Skin: an interactive hyperstereoscopic electro installation

Helen-Nicole Kostis[*a], Robert Kooima[a], John Kannenberg[b]

[a]University of Illinois at Chicago; 851 S. Morgan St. Rm 1120 SEO, Chicago, IL USA 60607-7053;
[b]Stasisfield.com, P.O. Box 257451, Chicago, IL 60625

ABSTRACT

It is the uniqueness of Virtual Reality as a medium that calls for the creation of hybrid realities which blur the finite boundaries between physical and digital existence. Virtual Reality's distinguishing features as an artistic medium embody a distinct form of aesthetics: it is a stereoscopic, immersive, interactive, performative, dynamic, and experiential medium. A Virtual Reality art piece manifests in multiple ways. It can present itself as an interactive virtual archetype, exploring concepts rendered from different perspectives, and as an impetus to challenge the platform's capabilities, not only theoretically as an artistic practice, but also by calling for the instantiation of authoring tools for the development of virtual reality experiences. The paradigm presented in this paper is a Virtual Reality art piece, called skin, 2006, developed on Electro, which is an open-source cross-platform development environment. *skin, 2006*, is an interactive hypersteroscopic high-definition audiovisual installation that explores a dialogue between physical and digital senses of "touch".

Keywords: Virtual reality, hyperstereo video, electro, open source

1. INTRODUCTION

Fig. 1. Real-time interaction with *skin, 2006*

1.1 Skin: Synopsis

skin, 2006 is a real-time interactive installation that visualizes a dialogue between physical and digital senses of "touch". The installation consists of a screen representing artificial skin, onto which hyper-stereo skin-related video imagery is

* eleni@evl.uic.edu, phone 1 312 375 2064; www.evl.uic.edu/eleni

projected, Fig.1. The imagery is accompanied by a generative composition of digitally manipulated and synthesized skin sources: porous breathing, the friction of skin on various surfaces, and synthesized representations of the electrical pulses inherent in skin at the molecular level. These skin-derived media serve as an exploration of a new territory: the intersection of the human body as a landscape/mindscape and the body's own traces of touch.

By reacting to the imagery, the participating audience deforms virtually the skin-derived media. Physical gestures are translated into accentuated and non-literal digital deformations of the projected imagery, while new sounds, both organic and synthetic, are added to the mix to accompany these deformations.

1.2 Description

The installation consists of a dark room and a floor-to-ceiling projection screen of a hyperstereoscopic high-definition video piece. The looped video is of non-narrative form. It consists of five scenes of "stillies" – slowly moving images of compositional body formations, as shown in Fig. 2. The skin-derived background sound composition is constantly being re-generated.

Visitors may walk close to the screen and put on the polarized glasses. The polarized glasses allow the audience to experience stereoscopic vision. The piece is also interactive. Participants can hold the Wanda™ [1], which is suspended from the ceiling. By means of a tracking antenna that is located on top of the screen and connected to a computer, and a tracking sensor that is embedded in the Wanda, physical motions and forward or backward gestures from the wand deform the skin imagery respectively inwards or outwards, as shown in Fig. 3. The gestures are accompanied by deformation sounds. The user interacts with the stereoscopic video imagery by performing touch-like motions. The magnitude of the deformation that is "sculpted" on the moving image grows larger as the participant moves the Wanda closer to or further away from the screen.

