Techniques for Interactive Audience Participation

Dan Maynes-Aminzade
Massachusetts Institute of Technology
monzy@mit.edu

Randy Pausch
Carnegie Mellon University
pausch@cs.cmu.edu

Steve Seitz
University of Washington
seitz@cs.washington.edu

Abstract

At SIGGRAPH in 1991, Loren and Rachel Carpenter unveiled an interactive entertainment system that allowed members of a large audience to control an onscreen game using red and green reflective paddles. In the spirit of this approach, we present a new set of techniques that enable members of an audience to participate, either cooperatively or competitively, in shared entertainment experiences. Our techniques allow audiences with hundreds of people to control onscreen activity by (1) leaning left and right in their seats, (2) batting a beach ball while its shadow is used as a pointing device, and (3) pointing laser pointers at the screen. All of these techniques can be implemented with inexpensive, off the shelf hardware. We have tested these techniques with a variety of audiences; in this paper we describe both the computer vision based implementation and the lessons we learned about designing effective content for interactive audience participation.

1. Introduction

Most research on interaction techniques focuses on single users or on small groups, possibly connected over a computer network. Systems enabling large audiences to interact offer numerous possibilities for entertainment, education, and team building. We present three interaction techniques that allow the members of a large audience to participate, either cooperatively or competitively, in shared interactive entertainment experiences. We implemented these techniques using a combination of real-time image processing, computer vision, and 2D and 3D graphics. We then tested them in 30 demonstration sessions on movie theater audiences of 150-600 college students.

Involving an audience in a shared experience is not a new concept in traditional media. Popular theatrical productions such as *Tony and Tina's Wedding* [5] encourage audience members to sing along with the actors or to join in a food fight. Films like the horror spoof *The Rocky Horror Picture Show* [22] have inspired spontaneous audience participation, and various attempts at "Interactive Cinema" have allowed audiences to choose the manner in which a prerecorded story is presented [8]. However, the potential for unscripted audience activities afforded by real-time computer vision techniques has been less thoroughly explored; this is the focus of our study.

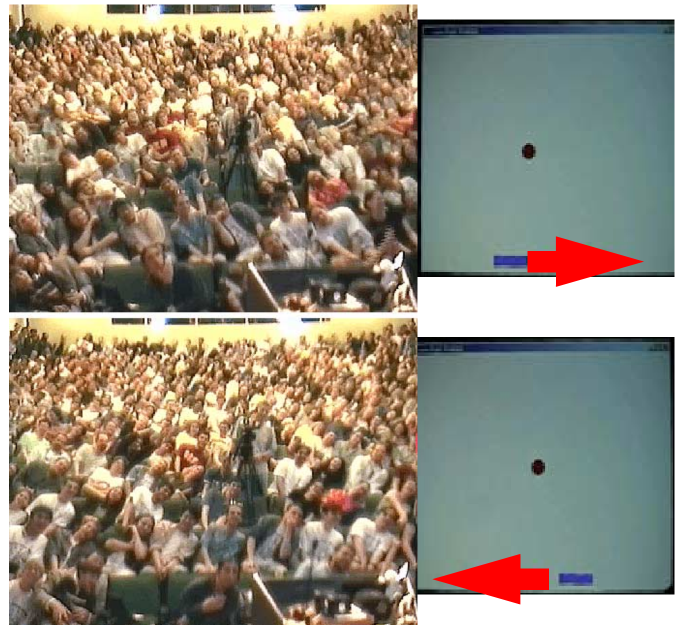

Figure 1: A large audience leans left and right to control an onscreen game of *Pong*.

The original inspiration for our work was the Cinematrix Interactive Entertainment System [4]. This system, which was presented at the SIGGRAPH Electronic Theater in 1991, employs computer vision and reflective paddles distributed to the audience as input devices. Cinematrix describes the system as a wireless "mouse for the masses" giving audiences real-time control over onscreen computer games. Audience members can participate interactively in activities such as maze navigation and opinion polling by displaying the red or the green side of their paddles.

Since relatively little progress has been made in the study of audience interaction since Cinematrix's debut ten years ago, we decided to take the ideas pioneered by Cinematrix a step further by introducing other forms of vision-based audience participation. Instead of creating artificial modes of interaction such as buttons on armrests [13], we attempted to develop techniques based on traditional crowd activities that require minimal apparent technology. In particular, we made use of three interfaces: audience movement tracking, object shadow tracking, and laser pointer tracking. This paper contains three contributions: (1) techniques for audience-scale interaction that expand on earlier methods, (2) a demonstration that these vision-based tech-

Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)
0-7695-1834-6/02 $17.00 © 2002 **IEEE**

niques can be implemented robustly and inexpensively using a single camera and PC when the tasks are properly constrained, and (3) the lessons we learned regarding interaction principles for designing effective, entertaining applications using our techniques.

