

THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

The Bathysphere

Motion Capture and Immersive Projection

Caitlyn Losee

12/1/2010

Under the direction of Professor Greg Welch

Abstract

The University of North Carolina at Chapel Hill's Institute for the Arts and Humanities works to promote collaboration across disciplines. In February of 2010, this institute coordinated the first CHAT Festival (Collaborations: Humanities, Arts and Technology), which involved several universities in the Triangle area; UNC-CH hosted the event (1). We created an engaging art installation for the festival where audience members stepped into a small building and entered an underwater world. The audience stood on the sandy sea floor, ocean stretching out into the distance around them. As they moved the physical objects in the space, sea creatures mirrored their motions in the underwater environment. This system employed motion capture technology to continuously measure and record the three dimensional position and orientation of the physical objects, a game development tool to move the models and render the environment, and multiple projectors to create the large-scale immersive environment. In this document, I describe from a technical perspective the motivation and evolution of the project as well as the final results presented at the festival.

Motivation

In February of 2010, the University of North Carolina at Chapel Hill hosted the CHAT Festival (Collaborations: Humanities, Arts and Technology). This festival was coordinated by the Institute for the Arts and Humanities, whose goal is to encourage cross-disciplinary collaboration and promote intellectual exchange. Several local universities participated in the festival, which included a variety of discussions, workshops, performances and exhibits (1). Professor Greg Welch of the UNC Computer Science department and Professor Francesca Talenti of the UNC Communication Studies department decided to team up to prepare an

installation for the festival. They secured a space, Gerrard Hall, for the installation (Figure 1). Gerrard Hall is a small rectangular building with two floors. The bottom floor is open, with a small alcove at one end, while the top floor consists of a balcony around three sides, overlooking the alcove (Figure 2). This small but open space was perfect for creating an immersive experience. Welch and Talenti knew they wanted to create an interactive experience that was engaging for the audience members controlling the interaction as well as the audience members observing. They also wanted the experience to be natural and intuitive, requiring as little explanation or instruction as possible. Since the installation would be in place for several days, it also needed to be robust.



Figure 1: Gerrard Hall, outside (2)



Figure 2: Gerrard Hall, inside (2)

Related Work

Artistic installations around the world involve using motion capture to create interactive art. The soundSpace at the North Carolina Museum of Life and Science uses multiple cameras (whose fields of view don't overlap) to track the movement of multiple users and generate audio dependent on location and amount and type of movement (3). There are also many installations that use the movements of the audience to alter imagery. Mobile Crash is an interactive installation where the direction of the users' movement within the performance space and the

action of pointing a finger plays a series of videos of technologic devices being smashed by a hammer (4). The Funky Forest is an “interactive ecosystem” where the growth of trees and the flow of water are dependent on the movements of children and the placement of moveable objects within the space (Figure 3) (5). Other installations try to incorporate their imagery into the physical space. Let’s Play with Time and the Weather was an installation at the 2009 Fête des Lumières in which the buildings were covered with projected imagery that caused the building to appear change shape, reacting to the effects of time and weather (Figure 4) (6). Out of Bounds was an interactive installation in which the beam of a flashlight allows the user to see through a small portion of a wall in the performance space (7). Our project used motion tracking to facilitate user interaction and alter the projected imagery, which was integrated with the physical performance space.



Figure 3: The Funky Forest (5)



Figure 4: Let's Play with Time and the Weather (6)

Technology Challenges and Methods Used

Motion Tracking

The process of measuring and recording three dimensional position and orientation of a person or object is generally referred to as *motion tracking* or *motion capture* (“mocap”). There are many different methods and technologies associated with motion capture, including optical,

inertial, acoustic, mechanical and magnetic approaches (8). The large size of our performance space limited our options. Magnet-based motion tracking doesn't work well over large areas because the field strength falls off relatively quickly. Acoustic methods also have size limitations, and don't tend to work well in places that echo. Mechanical approaches require the tracked object to be attached to a physical tracking apparatus, which would have limited the user experience. Consumer video game platforms including the Nintendo Wii and the PlayStation are releasing controllers that use inertial means of motion capture to track the physical motions of the users rather than using button presses or joystick movements. The Xbox Kinect takes an optical approach to motion capture, using cameras to gather depth information, which it uses to track the full body motions of multiple users. Given the size of our performance space and our desires for an interactive user experience, we chose to investigate inertial and optical systems.

