A Gesture-based Hyperrealistic News Space

Blake Dworaczyk, James Creel, Babatunde Azeez, Andruid Kerne
Interface Ecology Lab, Department of Computer Science
Texas A&M University
College Station, Texas USA
{blake, jsc6064, tuniks, andruid}@cs.tamu.edu

Brooke Beane
Department of Architecture
Texas A&M University
College Station, Texas USA
bbeane@viz.tamu.edu

ABSTRACT

In this paper, we describe an interactive installation that allows participants to use gesture-based movement to manipulate a recombinant information space consisting of news media. By recombinant information space, we mean a composition of media elements (text and image) from various sources with a navigable, visual representation. The participant experiences a hyperrealistic representation of current events. By a hyperrealistic space, we mean an experiential space in which digital representations take on significance which immerses the participant in a mediated reality. We employ the combinFormation project to retrieve and present semantically significant media elements. The participant interacts with the system by walking in a physical space that is mapped to the information space and gesturing with colored paddles. The system employs Max/MSP, Jitter, and a custom Max/Java patch to process video input, recognize gestures, and relay messages to the combinFormation system. By permitting participants to interact with visual compositions in a kinesthetic manner, the installation physicalizes and socializes the experience of authoring visual compositions with combinFormation. The installation draws audiences through the inherently social aspects of gesture and image. Participants can take turns manipulating the information space, allowing for collective authoring. The goal of our project is to encourage many participants to join together in a social setting to create collective meaning through visual composition.

Categories and Subject Descriptors

J.4 [**Arts and Humanities**]: Interactive Art

General Terms

Interactive Media, Semiotics, Interactive Art

Keywords

Recombinant Media, Information Spaces, Max/MSP, Jitter, News

1. INTRODUCTION

Our goal is to make visual composition a more engaging and intuitive experience by eliminating the need for traditional desktop interaction devices in the composition process. Additionally, we want the compositions to serve a social function by allowing an audience to observe the composition, encouraging participation. Furthermore, participants could be permitted to take turns modifying the composition.

We present an interactive installation that utilizes color tracking to receive inputs in the form of gesture-based motion, and use this movement to control a combinFormation recombinant information space. *combinFormation* provides a generative information space for browsing, collecting, and organizing information samples from the net [6]. By information space, we mean a place where you can arrange the samples in space, visually. By generative, we mean that there is an agent that works with you on developing the space. The agent tries to understand your interests, and work cooperatively with you on the space. To do this it needs to understand what you are interested in. We provide tools for expressing design and interest intentions at the same time.

More specifically, the participant can set a level of interest (positive, neutral, negative), and then express interest in media elements by manipulating them with combinFormation's tools. These tools allow the participant to remove and rearrange media elements within the information space. The agent responds to interest expression by refining the download and presentation of media elements.

We project the information space onto a surface for public viewing. The participant may experience temporary immersion in the information space as she *walks within it*. The visual surrogates serve as simulacra [1] for events and news. The end result is a visual composition, a simulacrum of the author's intended meaning. We observe layers of semiotic structures in our visual compositions within the combinFormation space. In this way human motion and public display can make the combinFormation experience more hyperrealistic [1]. News itself can be seen as a hyperrealism that distorts or embellishes actual events. Our installation provides a further hyperrealistic view into this hyperreality. Through this nested hyperrealism participants can experience multiple levels of immersion.

We intend our art to permit participants to express whatever viewpoint they desire on news topics. The installation is designed to provoke collective thought on the subject matter of compositions and to encourage people to vocalize and express their viewpoints in a public forum.