2. Audience Movement Tracking

Audience movement tracking allows audience members to control an onscreen game without the use of physical props. This approach borrows from the technological artist Myron Krueger, who pioneered the development of unencumbered, full-body participation in computer-created virtual experiences, coining the term “Artificial Reality” in 1973 to describe this concept [17]. William Freeman and his colleagues at the Mitsubishi Electric Research Lab have also done related work in developing low-cost, real-time, vision-based interfaces to computer games that allow a single unencumbered game player to use natural gestures and motions to affect the game [14][15]. Similar systems track small groups of 2-4 of people as they perform interactive music and dance activities [25][26].

In our system, we position a camera at the front of the auditorium and point it at the audience. We then encourage the members of the audience to move in certain ways, and analyze the streaming video to control an onscreen game. For example, audience members can lean left or right in their chairs to steer a race car or move a paddle in the video game *Pong*[1]. Although the computer vision literature contains many techniques for human motion analysis [9][11], applying these methods for interactive audience participation poses problems since all processing must work robustly in real-time, in uncontrolled environments. Current systems generally operate on a single person at a time, or on a small number of people. A contribution of our approach is to demonstrate that robust real-time analysis of raw video of crowds is possible using simple image processing algorithms that are carefully tailored to the interaction task and can operate on commonly available hardware. Another advantage of our system is that in under 30 seconds it can be trained to detect almost any type of audience gesture, making it possible for an operator to dynamically modify the interaction style.

Raw video-based input has been used in other domains before, including “magic mirror” environments [10][14][19], single-player computer games [18], and crowd games with reflective paddles [4]. However, previous systems have required estimating image features such as silhouettes [19], computing moments and orientation histograms with the assistance of specialized hardware [18], or placing reflective markers in the scene [4]. Instead, we adopt a template-matching approach: we begin by capturing reference images of the audience leaning left, right, and center, and we then compare each frame of input to these three templates to see which template provides the closest match. A sum of squared differences is used to generate one correlation coefficient corresponding to each template, and we take a weighted average of these coefficients to compute the continuous control value used as a parameter in the onscreen game.

People have a natural tendency to lean left and right to express a desire to move an object in a particular direction, just as a child at a bowling alley leans his body in a subconscious effort to steer the path of his bowling ball. We found that audiences immediately understood the leaning technique; when we provided an activity with a clearly defined goal, such as playing *Pong* or steering a car, the audience members cooperated to effectively achieve that goal. Leaning is also a successful technique when regarded from a social perspective, since it encourages, but does not require, closeness between theatergoing companions.

One weakness of this technique is that it requires a calibration phase, but this phase can be naturally integrated with the show, since an explanation of the game is clarified when audience members are asked to practice leaning before the game begins. An additional drawback of template-based motion tracking is that it does not adapt well to changing audiences. We found that when a significant number of latecomers joined the crowd after the templates had been captured, the quality of the motion tracking suffered as a result. We overcame this problem by giving our software the ability to capture new templates on the fly without interrupting the game activity; the game operator watches the audience and triggers the capture of a new template when he notices that the audience is leaning left or right. Unlike the Cinematrix technique, our technique does not uniquely identify each member of the audience. In fact, audience members have uneven degrees of control, since people closer to the camera have a more pronounced effect on the game. This imbalanced distribution of power could be addressed by varying the weight of different portions of the image, but since the audience is unaware of the imbalance, in our experience it has little effect on the quality of each user’s experience.

3. Beach Ball Shadows

The common crowd activity of batting a beach ball before a concert provided the inspiration for our second interaction technique. The closest related work was by the band D’Cuckoo [3]. To encourage audience participation at their shows, the band developed an interactive show toy called the “MIDI Ball,” an inflated ball outfitted with Musical Instrument Digital Interface (MIDI) triggers and a radio transmitter. The members of the audience were invited to “jam” with the band by batting the ball around the concert hall, and each time the ball was hit, a sampled sound or a computer visual was triggered.