The first motion capture system we considered was the MVN system by Xsens (9). It is a combination of several types of motion capture, using sensors containing accelerometers, gyroscopes and magnetometers to track position and orientation information. These sensors are wired together and worn as part of a Lycra suit and the data is transmitted wirelessly to a receiver (Figure 5, Figure 6). The entire system cost around \$75,000, with the sensors at about \$1,500 each. Using the sensors as part of a suit would limit audience interaction to one person at a time, with a long change over, thus effectively precluding audience participation. Instead, we considered using the sensors independently to track other objects, creating a more interactive audience experience. However, we would need to attach not only a delicate and expensive sensor to the object we wanted to track, but also the wireless transmitter. These restrictions meant that the MVN system was not a viable option for our project.

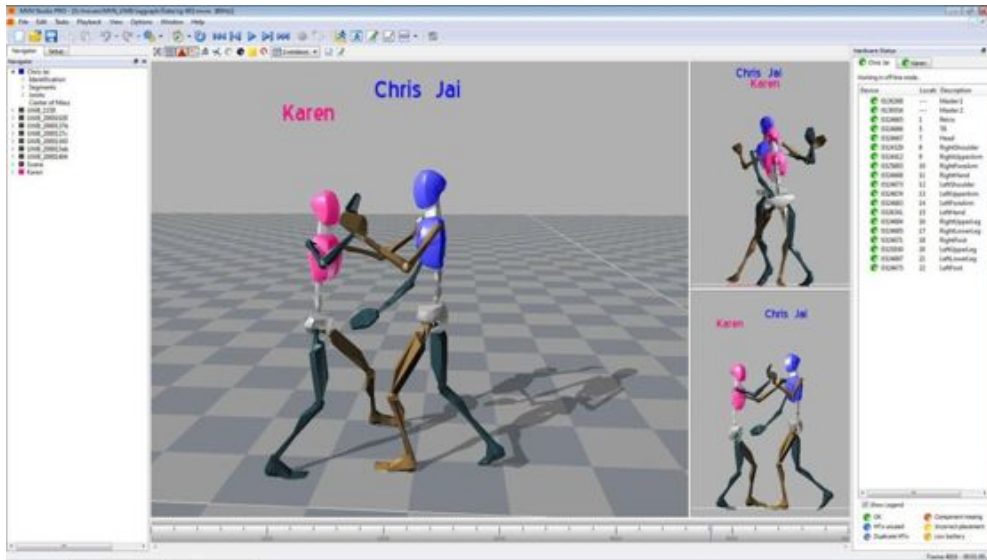


Figure 5: Software for MVN system (9)



Figure 6: MVN suit (9)

The other motion capture system we considered was the OptiTrack system by NaturalPoint (10). The system we needed was between \$10,000 and \$15,000, significantly less expensive than the MVN system. This was an optical system consisting of cameras and markers, rather than sensors. The markers were either small shiny foam balls or stickers, which we could easily attach to different objects. The cameras would emit bursts of infrared light and the retro-reflective markers would reflect this light back towards the camera. This allowed each camera to create a series of images with the markers appearing as bright white dots that were clearly distinguishable from the rest of the scene (Figure 7, Figure 8). The three dimensional area in which it was possible to do motion capture (the *tracking volume*) was where the fields of view of the cameras overlapped. In this space, the system could use images from multiple cameras to calculate the three dimensional position of the markers. Because of the precise nature of these calculations, the cameras and their tripods could not be moved or jostled without needing to recalibrate the system.