We performed two user studies and asked each user to use the installation to create a visual composition on current affairs. We communicated with the participants both during and after the session to obtain qualitative and quantitative data. Ultimately, the participants surprised us with many varied interpretations of the media collections.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

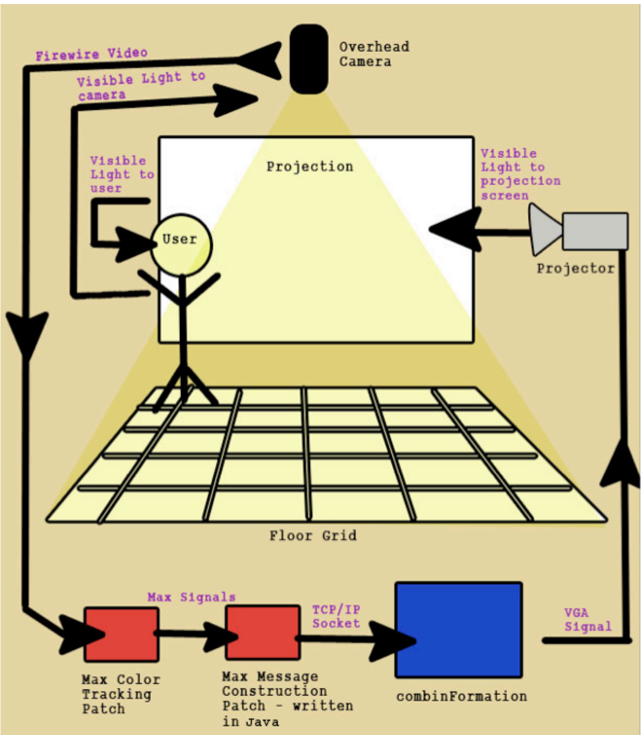


Figure 1. This diagram represents the flow of control between the participant and the installation components.

2. SYSTEM ARCHITECTURE

We use Cycling 74's Max/MSP [14], which provides the signal processing facility for gesture-based interactions. The Jitter [8] plug-in provides visual processing of participant movement by means of color tracking. Additionally, a Max patch written in Java enables the video input, as received by Max/Jitter, to be decoded and translated into gesture-based messages. These messages are then communicated to the combinFormation system. Max/Jitter captures movement with a ceiling-mounted camera that monitors a bright color on top of the participant's head. A delineated rectangle on the floor gives the participant orientation and allows them to understand the camera/information space boundaries (see fig. 1). There is a direct mapping between the rectangle on the floor and the visual composition projected in front of the participant. In this way, the participant takes the place of the mouse. They have the feeling of "walking" within the information space.

Max/MSP/Jitter

A ceiling mounted camera provides input to Max/MSP/Jitter, which tracks the position of various colored artifacts on the participant's person. Interactions are achieved with simple gestures. An orange dot on a hat corresponds to a cursor in the combinFormation window, which is projected on a screen. The orange dot is tracked by Jitter as the participant faces the projection, producing 2-dimensional coordinates corresponding to the participant's position in the information space. Since the participant's position directly moves the cursor/focus, moving left causes the cursor to move left, moving forward causes the cursor to move up, etc. In addition to orange, the patch tracks the colors red, green, blue, and yellow for gesture interpretation.

Max/Java

The Max/Java patch (henceforth the Java patch) is an object written in Java that is embedded within the Max/MSP/Jitter environment. Its execution life cycle is the same as that of the Max/MSP/Jitter patch which contains it.

The Java patch takes tracking data from the Max/MSP/Jitter environment and makes decisions about what information (messages) to relay to combinFormation. The Java patch opens a socket connection with the combinFormation remote daemon and sends regular tracking updates via TCP/IP to move the cursor and perform operations as the participant desires.

combinFormation

combinFormation [6, 9] is a tool for interactively creating recombinant information spaces and collections. In our system, these are ultimately visual compositions consisting of images selected by the program and presented to the participant. combinFormation offers a variety of facilities for interaction; we use a subset of these tools.

combinFormation uses an agent that learns the participant's interests in information elements through expressions of interest. Interest is enumerated as *negative, neutral,* and *positive*. The participant sets the interest level that she desires and interest is expressed automatically with each operation on a media element. The interest level can be changed at any time throughout the participant experience; in this way the participant trains and informs the agent about her preferences and intentions. The agent uses algorithms that use the participant's expressed interests to crawl and download appropriate web media.

