CHAPTER

# Computer Vision Interfaces for Interactive Art

# 2

**Andrew W. Senior**[1]
*Google Research, New York*

**Alejandro Jaimes**[2]
*Telefonica Research, Madrid, Spain*



## ABSTRACT

*This chapter reviews the use of computer vision as an interface for the arts. We provide a taxonomy of vision-based art works and describe the major paradigms they follow. We also examine how the availability of inexpensive hardware and powerful free software tools is fueling an explosion in the use of computer vision for art interfaces, driving experimentation with new interface technologies, and enabling a new expressiveness in both performance and audience-interactive art.*

[1]aws AT andrewsenior.com
[2]aj27 AT columbia.edu

*Key words*: computer vision, multimedia art, interactive art, tracking, performance, mirror interfaces, projector-camera systems.

## 2.1 INTRODUCTION

Vision is, for most people, the primary sense—our sense of presence in the world is most closely tied to the perceptions we receive through our eyes. Computer vision, by attempting to give computers the same sense of sight, opens up a new range of possibilities for sensing at a distance and for perceiving the world as humans do. For interactive art, vision is a unique modality by which a work can modulate its behavior according to conditions in its environment, whether by explicit user controls or by interpretation of visual occurrences without obvious causality.

This concept is directly related to the concept of ambient intelligence, with a twist: In art, reactions to the work need not be "intelligent" in the same sense that they might have to be, for example, in a digital home. "Ambient intelligence" reactions in interactive art are the artist's constructs and, as part of the work, can range from simple abstractions to complex physical movements of robots. Furthermore, such works are often implicitly human-centric from start to finish because they are conceived as part of an artistic concept. Computer vision in art is seldom discussed in ambient intelligence circles, but it is significant because some of the artworks and tools for art in many ways push the limits of technology in ambient spaces.

In this chapter we briefly introduce computer vision in the arts, with the goal of generating enough interest for the reader to explore further. There is a vast amount of information on the subject (from both art-critical and technical perspectives), as well as a large number of open-source software projects, artworks, tools, and performances that use computer vision. These "applications" of computer vision employ techniques ranging from simple background subtraction to state-of-the-art face detection, facial expression recognition, tracking, and many other technical subareas of computer vision. One key difference between the types of interactive arts we describe in this chapter and the technical work published in academic venues is that the systems/artworks are installed and "used" by many people (with no training) in public spaces, exhibitions, performances, and the like. This creates many real-world challenges for computer vision (and ambient intelligence) researchers, and great opportunities to "test" new technologies.

### 2.1.1 A Brief History of (Vision in) Art

Throughout history, art has been driven by technological change. New techniques and materials engender new styles of artistic expression. Computer vision is just one of the newest tools and is similarly powering changes in the arts as well as being a significant component in the burgeoning disciplines of new media and interactive art. As a sensing modality rather than a sensible medium, computer vision is not itself an art medium but rather a powerful tool that can be applied in many art forms, from photography and sculpture to dance and music.

One can find many precursors of computer vision art, such as the work of Muybridge [1], who used photography to understand human actions, and the original video games developed in the 1950s and 1960s. Electronic image sensors from the 1920s and 1930s through Bell Labs' CCD sensor introduced in 1969 laid the foundations for video art, such as the work of Nam June Paik. However, VideoPlace, introduced by Myron Krueger [2] in the 1970s, is often cited as the earliest true interactive video work of art (see Figure 2.1).

> *The VideoPlace System combines a participant's live video image with a computer graphic world. It also coordinates the behavior of graphic objects and creatures so that they appear to react to the movements of the participant's image in real-time. [3]*

Computer vision in art also has a strong historical link to performance—in particular, to the early happenings and multimedia performances of the 1960s (see Rush [4] for a general introduction to new media) or even to film (see Manovich [5] for a "film" perspective on new media). Vision in art can also be related to the concept of an active observer in "traditional" art, in which the main premise is that people "interact" with art (e.g., oil paintings) even if a work is not explicitly interactive.

