IASDR07
INTERNATIONAL ASSOCIATION OF SOCIETIES OF DESIGN RESEARCH THE HONG KONG POLYTECHNIC UNIVERSITY 12TH TO 15TH NOVEMBER 2007

Real-time Interactive Media Design with Camera Motion Tracking

Jieun Kwon[1] and Yuncheol Baek[2]

[1]MFA Computer Art, School of Visual Arts, New York, USA, daminian@gmail.com

[2]Divison of Computer Software, Sangmyung University, Seoul, Korea, ybaek@smu.ac.kr

ABSTRACT:

Interactivity using camera motion tracking is increasingly an active, effective and even entertaining way for users who use digital communication such as motion picture, internet and games in a multimedia era. The goal of this paper is to research the technology and to implement real-time interactive media with camera motion tracking and to apply this to design communication. In this paper we discuss the possibilities and advantages of using computer vision and the limitations and points of issue that the algorithm of camera motion tracking has still in various design fields.

KEYWORDS

Interactive design, interactive installation, camera motion tracking, computer vision

1. INTRODUCTION

The idea of motion tracking using a video camera has been around a couple of decades, and researched in a field known as "Computer vision". The creation of computer vision systems has been considered the exclusive domain of expert researchers and engineers in the fields of signal processing and artificial intelligence. Recently, however, new method of interaction has been approached by various areas of developing technology. Historically, Myron Krneger created some of the first body interaction based camera in design field. He used the computer and a camera to create a real-time relationship between the participants' movements and the environment. His project, "VIDEOPLACE(1975)" coordinated the movement of a graphic element with the actions of the audiences. It continues with Bregler(1997), Metaxas(1995), Rehg(1999), amongst others, who have produced 2D and 3D body tracking work towards a camera in order to understand the body in motion. (Flavia Sparacino 2001)

Tracking motion systems are surely an attractive mechanism to enable interaction, even in the design field. However still it is not exactly perfect, because the construction of vision is a complex subject. This paper focuses on real-time interactivity of the motion tracking for novice programmers and for users of camera-based detection techniques in interactive design. The first part of this paper introduces good examples of the ways in which motion tracking has been implemented in design applications. These examples are classified by analyzing the process of interaction with a camera. Secondly, basic techniques of motion tracking system are presented. The algorithms, physical environments and tools are demystified in the design field. Finally, we suggest capabilities and advantages of motion tracking. Furthermore, limitations and disadvantages of this system will be mentioned that will help a designer in designing the interaction with camera motion tracking.

2. ANALYZE INTERACTIVE PIECES

These examples are interactive pieces using camera motion tracking. We classified them by the way of interaction to human with regard to the object detection technique, and we analyzed their characteristics.

2-1. Detect location

(1) Reactrix – Commercial

Description

This interactive work is a place-based advertising medium made by Reactrix. These advertisements respond to the physical movements of the audience and engage them in true brand interaction. This system offers a shared experience between those interacting with the display and those watching in shopping centers, movie theaters and other public spaces. When a passenger passes by the projection screen on the floor, graphic elements of the commercial are changed depending on his or her movement. For example, the shape of the commercial product on the screen is distorted by passengers' location or the graphic ball is controlled by their movement.

Analysis

Most of this system displays the visuals on the floor in public place. A camera and a projection are hung on the ceiling to capture the moving passenger and show the visual. This system makes the customers easily recognize the commercial. This media provides them entertainment and a new, higher engagement through interaction.

Figure1, 2. Reactrix company's interactive commercial on the floor.

(2) Camille Utterback – "Text Rain (1999)"

Description

"Text Rain" is an interactive installation in which participants use their entire body to lift and play with falling letters that do not really exist. The audience interacts with the installation by standing or moving in front of a large projection screen. On the screen the text responds to motion and can be caught, lifted and released to fall again. If they accumulate enough letters along their outstretched arms, or along the silhouette of any dark object, they can read words and phrases formed by the falling letters.

Analysis

This installation uses a projection screen, computer and a camera on the center of the screen. It's very easy to draw participation from the audience because the video projection shows a mirrored reflection of themselves in black and white, combined with a color animation of falling letters. It doesn't matter how many people come into the view of camera. However, the installation is influenced by specific colors worn by the participants. For example, if the participant wears white clothing, the letters are not impeded by the body but fall through it on the video screen.

