Chapter 14
ZatLab Gesture Recognition Framework:
Machine Learning Results

André Baltazar
Catholic University of Portugal, Portugal

ABSTRACT

The main problem this work addresses is the real-time recognition of gestures, particularly in the complex domain of artistic performance. By recognizing the performer gestures, one is able to map them to diverse controls, from lightning control to the creation of visuals, sound control or even music creation, thus allowing performers real-time manipulation of creative events. The work presented here takes this challenge, using a multidisciplinary approach to the problem, based in some of the known principles of how humans recognize gesture, together with the computer science methods to successfully complete the task. This paper is a consequence of previous publications and presents in detail the Gesture Recognition Module of the ZatLab Framework and results obtained by its Machine Learning (ML) algorithms. One will provide a brief review the previous works done in the area, followed by the description of the framework design and the results of the recognition algorithms.

INTRODUCTION

Gestures are the principal non-verbal, cross-modal communication channel, and they rely on movements for different domains of communication (Volpe, 2005). Children start to communicate by gestures (around 10 months' age) even before they start speaking. There is also an ample evidence that by the age of 12 months children are able to understand the gestures other people produce (Rowe & Goldin-meadow, 2009). For the most part gestures are considered an auxiliary way of communication to speech, though there are also studies that focus on the role of gestures in making interactions work (Roth, 2001). We use our hands constantly to interact with things. Pick them up, move them, transform their shape, or activate them in some way. In the same unconscious way we gesticulate in communicating fundamental

DOI: 10.4018/978-1-5225-2589-9.ch014

ideas: stop; come closer; go there; no; yes; and so on. Gestures are thus a natural and intuitive form of both interaction and communication (Watson, 1993).

There is so much information contained in a gesture that is natural to think about using it besides simple human-to-human communication. However, the use of technology to understand gestures is still somehow vaguely explored, it has moved beyond its first steps but the way towards systems fully capable of analyzing gestures is still long and difficult (Volpe, 2005). Probably because if in one hand, the recognition of gestures is somehow a trivial task for humans, in other, the endeavor of translating gestures to the virtual world, with a digital encoding is a difficult and ill-defined task. It is necessary to somehow bridge this gap, stimulating a constructive interaction between gestures and technology, culture and science, performance and communication. Opening thus, new and unexplored frontiers in the design of a novel generation of multimodal interactive systems.

In this case one developed a framework to recognize human gestures automatically and use its information to enhance an artistic performance. This paper describes briefly the ZatLab Framework (for more detailed description please see (Baltazar & Martins, 2015)) and will focus on the empirical results of the Gesture Recognition Module that sustain it.

First one will review the literature on gesture research, followed by the framework proposal, implementation and results.

BACKGROUND

The field of human movements and gesture analysis has, for a long time now, attracted the interest of many researchers, choreographers and dancers. Thus, since the end of the last century, a significant corpus of work has been conducted relating movement perception with music (Fraisse, 1982).

Among the research community on this subject, there are works that stand out as important references on how video analysis technologies have provided interesting ways of movement-music interaction. Early works of composers Todd Winkler (Winkler, 1995) and Richard Povall (Povall, 1998), or the choreographer Robert Weschler work with Palindrome[1]. Also, Mark Coniglio continued development of his Isadora[2] programming environment, plus the groundbreaking work Troika Ranch[3] has done in interactive dance.

Other example of research in this field is the seminal work of Camurri, with several studies published, including an approach for the recognition of acted emotional states based on the analysis of body movement and gesture expressivity (Castellano, Villalba, & Camurri, 2007) and one of the most remarkable and recognized works, the EyesWeb software (Camurri et al., 2000).

Also, Bevilacqua, at IRCAM-France worked on projects that used unfettered gestural motion for expressive musical purposes (Bevilacqua, Müller, & Schnell, 2005; Bevilacqua & Muller, 2005; Dobrian & Bevilacqua, 2003). Chronologic speaking first development consisted of software to receive data from a Vicon motion capture system, translate and map it into music controls and other media controls such as lighting (Dobrian & Bevilacqua, 2003). Then this evolved to the development of the toolbox "Mapping is not Music" for Max/MSP, dedicated to mapping between gesture and sound (Bevilacqua et al., 2005). And in parallel (Bevilacqua & Muller, 2005) presents the work of the a gesture follower for performing arts, which indicates in real-time the time correspondences between an observed gesture sequence and a fixed reference gesture sequence.