In our system, we project a game on the front screen of a movie theater, and as the audience bats a beach ball into the air, the ball casts a shadow on the screen. We point a camera at the screen and use computer vision techniques to track the ball’s shadow. This tracking allows the audience to play interactive games using the shadow of the ball as a cursor. Though we could have tracked the ball instead, tracking the shadow has the advantage that it is easy to identify regardless of the type of ball used. Shadow tracking works well even in the presence of specularities and shading on the ball, as shown in Figure 2. Tracking the shadow also requires less calibration in order for the position of the game cursor to register precisely with the position of the shadow on the screen. The disadvantage of the shadow-tracking method is that it requires highly saturated color values in the projected onscreen game in order to establish the contrast between the game and the shadows.

Figure 2: Missile Command using beach ball shadow tracking. The beach ball (top left) casts a shadow (bottom right) which acts as a cursor.

We used the beach ball shadow technique to build a version of the video game *Missile Command*. Missiles streak down from the top of the screen, targeted at the cities below. The audience attempts to protect the endangered cities by destroying the missiles before they reach their targets. The missiles are destroyed when the audience maneuvers the shadow-cursor over them.

Shadow tracking is accomplished by detecting dark regions within the image. We use the image histogram to precompute a threshold value, and the threshold is used to divide the grayscale image into zeros and ones. We then employ standard blob-analysis techniques [16] to determine the longest horizontal segments. To speed up the shadow detection, we only look for shadows near the shadow locations in the previous frame, if such shadow locations existed. Using this approach, we have also tracked two balls simultaneously in real-time.

Like the leaning technique, beach ball shadow tracking is immediately obvious to audience members. Once the goal of the activity becomes clear, the audience adapts readily to the mode of interaction and acts cooperatively to achieve shared objectives. Although only several members of the audience control the game action at any given time, if the activity is chosen carefully the entire audience becomes emotionally involved. When this technique was tested on live audiences, audience members went to such lengths as throwing other objects around the theater in an attempt to generate additional shadows. Since we tracked only the largest of the onscreen shadows, throwing these haphazard objects had no effect on the game, but it did serve to demonstrate the extent to which the audience had become emotionally involved.

4. Laser Pointer Tracking

It is not uncommon for members of a large movie theater audience is to shine laser pointers at the movie screen before the film begins. We acknowledge that promoting this behavior has certain drawbacks; the irresponsible use of laser pointers can result in eye injury. However, by pointing a camera at the screen and tracking the dots, we can create compelling interactive entertainment experiences that encourage theatergoers to direct their laser pointers at the screen and not at other audience members. Admittedly, laser pointer tracking is a less intuitive and physical mode of interaction than leaning or batting a beach ball, but like these other methods we found that it was successful in engaging large crowds.

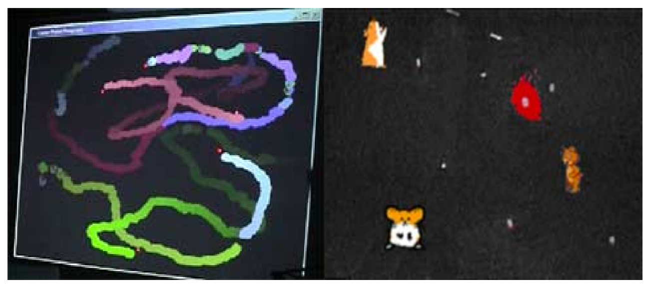

Figure 3: A laser pointer paint program (left) and laser Whack-a-Mole (right).

Single-dot laser point tracking is used in a number of systems [23][24], such as Richard Eckert and Jason Moore's "Interactive Learning Wall" [12], a Windows/PC-based virtual blackboard system that can be controlled remotely by a classroom instructor and students. The LumiPoint system, developed by Xing Chen and James Davis at Stanford University, can track a large number of dots, but requires a complex and expensive hardware architecture involving multiple cameras and multiple Silicon Graphics workstations [6].

A sample application of our system, shown in Figure 3, is a collaborative paint program that allows the audience to draw graffiti on the screen. Another example is a game that encourages rapid flocking of laser points to particular locations. Figure 3 shows one such game, a laser version of the carnival midway game "Whack-a-Mole." In this game, the audience members use their laser pointers to hit a sequence of targets. Figure 4 shows a different game in which the audience is encouraged to uncover a series of hidden images using their laser pointers. As the audience "scratches off" the black covering with their lasers, a humorous hidden image is gradually revealed. We built other games where the audience mixed audio samples like a disc jockey, formed particular shapes and figures, and played an interactive multi-user version of connect-the-dots.