Historical details are beyond the scope of this chapter. It may be sufficient to say that, although there are many resources on the history of new media, we are not aware of many that specifically focus on computer vision in the arts. (Levin's article [6], included in [7], is a notable exception.) Theoretical analysis of new media in art (including computer vision) is in the very early stages, and there are many active discussions on it in the arts community. As is natural, many of them occur online (see Section 2.6).

## 2.2 A TAXONOMY OF VISION-BASED ART

If we examine the work that uses computer vision techniques, we can see that a number of categories emerge. In this section we describe a taxonomy of computer-vision–based work and illustrate it with some notable examples. In doing so we identify several factors that we can use to categorize vision-based art. These are shown in Table 2.1.

The first division in our taxonomy is between works that use "live" vision and those that do not—that is, if images are captured in real-time and processed to generate an effect or whether existing images are somehow analyzed. (Examples of such noninteractive works are the *turbulence.org* commission "self portrait"[3] by Ham, which uses face recognition software to search for the artist's portrait on the flickr photo archive Website, and "Fingerbahn" (1998) by Levin,[4] which uses computer

[3] *transition.turbulence.org/Works/self-portrait/*.
[4] *acg.media.mit.edu/people/golan/photo/ps7/index.html*.

**Table 2.1** Taxonomic Categories for Interactive Computer Vision Art

| |
|---|
| Live/stored images |
| Local/remote |
| Pixels/regions |
| People/object |
| Active/passive |
| Public/performers |
| Model-based/model-free |
| Location/shape/motion features |

vision techniques to extract the ridges from a fingerprint still image and then adds virtual creatures that move along the ridges.

Among the works processing live video feeds, we can distinguish those that acquire the video locally—that is, in the same place as the viewer—from those that process a live feed from another location, whether outside the gallery or on the other side of the world. (For example "Suicide Box" by the Bureau for Inverse Technologies (Jeremijenko) placed a video camera below the Golden Gate Bridge to detect jumpers). In particular, among those with locally acquired live feeds are works that use computer vision to provide interactivity. We will examine these in more detail in this chapter.

We can distinguish levels of interpretation of the image. Some focus mainly on the pixel level, although these, such as "Light Tracer" by Willis (2005), could be said to use image-processing rather than computer vision algorithms. That work retains the maximum brightness of each pixel from a scene over time, producing an interface in which participants can draw—with their own bodies or with lights, cell phones, and so forth. "Couch Potato Farm" by Senior (2005)[5] uses motion from a live video feed together with pixel intensities in television signals to create artificial life forms in a hypothetical virtual ecosystem.

Like VideoPlace, some works extract edges or regions using simple thresholding or color keying. More complex algorithms such as background subtraction allow greater robustness to varying conditions. Finally, some works use high-level vision algorithms such as face detection or recognition. Vision-based motion capture technologies, commonplace in "offline" art production, also find their way into performance systems, allowing many degrees of freedom in human body poses to be captured, interpreted, and rerendered. Another way of looking at this

[5] *andrewsenior.com*.

**FIGURE 2.1**

Krueger's VideoPlace.

increasing complexity is according to the level of modeling used by the vision algorithm. More complex effects such as object detection and tracking may use sophisticated modeling of an object (face, person, vehicle) rather than more generic model-free processing like background subtraction or motion detection.

We can also categorize vision-based interfaces according to subject matter. The subject of most works is people, but some are designed to detect particular (e.g., color-keyed) objects. In one version of "Audience Interaction" by Maynes-Aminzad [8] a beach ball is tracked by a camera as it is tossed around by an audience. Its position controls a video game. In another version, the audience leans to control a game. Other works are agnostic about their subject matter, looking for silhouettes or other visual features wherever they may occur. Of those that look at people or human activity we can distinguish the ones that face "the audience" or gallery-goers from those in which the cameras are directed at the performers, although much of the motivation of interactive art is precisely to blur this distinction.