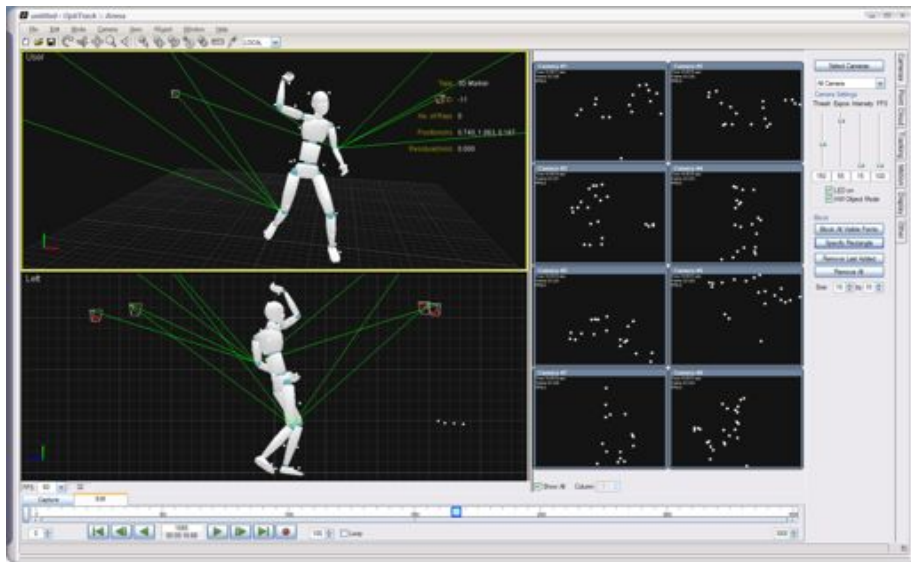


Figure 7: Software for OptiTrack System (10)



Figure 8: OptiTrack suit (10)

The system allowed grouping of markers into *rigid bodies*, but these rigid bodies need to have unique marker configurations for the system to be able to distinguish between them. It took us a few tries to find unique marker configurations that would allow the system to distinguish between our different objects. We created three rigid bodies, one for each of the objects we tracked: a fishing pole (Figure 12), an umbrella (Figure 14), and a mesh ball (Figure 15). The markers are clearly visible in Figure 14.

System architecture: modeling and rendering

We considered many different options for creating our imagery. MATLAB was briefly considered as a way to mathematically manipulate the motion capture data to generate abstract visuals (11). We also considered several Autodesk products, including Maya, which would produce beautiful complex imagery, but wouldn't usually work in real time (12). Additionally, they did not integrate well with the motion capture system. Max/MSP by Cycling '74 was another rendering option we considered (13). It allowed mathematical manipulation of data, as

well as image creation. It used a visual programming interface and had a large community of developers that openly shared their work for others to build upon. We eventually used Max/MSP to create a proof of concept where moving a motion-tracked sensor would cause an image of water to ripple.

When we decided that our installation would have gameplay-like components, rather than abstract imagery, we decided to consider options that allowed us to *model* our scene (place objects within the scene) in addition to *rendering* it (creating the image from the model). We considered game engines, in particular the Unity game development tool, for this purpose (14). It allowed the creation of an environment as well as individual objects and characters within the environment. It also allowed us to import models from Maya. Different characters could be controlled separately, so we could link each character to a particular physical object being tracked by the motion capture system. We also saw the potential of using event-based triggers to direct gameplay. We eventually chose Unity as our final solution because it also included several other features that helped us integrate it with the other pieces of our project. The ability to extend Unity's functionality with C based plugins gave us a way to import our motion capture data.

Unity also supported features designed for multiplayer gaming over a network, which allowed us to generate imagery of the same scene on multiple computers, something that was crucial for our immersive environment. Unity allowed networking with both distributed and centralized state. With the state distributed, individual instances can run faster, but it's challenging to keep the state consistent across all instances. Centralized approaches do not exhibit these problems, but they are slower (15). We decided that distributed state was best suited for our needs. A private local area network was set up to minimize contention for network

bandwidth and thus improve responsiveness. The final system used a single server that updated the state on four clients, each supplying imagery to a separate projector.

Display

There are several different kinds of standalone, relatively small, turn-key display systems that can be either purchased or built. A few classic examples include “caves” and “domes”. The CAVE (CAVE Automatic Virtual Environment), developed by researchers at the University of Illinois at Chicago, is a small room in which a user can stand surrounded by projected imagery on the walls and floor (16). The Showcase Dome, at the UNC RENCi Engagement Center, projects imagery on a large hemispherical dome that covers the peripheral vision of several users (Figure 9) (17). We wanted our display system to allow multiple users to view our imagery at once. Additionally, we wanted to integrate the imagery into the architecture of the building, rather than just projecting on screens. This meant that we needed to explore options beyond these turn-key systems.