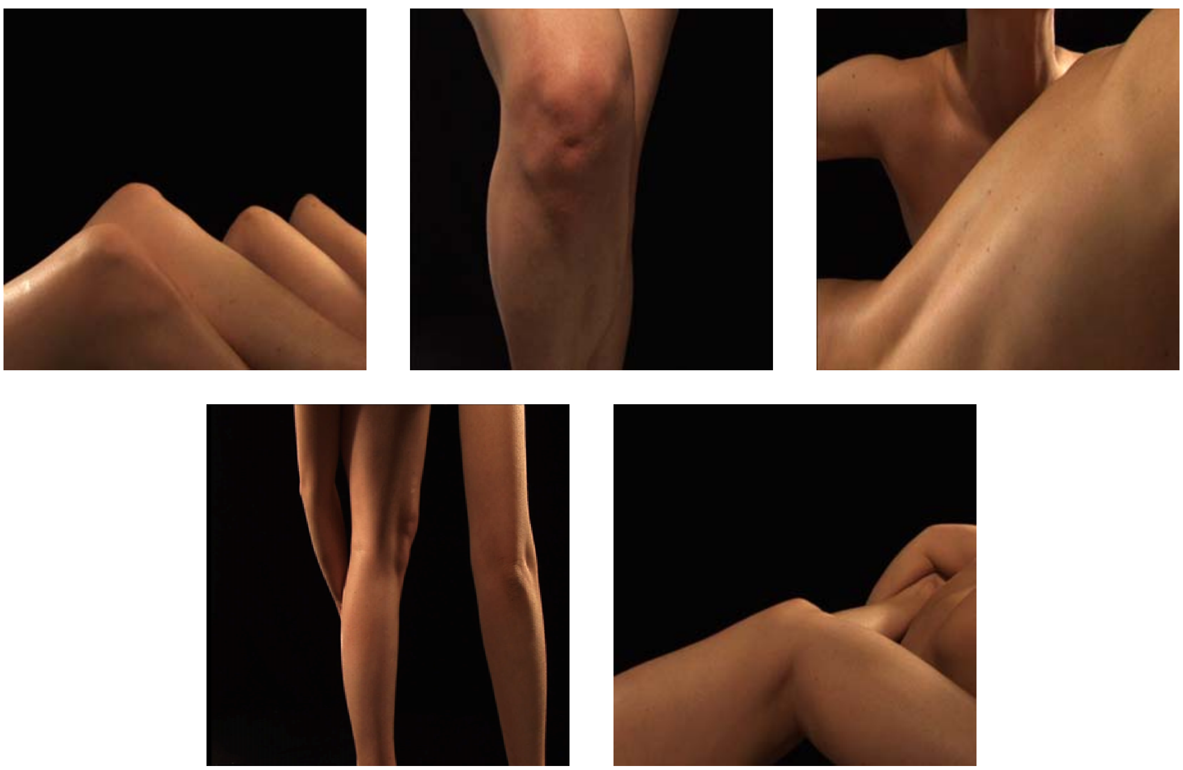

Fig. 2. Video stills of five scenes of skin, 2006

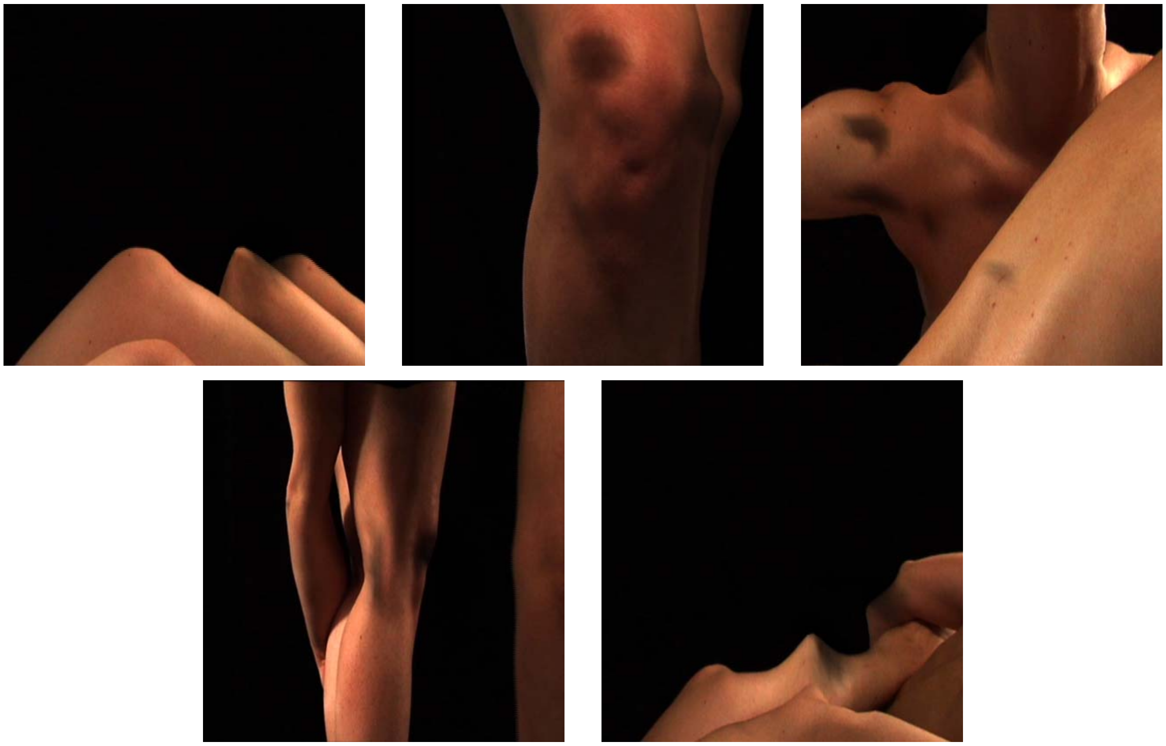

Fig. 3. Real-time screenshots of interaction/deformation of the five scenes of skin, 2006

2. RESEARCH, DEVELOPMENT & PRODUCTION

As the concept of the piece began to take form, it was clear that the following structural elements had to be researched and solved: Open-source license software supporting real-time interaction, sound and, most importantly, high-definition stereoscopic video playback; the process of shooting and producing stereoscopic video; interaction design; sound design and composition.

After careful consideration and research of available software, a recently released application development environment called Electro[2], was chosen as a software solution that met all the above requirements. In addition, it supported cross-platform implementation, independence from commercially-licensed software and support for input handling both from external devices and network. Robert Kooima, the author of Electro, unravels its architecture in subsection 2.1 and explains the interaction design technical aspects in subsection 2.3. Helen-Nicole Kostis provides information of the different levels and areas that had to be resolved for the stereoscopic video production in subsection 2.2 and John Kannenberg, the sound artist of the piece, describes his sound design process and composition in subsection 2.4.

2.1 Software: Electro

Electro is a multimedia development system designed for the creation of interactive applications for a wide variety of display devices and virtual environments. Electro combines a simple scripting language with a distributed scene graph data structure and a comprehensive display configuration system, enabling the rapid development of cross-platform interactive software. It supports the import and display of 3D models and 2D images and video, as well as positional audio playback.

Electro's scripting system uses the Lua programming language[3]. Lua is a small and efficient embeddable scripting language used in a wide variety of games, visualization tools, and other real-time interactive software. Lua's syntax is regular and easy to learn, making it useful to the beginning programmer, but it is powerful enough for use by the expert. It supports imperative, object oriented, and functional programming styles. Lua's data abstraction is uniform, providing a garbage-collected table structure that may be used as array, list, dictionary, object, and module. Lua is smoothly bound to a variety of programming languages, making it naturally embeddable. Lua forms a thin layer atop Electro's C implementation, abstracting its functionality and encapsulating its data types naturally.