Figure3. Camille Utterback, "Text Rain"

(3) Lofael Lozano-Hemmer – "Standards and double standards (2004)"

Description

This interactive installation consists of fifty fastened belts that are suspended at waist height from stepper motors on the ceiling. Controlled by 4 cameras tracking system, the belts rotate automatically to follow the public, turning their buckles slowly to face passersby.

Analysis

In this piece, the artist wants to visualize complex dynamics, turning a condition of pure surveillance into an unpredictable connective system. His installation consists of mundane objects, such as belts and motors. The audience can see the third object detect them and hear the rotating sound of the motor as a result of their action. However, when a single person comes to the space, the closest belts turn their buckles to face the individual. When many people are in the room their presence affects the entire group of belts, creating chaotic patterns of interference.

Figure4: Lofael Lozano-Hemmer, “Standards and double standards”

(4) Ruth West, Jeff Burke, Cheryl Kerfeld, Eitan Mendelowitz, Tom Holton, JP Lewis, Ethan Drucker, and Weihong Yan - Ecce Homology (2003)

Description

This project is an interactive environment presenting a novel visualization of genomic data and bioinformatic processes. Through slow and gestural whole-body exploration of the pictograms, a visitor performs a scientific experiment looking for evolutionary relationships between the human and model organism genomes. Results of visitors' interactions, which initiate the automated comparisons of the human and rice genomes, are shown through changes in the calligraphic figures. (Ecce Homology, http://www.insilicov1.org/default.htm,)

Analysis

"Ecce Homology" shows the visual result on a huge black surface from a front-projection. The images are presented by 5 projectors working with computers, servers and video cameras. The experience is responsive to the characteristics of the movement of each visitor. The more rapid one's motion, the more transient the luminous traces. Slower motions create a more sustained the luminous trace. This helps establish an aesthetic of slowness leading to a sense of presence and contemplation for visitors.

Figure5: Ruth West, Jeff Burke, Cheryl Kerfeld, Eitan Mendelowitz, Tom Holton, JP Lewis, Ethan Drucker, and Weihong Yan, " Ecce Homology".

2-2. Detect Figure (Silhouette)

(1) Scott Snibbe – "Shadow(2002-2003)"

Description

Scott Snibbe's work *Screen Series* allows viewers to create cinema with their bodies, either through reactive projections that respond to viewers or through porous projections that record viewers' movements. His works range from large-scale body-centric physical installations to interactive sculpture and screen.

Analysis

These works likewise emphasize viewers' shadows, rather than their exact representations. This emphasis on shadows paradoxically creates a stronger integration of viewers' bodies with the projections, since a picture of a viewer's shadow is almost identical to the shadow itself, while a picture of a viewer's body is less similar to their actual three-dimensional form. The contemporary combination of camera, projector, and computer echo early cinema cameras, with computer "processing" replacing chemical processing.

Figure6: Scott Snibbe, "Shadow."

(2) Daniel Rozin – "Wooden Mirror(1999)"

Description

Interactive installation, "Wooden Mirror" reflects on its surface any visitor that stands in front of the piece. A board is covered in hundreds of 1.5 inch square pieces of wood, and each is controlled by its own servo motor. The camera looks at the visitor, splits the image up into the wooden pixels that comprise the board, and the motors then adjust the wood squares to create the likeness of the person. The visitor hears rotating sound of the motors as well.

Analysis

This explores the line between analog and digital. It reflects any person in front of it by organizing the wooden pieces and moving the pieces fast enough to create live animation. The simple and immediate interaction between the viewer and piece harmonizes the viewer's interest and understanding. The sound is directly connected to the motion of the person. Mirror provides a pleasing secondary feedback to the image and helps the viewer's interaction. However, to make this installation, large wooden blocks are used to represent pixels. Some visitors who stand in front of the piece very closely can't recognize their figure on the big Mirror. If they stand some distance away from the piece, they realize their image is in the Mirror.

Figure7: Daniel Rozin, "Wooden Mirror."