Figure 9: The Showcase Dome (17)

The best way for us to create immersive environment in Gerrard Hall was to cover as much of the interior of the building as possible in imagery. The easiest way for us to accomplish this was using multiple projectors. We had a limited number of projectors at our disposal, and they had varying fields of view, brightness, and resolution. We did some initial planning and then visited the space before the festival and tried several different projector configurations to determine where to place the projectors, how to aim them to get the best coverage without distorting the images or overlapping the images from different projectors and how to avoid shadows from audience members. We eventually decided that given the available projectors, we would cover three walls and the floor with imagery (Figure 10). We put the projectors in the balcony on tall adjustable wooden mounts built by John Thomas; this kept the projectors out of harm's way and allowed us to project downward, casting audience shadows on the floor, rather than on the walls where most of the imagery would be projected. This caused us to project off-axis, but we were able to use the offset optics and the keystone correction on the projectors to compensate for the distortion this caused.

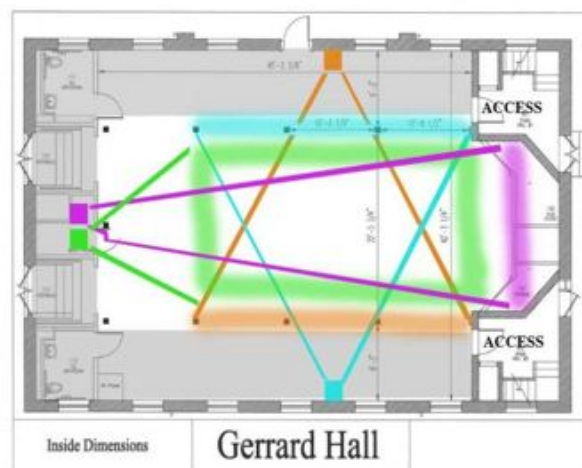


Figure 10: Projector Placement in Gerrard Hall

We considered using some form of *two-pass rendering*, where the desired image is rendered in the first pass and then altered in the second pass so that it looks correct when projected in the space (for another example of two-pass rendering, see the Office of the Future (18)). Using two pass rendering, we would have been able to render a black mask over areas of the imagery where the field of view of the projectors overlapped. While this approach would have eliminated the problem of projecting two different images on top of each other, it would have resulted in bright patches where the fields of view overlapped. We decided to simplify our computation using the built-in projector functionality, and then use physical masks on the projector lenses to eliminate the problem of bright patches (i.e. to blend the imagery).

Once we had determined the setup of the projectors, we had to determine the position and orientation of the *virtual viewing points* in the virtual environment. These points were used by the rendering software to generate images of the appropriate parts of the scene. The virtual and physical viewing points would need to align for the imagery to look correct in the space. We attempted to compute where these points should be based on measurements of the physical space, but, putting the virtual viewing points in the “correct” place did not always look right. In the end the best result was achieved by standing in the space and using the game development tool to nudge the points until they looked right.

Integration

The integration of the motion capture system and the rendering system proved to be a large technical hurdle. Initially, we wanted to share the motion capture data with several modeling programs which would modify or add to the data before passing it on to the rendering system (Figure 11). A fish flocking model would receive the position of a single tracked object

and generate the positions of many fish in a school that would follow the tracked object with flocking behavior (19). A water model would receive the positions of all the objects in the environment in order to generate caustics as well as surface waves. Unfortunately, we ran into time constraints trying to implement such a complex solution. However, we were able to use a dynamically linked library to gather tracking information from the OptiTrack system and make it directly available to Unity through a C based plugin.

