one around age 17-35. I feel that this is venue-specific and driven by the location of the art gallery on a college campus. During the installation at the music and arts library, for instance, many older adults participated.

**Behavior**: The most exhibited behavior was staying still (40%), however most users (58%) displayed some sort of active behavior. Among those who chose to be active, the predominant behavior was dancing (20%).

**Movements**: I also looked at the amount of movement, referring to the speed and extent of a user's motion. Low movement was assigned to individuals who stood still or created minimal movements, medium movement was assigned to individuals moving arms but
not legs, and big movement was assigned to individuals who moved the whole body throughout the recorded space. While the majority of users displayed low movement one third of the visitors moved their arms during the recording.

**Group Dynamics:** The analysis of group dynamics showed that out of our 100 users almost half of them engaged with VideoMob in small (pairs) or bigger groups.

**Emotions:** To gauge emotion and reactions to VideoMob we categorized our visitors using seven terms: awkward, confused, cool, engaged, happy, sad, and unaffected. Visitors roughly divided in two groups: 39% appeared unaffected by the camera, while another 46% reacted with positive attitude (happy or cool). Only a minority reacted negatively or by being uncomfortable with VideoMob (15% between awkward, confused and sad).

**Recordings:** The amount of time recorded per user was studied to discover whether visitors remained in the frame for the entire ten seconds of recording. We looked at events such as accidentally triggering recording by standing in front of the “waving hand,” as well as the amount of times visitors left the frame before the recording ended. About half of the recordings were purposeful and fully recorded for the duration of the recording. For the remaining half, it is interesting to see how only 8% of the visitors left the recording before the 10 seconds expired, while a good portion (26%) accidentally triggered recording, but resulted in being engaged with the system. I also looked at repetitive recordings: 17 of the one hundred analyzed videos where visitors who recorded themselves multiple times. In these cases, users either accidentally started
recording, or consciously played with the system with the goal of appearing multiple times beside themselves, or engaging in different kind of behaviors.

*Image Quality*: I also studied the quality of the videos recorded. In many cases (27%), our system masked out more than just the background leaving “image gaps” on important parts of the figure, like the face, during some frames. Also, when a user was too close to the Kinect Sensor (closer than 18 inches), the figure would be masked out of his own video causing the system to record empty frames. However, we still recorded 66% of good video, which is quite satisfying given the “in-the-wild” nature of the exhibition and the recorded data.

Based on these results I was able to compile subsets of the crowd and visualize them separated in different groups. Replaying the groups as stand-alone crowds enables further analysis and in-depth insights into behavior of the single visitor groups. Figure 25 shows eight of these *sub-crowds* represented statically for different age groups and body behaviors.
3.2 VISITOR INSIGHT

Since I wanted to also investigate the perceived impact on visitors of the VideoMob artwork, as well as their subjective reactions, I elicited direct feedback from them through an optional questionnaire. In the questionnaire I enquired about the kind of behavior visitors enacted and if they were influenced by previous behavior. Nineteen adults (age 18 to 35) and twenty-three youths (age 12 to 14) filled out our surveys. Table 3 presents a summary of the results in terms of the enacted behavior, while Table 4

Figure 25: TeleCrowd sub-groups videos as defined by a variety of the tags.
illustrates how visitors were influenced in their behavior. Both tables are divided by age group (adults vs. youths).

**Table 3:** *What kind of behavior did you enact while recording?*

<table>
<thead>
<tr>
<th></th>
<th>ADULTS</th>
<th>YOUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>4, waving</td>
<td>35% 8, dancing</td>
</tr>
<tr>
<td>15%</td>
<td>3, testing recording limits</td>
<td>26% 6, stood there/acted normal</td>
</tr>
<tr>
<td>15%</td>
<td>3, pretend fighting</td>
<td>22% 5, acted goofy/silly</td>
</tr>
<tr>
<td>10%</td>
<td>2, stayed still</td>
<td>13% 3, acted like others</td>
</tr>
<tr>
<td>10%</td>
<td>2, high five</td>
<td>4% 1, jumped</td>
</tr>
<tr>
<td>10%</td>
<td>2, rock paper scissors</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>1, dancing</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>1, pretending to have</td>
<td></td>
</tr>
<tr>
<td></td>
<td>conversation</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>1, jumping</td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>1, erratic, random</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4:** *Were you influenced by others recording before you?*