At the core of Electro lies a scene graph data structure. A scene graph is a hierarchy of entities representing textured 3D models, 2D images, cameras, and lights as well as entity groupings and transformations. The structure of this hierarchy reflects the structure of the virtual environment. Input events from mice, keyboards, joysticks, six-degree-of-freedom motion sensors, and timers trigger procedure callbacks within the application's Lua code. These procedures may then manipulate the scene graph as necessary, giving interactivity.

Much of the generality of Electro comes from its display configuration system. Display resolution, physical structure, and logical layout are defined using the same Lua API used to manipulate the scene graph. One or more rectangular display surfaces are positioned and oriented in space and a static view point is optionally defined. Motion tracking allows for the dynamic positioning of the user's head. Given display structure and user view position, view frustums for each display surface are computed for each rendered frame. Electro also supports a variety of stereo viewing mechanisms, including passive polarized GeoWall™ and C-Wall systems, active LCD glasses, and Varrier™ autostereoscopic display[4].

Originally designed to support cluster-driven tiled displays, Electro uses the Message Passing Interface (MPI)[5] for internal communication. Electro uses MPI to distribute the display configuration and scene graph data structure, and synchronize rendering, enabling clusters of rendering hosts to behave transparently as one. Electro's display abstraction allows applications to disregard the nature of the display device. Electro provides a few basic functions with which the application may query necessary display attributes, such as 3D volume and 2D aspect ratio, and adapt accordingly. For the most part, however, the properties of the display are immaterial to the application, and Electro guarantees that a script written for one display will run unmodified on another. This allows users to develop content on their personal computer and display on a large-scale device.