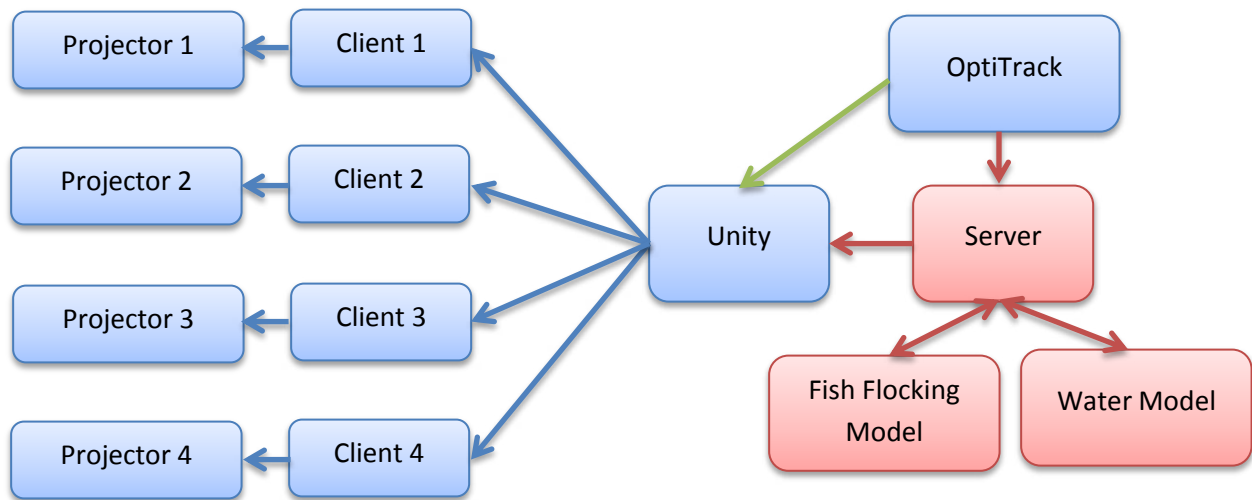


Figure 11: Project Architecture. Blue and red indicate the goal architecture; blue and green indicate the final architecture.

Results

Artistic Overview

The final installation was open in Gerrard Hall at UNC-CH for two to three hours a night every night of the CHAT festival, which lasted from Tuesday February 16th through Saturday February 20th 2010. When the audience arrived in Gerrard Hall, they saw what appeared to be the interior of the building. On the command of an operator, the opening animation sequence would begin. As the projected imagery changed, the building seemed to fill with water and the

walls fell away revealing an underwater landscape and the sounds of the ocean. Then the interactive portion of the installation began. We had three different objects for the audience to manipulate. The position and orientation of each object corresponded to the position and orientation of a particular sea creature. Each creature had a distinct melody that would play while the object was within the tracking volume. The umbrella corresponded to a manta ray, an octopus followed a mesh ball and a school of fish followed the tip of a fishing rod. On another command from an operator, the ending animation would begin. The water would begin to drain and the walls of the building would rise back into their original places.



Figure 12: Manta Ray and School of Fish (20)



Figure 13: During the Festival

Technical Overview

We set up six tripods holding two OptiTrack cameras each in the balcony of Gerrard hall. This setup made sure audience members didn't accidentally interfere with the carefully calibrated system. The system was aimed and calibrated so that the tracking volume was a few feet off of the ground on the first floor. We taped an outline of the capture volume on the floor and instructed the audience members to keep the objects inside the square. Some of this tape can be seen on the floor in Figure 14. We needed to carefully place the markers on our objects to ensure that they could be seen in this tracking volume by the cameras above. We used a mesh ball rather than a solid ball so the markers would be visible through the ball, and not occluded by the ball itself (Figure 15). We only placed markers on the top of the umbrella, because most people didn't try and turn this particular object upside-down (Figure 14). With the fishing rod we had to place the markers near the base because people would stand inside the taped square and wave the end of the fishing rod outside of the capture volume.



Figure 14: Manta Ray and umbrella with retro-reflective markers (21)