We also use laser pointer tracking to carry out live audience polls and trivia competitions. The audience members point at a series of choices with their laser pointers, and bar graphs corresponding to these choices continually update to show the audience's preferences until the time limit is reached. Audiences become very emotionally involved in polls and trivia, particularly when the topic of the poll is a highly contested issue, or when several audience members believe they know the correct answer to a trivia question, but the majority of the audience is choosing incorrectly.

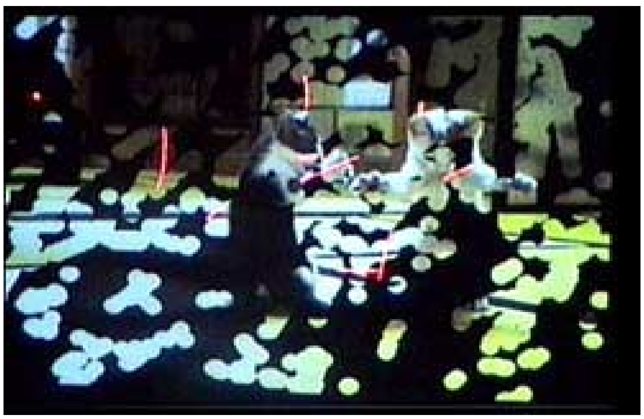

Figure 4: Using laser pointers to reveal a hidden picture.

Our laser point tracking uses a similar methodology as previous systems, but it uses only one Windows PC and one camera to track multiple laser dots. Laser dots are sufficiently bright that simple thresholding combined with a highpass saturation filter and a bandpass frequency filter is enough to distinguish a laser dot from the projected background, even when the projected image is quite bright. Lasers that move rapidly enough can create red streaks in the captured image, but we chose to ignore these dimmer streaks in order to encourage a slower and more deliberate movement of laser dots. This is advantageous because it requires the audience to focus on the careful strategic movement of laser points, as opposed to random, indistinct scribbling. When precise motion is compulsory, the activity becomes richer, since the audience members must attempt to coordinate their activities more carefully to accomplish their goal.

We use an association method [2] to distinguish the laser dots between frames. In every frame, we attempt to determine whether each observed point is the start of a new stroke, or the continuation of a stroke from the previous frame. We do this by checking whether the observed point lies near any point in the previous frame. Applying this technique is unnecessary for applications such as Whack-a-Mole, in which the idea of a stroke is irrelevant, but it is valuable for activities like the painting program, in which we used it to allow each person to paint with a different color. This association method is not wholly reliable, since strokes that cross each other can interchange colors and rapid motions segment strokes, but in practice we found that this did not detract from the enjoyment of the drawing activity.

One shortcoming of the laser-pointer tracking technique is that not every member of the audience is fully participating. We distributed laser pointers to the audience and found that although our system is capable of maintaining an interactive framerate while tracking 50 laser dots, attempts to play games with this many people generally result in onscreen chaos, since the players cannot tell their laser points apart. This shortcoming can be mitigated somewhat by creating activities that persuade the audience members without laser pointers to shout in order to give voice to their opinions, such as controversial polls and time-limited trivia questions.

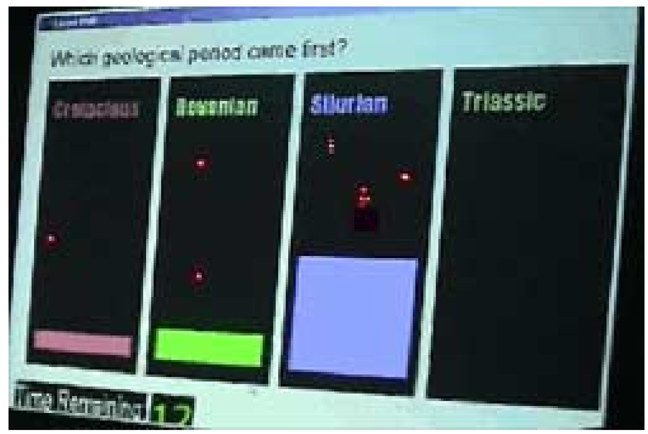

Figure 5: A laser pointer trivia question.

5. Design Principles and Conclusion

Over a period of eight months, we performed thirty tests of our audience interaction techniques on movie theater audiences ranging in size from 150 to 600. Through an arrangement with a university's student film organization, we conducted our trials as ten minute long interactive preshows for weekly film screenings. Our test subjects, primarily undergraduate college students, were not notified in advance that the movie they were attending would be preceded by an interactive show.