Since the beginning of the Electro project in early 2005, the goal of Electro has been to widen the user base of display technologies developed at EVL. These displays include not only the GeoWall[6] and C-Wall[7], but also the Varrier[4], the LambdaVision™ [8], and many others. These technologies, being new and unique, have necessarily high barriers to entry for the potential user. For most users, the need to master stereo rendering, cluster communication, or autostereo interleaving eliminates the potential exploitation of these displays for anything other than display research itself. Electro has proved an effective tool for use in these circumstances, and has resulted in a proliferation of content that otherwise would not have been generated.

2.2 High-definition Hyperstereoscopic Video

The production of a high-definition video played a critical role in the development of the piece. High-definition imagery was selected as it would provide superior visual content in terms of image quality and coloration, due to the subtleness of the subject matter – skin.

Stereoscopic video for the piece was shot using two Sony HD camcorders: models HDR-FX1 and HVR-ZIU, Fig.4. These cameras had recently been purchased from EVL and they provided a great learning platform for experimenting with producing stereoscopic video, using manual features to achieve high quality of color and motion. The size of the cameras and the distance of their lenses was proven to be larger than "normal" image separation, which is about 65 mm (2.5 inches) and is the average distance of human eyes; anything further apart than this distance is called *hyperstereo*. *Hyperstereo*, a term introduced by Charles Wheatstone in 1857, refers to a stereo image which uses a larger than normal stereo base (or separation).

Hyperstereo, as well as "normal" stereo, is obtained with one or two cameras for the production of still photographs or moving images. Depending on the subject, the distance between the lenses of the cameras recording the right and left images can vary from anything greater than 2.5 inches to tens and hundreds feet apart. Hyperstereo photography is usually employed for landscape stereo photography, like mountainous areas, city skylines, clouds, and in general very distant objects; it gives a stereoscopic depth effect to very distant objects that normally appear flat.

Fig. 4. Cameras, stereo base, bolts and level, 2006

In order to produce correct stereoscopic video a series of experiments took place in many different areas. The critical aspects of the production process that played a catalytic role were: working in a studio space that could be reconfigured freely to suit the needs of the piece and until the satisfactory result had been reached; access to two high-definition cameras, as well as access to tools and various types of equipment; proper alignment of both cameras using a finely constructed stereo base with alignment pins; tuning cameras in the appropriate manual setting to achieve the desired photographic look and film like motion; empirical methods and mathematical concepts of stereoscopy in order to understand, learn, and accomplish stereoscopic video production and proper frontal plane calculation; shooting; video editing left and right sequences; and exporting video sequences separately in frames to be read by Electro. Finally, in post production further treatment had to be made: color correction to fix coloration mismatch between left and right cameras, since both cameras had different chip sets; image processing; and then scripting to play back different video frame sequences in Electro. During the above stages it was very important for the artist to have the guidance and advice of her research advisor and mentor, Daniel J. Sandin[9] and of her thesis advisor, Daria Tsoupikova[10].

The stereoscopic video for the piece is overlaid on two digital meshes of n-number of vertices (where n is a variable), where each video sequence is attached and mapped, frame-by-frame, as a texture map. Electro keeps the playback of the two streams in-sync, which is important for experiencing correct stereoscopic visuals. The video sequences on the meshes are hard-coded[11] for alignment in order to achieve better fusion of the stereo.

2.3 Interaction Design

A critical aspect for the interactivity of the piece was the translation of the physical motion to the digital deformation of the image, as it is sketched in Fig. 5 and the creation of smooth subtle deformations that would follow the gestures and motions of Wanda sensor devices. In the below paragraphs the technical aspects of the interactivity are described.