(3) Jieun Kwon – "Your Destiny(2007)"

Description

This project entitled "Your Destiny" is an immersive, interactive installation based on tarot cards. When the audience comes into the room, they first see a projection of the backside of the tarot cards on the screen. A camera recognizes the visitor's movements. Tarot cards will flip over in the shape of the visitor's silhouette, revealing the front of the tarot cards. A pre-recorded voice will be heard by the audiences as the tarot is displayed, voicing the symbolic meaning of each revealed card. As the visitor moves and creates new silhouettes, a new display of front facing tarot cards and voices are revealed.

Analysis

Through programming, this work responds to the viewer's movements and form and shows the digital image randomly and vertically within their figure on the screen. It allows visitors various possibilities of interaction through image, animation, and sound. The audience enjoys colorful and varied visualizations as a result of their movement. In the physical environment, some lights are needed to capture the viewer's action by camera. This piece is very sensitive to lights. If the lighting condition changes during its operation, it has to be re-initialized.

Figure8: Jieun Kwon, “Your Destiny.”

(4) Animaatiokone Industries – “Kick Ass Kung-Fu(2003-2005”)

Description

Kick Ass Kung-Fu is an immersive game installation that transforms computer gaming into a visual, physical performance through movements like jumping and kicking. The participants can fight and defy gravity like kung-fu movie actors with real-time embodied interaction and virtual set technology.

Analysis

Animaatiokone industries explained that a players starts by standing in a 5 meter cushioned playfield and then does battle against virtual enemies. A video camera captures the image of the player in real-time and inserts this image on two screens in a virtual environment. This game immerses the participants in an active, not passive, performance. They can make unlimited actions and special poses because they don’t have to have any type of motion tracing devices connected to their bodies for the computer system to create a 3D image of the players.

Figure9: Animaatiokone Industries' Kick Ass Kung-Fu.

2-3. Sound Visualization

(1) Zachary Lieberman – RE:MARK

Description

This piece is an interactive installation of sound visualization with its users' speech. The installation screen shows a combination of textural and graphic shapes emerging from the shadow of the participants' head. The vocal sounds spoken into microphones are analyzed and classified by a phoneme recognition system. If the system's classifier does not recognize the participant's sound, then an abstract shape is generated. (Golan Levin, Zachary Lieberman 2004.)

Analysis

In this project, the sound analysis system works in tandem with a computer vision system, which estimates the location of a speaker's mouth by selecting a point somewhere above the center of their silhouette. This piece has an interactivity that combines sound with an animated visual. It allows limit how many people can participate because this interaction is connected with a specific person who speak into a microphone, not passive moving people in front of the screen.

Figure10: Zachary Lieberman, “RE:MARK.”

2-4. Dance Performance

(1) Troika Ranch

Description

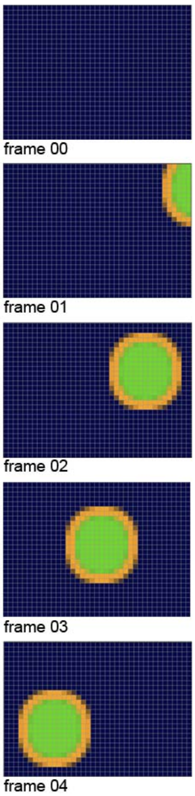
This event stages an interactive dance performance. When the performers in Troika Ranch move, their dancing costumes send the data to computer, triggering visual and musical accompaniment. This demonstration is a motion tracking system that enables the movements of dancers to generate, in real time, images, graphics and sounds that are projected on a screen behind the dancer.

Analysis

The stage works through projections and sounds that are created through a specific process of interaction between live performers and digital media. The result is a motion tracking system that has expanded to the stage design field. Especially, this work responds to the performance like professional dancer, not audience. The audience sees a performance that is an interaction between a dancer and the stage. Even though the interaction is a random response that depends on the dancer’s movement, the designer of the performance can plan for and expect a detailed interaction result.

Figure11: Troika Ranch's Dance Performance

3. TECHNOLOGY OF INTERACTIVE SYSTEM FOR CAMERA MOTION TRACKING



Camera motion tacking is founded on video and computer vision. It has to have subroutines that define the characteristics in the array of pixels that comes from a camera. A video frame is one of the many still images that compose the complete moving picture. The process of rendering each frame of video creates a sequence of images the viewer sees as one continuous motion. Each frame is composed by many pixels, which have color. Many computer vision algorithms are linked to the task of distinguishing pixels that is including color. Colors to be tracked need consistent lighting. The frame rate, the rate at which sequential frames are presented, are connected to the frequency of motion tracking.