For this installation, in order to stimulate social discourse on current events, we "seed" combinFormation with news-based information sources. Seeds give the *combinFormation* crawler initial pages to begin downloading and parsing. This allows the initial media elements and collections to be formed before the participant expresses interest in any particular element. In this way the seeds determine the initial framework that the *combinFormation* agent operates in. Our installation currently uses combinFormation to dynamically collect images about current affairs by starting with live RSS news feeds. Events in our world offer great potential for profound and far-reaching visual composition. Since the material is drawn from web sources with potentially wide-ranging political agendas, the combinFormation compositions can undergo various meaningful interpretations. Individual media elements and compositions as a whole form hypersigns through which hyperlinked material functions to create meaning [9].

We utilize two tools from combinFormation's toolset: grab and cut. The grab tool, as the name implies, allows the participant to move media elements within the information space. Media elements can be arranged and stacked in any way with the grab tool. The cut tool provides the means for eliminating unwanted elements. Both the cut and the grab operations express interest in their target media elements based on the current level of interest.

As an example, assume that the participant cuts an image with neutral interest. This expresses to the agent that the image was probably just unfavorable to the participant. On the other hand, had the participant cut the image with negative interest, the agent would then attempt to provide fewer images related to the disliked or irrelevant image. Along the same lines, if the participant expresses positive interest by cutting or dragging an image while in positive interest mode, the agent will compensate by finding and displaying more images of the favored topic.

With these tools the participant can use traditional collage techniques enhanced with improved visualization to develop her work. Additionally, the participant works with the combinFormation agent to determine future media topics. Through this process the participant and the agent collectively author the visual composition.

3. INSTALLATION FUNCTION

The participant is equipped with two paddles, each with a different color on either side, for a total of four different colors. One paddle is colored blue and yellow, and the other red and green. Functions are invoked by facing one colored side of a paddle upwards toward the overhead camera, obscuring the orange dot with the color. The interaction is pictured in figure 2.

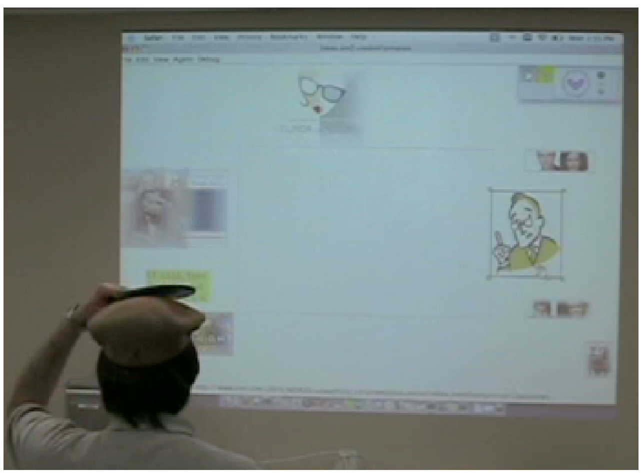


Figure 2. A participant expresses negative interest in a media element.

The participant may increase the current level of interest (from negative to neutral or from neutral to positive) by placing green over the orange dot. The participant may decrease the current level of interest (from positive to neutral or from neutral to negative) by placing red over the orange dot. The participant may cut the object under the cursor by placing yellow over the orange dot. Finally, the participant may grab and move objects by placing blue over the orange dot, and walking with the color held there, removing the blue when the object has been moved into the desired position.

The mere colors of our paddles did not offer adequate affordances [15] for the intended interactions, because the association between each colored paddle and its corresponding interactive operation was not clear to participants. Therefore, we developed a combinFormation interface in which the tool icons are colored in correspondence to the colored paddles. This 1-1 mapping informs participants about their capabilities with minimal explanation. In addition, we attempted to make the color mappings as intuitive as possible. We take green, which enables positive interest expression to be symbolic of “Go.” We take red, which enables negative interest expression to be symbolic of “Stop.” Pink is substituted for red in the present version to eliminate conflicts with other colors that the tracking algorithm recognizes. We desired to symbolize the cut operation with a warm color [7], so we chose yellow, which was the remaining warm primary color. We desired to symbolize the grab operation with a cool color, so we chose blue, which was the remaining cool primary color.