The skin Electro scene graph consists of two illuminated 3D mesh entities filling the area of the display, one for the left and one for the right eye. In the current implementation, the meshes are 50-by-50 grids of vertices, tessellated with triangles. Stereoscopic video is applied to the meshes as texture maps, one for the left and one for the right eye. The Electro timer callback is invoked 60 times per second. At each update the position of the user's hand is read from the motion sensor, and checked for collision with the mesh. On collision the mesh is deformed either inward or outward depending on the position of the hand. This deformation has a Gaussian distribution with a radius of 6 vertices, giving the interaction a maximum influence over about a quarter of the width of the mesh. The motion of mesh vertices is low-pass filtered, slowing the deformation and allowing the effects of interactions to accumulate. Where no hand-mesh collision is occurring, the mesh seeks to return to its default flat state. This motion is also low-pass filtered, causing the history of interactions to fade slowly. The normal vector of each mesh vertex is recomputed each frame from the positions of neighboring vertices. This alters the influence of the illumination on the mesh, enlightening deformations from above, darkening them below, and enhancing the presence of the 3D virtual surface. The video applied to the mesh is modulated by this illumination, effectively composing the illumination of the mesh into the scene portrayed by the video.

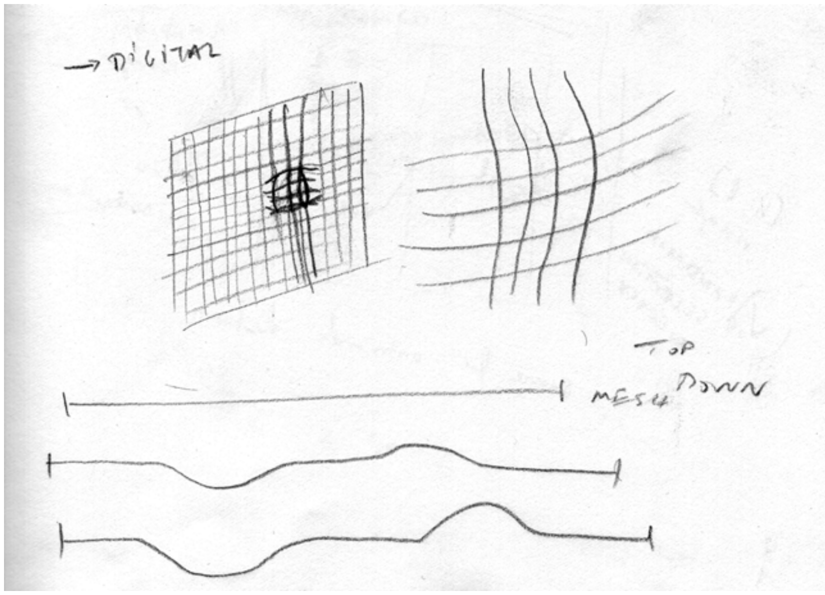

Fig. 5. Interaction design sketch, 2005

The real-time display of two simultaneous streams of high definition video is a performance bottleneck. Update rate is key to any real-time interactive 3D application, and is especially critical in viewer-centric virtual reality. Video must be heavily messaged in order to manage the fairly extreme data requirements of this application and make interaction possible given limited graphical and computational resources. Each frame of video is individually decoded, padded to a power of two in size, and recoded in DXT[12] format. DXT is a compression scheme used in real-time 3D graphics, and is supported directly by modern graphics acceleration hardware. It uses a deterministic decompression algorithm and provides a 6-to-1 compression ratio. Since DXT is natively supported by the hardware, it allows frames of video to be paged directly from the hard disk to the VRAM of the graphics adapter without incurring the CPU overhead of decompression.

2.4 Sound Design & Composition

In keeping with the visual element of the installation's macro-view of oversized skin imagery, the sound accompanying the visuals also exaggerates scale. Contextually small, quiet sounds are amplified and time-stretched to give them a larger presence within the installation's narrative while coexisting with a composed "background" sound bed.

This background of sound is created using simple generative techniques, with sound files of differing lengths playing simultaneously; once they begin to loop, the sounds enter a state of constant reconfiguration and re-contextualization. These sounds and their constantly shifting relationships to one another create a subtle, malleable sense of closeness to skin.

Three types of sounds combine to create the generative background sound. The first type of sound, repetitive recordings of breathing represents the porous nature of skin. The overlapping images in the hyperstereo video suggested the inclusion of the second sound type, which represents friction. Skin rubbing against skin, hair and clothing was recorded and time-stretched and added to the background mix. Finally, purely synthesized tones were composed to add a third type of sound meant to evoke the electrical impulses that inhabit skin at the molecular level.

