Tangible Comics:
A Performance Space with Full-body Interaction

Özge Samanci, Yanfeng Chen, Ali Mazalek
Synaesthetic Media Lab, GVU Center
Georgia Institute of Technology
Atlanta, GA 30332 USA
+1 (404) 457-0961
{gth760k, ychen, mazalek} @gatech.edu

ABSTRACT
Tangible Comics is a computer vision based full-body interactive storytelling environment that also functions as a comics generator. Prevailing applications of full-body computer vision have not utilized the full storytelling or performance potential of these environments. Our aim is to produce an environment that can create a space for redefining the conventions of comics, performance, film, photography, and animation. In relation to that, we are exploring the design problems that can arise when computer vision technology is contextualized in an interactive story telling environment.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces - input devices and strategies, interaction styles; I.5.4 [Pattern Recognition] Application - Computer Vision; J.5 [Arts and Humanities]: Fine arts, Performing arts.

General Terms
Design, Human Factors

Keywords
Tangible interface, direct manipulation, interactive video installation, full body interaction, gesture-based interaction, computer vision, interactive storytelling.

1. INTRODUCTION
Tangible Comics is an attempt to redefine the comics and animation conventions for a narrative performance in a full-body interaction environment. It seeks ways of taking digital comics out of the computer screen and merging comics and animation conventions into three-dimensional real space. This installation utilizes the qualities of digital games, interactive narration, digital art, performance, and augmented reality. The installation has been designed with narrative, conceptual, and interaction/technical goals as follows:

ACE'07, June 13–15, 2007, Salzburg, Austria.

1.1 Narrative Goals
- To design a computer vision environment for storytelling purposes rather than for a momentary-attention grabbing event.
- To utilize the environment as a comics generator by periodically taking snapshots of the happening.
- To create a dynamics of collaboration and rivalry between the performer and the audience by giving enough agency to the performer and audience.

1.2 Conceptual Goals
- To provide self-identification with the performer's own self-projection by using extreme abstraction.
- To explore ways of producing new conventions and a new vocabulary for a set of art forms, especially for comics, performance, film, photography, and animation when they are contextualized in a full-body interaction environment.
- To alter the classical male-centered discourse of reproduction narratives such as active sperm-passive egg.

1.3 Interaction / Technical Goals
- To create an interaction space for a performance-based activity without using wearables or instrumenting the performer in any way.
- To enable the performer to understand the interaction pattern without using explicit textual or audio explanations.
- To create short, immersive and physically active experiences by combining virtual and real objects.

Even though there are few examples such as *The KidsRoom* [1] we realized that the majority of prevailing computer vision installations have not been designed for storytelling purposes. We aim to contribute to the closing of this gap and to establish conventions for storytelling with full-body interaction. This paper describes the Tangible Comics installation, including the technology used, the design problems identified and the solutions developed.

2. RELATED WORK
While investigating the prior art which utilizes computer vision technology, we came across a set of installations that are not designed for storytelling. The majority of computer vision based art creates instantaneous, non-narrative, entertaining or challenging single-gag-based inventive moments around a concept. They resemble single frame caricatures with a momentary quality. Worthington's *Shadow Monster* [11], Khan's *Finger Prints* [3], Hieronymi's *Pillow Paint* [2], Levin and Lieberman's *Messa Di*

Voce [5], Snibbe's *Body, Screen and Shadow* [10], Lozano-Hemmer's *Standards and Double Standards* [6], Rokeby's *Watched and Measured* [9], Moeller's *Cheese* [8] and *Electronic Mirror* [7] all take the direction of using computer vision for non-narrative interaction. Among all art work that we looked into, only Bobick's *KidsRoom* [1] attempted to use the storytelling potential of computer vision based full-body interaction environments. The target participants of *KidsRoom* are children. We are interested in developing a storytelling environment that is functional both for children and adults. Moreover, In *KidsRoom*, if children felt confused about the interaction pattern, a teacher-like voice would whisper the instructions to them, such as "stay on the rug". Even though this direction is very functional for an application that targets children, we tried to avoid explicit explanations during the participant's exploration of the story world.