Figure12: The process of rendering video frames.

The camera captures the scene including the audiences and this data is input to the computer. This video data is organized as frames and are distinguished by a programming tool. The motion is tracked by distinguished frame and presented as a specific visual by designer's idea, and output on a screen or monitor.

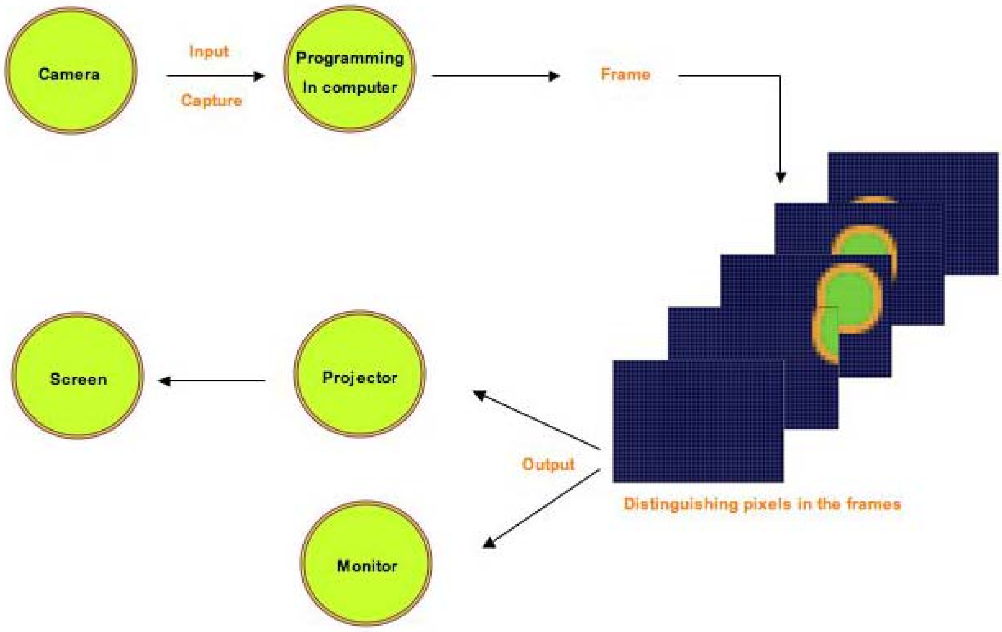


Figure13: The process of camera motion tracking.

When creating motion tracking applications, that designer may apply the following methods: object tracking, frame differencing, background subtraction and brightness thresholding. These are basic detection methods and have been used in the examples that I mentioned in first section.

3-1. Basic Image Processing Algorithm

(1) Frame differencing

Moving people may be tracked using frame differencing. This technology continuously compares the color value of each pixel between a certain video frame and the subsequent frames. The difference in color value between these two pixels is a measure for the amount of movement in that particular location. This value can be calculated by color value like RGB or HBS, or only by brightness. The frame differencing application depends on a stable lighting environment in the working space because the color is derived from the light interacting in the eye. It means that the camera tracking system always needs suitable light even when the projection screen is presented properly under the dark condition.

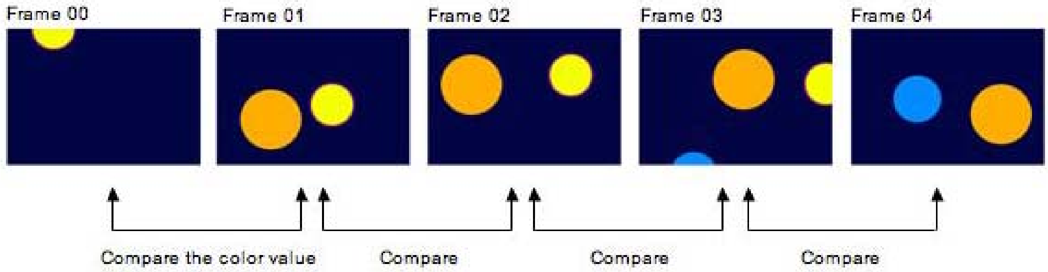


Figure14: Frame differencing.