As the audience's physical interactions with the screen manipulate the video, they also produce a similar reaction sonically. Both organic and synthesized sounds accompany these visual deformations, some of which are sonically related to the background sounds while others are strikingly different. Some of the deformation sounds are once again sourced from recordings of skin or other body-related sounds, particularly swallowing. However, to inject a non-human element into the deformation sounds (to emphasize the installation's deconstruction of the human body into something impurely human), a contact microphone was placed onto a cat in order to capture purring sounds that were then time-stretched and processed. To further advance the theme of molecular electrical impulses, manipulated recordings of shortwave radio static were also included in the deformation sounds.

Sounds for the project were derived from recordings of the composer's skin as well as purely synthesized sounds created using modified synthesizers in Propellerhead's Reason software. The sounds were edited using Bias' Peak editor as well as Apple's Soundtrack.

3. INSTALLATION & TECHNICAL SETUP

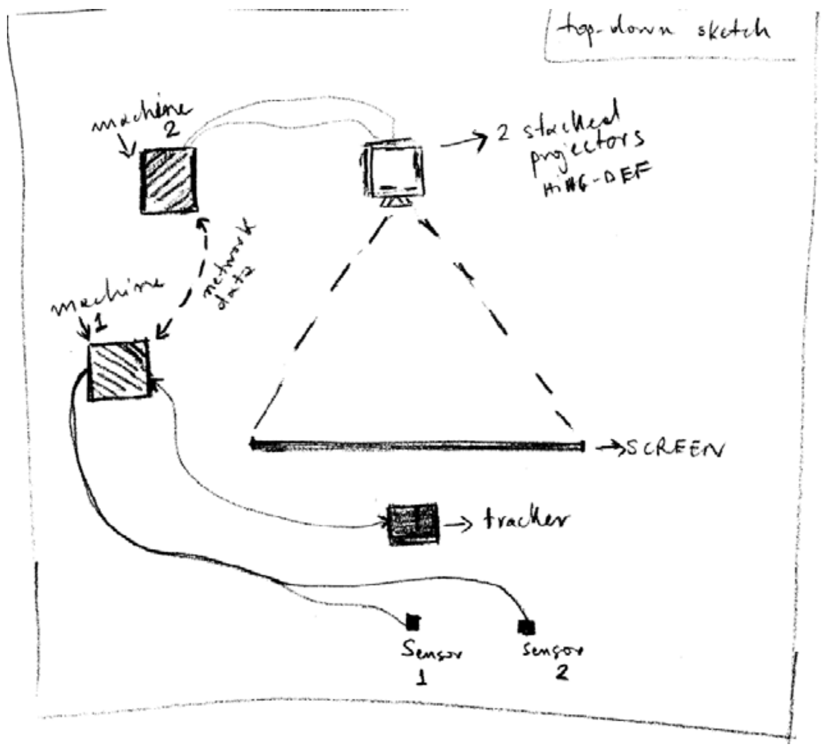

Fig. 6. sketch of technical setup, 2006

skin, 2006 is an interactive audiovisual installation that has been exhibited and runs on stereoscopic systems, such as the C-WALL[7]. The CWALL system is a site-specific, high quality circularly polarized stereo wall with tracking. It functions as an one-wall CAVE®[13] room. Since C-WALL is a site-specific system, its setup varies on equipment availability, costs and location.

In general, the technical requirements for *skin, 2006,* follow the requirements of a C-WALL setup, as shown in Fig. 6. The CWALL setup employed for *skin, 2006,* utilized two high-definition projectors with circularly polarized filters mounted on the front of the projector lenses. The projectors are set for rear-projection in a polarization-preserving *Black Screen* from Stewart Filmscreen Corporation. The screen is frame mounted and, depending on the architecture of the exhibition space, is either embedded in a wall or in a cloth-covered wooden scaffold in lieu of a wall, as shown in Fig. 7. A tracking system is utilized, using either a SpacePad[14] antenna or the Flock of Birds[15] tracking system. Two receiving antenna provide the tracking for the head sensor and the Wanda controller. An additional receiving antenna can be used to provide tracking for a second Wanda controller. All tracking-device cables are inserted in black spiral tubes for aesthetic purposes. The stereoscopic imagery, sound, and interactivity are driven by two machines: a tracking machine that runs the Trackd®[16] server (*machine 1*) and a GNU/Linux desktop computer (*machine 2*) which runs Electro and a Trackd client in order to accept the tracking data from the Trackd server. *Machine 2* should have a graphics card that can support stereo output. In addition, two speakers are attached to a mixer that is connected to *machine 2*, in order to provide the generative audio for the piece.