<table>
<thead>
<tr>
<th></th>
<th>ADULT</th>
<th>YOUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>8, 42%</td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>3, 16%</td>
<td>No</td>
</tr>
<tr>
<td>Don’t know</td>
<td>8, 42%</td>
<td>Don’t know</td>
</tr>
</tbody>
</table>

Subjective results illustrate how most younger people reported to dance, and act goofy or silly in front of the camera (total of 61%), while adults generally had a less interactive experience, often characterized by still recording, or mostly reduced to just waving and predominantly using their upper body (total of 80%). Most participant across both groups (42% and 52% respectively) indicated that their behavior was influenced by what was already visible on the screen.
CHAPTER 4: DISCUSSION

Visitors responded in different ways to being asked to participate amongst a crowd of creative performances. Historically, it is difficult to get visitors – with the exception of children – to participate in public participatory installations because of their skepticism toward surveillance, their dislike toward art interventions interrupting the norm, or an engrained “do not touch” attitude (Ranzenbacher 2001). However, with VideoMob, many users were willing to participate, not only children. In many cases, a single user with no connection to the authors would become a “guide” to the system: with no prompting this user would invite strangers to pose and explain how the system works. Also, because the videos were displayed in large groups where no one was singled out, users felt they could “blend into the crowd.” Because of some users extending their participation to become guides for others, and because the system output was visible (users knew what they were recording themselves for) users felt comfortable recording a video.

People interacted with VideoMob in different ways. Some people viewed the work without recording a video. Others watched the videos, recorded their own, and then quickly noticed it in the projection before leaving. Some participants recorded a video, then studied the projections, and then recorded more videos. Some users directly interacted with their own videos or the videos that would surround theirs, creating a series of actions that relate and interact with earlier actions. For example, one user turned his back to the camera and began pointing and talking in every direction. In the projection, it appeared he was engaged with the users above, below, and to the sides of him.
When recording, visitors could choose to record individually or in a group. Although many users chose to record with another individual, the way they behaved did not seem to be determined by the number of people recording. Actions like dancing, kissing, and fighting, which tend usually to be done in pairs, were acted out individually as well (we viewed recorded videos where a user kissed himself by recording twice and another who kisses his virtual neighbor by “air smooching” during the recording.) These virtual interactions added richness and new dimensions to the VideoMob world.

In order to understand how people behave in interactive environments, I studied recorded videos of their interactions. Also, I was present during a large part of the VideoMob installation periods. It appeared that while many users were able to easily interact with the system, some users would try to touch the hand on the screen instead of holding their hand in line with the start button. The system was only able to start by recognizing that there was person in front of the Kinect sensor and by tracking that they were in line with the start shape. Although I did not expect this behavior, I observed how visitors quickly learned how to interact with VideoMob. This re-enforced my feeling that this kind of interactive artworks can easily be integrated in every public or semi-public environment as a vehicle to engage visitors in the artwork itself, while also helping close the gap between interactive art and cutting-edge sensing technology.

4.1: FUTURE WORK

Even though VideoMob is currently based on simple capturing of user’s movement, without any real-time interpretations, it could benefit from a more in-depth understanding of users’ behavior. For example, it could make use of activity recognition or pattern recognition to cluster people doing similar activities (e.g. dancing, jumping, running, looking) or wearing similar clothes (college t-shirts, sports clothes, etc.) in order
to personalize the result displayed to certain group of users. Technologies such as RGB-HuDaAct (Ni 2013) or algorithms that recognize movements of 3D joints using histograms (Xia 2012) can be used for activity recognition based on data captured with a Kinect sensor.

Another interesting development for VideoMob would be to combine personalization with persuasion (Berkovsky 2012). For instance, VideoMob could personalize the crowd displayed, as well as the way visitors are hooked to the display to certain users or user groups in order to motivate them to record a video or repeat an activity that is shown on the screen.