(2) Background subtraction

The background subtraction algorithm compares the frame with the presence of people or other object and a specific frame stored as a background. This algorithm needs to capture at a point in time when the scene was known to be empty first and then subtract it from each subsequent frame. If the absolute value of the difference between a pixel in the background image and the corresponding pixel in the current frame is over a given threshold, it is considered movement. Using this technique, heterogeneous background environments can be used and stored by the algorithm as long as the camera remains stationary and the background is kept in the same lighting conditions during its working. If moving people or objects have the same color as the background, this algorithm cannot detect the motion.

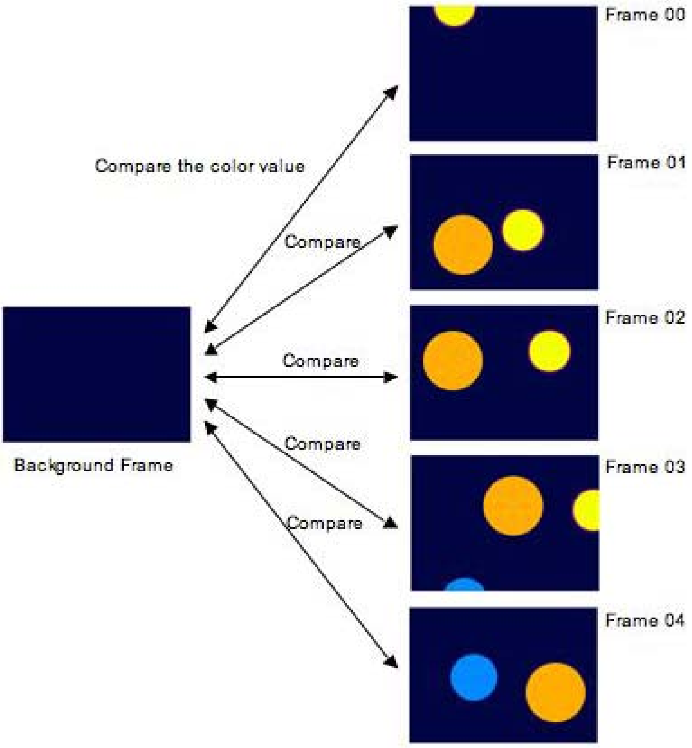


Figure15: Background subtraction

(3) Brightness thresholding

Thresholding is the simplest method of image segmentation. Typically, an object pixel is given a value of “1” while a background pixel is given a value of ”0.” Brightness thresholding distinguishes considerably lighter or darker values. In this method, the illumination of backlighting or surface treatments such as retroreflective surfaces is effective in ensuring the difference of the brightness values.

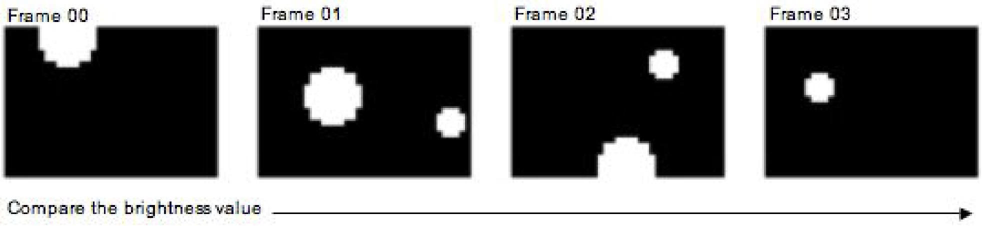


Figure16: Brightness thresholding

3-2. Physical circumstances

These image processing algorithms used to detect motions are applied to the physical world. Each detection algorithm is critically dependent on certain unique assumptions about the analog scene. That's why we have to consider the circumstances in which image processing algorithm is used. Some of them have the possibility of being unstable or can fail to detect motions depending on physical circumstances.

(1) Lighting

Cameras work with the light of the visible spectrum. If there is no considerable light, a camera cannot recognize objects in the scene. However, excessive lighting can disrupt a projection screen or monitor. The visual result depends on the quality of the projection or monitor's type of camera used. Appropriate light condition keep motion detected elements from breaking up and help remove the surrounding area of working piece.

