ZatLab Gesture Recognition Framework: Machine Learning Results
André Baltazar, Catholic University of Portugal, Center for Science and Technology in the Arts, Porto, 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.

KEYWORDS
Computer Vision, DTW, Gesture Recognition, HCI, HMM, Interactive Performance, Kinect, Machine Learning

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 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 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.

DOI: 10.4018/IJCICG.2016070102

Copyright © 2016, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
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 Palindrome1. Also, Mark Coniglio continued development of his Isadora2 programming environment, plus the groundbreaking work Troika Ranch3 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.

Likewise, Nort and Wanderley (Nort, Wanderley, & Van Nort, 2006) presented the LoM toolbox. This allowed artists and researchers access to tools for experimenting with different complex mappings that would be difficult to build from scratch (or from within Max/MSP) and which can be combined to create many different control possibilities. This includes rapid experimentation of mapping in the dual sense of choosing what parameters to associate between control and sound space as well as the mapping of entire regions of these spaces through interpolation.

Another important work, published in 2011, is the one of Gillian (Gillian, Knapp, & O’Modhrain, 2011). He presented a machine learning toolbox that has been specifically developed for musician-computer interaction. His toolbox features a large number of machine learning algorithms that can be used in real-time to recognize static postures, perform regression and classify multivariate temporal gestures.

Also in 2009, the author made part of the project “Kinetic controller driven adaptive and dynamic music composition systems”4. One of the aims of the project was to utilize video cameras as gestural controllers for real-time music generation. The project included the development of new techniques and strategies for computer-assisted composition in the context of real-time user control with non-standard human interface devices. The research team designed and implemented real-time software that provided tools and resources for music, dance, theatre, installation artists, interactive kiosks, computer games, and internet/web information systems. The accurate segmentation of the human body was an important issue for increased gestural control using video cameras. In the International
Conversation China: Serendipity on a Plate
www.igi-global.com/article/conversation-china/147397?camid=4v1a