Based on the successes and failures of our repeated trials and the interviews we conducted with audience members, we present a set of design principles for interactive crowd activities. These principles follow Thomas Malone's guidelines for designing entertaining and intrinsically motivating educational experiences[20], but focus on audiences rather than on single users. We examine three areas of design: system design, game design, and social factors.

IEEE COMPUTER SOCIETY

5.1 System Design

- **Focus on the activity, not the technology.** While people are initially amazed at the technology allowing the interaction to occur, within 30 seconds they lose interest if the activity is not inherently entertaining.
- **You do not need to sense every audience member.** What matters is what the audience thinks is going on, not what is really going on. A weakness of our motion detection system, as compared to the Cinematrix system, is that we cannot identify an individual audience member's behavior. However, we have found that even those audience members who happen to be outside the camera's field of view participate fully. We have had repeated success with only 70 percent of the audience in view of the camera.
- **Make the control mechanism obvious.** Although the underlying technology need not be exposed, it is important that audience members understand how their actions affect the game activity. When leaning to steer a car, our audiences sometimes coordinated an effort to veer off the road in order to prove to themselves that they were indeed in control. Audience members will not continue to participate in an activity if there is no immediately clear indication that they are affecting the gameplay.

5.2 Game Design

- **Vary the pacing of the activity.** The leaning and beach ball games work best with a pacing that alternates moments of intense activity with periods of relaxation. When using the leaning technique, *Pong* is a much better activity than controlling a race car; while *Pong* provides sudden deadlines separated by rest periods, race car driving requires a more sustained behavior to keep the car on the road. The punctuated deadlines give the audience a chance to succeed or fail; the rest periods give them a chance to cheer, applaud themselves, and prepare for the next moment of tension.
- **Ramp up the difficulty of the activity.** We found that our games did not require an explicit tutorial if they were presented properly. Rather than asking the audience to play Missile Command as soon as we tossed out a beach ball, we first demonstrated the shadow tracking by displaying a circle on the screen that followed the ball's shadow. By gracefully scaling up the complexity of the activity presented to the audience, we avoided a tedious training phase.

5.3 Social Factors

- **Play to the emotional sensibilities of the crowd.** Social involvement is more important than technological involvement. When using laser pointers, our best shows were those that generated shouting by audience members without laser pointers, not the shows where everyone had their own laser pointer. With the beach ball, the lottery effect ("I might be next!") and the cheering or booing of one another fully engages all of the members of the audience, even though technically only one or two out of 500 people were directly participating.
- **Facilitate cooperation between audience members.** Laser pointer games are more engaging when they foster a sense of camaraderie between audience members. In a game like Whack-a-Mole, each audience member is involved in the activity for himself, much like a game of soccer played by young children in which everyone clusters around the ball. Connect-the-dots required each audience member to position his laser over a different dot, and since it required the audience to cooperate in order to succeed it was a more social game.

In 1967, visionary media guru Marshall McLuhan lamented the inactive nature of group entertainment, simultaneously acknowledging the power of audience participation [21]: "Though the mass audience can be used as a creative participating force," he wrote, "it is, instead, merely given packages of passive entertainment."

Cinematrix has continued to produce commercial shows since its watershed event in 1991, but there has been little other development in the realm of audience interaction technology in the last decade. This is surprising given the large number of venues in which audience participation techniques could be applied, including movie theaters, science museums, and theme parks. Many techniques, like the ones described in this paper, can be implemented robustly and inexpensively, and can meet with success when presented to live audiences. In our experience, the greatest challenge lies not in developing the technology for audience interaction, but in designing engaging activities.

6. Future Work

Template-based audience movement tracking works well with activities other than leaning. One example application, currently under development, is a pinball game in which audience members control the pinball paddles by raising and lowering their arms. Another application idea is a racing game in which the left side of the audience competes against the right side of the audience, moving their bodies to control the jumping of a hurdler.

Shadow tracking can be equally effective with objects other than beach balls. We are currently building a game that allows the audience to raise and lower an onscreen object by "holding it up" with shadows. When this game is shown in a large auditorium, the audience members in the front row are given pool noodles (long, cylindrical Styrofoam rods) that they hold in front of the screen to create shadows. The top of each shadow is used as the control point on a spline defining a 2D surface. As a ball rolls and bounces on this surface, the audience attempts to modify the contours of the surface so that the ball hits a moving target.

Augmenting laser pointers with twist-on lenses allows them to display a variety of interesting laser patterns. These patterns could be used as the basis for a glyph recognition system that distinguished between laser pointers, allowing different members of the audience to play different roles in an interactive game. Alternatively, laser dots could be distinguished by using lasers at different frequencies, such as green lasers and infrared lasers.