3. ENVIRONMENT SETUP

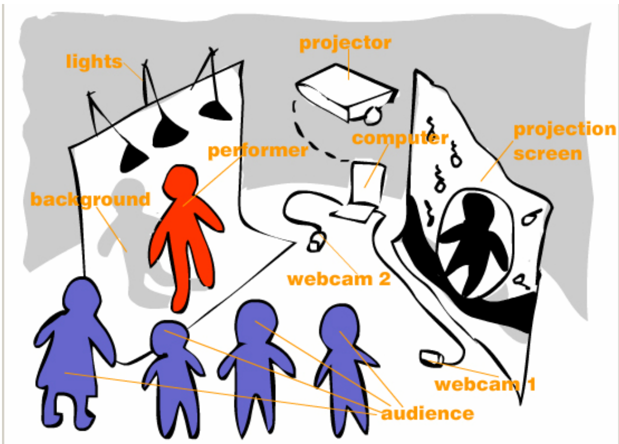

Figure 1. Tangible Comic Environment Setup Plan

Figure 1 shows our system's running environment. The environment consists of two white screens (one for background, one for projection), a set of lights, two webcams, one computer, and a projector. The first webcam captures the performer's actions. The second webcam captures the audiences' actions. Additional webcams can be added to capture the performer's action from different angles of view. The performer's image is distorted, combined with virtual objects via Java code, and reflected to the projection screen. Eventually, the acoustic and kinetic participation of the audiences will have an impact on the performer's distorted projection. Also, the audience's hands and arms will be projected to screen. As the performer moves in the real space, he or she manipulates the virtual objects surrounding him or her.

We use a background screen and a set of lights to create a bright background behind the performer. By adjusting the webcam to use a relatively fast shutter speed, the performer will look quite dark in the camera, regardless of what kind of clothing s/he is wearing. At the same time, the background screen is almost pure white in captured images, which creates a sharp contrast between the background and the performer (figure 2). This kind of setup helps reduce the computing burden and interference in background subtraction as well as blob detection.

Figure 2. Environment Setup – Background

The interaction design aims to supply an arc of excitement/revelation by utilizing and redefining the aesthetics of the comics/animation medium. The participants' simultaneous distorted self-projections (their silhouettes) are the characters of this dramatic experience. The participants' masked presence on the screen triggers a set of decision points as a consequence of the participants' gestures tied to virtual objects. While the participants are experiencing the narration, they are both the object and the subject of their own gaze. At the end, periodically taken snapshots of the happening supply a traditional comic strip version of the entire experience. Overall, regardless of the participants' ability to draw or design a comic strip, the experience turns out to be a comics generator by supplying an improvisational space.

4. THE STORY

The story of Tangible Comics brings a feminist perspective on procreation. The participants, regardless of their gender, play out the story of the female egg in the reproductive system. Each participant, by collaborating with the audience or acting as a rival to them, tries to reach the fertilization site and select the best profile among the sperm that are rushing to the egg. During this journey, the performer can make use of a real stopper, an umbrella, and a flashlight. This journey is described in the text below, illustrated with storyboard images.

4.1 Discovery of the story location and journey from ovary to fallopian tube

The interaction starts when the performer enters the performance space. The performer sees his or her distorted silhouette standing on a soft-jiggling ground.

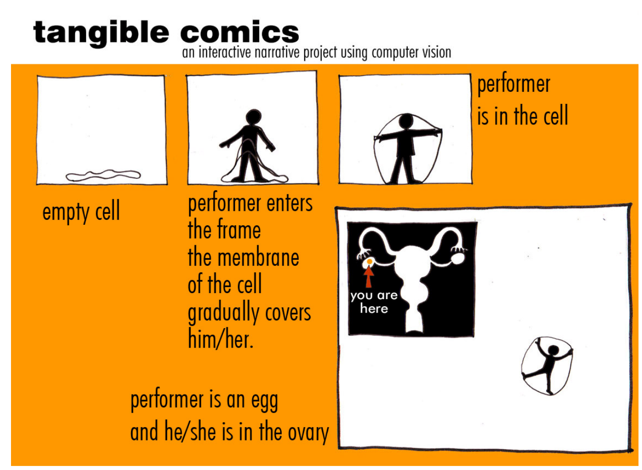

Figure 3. The Performer in the Egg and Navigation Map

The performer is then gradually covered by the membrane of an egg. A navigation map appears, enabling the performer to comprehend their location. The performer is actually in the reproduction system of a female body. The arrow on the navigation map points to the ovary with the explanation "You are here." At this point, the performer needs to jump from ovary to fallopian tube.