Due to the immediate visual feedback, in which the cursor moves when the participant moves across the space, interaction offers participants immediate gratification and encourages participation. Furthermore, the kinesthetic mode of interaction results in the social engagement of participants and audience. We have observed that they enjoy the kinesthetics of physical and visual movements, feedback, vivid colors, and progression of the visual composition.

4. PRIOR WORK

Computational and electronic advances have afforded a new set of paradigms for meaningful interactive media. New forms of artwork allow for human-computer interaction that not only deals with the aesthetics of a piece, but also with the relationships that are formed between the participant and the medium.

Cornock and Edmonds have classified these relationships [4]; in their schema, interactive artwork falls into three categories. The first category, *static*, is described as art that does not change or react to the viewer or its environment; however the participant may experience “personal psychological or emotional reactions.” The d*ynamic-passive* category involves art programmed by the artist to react to changes in environment factors such as temperature, sound or light. This interaction is a passive one because the human watches the interaction take place but is not actually part of it. *Dynamic-interactive* art involves the viewer playing an active role in composing the artwork; our installation falls under this category. There is a constant feedback loop between the participant and the installation.

Creators of dynamic-interactive art pursue goals of using technology and art in unison. According to Myron Krueger [10], interactive art requires computational perception of human expression. Such art, then, is not composed solely by the artist, but also by the participant. The meaning of interactive art derives from the framework imposed by the artist and the creative expression of the participant. Human motion is of particular interest in our form of interactive art. This involves the use of gestures and physical movement within the environment. Through algorithmic tracking of color, motion, and gesture, the system recognizes the participant’s expressions.

Human motion is explored in contexts of dance and performance. Rudolf Laban [11], a notable dancer and dance theorist, became interested in the study of human motion and its relationship to the surrounding environment. After creating a system that annotates movement using abstract symbols called Labanotation, he redirected his attention to studying the amount of effort exerted during movement. His goal during these studies was to provide methods of movement that eliminated any unnecessary movements that wasted time and effort, and enhanced those movements that would accomplish a task faster. Derived from his theories on

Figure 3. A number of stories created by one study participant through the juxtaposition of images and text.

effort and shape, he created Laban Movement Analysis, a system that interprets and documents all forms of human movement. This type of analysis allows society to understand meanings behind movement and physical expression.

Laban Movement Analysis consists of five major components: Body, Space, Shape, Effort and Relationship. As a unit, these components form the basis for describing movement. Body corresponds to parts of the body that are used to initiate motion. Space describes the location, direction and path of the movement. Shape refers to the undulating forms that the body creates within the space. Effort correlates to the amount of energy the body exerts while performing such movements. Relationship describes the interactions generated between the environment, the participant, and audience members.

Laban's form of analysis has traditionally been applied to dance, but is generally interesting when applied to any human motion, gesture, or body language. Of the components of Laban Movement Analysis, we consider Space, Effort, and Relationship to be most relevant to the function of our installation. This is because we are concerned with the locations and paths, i.e. the Space of the colors tracked by Max/Jitter. We seek to invest participants more deeply in their compositions by encouraging them to interact kinesthetically and with more Effort than would be the case with a mouse. Finally, we intend for audience members to freely advise participants, and to discuss the composition amongst themselves. Likewise, we intend for participants to effectively modify their environment and thus influence the audience; these activities embody Laban's concept of Relationship.

More recently, color tracking, motion tracking, and more sophisticated statistical methods have provided precise information about participants' real-time movements, including gestures. Such systems have typically given participants access to certain hyperrealities. For example, Camille Utterback [16] has created an interactive piece called *Untitled 5* wherein the participant moves about beneath an overhead camera. A visual projection is formed with simulated brush strokes that respond fluidly and intriguingly to people's motion. On the other hand, Eckel [5] appeals to hearing rather than sight. His *LISTEN* project provides a system that tracks users' movement to generate immersive soundscapes.