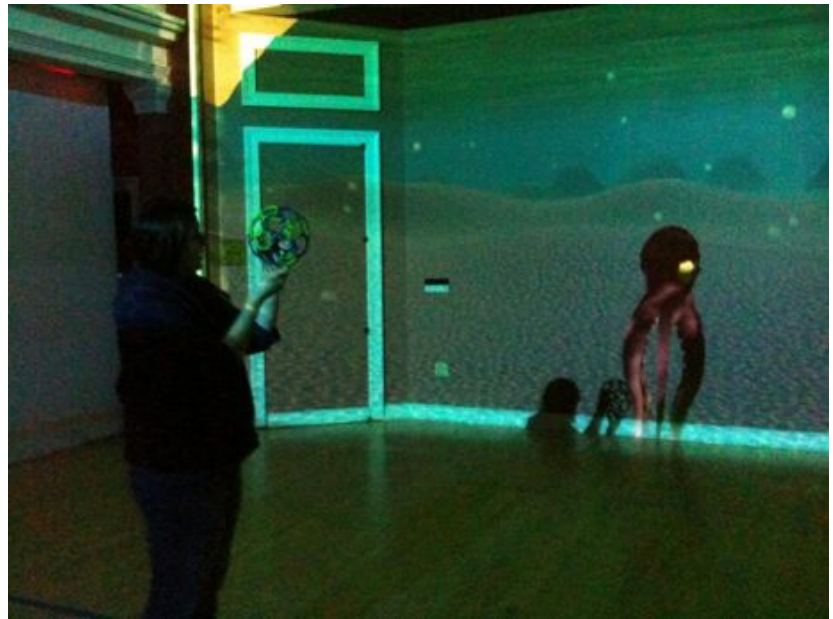


Figure 15: Octopus

We used a dynamically linked library to gather the motion capture information from the OptiTrack system, and passed it to Unity through a C plugin. The environment, the character models and the animations all existed within and were controlled by Unity. We were able to use Unity's built in features for distributed multiplayer networking to allow different computers, each with a synchronized copy of the environment, to generate projected imagery. We had four projectors, also in the balcony, covering three walls and the floor. We used a combination of the built in keystone correction on the projectors and paper masks on the lenses to correct for the off-axis projections and blending issues.

Discussion

The lack of clear definition of the tracking volume was one of our biggest obstacles. Sometimes audience members would not notice the taped square on the floor. Other times they would see the square and stand inside it while holding the objects outside of the capture volume.

The installation attracted a lot of children, but with the tracking volume calibrated to be a few feet off the floor, some of our younger audience members would hold the objects below the bottom of the capture volume. Future installations would benefit from adjusting the placement of the tracking volume.

When the installation had more than one or two people participating, audience members did not notice that each creature had its own melody. People also tended to manipulate each creature independently rather than attempting to bring the creatures closer together or try to get them to interact. By changing elements of the gameplay, like how close together the creatures need to be to trigger an event, future installations could improve the user experience.

What's next

In the future, the Bathysphere may be improved with more responsive characters and easier interaction between characters. The gameplay could be extended to allow for more complex interactions as well as multiple levels (and underwater level, a level on land, etc). The technological package could be refined further for use with similar applications that require motion tracking and an immersive environment.

The paradigm of multi-participant motion capture and interaction in an immersive environment could be used in many other ways as well. Using these technologies, we could create an illusion of different gravities, speeds, or motions. By projecting the imagery directly onto the space rather than onto screens, we move into the area of *Augmented Reality* (22). These possibilities would pose similar technical challenges of creating and displaying the imagery. Depending on the application, different types of motion capture or different motion capture systems might be more applicable.

Acknowledgements

We would like to thank the many people involved in creating and presenting the installation. Francesca Talenti and Greg Welch were the principal investigators on the project. Tao Li helped with the motion capture. Zhen Shao designed the client/server architecture. Chiun-Yi Tseng implemented the fish flocking model. Mingsong Dou and Ying Shi created the water model. Darryl Dunn composed the music. Madison Bullard incorporated the audio into the game environment. Marc Russo modeled the character and created the character animations. Matt Harris and Charlie Hicks created the model of Gerrard Hall. Stephen Roller implemented the gameplay and modeled the environment. Bil Hays, John Thomas, and Herman Towles helped with the logistics of moving into the performance space and setting up the network. This work was supported by a UNC-Chapel Hill grant “Scholarship, Creative Activity or Research in the Humanities and Fine Arts” (Talent and Welch). It was also supported in part by an NSF CRI:IAD equipment grant “Integrated Projector-Camera Modules for the Capture and Creation of Wide-Area Immersive Experiences” (Welch & Fuchs et al.). This work was also supported in part by a gift from Mitsubishi Electric Research Lab (MERL), Cambridge, Mass.

Works Cited

1. **Institute for the Arts and Humanities.** *C.H.A.T. A Digital Arts & Humanities Festival*. [Online] [Cited: December 5, 2010.] <http://www.chatfestival2010.com/>.
2. **Carolina Performing Arts.** Carolina Performing Arts. *Gerrard Hall*. [Online] [Cited: December 12, 2010.] <http://www.carolinaperformingarts.org/aboutus/gerrard>.
3. **Museum of Life and Science.** soundSpace: Hear Motion. *Museum of Life and Science*. [Online] [Cited: December 5, 2010.] <http://ncmls.org/visit/campus-and-exhibits/exhibits/soundspace>.
4. **Bambozzi, Lucas.** Lucas Bambozzi. *Mobile Crash*. [Online] [Cited: December 11, 2010.] <http://bambozzi.wordpress.com/projetosprojects/mobile-crash/>.

