(LAND)MOVES

Bruno Zamborlin [*]
Goldsmiths/IRCAM
SE146NW London UK
bruno.zamborlin@ircam.fr

Giorgio Partesana [†]
http://glp.oivil.eu
gioparte@gmail.com

Marco Liuni [‡]
IRCAM
1, place Igor Stravinsky -
75004 Paris, France
marco.liuni@ircam.fr

ABSTRACT

(land)moves is an interactive installation: the user's gestures control the multimedia processing with a total synergy between audio and video synthesis and treatment.

Keywords

mapping gesture-audio-video, gesture recognition, landscape, soundscape

1. INTRODUCTION

The project **(land)moves** is an installation where sounds and images evolve according to the same set of control parameters. Such parameters are deduced from the analysis of the user's gestures, and are interpreted in real time by the audio and video engines: human gestures influence the evolving audiovisual landscape. By interacting with a flock of polygons in the video foreground, the user gradually learns to use the device and to influence the evolution of the whole audio-video environment. Gestures, sounds and video create objects with a joint identity: we refer to these entities as a *move*. The user interacts with the multiple media of each move depending on his will and attitudes, but still perceiving a coherent reaction to his movements.

In the next section we describe the artistic foundation of the installation and the user's experience that we aim to exploit. The third section is a summary of the features exchanged between the gestures analysis and recognition engine and the sound and video processing.

2. ARTWORK EXPERIENCE

On the one hand there is the user, on the other an audiovisual landscape: their relation consists in a multimodal interaction. The user is responsible for the landscape modeling through his gestures (see http://www.youtube.com/watch?v=CfKQCAxizrA for the video game *From Dust*, exploiting the idea to model planet Earth). On the other hand, the landscape also conveys a mood (see an example in the opening scene of *Gerry* by Gus Van Santis http://www.youtube.com/watch?v=-_JiB4N-0Ro); therefore, certain features of our landscape are modeled by the user's gestures, which are the expression of an emotional state partially determined by the landscape itself.

2.1 Scenario

The system reproduces a planet that is continuously changing, with a time dimension consisting in cycles of days and nights. The gesture analysis is performed according to two time-scales, which control two distinct but dependent audiovisual levels: the first, constituting the visual foreground, consists of a flock of polygons flying over the landscape; the second background level is the evolving land (see figure 1). The division between the foreground and the background affects the audio domain as well: different sounds and treatments contribute for creating a perception of distance and motion, as detailed in the third section. Some features of user's actions have immediate consequences, while the whole performance of the user is analyzed and affects the whole system in a long-term scale.

This first realization of **(land)moves** is based on two classes of moves: every performed gesture is placed at a certain distance from these classes, which acts as a parameter for the audio and video engines. This classification is performed by the gesture recognition algorithm detailed in section 3.1.

2.2 Space and Interface

The artwork is controlled by a single user, standing in front of a vertical projection of approximately 4 meters width and 2 meters height based on the floor (we are considering the possibility of a multi-users mode for future versions); other visitors can nevertheless access the space. This placement is intended to provide the user for a natural feeling of a landscape. There are no devices to handle, the motion is captured by a Microsoft *Kinect* sensor device.

3. REAL-TIME GESTURE BASED MULTIMEDIA PROCESSING

The implementation consists in three separated applications which communicate through the OSC protocol: the gesture analysis algorithm generates the control messages, which are then sent to the audio and video engines.

3.1 Gestures Analysis

We describe here how the arm movements of the user are represented and translated into control messages. The system takes advantage of the Microsoft Kinect which interprets 3D scene information from a continuously-projected infrared structured light [2]. The main purpose in using this device is to analyze arm gestures without asking to the user to hold a physical device. The PrimeSense OpenNI software framework (http://www.openni.org/) has been used for robustly tracking the hand position providing the first

[*] Gesture analysis and recognition.

[†] Visual Artist.

[‡] Sound Artist.