## 2.3 PARADIGMS FOR VISION-BASED INTERACTIVE ART

Although we can classify individual works according to the preceding taxonomy, it is also possible to identify several major paradigms for computer-vision–based interactive art:

- Mirror interfaces
- Projector-camera systems
- Physical actuation
- Dance performance
- Musical performance

In this section we describe these paradigms, illustrating them with notable works.

### 2.3.1 Mirror Interfaces

A principal paradigm in interactive art is the presentation of a mirror-like surface to the viewer, representing the image seen by the camera. A striking series of mirror works devised by Danny Rozin[6] uses a variety of physical effects to produce light and dark pixels—from tilted wooden tiles in "Wooden Mirror" (1999) to retractable ball bearings in "Shiny Balls Mirror" (2003) (Figure 2.2). Rozin also produced a number of software mirrors that represent the image of the camera in a variety of ways. These simple mirror interfaces are, strictly, based on image processing rather than computer vision, since they use low-level, pixelwise operations. A work that uses a mirror paradigm while exploiting computer vision techniques is "Magic Morphin' Mirror" (1997) by Darrell et al. [9][7] (Figure 2.3). This presents the image captured by the camera as it would appear in a mirror except for faces, which, as one watches, begin to distort and warp. Here the faces are detected and tracked by a face detection algorithm, applying a complex pattern recognition system to control the mirror interface.

**FIGURE 2.2**

Rozin's "Shiny Balls Mirror."

[6] *www.smoothware.com/danny/*.
[7] *people.csail.mit.edu/trevor/Hallucinations.html*.

**FIGURE 2.3**

Darrell et al.'s "Magic Morphin' Mirror."

Another paradigm for visual art works is the camera-projector system. Such systems have been studied for a number of years as an interaction paradigm for computers [10]. They are often configured as augmented reality environments in which the projector projects an image into the space of the participants, which is the same area viewed by the camera. One early example of this is "Text Rain" (1999) by Utterback and Achituv,[8] in which falling letters gather like snow on the silhouette of the participants (Figure 2.4). A more recent work with a similar format is "Messa di

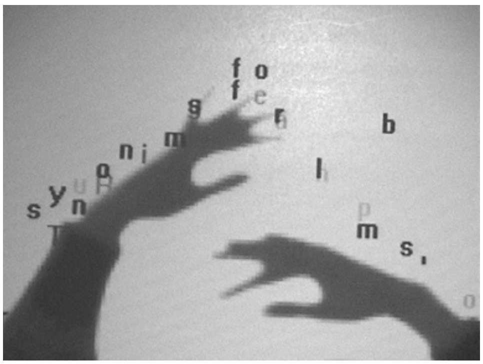

**FIGURE 2.4**

Utterback's "Text Rain."

[8] *www.camilleutterback.com/textrain.html*.

Voce" (2003) by Levin and Lieberman, with Blonk and La Barbara,[9] in which sounds from a microphone are visualized, for instance, as bubbles rising from the actors' heads which are tracked visually (Figure 2.5).

Camera-projector systems can be oriented horizontally as well as vertically. For example, "Boundary Functions" (1998) by Snibbe[10] (Figure 2.6) detects people with a camera and projects boundaries between their locations according to a Voronoi tessellation of the space. As pairs of participants move in the space, a line always

**FIGURE 2.5**

Jaap Blonk performing "Messa di Voce" by Levin et al.

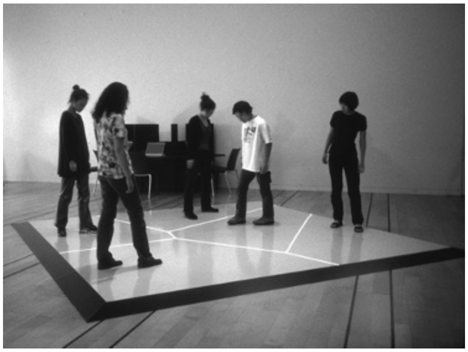

**FIGURE 2.6**

Snibbe's "Boundary Functions."

