A tool for analysis of expressive gestures: The EyesWeb Expressive Gesture Processing Library

Antonio Camurri, Barbara Mazzarino, Gualtiero Volpe

InfoMus Lab (Laboratorio di Informatica Musicale)
DIST – University of Genova
Viale Causa 13, I-16145 Genova, Italy
http://infomus.dist.unige.it
{toni, bunny, volpe}@infomus.dist.unige.it

Abstract. This paper presents some results of a research work concerning algorithms and computational models for real-time analysis of expressive gestures in human full-body movement. As a main concrete result of our research work, we present a collection of software modules for the EyesWeb open architecture (distributed for free at www.eyesweb.org). These software modules, collected in the EyesWeb Expressive Gesture Processing Library, have been used in real scenarios and applications, mainly in the field of performing arts. The work has been carried out at the DIST - InfoMus Lab in the framework of the EU IST Project MEGA (Multisensory Expressive Gesture Applications, ww.megaproject.org). The MEGA project is centered on the modeling and communication of expressive and emotional content in non-verbal interaction by multi-sensory interfaces in shared interactive Mixed Reality environments. Analysis of expressiveness in human gestures can contribute to new paradigms for the design of interactive systems.

1. Introduction

Our research is focused on the design of interactive systems mainly for theatre and performing art applications, explicitly considering and enabling the communication of expressive, emotional content. Such a research work includes (i) the analysis and classification of expressive gestures in music (audio) and movement (video), (ii) the real-time generation of audio and visual content depending on the output of the analysis, (iii) a study of the interaction mechanisms (mapping strategies) enabling the results of the analysis to be employed (transformed) in automatically generation of audio and visual material.

In this paper we focuses on the first aspect, and in particular we address algorithms and computational models for the extraction of a collection of expressive features from human movement.

Since the particular interest in interactive systems for performing art, and since it can be considered as the artistic expression of human movement, dance has been chosen as a particular test-bed for our research.

The generation of a particular output (e.g., a sound, a colored light) can directly depend on low-level motion features (e.g., position of a dancer on the stage, speed of the detected motion), or can be the result of the application of a number of decision rules considering the context, the history of the performance, the information about the classified expressive intention of a dancer (e.g., in term of basic emotions: joy, grief, fear, anger). In order to extract and provide such a variety of possible expressive cues, a layered approach [1] has been adopted to model expressive gestures, from low-level physical measures (e.g., in the case of human movement, position, speed, acceleration of body parts) toward descriptors of overall (motion) features (e.g., fluency, directness, impulsiveness).

Models and algorithms will be presented with reference to a concrete output of the research process: the EyesWeb Gesture Processing Library, a collection of software modules for the EyesWeb open software platform (distributed for free at www.eyesweb.org)

2. The EyesWeb Expressive Gesture Processing Library

The *EyesWeb Expressive Gesture Processing Library* includes a collection of blocks (software modules) and patches (interconnections of blocks) contained into three main sub-libraries:

- *The EyesWeb Motion Analysis Library*: a collection of modules for real-time motion tracking and extraction of movement cues from human full-body motion.
- *The EyesWeb Space Analysis Library*: a collection of modules for analysis of occupation of 2D (real as well as virtual) spaces. If from the one hand this sub-library can be used to extract low-level motion cues (e.g., how much time a dancer occupied a given position on the stage), on the other hand it can also be used to carry out analyses in semantic spaces.
- *The EyesWeb Trajectory Analysis Library*: a collection of modules for extraction of features from trajectories in 2D (real as well as virtual) spaces. These space may be physical spaces or not such as semantic and expressive spaces.

2.1. The EyesWeb Motion Analysis Library

The EyesWeb Motion Analysis Library applies computer vision, statistical, and signal processing techniques to extract expressive cues from human full-body movement.

A first task consists in individuating and tracking motion in the incoming images. Firstly, background subtraction is used to segment the body silhouette. Algorithms based on searching for body centroids and on optical flow based techniques (e.g., the Lucas and Kanade tracking algorithm [5]) are available.

Starting from silhouettes and tracking information a collection of expressive parameters is extracted. Three of them are described in the following.

- *Quantity of Motion* (QoM), i.e., the amount of detected movement. It is based on the Silhouette Motion Images. A Silhouette Motion Image (SMI) is an image carrying information about variations of the silhouette shape and position in the last few frames. SMIs are inspired to motion-energy images (MEI) and motion-history images (MHI) [4][7]. They differ from MEIs in the fact that the silhouette in the last (more recent) frame is removed from the output image: in such a way only motion is considered while the current posture is skipped. QoM is computed as the area (i.e., number of pixels) of a SMI. It can be considered as an overall measure of the amount of detected motion, involving velocity and force.