The *Screen* system [3] enables bodily interaction with pieces of text in a virtual 3-D environment. This interaction occurs in three stages. First, the participant reads a composition of text appearing on the virtual wall. Words from the text begin to fly off the wall, accompanied by a ripping sound. The participant may interact with the words as they swirl about by using a glove or wand. When a word is "hit" it returns to the wall at a random free location. Similarly to our installation, the participant's bodily motions contribute to a symbolic composition.

Lenman et al in [12, 13] present a project based on hand gestures for interaction. Their goal is to replace remote control functions with gestures for home electronic devices like TV. They use multimodal user interfaces, a scheme where hand poses and specific gestures are used as commands in a language for manipulating and controlling actions. The system uses an algorithm that tracks and recognizes the hand poses based on a combination of multi scale color feature detection, view-based hierarchical hand models, and particle filtering. Here the hand poses are represented as hierarchies of color images. Each image is then detected and tracked based on gestures extrapolated from particle filtering. A hand pose with index finger and thumb outstretched is used for activating the menus while a hand with five fingers outstretched is used for selecting a choice. In contrast, we use color tracking to simplify the recognition of gestures for interacting with information.

Wren, et al developed Pfinder (Person finder) [17], which enables real-time tracking of the human body in motion. It uses a statistical model of color and shape to recognize 2D blobs of heads and hands. This has enabled interaction for performance spaces, video games, and virtual reality. In [18] the authors describe how the interface has been used to navigate a 3D virtual game environment. By estimating positions of body parts, the system is able to create convincing shared virtual spaces. Similarly, we offer a physical space with direct mappings to a virtual space and the ability to navigate it.

Generally, previous work involving human gesture has provided the potential to produce art of the dynamic-interactive category. This approach has been used to immerse participants in spaces of sound, language, image, or virtual

Figure 4. A subsequent participant worked from the composition in figure 3 to create a more unified assemblage.

reality. Our project extends this approach to information spaces.

5. EVALUATION

In order to address the intersections and discrepancies between the processes of developing art and human-computer interaction, we employed evaluation processes with the goal of measuring both the efficacy of technology and the depth of creative experience. Thus, our evaluation was both process- and outcome-based. The process-based evaluation was an observational experiment where interaction was viewed and analyzed as it occurred, while the outcome-based evaluation involved feedback from participants after their experience.

The sessions were organized using techniques suggested by Boal [2], to promote effective administration of the evaluation process. During these evaluations, a member of our team played the role of "greeter," providing orientation and instruction to participants, while another member played the role of "observer," recording facts of interest during the evaluation.

During early phases of development, a prototype evaluation was conducted in which several participants interacted with a mock-up version of the installation. The results from this evaluation assisted in the further development of the installation.

For our final evaluation, we recruited 8 participants from several academic departments at Texas A&M University to interact with the installation while we observed and analyzed them. The participants were each willing to fill out a brief questionnaire after their sessions. Prior to the interaction, the participants were briefly familiarized with the concept of interactive art installations and the purpose and function of our installation, including the interaction techniques, and the role of combinFormation as a recombinant environment.

Prototype Evaluation

In order to evaluate prototypes of the installation, we ran mock-sessions in which a participant played Boal's role of "computer," by simulating the activity of the program.

In this context, process-based evaluation involved both physical observation and recording of the participants' interaction with the installation. We wanted to observe how the participants moved around within the installation setup. Therefore, notes on participants' actions were taken during the prototype evaluation. We emphasized information regarding participants' emotional states and degree of cognitive and physical effort during their introduction to the system and their subsequent interaction. The outcome-based evaluation involved a series of verbal qualitative questions that captured the participants' initial experience with the installation.

During the process-based evaluation of the prototype, we observed that participants and the team member playing the role of "computer" had trouble remembering the mappings between color and tool. Participants repeatedly asked about color mappings, even those recently invoked.

Essentially, we concluded that color-only mappings were not suggestive of the corresponding interactions. This prompted us to write the custom *combinFormation* interface with modified tool colors described previously.