Especially, in the background subtraction, if one fails to keep the same lighting condition while the piece is running, in the case that the background scene's color may change and differ from the original background color that the program had originally memorized, the program may need to be initialized again to capture a new background frame.

(2) Camera

We can use a general video camera, web camera or any special camera (which may have a special function like an infrared camera), or a surveillance camera to capture the images. They can be connected to computer directly by fire-wire cable. Considerations must be made on the kind of computers and devices to be used when designing the concept and the presentation.

An infrared camera, which can detect infrared rays that is invisible to the human eye, can supplement the light detected by conventional black and white security cameras. It enables capturing the scene of a low light environment or even totally dark environment because ambient light level does not matter. This camera is useful for tracking the location of moving people, even in the darkness, however, it is not good at detecting the exact silhouette of objects because its resolution is considerably lower than an optical camera's.

(3) Display

Finally, tracked motion is displayed on the screen, or some surface such as floor or wall. It is geared with the light because visual result is shown by it as well. A monitor is disrupted less by ambient light than a projection screen and has good resolution, but it has limited size. A projection can create a bigger screen size, however, it's possible to see the visuals not very well because of bright ambient background lights. In this case, we can use the high illuminating projector or special screen material, which allows good quality image or video. If the camera is placed in the opposite end of the front projector, the motion capture may be impeded by the flickering projector lights. A rear projection screen helps the camera capture the participants as escaping the projection light that interrupt.

Retroreflective materials are remarkably efficient at reflecting light back towards their source of illumination, and are ideal aids for ensuring high-contrast video of tracked objects. (Golan Levin 2004)

3-3. Programming

Any programming language including C++, Java, Lingo and Action Script used to make motion tracking ask the programmer to provide direct read-access to the array of video pixels obtained by the computer's frame-grabber. Many programming languages can be used to do motion detection but, in our experiences, visual programming environments such as Max/MSP/Jitter, Isadora or Processing make it easy to understand and implement for a designer, not programmer. For instance, Processing provides a video library to track motion. The designer can apply this library to present his or her ideas. Max/MSP is a graphical development environment for music and multimedia developed and maintained by the company Cycling'74. It is very powerful in real-time video performance.

3-4. Applied Example

In the installation "Your Destiny", the concept involves detecting the silhouette of moving people and shows tarot card images within that silhouette. For this idea, background subtraction algorithms and Processing was effective for building the interaction between the audience and the camera motion tracking. This algorithm is useful in storing the background in any place where this piece is running. Even though it becomes unstable during certain circumstances such as changing lighting conditions, it is still easy to initialize background subtraction. Processing is a powerful tool capable of creating multiple concurrent interactive images.

Physically, Mac, fire-wire web camera, and front projection screen were used. Web camera has high quality enough to capture silhouettes and it is good to use because of their inconspicuous size. In the case of using front projection, however, we should be careful about the position of the camera because it may be affected by the flicking lights from the projection. In this piece, the camera was placed in a low position, and not in the middle of the screen so as not to bother the projector's field of view. If we used the rear projection screen, camera placement would be more flexible. The light position and brightness are also important in this case. We used two spotlights on both sides of the screen, one toward the audiences and sub-light to remove the shadow of people on the back side wall.