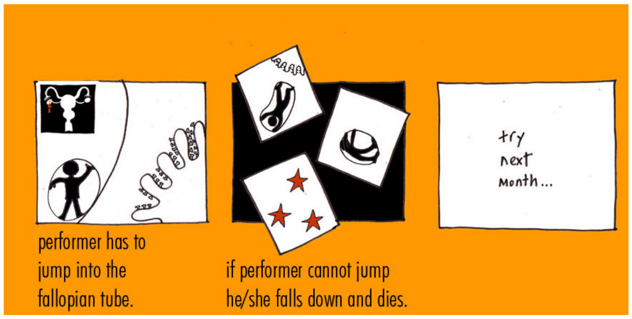

Figure 4. A Jump from Ovary to Fallopian Tube

The accomplishment of this step requires an actual physical jump. The performer should jump far enough until their projected image falls into the borders of the fallopian tube. Otherwise the performer will see themselves fall down, and will need to try "next month." Since "try next month" is a self-conscious gag based on the verisimilitude of the environment, the performer can immediately start his or her interaction from the very beginning. In this context, the jiggling-soft ground appears as a very fragile tissue. The performer cannot stand straight on the same location for a long time, and instead must move around constantly. Otherwise his or her weight will tear apart the tissue and the performer will see him or herself fall down.

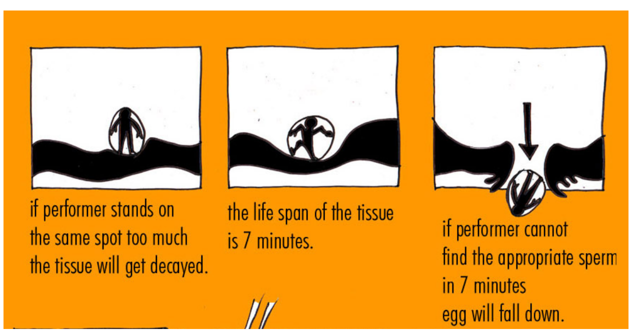

Figure 5. Fragile Tissue

The implementation of the falling down scenes is based on prevailing comics conventions. The snapshots of the performer will be scaled down, tangled or turned upside down, and placed into dynamic frames with stars that represent the plummet.

4.2 From the fallopian tube to the site of fertilization

The performer who has succeeded in jumping into the fallopian tube needs to travel to the fertilization site. At this point, the audience, by waving their arms, can hold, roll, and move the egg forward.

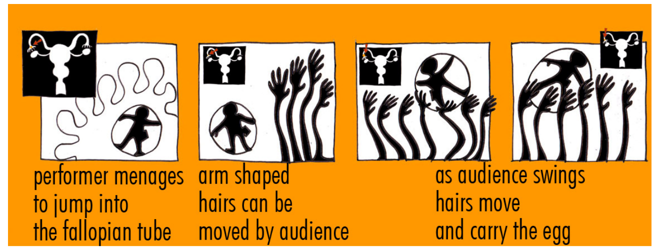

Figure 6. Audience Interaction

The distorted image of audience's arms and hands will be projected to the screen. The projected image of the audiences' waving arms and hands will resemble the hairy structures in an actual fallopian tube. As soon as the egg arrives to the site of fertilization, a set of sperm rush together to the egg.

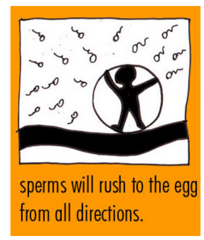

Figure 7. Rushing Sperms

If the egg tries to escape, the sperm follow close behind. The performer can hold any sperm with his or her hand and see the possible child that the egg and that particular sperm can produce.

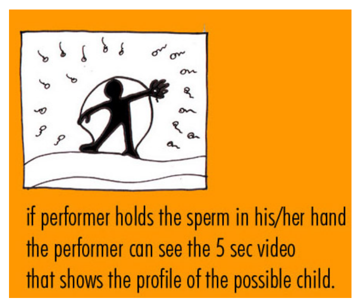

Figure 8. Seeing the Profile of the Possible Child

By checking the sperm one by one, the performer can find out the most appropriate option and pull the decided sperm into the egg. In this case, the performer sees the profile of the possible child as a grown up. This action completes the interaction and another performer may come up from among the audience.