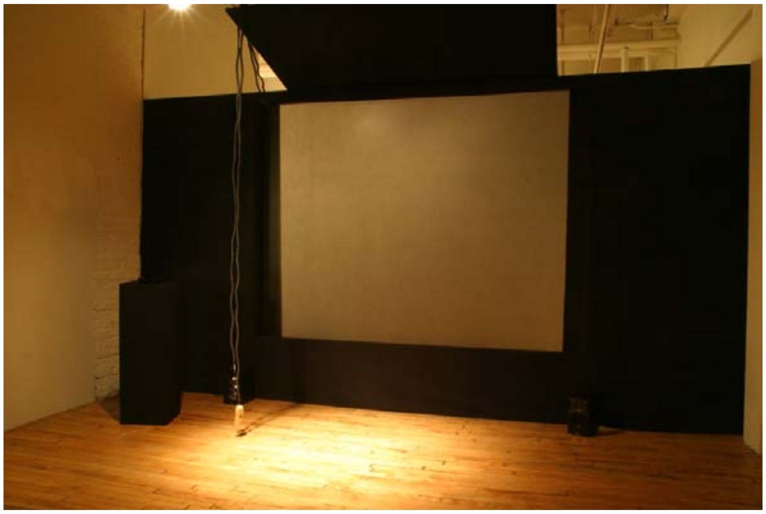

Fig. 7. C-WALL setup, Great Space Gallery, Chicago, IL, 2006

4. CONCLUSIONS

The purpose of this paper is to demonstrate the creative process from conception to production: it discusses and sheds light on the creative, experimental, and analytical processes of transforming and articulating visions to realities. Furthermore, it reveals processes of self-exploration for meaning and delineates the associated concepts of outcome and destination.

The piece contains also performance elements in different layers. The content of the stereoscopic video is a series of performative actions of abstracted body formations in slow-motion, while at the same time is trying to retain and position these formations for the production of good stereo. The interaction of the audiences with the piece is also performative. Audience interaction and treatment of the piece contain elements that would be interesting for further study. The interactive behavior varies for each individual, and in the end everything relates to what the particular piece brings out on each participant. In some cases people described their interaction as caresses of bodies with subtle motion and hidden poetics, others expressed their experience as dark, while others, being contradicted with ambiguity regarding the implied thematics and politics of the piece, such as eroticism, pain, violence and beauty, either transcended their emotions and experiences along with the piece or decided not to participate in the interactive part, feeling more comfortable with simply experiencing it as a stereoscopic video artwork.

During exhibitions of the piece, audience members noted the new type of aesthetics introduced, especially in comparison with, and in contrast to, VR artworks developed using 3D modeling and geometry. Furthermore, the majority of the audience did not realize that the stereo imagery was recorded in video, and some described it as “stereo paintings”. Finally, the landscape nude imagery in collaboration with the organic sound composition, along with the smooth and subtle deformation with touch-like gestures produced individual, private-like encounters experienced in a public setting, such as the gallery space.

Future work may incorporate the addition of different forms and types of interactivity as alternatives to the Wanda, either by using a haptic/tactile interface or by tracking via pattern recognition. In terms of stereoscopic video, areas of interest could include shooting stereo with cameras in motion, in order to investigate the challenges of simultaneously retaining content suitable for stereo on both cameras, and experimenting with depth-of-field during close-ups of skin imagery. Other directions relating to the subject of skin may include the use of touch screens as a proxy for skin-to-skin contact or of the human skin/body as the functional interface for interactivity.

ACKNOWLEDGEMENTS

Helen-Nicole Kostis would like to acknowledge all those who provided exceptional support, guidance, help and advice in the production of this artwork: her thesis committee comprising Daria Tsoupikova (chair), Dana Plepys, Sabrina Raaf, Daniel Sandin; especially important has been the contribution of John Kannenberg with his sound design and composition; Robert Kooima with his passionate support, technical guidance and wonderful help to the evolution of the Electro framework; Stephen Cady with his assistance on lighting techniques and tips; Anastasios Golnas and Kimon G. Kostis for their multi-faceted support; Tim Loucopoulos, Luc Renambot, Alan Verlo, Lance Long, Pat Hallihan, the Design Visualization Laboratory (DVL), UIC and the School of Art & Design, UIC.
All of us give special thanks to EVL, UIC for its generous financial and technical support and for providing necessary equipment for the exhibition of the artwork at SPIE Electronic Imaging 2007 conference.

The Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago specializes in the design and development of high-resolution visualization and virtual-reality display systems, collaboration software for use on multi-gigabit networks, and advanced networking infrastructure. This material is based upon work supported by the National Science Foundation (NSF), awards CNS-0224306, CNS-0420477, OCI-0229642 and OCI-0441094, as well as the NSF Information Technology Research (ITR) cooperative agreement (OCI-0225642) to the University of California, San Diego (UCSD) for "The OptIPuter." EVL also receives funding from the National Institutes of Health, the State of Illinois, the Office of Naval Research on behalf of the Technology Research, Education, and Commercialization Center (TRECC), and Pacific Interface on behalf of NTT Optical Network Systems Laboratory in Japan. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies and companies.

REFERENCES

1. The Wanda™ is the major input device used to interact with and control a virtual reality experience in the CAVE, ImmersaDesk and other VR systems. It is essentially a 3D mouse, with a receiving antenna attached which provides the computer with information about the Wanda's position and orientation. Wanda is commercially available from Ascension Technology Corporation. For more information please visit the following sites: http://www.evl.uic.edu, http://www.wandavr.com
2. Electro, http://www.evl.uic.edu/rlk/electro
3. Lua, http://www.lua.org
4. Autostereoscopic systems display three-dimensional images that can be viewed without the need to wear special glasses on the part of the user and afford an effective sense of immersion. The Varrier™ Autostereoscopic Virtual Reality Display, http://www.evl.uic.edu/core.php?mod=4&type=3&indi=271
5. The Message Passing Interface (MPI) standard, http://www-unix.mcs.anl.gov/mpi/
6. GeoWall™ is a low-cost, non-tracked, passive-stereo system that allows distributed audiences to view and interact with 3D immersive content, http://www.evl.uic.edu/core.php?mod=4&type=1&indi=233
7. C-Wall, http://www.evl.uic.edu/core.php?mod=4&type=1&indi=234
8. LambdaVision™ is an ultra-high-resolution visualization and networking instrument for research and education in geoscience, computer science and other research disciplines. LambdaVision consists of 55 LCD panels tiled to produce a 100Mpixel display, http://www.evl.uic.edu/core.php?mod=4&type=2&indi=273
9. Sandin, D.J., Director Emeritus, Electronic Visualization Laboratory. Professor Emeritus, School of Art & Design, http://www.evl.uic.edu

10. Tsoupikova, Daria, Electronic Visualization Laboratory Faculty. Assistant Professor, School of Art & Design, http://www.evl.uic.edu/core.php?mod=4&type=5&indi=222
11. Each sequence of the video is aligned specially, in order to have better stereo. Different scenes were shot in different days or instances. That might have resulted in small separation of both cameras.
12. DXT, http://oss.sgi.com/projects/oglsample/registry/EXT/texture_compression_dxt1.txt
13. CAVE® is a multi-person, room-sized, high-resolution 3D video and audio environment invented at EVL in 1991. Graphics are projected in stereo onto three walls and the floor, and viewed with active stereo glasses equipped with a location sensor. As the user moves within the display boundaries, the correct perspective is displayed in real-time to achieve a fully immersive experience. For more information, please visit: http://www.evl.uic.edu/core.php?mod=4&type=1&indi=161
14. The SpacePad anntena, consists of three wire loops mounted on cardboard panel and configured on a flat plane that is inclined, and held in place above the C-Wall. For more information, please visit: http://www.evl.uic.edu/core.php?mod=4&type=1&indi=180
15. Flock of Birds Tracking System, http://www.evl.uic.edu/core.php?mod=4&type=1&indi=179
16. The Trackd® is the standard device software for VR environments and input device manufacturers in the immersive display industry. The Trackd is a small “daemon” application that takes information from a variety of tracking and input devices and makes that information available for other applications to use. For more information, please visit: http://www.vrco.com