5. **Watson, Theo.** Funky Forest Moomah Edition. *TheoWatson*. [Online] [Cited: December 5, 2010.] http://www.theowatson.com/site_docs/work.php?id=45.
6. **Fetes des Lumieres.** Fete des Lumieres. *Let's Play with Time and the Weather*. [Online] [Cited: December 11, 2010.] http://www.fetedeslumieres2009.lyon.fr/lumieres/sections/en/event_view?id=643.
7. **O'Shea, Chris.** Chris O'Shea. *Out of Bounds*. [Online] [Cited: December 11, 2010.] <http://www.chrisoshea.org/out-of-bounds>.
8. *Motion Tracking: No Silver Bullet, but a Respectable Arsenal.* **Welch, Greg and Foxlin, Eric.** 2002, IEEE Computer Graphics and Applications, pp. 24-38.
9. **Xsens.** Xsens MVN - Inertial Motion Capture. *Xsens*. [Online] [Cited: December 5, 2010.] <http://www.xsens.com/en/general/mvn>.
10. **Natural Point.** *OptiTrack Optical Motion Capture Solutions*. [Online] [Cited: December 5, 2010.] <http://www.naturalpoint.com/optitrack/>.
11. **The MathWorks, Inc.** MATLAB - The Language of Technical Computing. *MathWorks: Accelerating the pace of engineering and science*. [Online] [Cited: December 5, 2010.] <http://www.mathworks.com/products/matlab/>.
12. **Autodesk, Inc.** *Autodesk*. [Online] [Cited: December 5, 2010.] <http://usa.autodesk.com/>.
13. **Cycling '74.** Max. *Cycling '74*. [Online] [Cited: December 5, 2010.] <http://cycling74.com/products/maxmspjitte/>.
14. **Unity 3D.** [Online] [Cited: December 5, 2010.] <http://unity3d.com/>.
15. **Neumann, Christoph, et al., et al.** Challenges in peer-to-peer gaming. *ACM SIGCOMM Computer Communication Review*. January 2007, Vol. 37, 1.
16. *Surround-screen projection-based virtual reality: the design and implementation of the CAVE.* **Cruz-Neira, Carolina, Sandin, Daniel J. and DeFanti, Thomas A.** s.l. : ACM, 1993. Proceedings of the 20th annual conference on Computer graphics and interactive techniques (SIGGRAPH '93). pp. 135-142.
17. **RENCI.** Showcase Dome. *Renaissance Computing Institute at UNC Chapel Hill*. [Online] [Cited: December 5, 2010.] <http://unc.renci.org/resources/visualization-resources/showcase-dome>.
18. *The Office of the Future: A unified Approach to Image-Based Modeling and Spatially Immersive Displays.* **Raskar, Ramesh, et al., et al.** Orlando : ACM, 1998. Proceedings of the 25th annual conference on Computer graphics and interactive techniques (SIGGRAPH '98). pp. 179-188.
19. *Flocks, herds and schools: A distributed behavioral model.* **Reynolds, Craig W.** New York : ACM, 1987. Proceedings of the 14th annual conference on Computer graphics and interactive techniques (SIGGRAPH '87). pp. 25-34.

20. **Talenti, Francesca.** *Francesca Talenti: Media Matador.* [Online] [Cited: December 5, 2010.] <http://www.ibiblio.org/talenti/>.
21. **Institute for the Arts and Humanities.** C.H.A.T A Digital Arts and Humanities Festival. *Photos from Faculty Exhibitions.* [Online] [Cited: December 12, 2010.] <http://www.chatfestival2010.com/faculty-exhibitions-photos.html>.
22. *A Survey of Augmented Reality.* **Azuma, Ronald.** 4, s.l. : MIT Press, 1997, Presence: Teleoperators and Virtual Environments, Vol. 6, pp. 355-385.
23. **Institute for the Arts and Humanities.** The Bathysphere: Motion Capture as Art. *C.H.A.T. A Digital Arts & Humanities Festival.* [Online] [Cited: December 5, 2010.] <http://www.chatfestival2010.com/bathysphere.html>.
24. —. *Institute for the Arts and Humanities.* [Online] [Cited: December 5, 2010.] <http://iah.unc.edu/>.
25. —. Collaboration, Innovation, Conversation: What We're All About. *Institute for the Arts and Humanities.* [Online] [Cited: December 5, 2010.] <http://iah.unc.edu/news/newsarchive/2010/chatfestival/?searchterm=CHAT>.