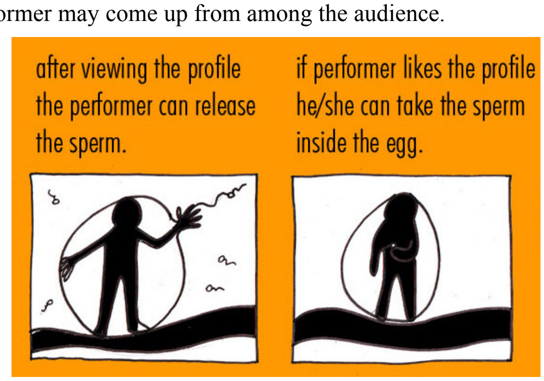

Figure 9. Releasing a Sperm or Taking the Sperm in

A sperm can also penetrate the membrane if it touches the membrane for more than 5 seconds. If the performer lets a sperm penetrate the membrane, the performer will not be able to make a choice.

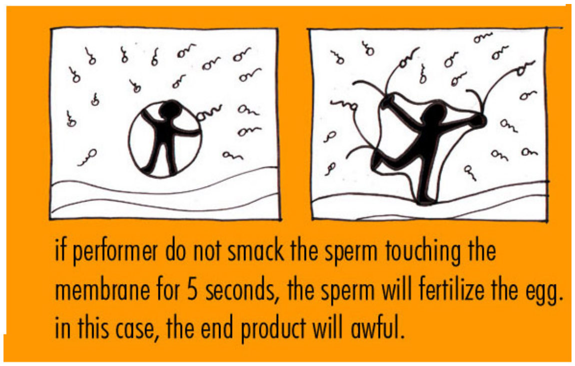

Figure 10. Smacking and Kicking the Sperms

In this case, the sperm makes the choice and completes the interaction. In order to prevent this unwanted penetration, the performer should smack and kick the rushing sperm. Sperm will bounce back if the performer hits them. If the performer holds one sperm, the other rushing sperm will suspend in the air and let the performer view the animation that exhibits the profile of a possible child.

4.3 Possible dangers

While the egg is at the site of fertilization, an artificial insemination tube may appear and try to suck the egg in.

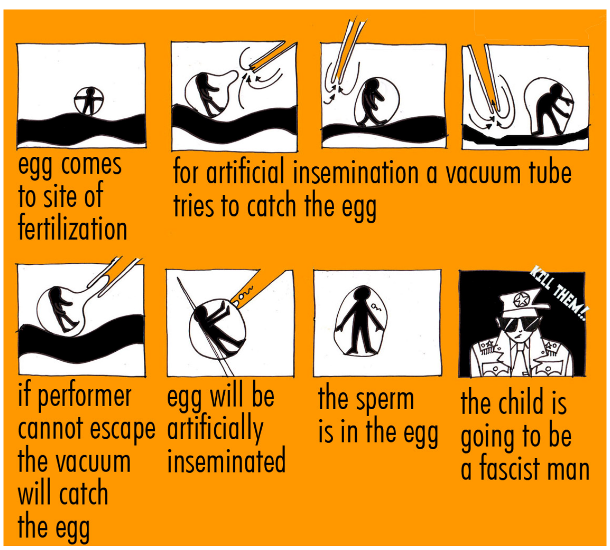

Figure 11. The Danger of Being Sucked in to an Artificial Insemination Tube

If the performer cannot escape from the tube, it will suck the egg in and the egg will be artificially inseminated. Since this process does not contain a selection action taken by the performer, the end product of this event will give a profile of a person who has fascistic traits. The performer can take a real physical stopper from their environment and try to plug the entrance of the virtual artificial insemination tube.

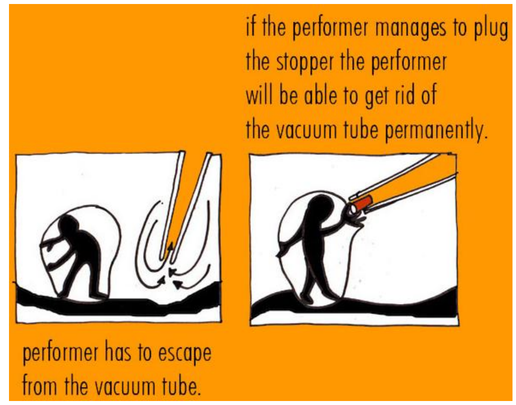

Figure 12. The Use of First Object, Stopper

If the performer manages to plug the physical stopper into the projected virtual tube, they will eliminate the tube permanently. Otherwise, the danger of being sucked into an artificial insemination tube will reappear from time to time. Audiences can increase the number of sperm in the environment by clapping their hands and thereby increase the challenge for the performer. If the number of sperms becomes enormous, the environment will automatically send the sperm killer rain. This virtual chemical fluid descends from the top of the frame and kills the sperm it encounters. In order to prevent the loss of all sperm in the environment, the performer can take a real physical umbrella from their space, open it up, and protect some of the sperm by covering them with the umbrella.