[9] *tmema.org/messa/*.
[10] *www.snibbe.com/scott/bf/index.htm*.

separates them, evoking the metaphorical boundaries that separate people and providing a sharp visualization of the idea of "personal space."

"Diorama Table" (2005) is a whimsical work by Takahashi and Sasada [11] that also uses a camera pointed at a surface on which projections are based on what the camera sees (Figure 2.7). In this case users place objects (plates, cutlery, and so forth) on a table; virtual objects such as trains are then projected on the table, circling the objects. Users can rearrange the physical objects to redirect the behavior of the virtual objects. "Shadow Monsters" by Worthington,[11] a similar work, augments hand silhouettes with spines to turn them into monsters. Here, though, the coupling between sensing and projecting space is not necessary. "Subtitled Public" (2005) by Lozano-Hemmer[12] extends the camera-projector system by projecting words onto the people being tracked, while "Under Scan" by the same artist (2005) projects videos of faces into the shadows of pedestrians. "Ghost Pole Propagator" (2007) by Levin[13] uses a fundamental computer vision algorithm, known as "skeletonization," that reduces regions of a silhouette to a linear armature. In this case, the "skeletons" of different viewers are projected onto a nearby wall.

Other works employ computer vision less directly. Appropriating the technology of video surveillance systems, "You are here" by Snibbe[14] (2004) tracks gallery visitors across multiple cameras and presents their paths (as in Figure 2.8) in an interactive display, highlighting their own position. Video surveillance, with its privacy concerns and hints of Big Brother has been the theme of many art works [12] but few so far have used computer vision despite the extensive use of vision technologies in surveillance systems.

**FIGURE 2.7**

Takahashi et al.'s "Diorama Table."

[11] *www.worthersoriginal.com/*.
[12] *www.lozano-hemmer.com/*.
[13] *www.flong.com/projects/gpp/*.
[14] *www.snibbe.com/scott/public/youarehere/index.html*.

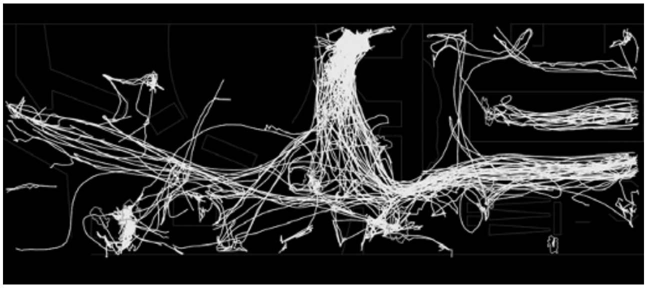

**FIGURE 2.8**

Snibbe's "You are here."

Yet another category is works that use computer vision to drive physical effects. Lozano-Hemmer's "Homographies" (2006) and "Standards and Double Standards" (2004) detect viewers and move physical objects (fluorescent lights and suspended belts, respectively) to reflect their presence, giving an uncanny sense of being watched, in the latter work, as the belts turn to "face" the viewer.

A particular variety of physical effects is the control of a robot. "Double-Taker (Snout)" (2008) by Levin, Hayhurst, Benders, and White[15] steers a huge eye on a robotic stalk (Figure 2.9) to stare at passersby, using visual detection and tracking to choose a subject. "Skeletal Reflections" by MacMurtrie (2000)[16] analyzes the posture of a viewer using two cameras, and matches the pose in a database of "art historical classical poses." A skeletal metal robot (Figure 2.10) then enacts the pose. "POMONA" by Senior[17] (2007) explores the domains of bioengineering, hybrids, and biofuels. In this work robotic–plant hybrids are steered by computer vision algorithms to maximize their exposure to the sun or to lights controlled by onlookers.

