Continuous realtime gesture following and recognition

Frédéric Bevilacqua, Bruno Zamborlin, Anthony Sypniewski, Norbert Schnell, and Fabrice Guédy

IRCAM, CNRS - STMS,
1 Place Igor Stravinsky, 75004 Paris, France
{Frederic.Bevilacqua}@ircam.fr

Gesture recognition systems have been successively developed based on methods such as Hidden Markow Models, finite state machines, template matching or neural networks. In most cases, gestures are considered as units that must be recognized once completed. Therfore, these systems output results at discrete time events, typically at the end of each gesture. Motivated by applications on expressive visuals and sound control, we present here a different paradigm for online gesture analysis : the recognizing system outputs "continuously" (i.e. on a fine temporal grain) parameters characterizing the performed gesture. These parameters are obtained by online comparison with temporal shapes stored in a database. Precisely, two types of information are continuously updated. These are probabilistic estimations of 1) the similarity of the performed gesture to prerecorded gestures (likelihood) and 2) the time progression of the performed gesture. The first type of information allows for the selection of the likeliest gesture at any moment and the second type of information allow for the estimation of the current temporal index inside the gesture, which referred here as "gesture following". These continuous output allows for both the selection and the synchronization of continuous visual or sound processes to gestures.

This framework for gesture characterization relies on a precise temporal modeling of the gesture. Thus, our approach is especially suited for cases where the temporal profiles are intrinsically important and remains consistent, as generally found in performing arts (dance or music performances). Importantly, we will also show that this temporal modeling allow for the prediction of the evolution of a performed gesture. This feature can be important in several applications in human computer interaction.

The algorithm fundamentally works with any type of multidimensional data flow, sampled at a regular sampling rate. Our system has been developed with the constraint generally found in performing arts that only few examples will be available, as also discussed in Ref.[1]. Such constraints are generally incompatible with established algorithm, such as standard implementations of Hidden Markov Models (HMM) [2]. A first version of our algorithm has been reported in [3]. Itis based on Hidden Markov Models but with a modified learning schema. We report here improvements of this first version of the algorithm. Synthetic data were created in order to asses quantitatively the algoritm in various conditions.

First, we incorporated a sliding window mechanism operating on the reference gesture. This enables us to use the algorithm for very long reference gestures (typically several minutes of recording at 200 Hz sampling rate) while keeping a CPU usage constant (which depends on the actual size of the window, since the algorithm operates only on the window). This feature allows for the use of the system as a "gesture follower" operating similarly to a score following system.

Second, we developed methods to improve the overall algorithm robustness by implementing different approaches for the observation probability computation defined in the Hidden Markov Model. Originally, this observation proabablity was computed using only the current dataframe. Simulations demonstrated that improved results can be achieved using buffered data.

Third, we implemented a feedback loop on the transition probability matrix of the HHM, based on a measurement of the local speed of the performance. With such a feature, the algorithm can be less sensitive to speed variation of the performance.

This system has been implemented as a collection of Max/MSP modules called the "gesture follower". Recently, the core algorithm has been implemented as a C++ library and can therefore be implemented in other environnements.

Several applications have been developed using this system. For examples, the following feature has been used to synchronize precisely musical tracks (by time stretching audio buffer) to musician gestures. Such a application was experimented in a music education context. Case of "virtual" conducting and violin playing with sensors attached to the bow were successfully experimented. In other contexts, gesture sonification were experimented using both the following and the recognition features. This allows the user to select and control the time pace of sound processes.

Acknowledgement

We acknowledge partial support of the following projects: The i-Maestro project (IST-026883, www.i-maestro.org), the EU-ICT Project SAME (Sound And Music for Everyone Everyday Everywhere Every way, http://www.sameproject.eu/) and the ANR project EarToy.

References

1. Rajko, S., Qian, G., Ingalls, T., James, J.: Real-time gesture recognition with minimal training requirements and on-line learning. In: CVPR07: IEEE Conference on Computer Vision and Pattern Recognition. (2007) 1–8
2. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE **77**(2) (1989) 257–286
3. Bevilacqua, F., Guédy, F., Schnell, N., Fléty, E., Leroy, N.: Wireless sensor interface and gesture-follower for music pedagogy. In: NIME '07: Proceedings of the 7th international conference on New interfaces for musical expression, New York, NY, USA, ACM (2007) 124–129