Final Evaluation

During the process-based final evaluation sessions, at least one observer took notes on the participants' interactions and audience responses. Audiences consisted of 5-10 people in the vicinity of the installation. We observed that composition proceeded largely as we had intended, though some work could be done to refine the interaction. Specifically, we would like to make the cursor less jumpy and more responsive in the corners of the interaction area. We consider this problem to be largely environmental due to the inconsistency of lighting. Although participants exhibited some initial difficulties learning the gestures, these difficulties were generally overshadowed by interest in the composition. We observed such interest on the part of both participants and audience members. Audience members were usually interested and occasionally outspoken in their opinions on various media elements and juxtapositions. Audience suggestions typically involved interest expression operations and occasionally started conversations or even debates on the content.

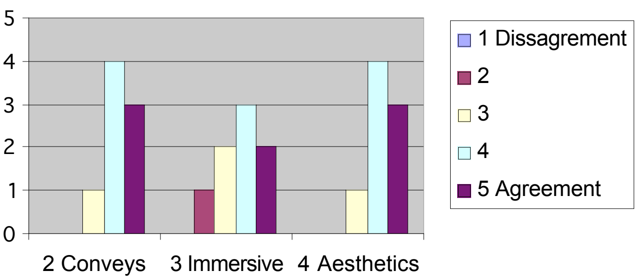
For our outcome-based evaluation, participants were asked to take the following questionnaire. In questions 2-4 a response of 1 indicates strong disagreement and a response of 5 indicates strong agreement.

1. What was the intended meaning of your composition?
2. The composition effectively conveys my intended meanings: 1 2 3 4 5
3. The mode of interaction made me feel immersed in the composition: 1 2 3 4 5
4. The installation encourages the generation of aesthetics and meaning: 1 2 3 4 5

The responses to the first question are as varied as the news itself. These responses include:

"Composition surveying world architecture," "major Star Wars characters," "public disobedience," "putting related elements together, cutting related elements" (this made in reference to a composition of female celebrities and fashion icons), "criticize Slobodan Milosevic," "demonstrate popularity of Star Wars in our culture," "I tried to drag photos of things I like to the middle of the screen" (this made in reference to a composition about civil unrest), and "destruction toward political aspects of the world."

These responses indicate a wide range of intentions on the part of the participants, ranging from disruptive to thoughtful to lighthearted. Furthermore, the participants reported success in realizing their intentions.



Graph 1: Histogram of final evaluation results.

To question 2, the mean of the responses was 4.25, with a standard deviation of .707.

To question 3, the mean of the responses was 3.753, with a standard deviation of 1.035.

To question 4, the mean of the responses was 4.25, with a standard deviation of .707.

Analysis

Analysis of the recorded data reveals that our installation is successful in allowing participants to create meaningful compositions in a dynamic and immersive manner, yet has minor initial usability issues. Participants' and audience members' facial and verbal expressions during interactions lead us to conclude that the installation is interesting and socially engaging. Audience members respond readily to meaningful media elements and provocative juxtapositions, thereby augmenting participants' enthusiasm.

Participants produced meaningful and aesthetically pleasing compositions with relative ease. This facility derives from the recombinant capabilities of the combinFormation engine [9].

In Graph 1, we present a histogram of the final evaluation results. These results indicate that participants found the installation immersive and the compositions meaningful. Participants found the installation less immersive than we initially hoped, though. We attribute much of this to testing environment constraints, which were not ideal for our installation. Participant's garments and inconsistent lighting caused occasional spurious operations that detracted from the immersion.

A composition authored by several successive participants offers evidence of the social aspects of our interaction scheme. In figure 3, we see a number of stories created by one participant through the juxtaposition of images and text. Tales of kidnapping and mayhem appear in proximity to photographs of hostages and ruined automobiles.

Another participant continued to develop the composition created by the prior participant (see figure 4). In comparison to the previous participant, loosely related elements have been eliminated, to produce a more consistent theme. The emphasis of the composition shifts from a broad and varied symbolism, to more specific symbols of the land and people of Iraq.

In figure 5, we see an example of a participant's interaction during a different session. This participant chose to draw exciting and incendiary images toward the crowded center of the composition. The participant depicts a variety of conflicts in which humans engage.