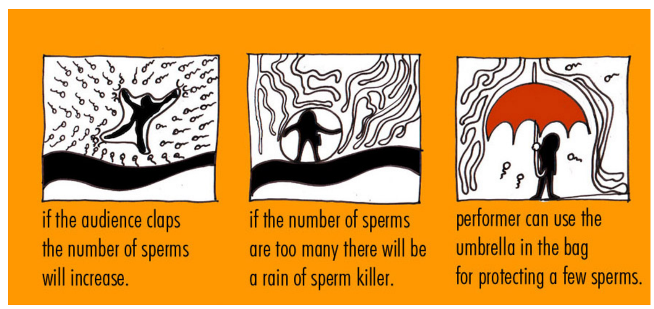

Figure 13. The Use of Second Object, Umbrella

The performer can also use a real flash light to deactivate some of the rushing sperm. Sperm that face the light beam become suspended in the air like a deer in headlights.

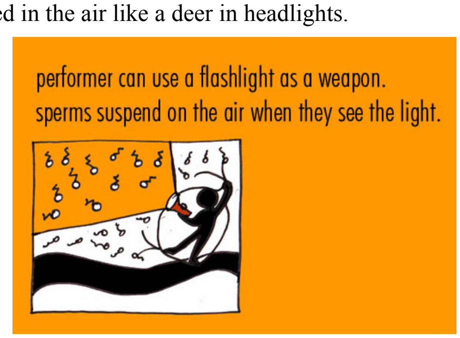

Figure 14. The Use of Third Object, Flashlight

5. HUMOR STRATEGIES AS COMMUNICATION TOOL

The humor strategies that we have developed for Tangible Comics aim to jolt the established male-centric discourse of reproduction narrations. In these conventional narrations, the human egg has been represented as a stable, slow, passive, and conceiving target, while the sperm take the role of aggressive-active agents who can compete with each other and make the selection. We aimed to alter this classical sexist discourse by blurring it with humor strategies. The performer can take the egg's point of view regardless of their gender. Also, in this narration, the egg holds the agency. It jumps, floats, takes a journey and finally makes the selection. The performer's ability to smack the sperm alters the established hierarchy between egg and sperm.

Figure 15. Profile 1 - Child Version

The animated profiles of possible children establish the second level of our humor strategies. The distinction between the child and adult versions of a profile will highlight and play with the human tendency towards stereotypical thinking. Most of the time, the performer will be disappointed or surprised when they see the result of their choice. The beautiful, blond, white, well dressed, healthy, cute, playing girl will be a good target for selection (fig. 15). However, if the performer falls in the trap and selects this profile, the grown up version of the same child may appear as an alcoholic young woman who has an enormous credit card debt and is asking her mother (the performer) to cover it.

Figure 16. Profile 1 – Adult Version

As another example, a naughty boy who is about to put his cat into washing machine initially appears as a bad target, but his grown up version exhibits a seemingly intelligent young man who has invented the "gravitron" ride for amusements parks.

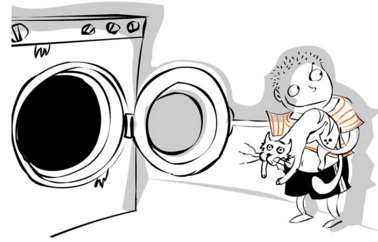

Figure 17. Profile 2 – Child Version

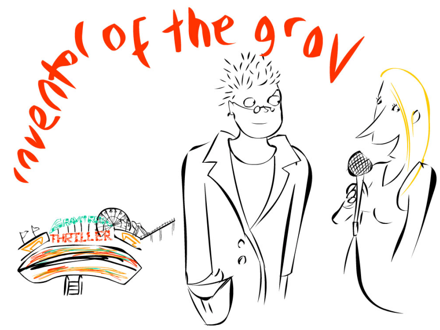

Figure 18. Profile 2 – Adult Version

Although these examples of an alcoholic blond woman in debt and a smart inventive man reveal another form of stereotypical thinking about man and woman, they do no seem like stereotypical cases when viewed in the context of the numerous profiles collected in the database. Overall, the database reflects a broad range of character types, both male and female.