The biggest constraint of these kind of real-time and online analyses is the processing power required to run most of the described activity recognition algorithms, making it only possible with computers that are more powerful, bigger and less mobile than the ones we used for setting up a VideoMob in the wild.

Although online processing of data might be challenging given the constraint of a flexible and mobile installation, data about visitors’ activity and behavior can still be used for off-line analysis of behavior. Given our growing interest in the dynamics of movements in front of the VideoMob installations we decided to add tracking of body joints for visitors interacting with VideoMob. By exploiting Kinect Skeleton tracking functionality VideoMob is now able to record high-resolution data of the movements of up to two people’s bodies in front of it. This data is composed by a set of three-dimensional points for each of the hands, elbows, wrists, arms, shoulders, knees, ankles and feet along with points for hip, head, spine, and shoulder-center (neck). Future releases of VideoMob will allow us to further study the behavior and activity of our visitors through interactive visualization of VideoMob body data. I would like to discover
how I can explore the abstraction of a visual digital crowd as a way to represent group’s behaviors and feelings. I hope that these novel ways of visualizing behavior of crowds will inform new developments in terms of interactive art installations.

I also see potential bringing VideoMob beyond the online and offline tracking of behavior and activity. We are currently starting to setup a future release of VideoMob, in which the two linked computers would be located in two geographically distinct places, for instance in San Diego, California and Campinas, Brazil. The image shown on each screen would be the image seen from the camera of the faraway computer, instead of acting as a high tech mirror. A user in Campinas would see the person in San Diego; the user in San Diego would see the person in Campinas. They could talk to each other. Google Voice (http://google.com/voice) and Google Translate APIs (http://developers.google.com/translate) could be integrated to create captions from Portuguese to English and from English to Portuguese. Like the screens in TeleCrowd, the Campinas projection would face the San Diego projection. Audio would be layered and played.

4.2: LIMITATIONS

The process of creating VideoMob has been a successful step toward realizing my goals of creating meaningful connections among strangers, exposing art to those who would normally not be interested, and studying how people behave when invited to participate in an artwork using new sensing technology and based on human-computer interaction. VideoMob was limited by a few technological impediments. First, the Kinect Sensor has a lower resolution RGB cameras than many small USB cameras so our collected videos do not compare in quality to affordable and easily accessible webcams. Secondly, using the depth data to create a “green screen” effect is less accurate than
traditional chroma-key methods. Thirdly, the Kinect Sensor works poorly in sunlight limiting the location where we set up VideoMob. Finally, the program uses a large amount of processing power impeding us from running multiple programs that may contribute to giving the user a deeper experience (like Skype or other live communication software). We are in the process of integrating Kinect v2, which would help mitigate some of these problems.

My study of VideoMob data was also limited by some important factors. Because of the system’s high use of processing power, VideoMob would crash when we attempted to run tracking data to record skeleton coordinates and live data analysis. Also, manually tagging the data was time consuming, and the authors were unable to record data of each video. Lastly, the person doing the video tagging frequently needed to make subjective decisions throughout the process. We hope that adding facilities for automatically recognizing behavior and activity will help further understanding how VideoMob and other similar artworks can engage visitors and how their behavior is shaped by the visiting experience.
CHAPTER 5: CONCLUSION

VideoMob is a new interactive artwork that employs Kinect-based recordings of visitors and automatic placement in virtual crowds. It is successful in engaging a large number of participants in a very diverse range of settings, as well as in creating a novel way of connecting strangers over time. The behavior analysis highlights how participants are engaged with the artwork and exhibit a range of different activities. Group-based activities were mostly playful or exuberant, such as dancing, while single activity varied. Independent of age or social context, as well as the environment chosen to deploy the work, VideoMob acts as a catalyst for passing visitors, who can – sometimes for the first time – experience interaction with a digital environment in an easy and joyful way. I hope that continued efforts will further enable visitors to experience social and digital connectedness in a variety of different and novel dimensions, such as asynchronous time, or geographically separated environments -- further exploiting the power of inclusive interactive art such as VideoMob.
REFERENCES