6. CONCLUSION

Participants were able to personalize the meanings they expressed, while preserving previous themes if desired. The

compositions emotionally influenced the audience members, who in turn communicated their own interests in particular media elements. Thus, the Relationship aspect of Laban's Movement Analysis is manifested in the participants' expression and interaction through gesture.

The combinFormation experience becomes more hyperrealistic [1] as the participant "walks" within the information space, forming collage as simulacra for world events and intended meanings. The information space has a physical mapping to the participant's space of interaction, making the experience of the virtual space more real.

Our decision to use news media in our installation proved successful in permitting varied and provocative participant responses. Furthermore, the installation facilitates composition in a manner suitable for collective consideration. Because many people are aware of current news topics, elements of news information can serve to afford social interaction. Participants take positions and become involved in discussions.

The ease of processing video information in Max/Jitter permitted us to avoid many of the complex issues of video processing addressed by the statistical techniques used by Lenman [12, 13] and many others to recognize gesture. In addition, as Boal [2] suggests, we have enabled audience participation in the composition of the participant's (and our) art, creating a kind of active, participatory theater.

REFERENCES

[1] Baudrillard, Jean. *The Ecstasy of Communication*, *in Hal Foster, The Anti-Aesthetic: Essays on Postmodern Culture*. New York: New Press, 1983.

[2] Boal, A. Games for Actors and Non-Actors, New York: Routledge, 1992.

[3] Carroll, J., Coover, R., Greenlee, S., McClain, A., Wardrip-Fruin, N. *Screen: Bodily Interaction with Text in Immersive VR*. In Proceedings of SIGGRAPH 2003, 1-1.

[4] Cornock, S. and Edmonds, E. *The Creative Process where the Artist is Amplified or Superseded by the Computer*. Leonardo, vol. 6, 11-16. Pergamon Press, 1973.

[5] Eckel, G. *The Vision of the LISTEN Project*. Seventh International Conference on Virtual Systems and Multimedia (VSMM'01) 2001, 393-393.

[6] Interface Ecology Lab, *combinFormation*, http://ecologylab.cs.tamu.edu/combinFormation

[7] Itten, Johannes. The Art of Color. New York: Wiley, 1997.

[8] Jitter. http://www.cycling74.com/products/jitter.html.

[9] Kerne, A., Mistrot, J.M., Khandelwal, M., Sundaram, V., Koh, E., Using Composition to Re-Present Personal Collections of Hypersigns, *Proc CoSIGN 2004*.

[10] Krueger, Myron. *Responsive Environments*. AFIPS 46 National Computer Conference Proceedings, 1997, 423-433.

[11] Laban, Rudolf. The Mastery of Movement. Northcote House, London. Revised by Lisa Ullman, 1988.

[12] Lenman, S., Bretzner, L., Thuresson, B. *Computer Vision Based Recognition of Hand Gestures for Human-Computer Interaction*. Technical report TRITA-NA-D0209, CID- report, June 2002.

[13] Lenman, S., Bretzner, L., Thuresson, B. Using Marking Menus to Develop Command Sets for Computer Vision Based hand Gesture Interfaces, *Proceedings of the second Nordic Conference on human-Computer Interaction*, pp. 239-242. Aarthus, Denmark 2002.

[14] Max/MSP. http://www.cycling74.com/products/maxmsp.html.

[15] Norman, D. The Design of Everyday Things. Basic Books, 1988.

[16] Utterback, Camille. http://www.camilleutterback.com/.

[17] Wren, C., Azarbayejani, A., Darrel, T., Pentland, A. Pfinder: *Real-time tracking of the human body*. IEEE Transactions on Pattern Analysis and Machine Learning, vol. 9, no. 7, 1997, 780-785.

[18] Wren, C., Sparacino, F., Azarbayejani, A., Darrel, T., Starner, T., Chao, K.A.C., Hlavac, M., Russel, K., Pentland, A. *Perceptive Spaces for Performance and Entertainment: Untethered Interaction Using Computer Vision and Audition*. Applied Artificial Intelligence 1997, vol. 11, no. 4, 267-284.