6. IMPLEMENTATION

We use webcams to capture the user's body movement. Our system analyzes video streams frame by frame. After being processed by computational visual algorithms, the images are then converted into an internal model, reconstructed and finally blended into the current scene of the story.

The software is written in Java, and includes modules handling image capture and processing, body part recognition, interaction logic, 2D rendering and a game/story engine.

Various computer vision technologies have been employed in the image processing and body part recognition. We applied the following algorithms in sequence to each single frame captured by the camera (Fig. 19):

Figure 19. Original image captured by webcam

1. Background subtraction: used to separate the performer from the background and obtain a clean silhouette of the performer. Because we used static lighted white background,

we can assume that the lighting condition and background do not vary over time. The current camera image is compared to the background captured apriori. The pixel is determined to be part of silhouette if the difference is over a preset threshold;

2. Blob/edge detection: used to convert the shadow blob into vector data for subsequent processing (Fig. 20).

Figure 20. Blob Detection

3. Curvature analysis and image abstraction: used to analyze the shape of the human body to determine key vertices (Fig. 21). Curvatures are continuously computed based on the spatial relationship among neighboring pixels along blob edge;

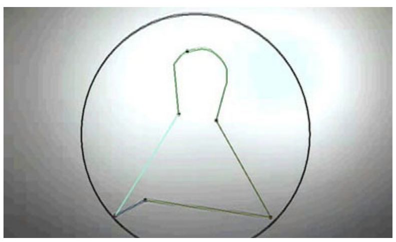

Figure 21. Image Abstraction

4. Noise reduction: in order to simplify and speed-up curvature analysis, we get rid of the high frequency variation along blob edge (noise) and reduce other insignificant detail in blob shape before curvature analysis.
5. Heuristic body-part identification: used to determine body parts based on a heuristic model of the human body.

7. INTERACTIONS

An exciting property of movement and gesture based input is that it can provide a new kind of user experience in digital interactions. By combining different parts of the human body (head, body, arms, legs, hands, feet, fingers), and various degrees of freedom, there is an enormous number of possibilities in how the user can make an input. However, in real world situations, not all of them are easily detectable and/or make sense in particular contexts.

To accomplish the storytelling and interaction goals of Tangible Comics, we needed to look at the natural movements that people make in everyday life, particularly those with relevance to the unfolding narrative: jumping, hitting/knocking away, pushing, reaching, grabbing, pulling, kicking, etc. We implemented these by combining various standardized input variables from image processing and recognition. In addition, some non-technological everyday artifacts can also be used to enhance interaction.

7.1 2D Positioning

2D positioning is one of the most common methods of interaction in traditional input devices, such as the mouse or light pen. By moving a mouse or pointing a light pen in a different direction, the computer responds by moving a cursor on the screen. Combined with other buttons or modifier keys, it can also accomplish more complex tasks. There are two possible implementations of positioning: absolute (e.g., light pen or tablet) and relative (e.g., mouse or track ball).

In full body interaction, the central task of implementing 2D positioning is body part identification. If we can identify different parts of the human body such as head, feet and hands (or fingers if resolution is high enough), we can combine them to provide multiple sources of positioning input. A typical example is bi-manual input – using two hands to manipulate elements on the screen. If finger recognition is feasible (camera resolution is high enough and/or hand is wide-open), we can simulate multi-touch, as in some touch-screen input devices.

Although both methods (absolute or relative) of positioning are possible in our implementation, they are not suitable in all situations. In our context, the human body is reconstructed on the screen, and the topology and scaling of the human body does not change during the interaction. So absolute positioning will be suitable for most performer interactions. However relative positioning, which can cause some body parts to look disproportionate or disjointed, can still be employed for certain special effects based on the context, e.g. an extensible arm or a flying fist. These effects could be used to achieve certain types of comics conventions, such as those which exaggerate characters' body parts (e.g. when a comics character punches something, and the fist flies out farther than a normal human fist would).