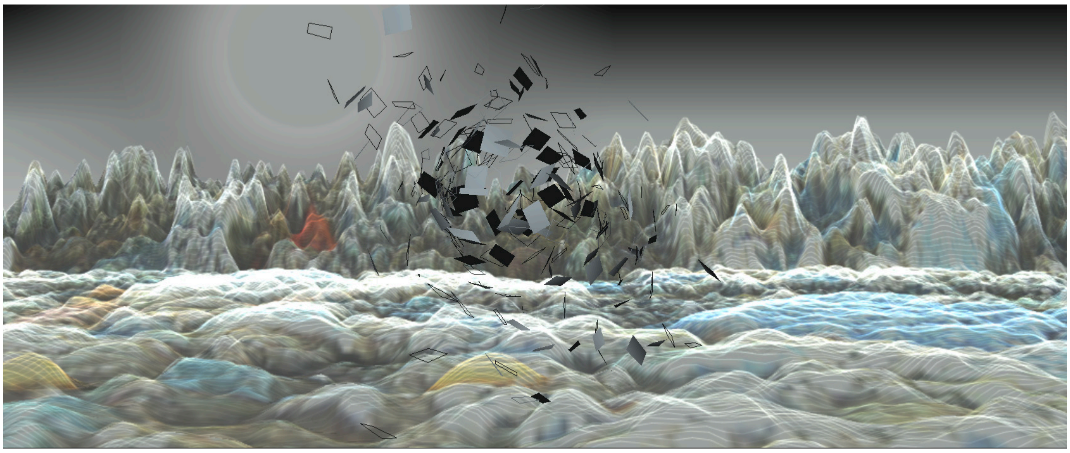

Figure 1: A screenshot of (land)moves .

control signal as 3D position. Based on this information, the quantity of movement of the user is calculated. Two types of quantity of movements are given: *instantaneous* and *global*. The first one is computed over a sliding window of few hundreds of milliseconds and it is used to control parameters that need to react rapidly to user behavior. The second one is computed over a larger scale and gives information related to the global activity of the user's performance. In a similar way, other descriptors based on the symmetry and the velocity of the gesture are calculated. The utilization of these two pieces of information is described in sections 3.2 and 3.3.

Furthermore, the system uses machine learning techniques to estimate a high level quality of movement, called the *roundness/sharpness ratio*. We consider here different gestures, each one assigned to a different value of this movement quality. We use the HMM-based Gesture Follower [1] for continuous likelihood recognition of these gestures.

The final information that is given to the audio and video engines is a high level description of the roundness/sharpness ratio information. This quality reaches here a human meaning, and defines two reference classes of moves: *round* and *sharp* gestures. As the Gesture Follower continuously returns a likelihood estimation referred to each one of these gestures, the final information is continuously interpolated and intermediate values between different signs are possible.

3.2 Video Synthesis

Quartz Composer's Particle System is used to create the foreground objects which are more reactive and engage the user in a direct interaction. The landscape in background, which slowly but constantly mutates, is created by extruding and color correcting photographic textures. This process reproduces some morphologic treats of real world despite the digital polygonal feeling, creating a troubled perception of an imaginary planet. Textures are chosen among an indexed database according to gesture characteristics on a long term analysis.

On the foreground level, gesture features such as velocity and amplitude influence the movement of the flock, making the polygons fly faster or wider following the movements of the arm. Other examples of the mapping for the landscape in background use the energy and the roundness quality defined in section 3.1 of gestures to determine whether the land shapes as gentle hills or spiky mountains. A further element which defines the image is light; the lighting system consists of three elements, the sun/moon, the sky and the ambient light.

3.3 Audio Treatments

The audio feedback is generated from both pre-recorded materials and real time processing: electric guitar samples are used to provide the device for a concrete instrumental feeling. The whole sound engine is driven by gestural descriptors, at different time levels: the sounds in the foreground react to short term descriptors of energy and roundness/sharpness ratio, while those in the background are treated according to features on a longer period. Sounds are classified according to the two classes of moves defined in 3.1. With this choice, we aim to establish two distant points in an appropriate space of timbres, which represent the two classes of gestures, as well as the two possible shapes of the flock of polygons used in the video feedback: all the transitions in between are realized with a source-filter technique based on the *SuperVP* phase vocoder [3].

4. ACKNOWLEDGMENTS

The work described in section 3.1 has been realized in collaboration with Parag Mital from the Goldsmiths EAVI group pkmital.com

Gesture Follower has been developed and is currently maintained by the IMTR team at IRCAM. More information at http://imtr.ircam.fr/imtr/Gesture_Follower

The gesture analysis framework described in 3.1 takes full advantage of FTM. See ftm.ircam.fr for details

SuperVP is developed by Axel Roebel and the Analysis/Synthesis team at IRCAM

Kineme comunity for the development of the particle system http://kineme.net/

Vade for Rutt Etra plugin. v002 Rutt/Etra attempts to emulate the Rutt/Etra raster-based analog synthesizer http://v002.info/?page_id=19

5. REFERENCES

[1] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana. Continuous realtime gesture following and recognition. *Gesture in Embodied Communication and Human-Computer Interaction, Springer*, pages 73–84, 2010.

[2] P. MS. Primesense supplies 3-d-sensing technology to project natal for xbox 360. *MsPress*, 2010.

[3] A. Roebel. A shape-invariant phase vocoder for speech transformation. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, September 2010.