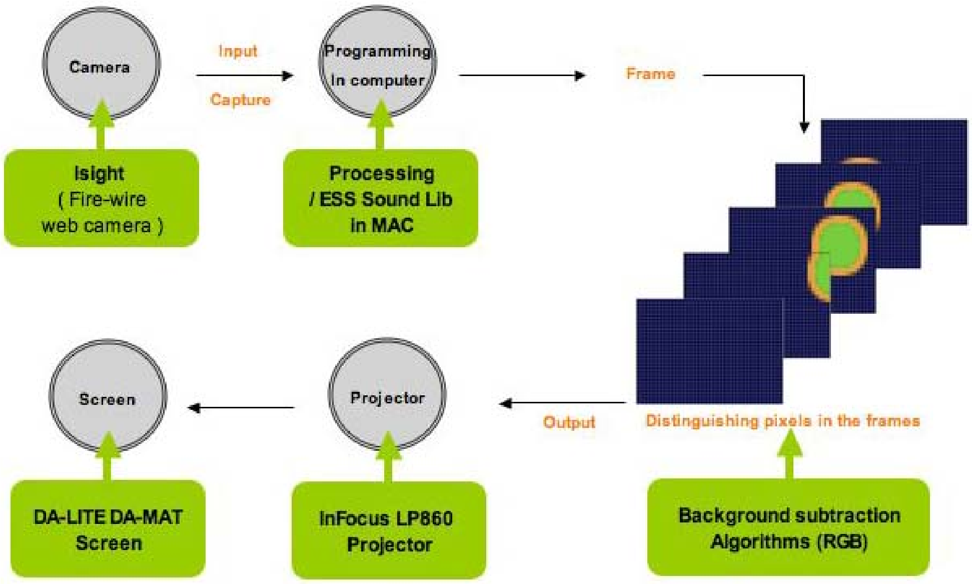


Figure17: Applied component of motion tracking in "Your Destiny."

4. CONCLUSIONS

Camera motion tracking allows single or multiple visitors to participate simultaneously. It provides real-time detection and tracking through interactive communication between the designed piece and user. Tracking enables a passive and active interaction with the audience. Increasing the participation and the real-time interaction helps bring interest and creates an impression with the audience. The audience's interactive experience is linked to the designer's goal and idea directly.

Even though a motion tracking system has attractive factors and benefits to make an interesting interactive design, it is still not perfect. With some unexpected variables, it can be unstable in the working space. There are limitations which include the state of the running environment, and the type of hardware, software, and coding used in programming. Participants' movement space is confined by the locations of camera and projector. These two devices are always interrelated. The space is also limited by lighting conditions because image processing techniques need a stable background in which to operate.

Therefore, we have to consider some issues before designing or creating a camera motion tracking system.

(1) Consider the suitable physical circumstance and scale to express the idea and concept for design.

(2) Select the tracking algorithm to best negotiate physical conditions presented in real space.

(3) Predict cases of fallibility and make plans to reduce it.

Accordingly, if a designer takes advantage of motion tracking with camera and makes considerate observations, in our opinion, it creates interesting interactive effects and enhances the user's experience.

REFERENCES:

Computer Vision for artists, http://webcamxtra.sourceforge.net/index.shtml

De Oliveira, Nicolas, Nicola Oxley and Michael Petry(2004). Installation Art in the New Millennium: The Empire of the Senses. London: Thames & Hudson, 2004.

Ecce Homology, http://www.insilicov1.org/default.htm.

Flavia Sparacino (2001) (Some) computer vision based interfaces for interactive art and enterainmnet installations, Paris, France, 2001.

Golan Levin (2004) Computer Vision for Artist and Designers: Pedagogic Tools and Techniques for Novice Programmers, Pittsburgh, USA, December 30, 2004.

Golan Levin, Zachary Lieberman (2004) In Situ Speech Visualization in Real-Time Interactive Installation and Performance, USA 2004.

Jonah Warren (2003) Unencumbered Full Body Interaction in Video Games, New York City, USA, April 2003.

Manovich, Lev(2002) The Language of New Media. Cambridge, MA: The MIT Press, 2002.

Paul, Christiane(2003). Digital Art. London: Thames&Hudson, 2003.

Quoc Nguyen, Scott Novakowski, Jeffrey E. Boyd, Christian Jacob, Gerald Hushlak (2006) Motion Swarms: Video Interaction for Art in Complex Environments, 2006.

Rafael Lozano-Hemmer, http://www.lozano-hemmer.com/eprlh.html

Rozin, Daniel, http://www.smoothware.com/danny/

Reactrix, http://www.reactrix.com

Snibbe, Scott. 'Screen Series'. Scott Snibbe. Retrieved 2006, http://www.snibbe.com/

Utterback, Camille. 'Text Rain.', Camille Utterback. Retrieved 2006, http://www.camileutterback.com

Wands, Bruce(2002). Digital Creativity: Techniques for Digital Media and the Internet. New York: John Wiley & Sons, Inc., 2002.

Wardrip-Fruin, Noah and Nick Montfort(2003), The New Media Reader. Cambridge, MA: The MIT Press, 2003.

Wilson, Stephen(2003). Information Arts : Intersections of Art, Science, and Technology. Cambridge, MA: The MIT Press, 2003.