2D positioning can also be combined with other input methods to simulate some actions like hitting, bumping, catching, grabbing, pulling and etc.

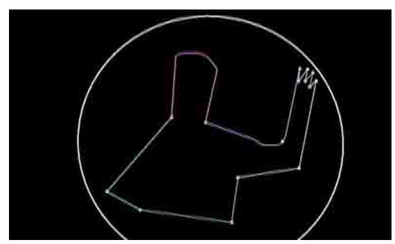

Figure 22. Finger Detection

7.2 Button Input

Unlike traditional input devices, with full-body interaction it is difficult to obtain on/off states like with mouse buttons or keys, because it is hard to find intuitive on/off data from silhouette analysis. In our context, this is important because sometimes we need to distinguish between plain movement and grabbing/dragging an object.

In order to implement dragging by hand, we added an open/close status of the hands to simulate button state input. This is a natural way to implement grabbing since in order to grab an object, a human first needs to open his or her hand then close it.

7.3 Interaction with On-Screen Elements

Similar to the tool bar in WIMP interfaces, we place images or icons of interactive items on screen so that the user can choose one to pick up. They can be presented as hanging in the air, laying on the surface, or even held by a non-player character.

Based on the way they are placed in the scene, users need to take different actions to acquire the items.

When an item is picked up, the user can then use it to perform special jobs based on the story context. The user can also drop it by reopening his/her hand. For example, in our story the sperm are interactive items that the user might want to select, hold, drag or reject based on the displayed character profile.

7.4 Gesture

In a WIMP environment, the user can sometimes maneuver the mouse cursor in a special way (i.e. making a gesture) in order to perform a certain task without selecting a tool. The same rule can apply to full-body interaction. Sometimes it is not feasible to display all of the possible interactive elements in the scene for the user to pick up. For most foreseeable tasks, we allow the user to perform them simply by making a gesture with their hands, fingers or full body. For example, a user can gesture a cross to wipe unwanted objects or writing from the wall. In our story, if the user cannot find the actual physical umbrella to open when the killer rain comes, they can instead obtain a virtual umbrella simply by holding their fist high and then pulling it down violently. Since gestures are internally defined and not necessarily obvious to a first-time user, floating hints could be provided at appropriate times in the story, so that users could learn how to make the necessary gestures when needed. Nevertheless, we strive to use natural human gestures from everyday life, so that users can discover them on their own simply by responding in a natural way to the story events on screen.

7.5 Velocity-based Input

In the physical world, the speed of his or her movements affects a person's interactions with surrounding objects. For example, if a person is trying to reach for something, but accidentally hits it too fast and hard, the target will be knocked away. If they act subtly, they should instead be able to reach and grab hold of the object. A similar rule also applies to creatures resting nearby: quick actions would most probably scare them away. Velocity is measured by comparing position of body parts frame by frame.

8. PARTICIPANT FEEDBACK

We have demonstrated our project as an installation piece during two demo open houses, and received a variety of feedback from users. We also presented Tangible Comics as a technology-based humor project at the American Humor Studies Association Conference in New Orleans in December 2006. The sense of humor grounded in the story world was easily understood and well received. Participants gave enthusiastic reactions to the following ideas: the possibility of sharing the egg's point of view regardless of their gender, the opportunity of selecting a possible child profile and the surprise in seeing the corresponding adult profile, the interaction pattern based on kicking and smacking sperm, and the use of tangible objects such as stopper, flashlight, and umbrella in unexpected ways. Seeing the image of the participant fall down followed by a flashing "try next month" sign, and the result of a participant being sucked into an artificial insemination tube were the two moments that resulted in the most laughter from the audience.

In terms of demographics of the participants, women and younger generations showed more enthusiasm for participation. A motivating common feedback from users was that they felt the piece was quite playful and immersive. Participants enjoyed exploring their spatial freedom. Some young participants jumped, crawled on the floor, and entered the scene by running. Some groups tried to break the single participant rule and embodied the egg with two people, collaborating to smack away the sperm. These two-person-in-the-egg scenes gave us ideas for future directions, since we had not anticipated multi-performer interactions. We are considering developing gags based on twins and triplets as a consequence of such cases. Producing animations as a procedural automated process was another suggestion regarding the current flash animations that we use. Since procedural animations may have a cold look, we are hesitant about pursuing this direction. From an aesthetic perspective, certain users said they would like the egg and silhouette to have a more fluid and organic visual look.