### 2.3.2 Performance

In recent years, as processing power increased at a reduced cost, it became more commonplace for vision to be used in live interactive performance. Groups that use computer vision in their performances include Dumb Type, Cie Incidents Mémorables, and Troika Ranch. For example, Winkler [13] uses vision-based motion tracking in live performance. Motion capture using markers has also been explored; one example is the work of Cunningham and Jones [14]. Audience participation, as described by Maynes-Aminzade et al. [8], constitutes one area of opportunity, although most works focus on the performers themselves, at different

[15] *www.flong.com/projects/snout/.*
[16] *amorphicrobotworks.org/works/skelli/fabrication.htm.*
[17] *andrewsenior.com/gallery/electronica/pomona.*

**FIGURE 2.9**

Levin's "Double-Taker (Snout)."

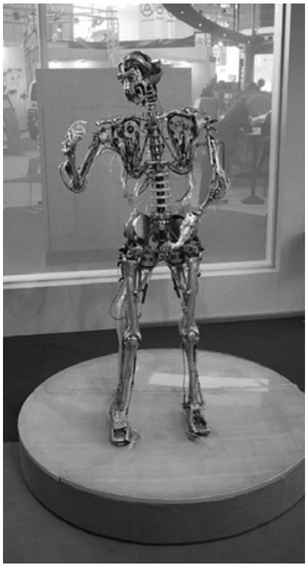

**FIGURE 2.10**

MacMurtrie's "Skeletal Reflections."

levels. In the work of Lyons et al. [15], a wearable camera pointing at the wearer's mouth interprets mouth gestures to generate MIDI sounds. The "mouthesizer" as a tool has been used in performances not only by itself (the performer moving her mouth to generate sounds) but also in combination with other instruments since the performer's hands remain free. The mouthesizer uses simple techniques to detect the shape of the performer's open mouth, while works such as that of Valenti et al. [16] use more sophisticated techniques (facial expression recognition) to sonify facial movements. In contrast, Hornof et al. [17] track eye movements to generate sounds, although it is not clear if their system has been used in a performance.

In the work of Paradiso and Sparacino [18], the camera tracks limbs so that the movements generate music, and in the work of Jean et al. [19] cameras track foot movement to generate sounds on a virtual keyboard.

Some of these systems are actually close to being additional tools for the performer, as well as an intricate part of the performance itself. In the "Live Cinema" work of Lew [20], computer vision is used to track fingers on a rear-projection screen: Multiple sensors are integrated so that finger motions on the screen, combined with other activities, determine what a VJ is projecting as a film. The system in this case is both a tool for the VJ and a performance element. "Audio d-touch" by Costanza et al. [21] uses computer vision to recognize basic patterns in small blocks placed on a table. As the performer moves the blocks, different sounds/variations are generated. This tool has also been used in combination with other instruments played by professional musicians. Some compositions explore intimate performance: In a body-drawing communicator for instant partners,[18] vision and a projection system are combined to allow two remote participants to draw on each other's bodies while lying in bed. Drawings are transmitted "live" between the two beds with the goal of creating a sense of communication that leverages the emotional quality of physical gesture.

## 2.4 SOFTWARE TOOLS

In this section we review some of the software tools that are available for artists who wish to use computer vision to provide interactivity in their work.

### 2.4.1 Max/MSP, Jitter, and Puredata

Max/MSP was originally created at IRCAM in the 1980s and has been in development since, in a commercial distribution (Max/MSP) and in a similar open-source "version" developed originally by the author (Puckette). Both environments use a visual programming interface for multimedia, music, and inputs from various sensors

[18] *www.distancelab.org/projects/mutsugoto/*.

including cameras (using plug-in modules such as Jitter for Max/MSP and framestein for puredata). The two software families have been used extensively over the years by musicians, designers, artists, and engineers in a wide range of applications, including performance art and installation.

### 2.4.2 EyesWeb

EyesWeb is an open-source software platform designed to support development of real-time multimodal distributed interactive applications. It was developed at the InfoMus Laboratory (at the Universidad de Genoa, where a project by the same name investigates real-time multimodal interaction) and has been used extensively in interactive performances and installations. The platform is similar to the Max/MSP family in that a visual programming language is used. However, while the original focus of Max/MSP was on music, EyesWeb was developed from the start with real-time multimodal interaction in mind.