Real-time sound processing [7] offers numerous possibilities for interactive audience participation. By positioning microphones throughout the auditorium and applying techniques of sound localization, pitch and beat detection, and amplitude measurement, various interaction techniques could be devised based on levels of applause, locations of loudest applause, and timing and pitch of audience shouts and singing. This could lead to call-and-response style games, audience polling, sing-along games, and “pep rally” activities that encourage crowd enthusiasm. This approach would be similar to the applause meters at sporting events that measure the noise of the crowd in decibels.

Our applications were primarily entertainment based and were focused at audiences of college students, but we believe that methods for audience interaction can be broadened to a wide variety of educational and professional applications. Cinematrix has successfully used its system at planetarium exhibits, trade shows, and company seminars. We would like to see further development of audience interaction techniques in these domains.

References

[1] Baer, R., Television Gaming and Training Apparatus. US Patent #3728480, 1973.

[2] Bal-shalom, Y. and Fortmann, T.E., Tracking and Data Association, *Academic Press*, 1988.

[3] Blaine, T., The Outer Limits: A Survey of Unconventional Musical Input Devices, *Electronic Musician*, August 2000.

[4] Carpenter, L., Cinematrix, Video Imaging Method and Apparatus for Audience Participation. US Patents #5210604 (1993) and #5365266 (1994).

[5] Cassaro, N., Tony and Tina’s Wedding (Theatrical Production), New York, 1985.

[6] Chen, X. and Davis, J., LumiPoint: Multi-User Laser-Based Interaction on Large Tiled Displays, *VisSym 2001*.

[7] Dannenberg et al, Stadium Games, Method and Apparatus for Interactive Audience Participation by Audio Command. US Patent #798382, 1997.

[8] Davenport, G., Seeking Dynamic, Adaptive Story Environments, *IEEE Multimedia*, vol.1, no.3, Fall 1994.

[9] Aggarwal, H.K. and Cai, Q. Human Motion Analysis: A Review, *Computer Vision and Image Understanding: CVIU*, 1999.

[10]DeLeeuw et al, Reality Fusion, Method of Using Video Reflection in Providing Input Data to a Computer System. US Patent #6088018, 1998.

[11]Gavrila, D.M., The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding: CVIU*, 1999.

[12]Eckert, R.R. and Moore, J.A., The Classroom of the 21st Century: The Interactive Learning Wall, *SIGCHI Bulletin*, 2000. 23(2): pp. 33-40.

[13]Evans & Sutherland StarRider Digital Dome theater system, used in Chicago’s Adler Planetarium. Described in Evans & Sutherland Computer Corporation, SEC Form 10-K Annual Report for the fiscal year ending December 31, 2000.

[14]Freeman, W. T. et al, Computer Vision for Interactive Computer Graphics, *IEEE Computer Graphics and Applications*, Vol. 18, No. 3, May-June 1998

[15]Freeman, W. T. et al, Computer Vision for Computer Games, *2nd International Conference on Automatic Face and Gesture Recognition*, Killington, VT, USA.

[16]Jain, R. et al, Machine Vision, McGraw-Hill, 1995.

[17]Krueger, M., Artificial Reality II, Addison Wesley, l991.

[18]MacDougall, F., Vivid Group, Video-based Interactive Gesture Control of Computer Processes. US Patent #5534917, 1996.

[19]Maes, P. et al, A., The ALIVE System: Wireless, Full-Body Interaction with Autonomous Agents, *ACM Multimedia Systems*, 1996

[20]Malone, T., What Makes Things Fun to Learn: A Study of Intrinsically Motivating Computer Games, Ph.D. Dissertation, Stanford University, 1980.

[21]McLuhan, M., The Medium is the Massage, Bantam Books, 1967.

[22]O’Brien, R., The Rocky Horror Picture Show (Film), 20th Century Fox, 1975.

[23]Olsen, D. and Nielsen, T., Laser Pointer Interaction, *ACM SIGCHI Proceedings*, 2001.

[24]Quay, A. et al, GhostWriter System, *International Symposium on Wearable Computers*, 2000.

[25]Reynolds et al, An Immersive, Multi-User, Musical Stage Environment, *ACM SIGGRAPH Proceedings*, 2001.

[26]Ulyate, R. and Bianciardi, D., The Interactive Dance Club: Avoiding Chaos In A Multi Participant Environment, *CHI 2001 Workshop on New Interfaces for Musical Expression*, 2001.