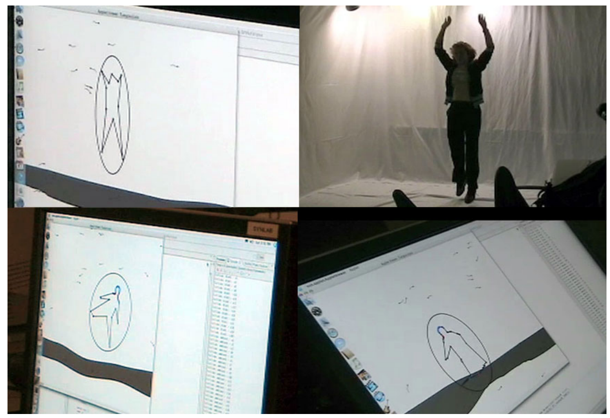

Figure 23. Recorded Performance

Some users were interested in how the system was implemented from a software and hardware perspective. Since our system captures and processes a live video stream in real time, they generally expected that it would require a fast processor and hardware-based video acceleration. Actually, the demo system ran on a single-CPU laptop with no significant video acceleration or special hardware. When experiencing the 2D line-drawing animations, most people felt it was quite responsive and smooth even though the frame rate was only 15 frames per second. Since the system discards most image detail and only handles silhouette recognition, the image processing does not take a lot of CPU time. We are thus able to handle a 320x200x15fps video input stream with the hardware mentioned above, which is sufficient for the Tangible Comics application.

9. CONCLUSION AND FUTURE WORKS

In this paper, we have presented Tangible Comics – an interactive storytelling video installation. The system uses computer vision based full-body tracking technology for interactions in a narrative piece. The platform provides a rich set of interactions, and supports simultaneous multi-point and multi-user input. Thus far, we have demonstrated how a performer can play-out the humorous story in front of an audience. We are still implementing the functionality to provide performers with a generated comic-strip version of their performance at the end, and there are also some scenes of audience interaction in the storyboard that remain to be implemented, such as the waving of the arms to help the egg move forward at the beginning. In response to participant feedback, we also plan to extend the story to allow multiple performer scenarios. Overall, the piece has proven to be an engaging and humorous story experience for performers and audiences alike, and we are motivated to try other narrative scenarios in our system as well.

10. ACKNOWLEGEMENT

We would like to thank Janet H. Murray for her advice during the development of story and interaction methods. We would also like to thank the members of the Synaesthetic Media Lab at Georgia Tech for their input.

11. REFERENCES

[1] Bobick, A. and et al. The KidsRoom: A Perceptually-Based Interactive and Immersive Story Environment. In *PRESENCE: Teleoperators and Virtual Environments, 8(4)* (Aug. 1999), 367-391.

[2] Hieronymi, A., Pillow Fight. *Web: http://classes.design.ucla.edu/Winter04/256/projects/andrew/body.html*, 2004

[3] Khan, O. Fingerprints. *Web: http://www.osmankhan.com/fingerprints/index.html*, 2003

[4] Konrad, T., Demirdjian, D., and Darrell, T. 2003. Gesture + play: full-body interaction for virtual environments. *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003).

[5] Levin, G. and Liberman, Z. Messa di voce. *Web: http://www.tmema.org/messa/messa.html*, 2004

[6] Lozano-Hemmer., R. Standards and Double Standards. *Web: http://www.lozano-hemmer.com/video/basel.html*, 2004

[7] Moeller, C. Electronic Mirror. Interactive computer installation at the Institute for New Media, Frankfurt 1993. *Web: http://www.christian-moeller.com/display.php?project_id=48&play=true*

[8] Moeller, C.. Cheese. Video installation at the Williamson Gallery, Art Center Pasadena, 2000. *Web: http://www.christian-moeller.com/display.php?project_id=36&play=true*

[9] Rokeby, D. Watched and Measured. Installation in Science Museum, London, UK, 2000. *Web: http://homepage.mac.com/davidrokeby/wm.html*

[10] Snibbe, S. Body, Screen and Shadow. In *San Francisco Media Arts Council (SMAC) Journa*l (San Francisco, CA, Jan 2003)

[11] Worthington, P. Shadow Monster. *Web: http://www.worthersoriginal.com/viki*, 2005