### 2.4.3 processing

processing[19] by Reas and Fry [7] is a simplified programming environment modeled on the Java language (and compiled into Java to allow platform independence). It was designed for computer-generated video art around a frame-by-frame redrawing paradigm, but has built-in support for video and add-ons for sensors, including JMyron[20] by Nimoy for video capture.

### 2.4.4 OpenCV

OpenCV is an open-source library of computer vision algorithms, for users with significant programming experience. It is written in C++ and used for industrial computer vision applications. However, it also provides complex vision algorithms, including face detection, which run at high speed.

## 2.5 FRONTIERS OF COMPUTER VISION

The field of computer vision research has expanded tremendously in recent years as cameras, processing power, and storage have become abundant and affordable. Actual industrial systems have grown from those that perform highly constrained machine inspection to those that operate in the real world in domains as diverse and uncontrolled as video surveillance, biometric identification (particularly face recognition), fruit identification, and video indexing. The ubiquity of the Internet

[19] *processing.org.*
[20] *webcamxtra.sourceforge.net.*

and the open-source movement have led to wide sharing of research algorithms, putting them in the hands of adventurous and technically savvy artists. The breaking down of art/technology barriers in the field of new media has spawned artist-technologists as well as many collaborations between artists and technologists.

We expect this transfer and appropriation of technology to continue and accelerate as artists thirst for new and more complex effects. Recent developments of particular interest, which we predict will increasingly be seen in arts, are as follows:

- *Object detection*. Effective feature extraction and indexing have led to powerful algorithms to detect classes of objects (cars, bicycles, animals, etc.).
- *Mobile platforms*.
- *Face detection, identification, and analysis*. As algorithms for face processing improve, many new artistic effects based on recognizing people and understanding their facial expressions will be achievable.
- *Human tracking*. Works will harness improved tracking capabilities and will begin to extract more information more accurately about pose, whether of dancers or gallery-goers.
- *Video indexing and understanding*. Artists will use technologies to appropriate and mash up video sources from surveillance to home movies to television broadcasts.

The new technologies will spread from static installations to mobile devices as smartphones become widespread. One example is the Spellbinder system,[21] a gaming platform in which users take photographs of people, places, or objects with camera phones which are recognized on a server that controls game play.

## 2.6 SOURCES OF INFORMATION

Although many artists do not document their work in article form, many print and online publications provide not only technical details but also audio-visual documentation of performances and installations. In this section we list some of these resources, although they are not specific to computer vision.

- *Neural.it*.[22] An online and print magazine covering many types of work.
- *runme.org*.[23] An online user-contributed collection of software for art and software art.
- "Networked Performance."[24] A blog of *turbulence.org* that mainly deals with performance.

[21] *news.bbc.co.uk/2/hi/technology/6938244.stm*.
[22] *www.neural.it*.
[23] *www.runme.org*.
[24] *www.turbulence.org*.

- *Rhizome.org*.[25] An online community that includes a large database of artworks, documentation, interviews, and other texts.
- "we-make-money-not-art."[26] A blog about technology, art, and design.
- *Leonardo* and *Leonardo* Music.[27] Journals about art and technology.
- ACM Multimedia Interactive Art program.[28] A yearly technical multimedia conference that has an art track and exhibition.
- New Interfaces for Musical Expression (NIME) conference.[29] A conference on novel interfaces for musical expression.

## 2.7 SUMMARY

Ambient intelligence is an emerging field that promises to integrate many technologies that have been under development since the early days of computing (including algorithms in machine learning and other areas). A key component of an ambient intelligence platform is how it relates to humans and how humans relate to it. Computer vision can play a major role because it is unobtrusive (in relation to wearable sensors) and can sense a wide range of human actions and events. In the arts, computer vision has grown tremendously in recent years, precisely for these reasons.

In installation art, it is usually impractical for viewers (e.g., in a gallery or a museum) to wear specialized sensors. At the same time, the recent availability of feature-rich open-source (and affordable) computer vision toolkits for off-the-shelf hardware has allowed artists with few resources to create interesting compositions that run on basic laptops or PCs. For this reason, many artworks and interactive performances use computer vision techniques (often in combination with other sensors) for a wide range of purposes. These include detecting simple background changes and recognizing gestures, body parts, and objects. All of this is accomplished in the toughest conditions for a real-world application: before live audiences or in interactions with off-the-street "participants," and considering sociocultural issues that include the impact of technology (what may in part differentiate art from nonart). Thus, computer vision in the arts is perhaps the single most important ambient intelligence application to date, and in particular the most important example of human-centric interfaces for ambient intelligence.

## ACKNOWLEDGMENTS

The authors would like to thank the artists who granted us permission to use images of their work in this chapter.

[25] *www.rhizome.org*.
[26] *www.we-make-money-not-art.com*.
[27] *www.mitpress.com*.
[28] *www.sigmm.com*.
[29] *www.nime.org*.

## REFERENCES

[1] Muybridge E. The Human Figure in Motion. Dover; 1955.

[2] Krueger M. Artificial Reality. Addison-Wesley; 1983.

[3] Krueger M, Gionfriddo T, Hinrichsen K. VideoPlace—an artificial reality. ACM SIGCHI Bulletin 1985;16(4):35–40.

[4] Rush M. New Media in Art. Thames & Hudson; 2005.

[5] Manovich L. The Language of New Media. MIT Press; 2001.

[6] Levin G. Computer vision for artists and designers: Pedagogic tools and techniques for novice programmers. Journal of Artificial Intelligence and Society 20(4).

[7] Reas C, Fry B. Processing: A Programming Handbook for Visual Designers and Artists. MIT Press; 2007.

[8] Maynes-Aminzade D, Pausch R, Seitz S. Techniques for interactive audience participation. ICMI, IEEE; 2002.

[9] Darrell T, Gordon G, Woodfill J, Harville M. A virtual mirror interface using real-time robust face tracking. Workshop on Face and Gesture, No. 3. IEEE; 1998. p. 616–21.

[10] Workshop on Projector Camera Systems. *procams.org*.

[11] Takahashi K, Sasada S. Diorama table. ACM Multimedia 2005;1077–8.

[12] Levin T, Frohne U, Weibel P. CTRL [SPACE]: Rhetorics of Surveillance from Bentham to Big Brother. MIT Press; 2002.

[13] Winkler T. Fusing movement, sound, and video in falling up, an interactive dance/theatre production. In: Proceedings of the Conference on New Interfaces for Musical Expression. 2002.

[14] Dils A. The ghost in the machine: Merce Cunningham and Bill T. Jones. PAJ: A Journal of Performance and Art 2002;24(1):94–104.

[15] Lyons M, Haehnel M, Tetsutani N. Designing, playing, and performing, with a vision-based mouth interface. In: Proceedings of the Conference on New Interfaces for Musical Expression. 2003.

[16] Valenti R, Jaimes A, Sebe N. Facial expression recognition as a creative interface. In: Proceedings of the International Conference on Intelligent User Interfaces. 2008.

[17] Hornof A, Rogers T, Halverson T. Eyemusic: Performing live music and multimedia compositions with eye movements. In: Proceedings of the Conference on New Interfaces for Musical Expression. 2007. p. 299–300.

[18] Paradiso J, Sparacino F. Optical tracking for music and dance performance. In: Gruen A, Kahmen H editors. Optical 3-D Measurement Techniques IV. 1997. p. 11–8.

[19] Jean F, Albu A, Schloss W, Oriessen P. Computer vision-based interface for the control of meta-instruments. In: Proceedings of 12th International Conference on Human-Computer Interaction. ACM Press; 2007.

[20] Lew M. Live cinema: Designing an instrument for cinema editing as a live performance. In: Conference an New Interfaces for Musical Expression. 2004.

[21] Costanza E, Shelley S, Robinson J. Introducing audio d-touch: A novel tangible user interface for music composition and performance. In: Proceedings of the 6th International Conference on Digital Audio Effects. 2003.