- *Silhouette shape/orientation of body parts*. It is based on an analogy between the image moments and mechanical moments: in this perspective, the three central moments of second order build the components of the inertial tensor of the rotation of the silhouette around its center of gravity: this allows to compute the axes (corresponding to the main inertial axes of the silhouette) of an ellipse that can be considered as an approximation of the silhouette: orientation of the axes is related to the orientation of the body [8].

Figure 1 shows the ellipse calculated on a reference dance fragment.

Fig. 1. Silhouette shape and orientation

By applying the extraction of the ellipse to different body parts, other information can be obtained. For example, by considering the main axis of the ellipses associated to the head and to the torso of the dancer, it can be possible to obtain an estimate of the directional changes in face and torso, a cue that psychologists consider important for communicating expressive intention (see for example [9]).

- *Contraction Index*, a measure, ranging from 0 to 1, of how the dancer's body uses the space surrounding it. It is related to Laban's "personal space"(see [2][3]). It can be calculated in two different ways: (i) considering as contraction index the eccentricity of the ellipse obtained as described above, (ii) using a technique related to the bounding region, i.e., the minimum rectangle surrounding the dancer's body: the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette. Intuitively, if the limbs are fully stretched and not lying along the body, this component of the CI will be low, while, if the limbs are kept tightly nearby the body, it will be high (near to 1).

The EyesWeb Motion Analysis Library also includes blocks and patches to extract measures related to the temporal dynamics of movement. A main issue is the segmentation of movement in pause and motion phases. A motion phase can be associated to a dance phrase and considered as a gesture. A pause phase can be associated to a posture and considered as a gesture as well. For example, in a related work ([6]) the QoM measure has been used to perform the segmentation between pause and motion phases. In fact, QoM is related to the overall amount of motion and its evolution in time can be seen as a sequence of bell-shaped curves (*motion bells*). In order to segment motion, a list of these motion bells has been extracted and their features (e.g., peak value and duration) computed. Then, an empirical threshold has been defined: the dancer is considered to be moving if the area of the motion image (i.e., the QoM) is greater than 2.5% of the total area of the silhouette.

Several movement cues can be measured after segmenting motion in motion and pause phases: for example, blocks are available for calculating durations of pause and motion phases and inter-onset intervals as the time interval between the beginning of two subsequent motion phases. Furthermore, descriptive statistics of values of extracted cues can be computed on motion phases: for example, it is possible to calculate the sample mean and variance of the QoM during a motion phase.

2.2. The EyesWeb Space Analysis Library

The EyesWeb Space Analysis Library is based on a model considering a collection of discrete potential functions defined on a 2D space [10]. The space is divided into active cells forming a grid. A point moving in the space is considered and tracked. Three main kind of potential functions are considered: (i) potential functions *not* depending on the current position of the tracked point, (ii) potential functions depending on the current position of the tracked point, (iii) potential functions depending on the definition of regions inside the space.

Objects and subjects in the space can be modeled by time-varying potentials. For example, a point moving in a 2D space (corresponding to a stage) can be associated to a dancer. Objects (such as fixed scenery or lights) can be modeled with potential functions independent from the position of the tracked object: notice that "independent from the position of the tracked object" does not mean time-invariant. The trajectory of a dancer with respect to such a potential function can be studied in order to identify relationships between movement and scenery. The dancer himself can be modeled as a bell-shaped potential moving around the space by using the second kind of potential functions. Interactions between potentials can be used to model interactions between (real or virtual) objects and subjects in the space.

Regions in the space can also be defined. For example, it is possible that some regions exist on a stage in which the presence of movement is more meaningful than in other regions. A certain number of "meaningful" regions (i.e., regions on which a particular focus is placed) can be defined and cues can be measured on them (e.g., how much time a dancer occupied a given region).

This metaphor can be applied both to real spaces (e.g., scenery and actors on a stage, the dancer's General Space as described in [3]) and to virtual, semantic, expressive spaces (e.g., a space of parameters where gestures are represented as trajectories): for example, if, from the one hand, the tracked point is a dancer on a stage, a measure of the time duration along which the dancer was in the scope of a given light can be obtained; on the other hand, if the tracked point represents a position in a semantic, expressive space where regions corresponds to basic emotions, the time duration along which a given emotion has been recognized can also be obtained.

The EyesWeb Space Analysis Library implements the model and includes blocks allowing the definition of interacting discrete potentials on 2D spaces, the definition of regions, the extraction of cues (such as, for example, the occupation rates of regions in the space). For example, Figure 2 shows the occupation rates calculated on a rectangular space divided into 25 cells. The intensity (saturation) of the color for each cell is directly proportional to the occupation rate of the cell. The trajectory of the tracked point is also displayed.

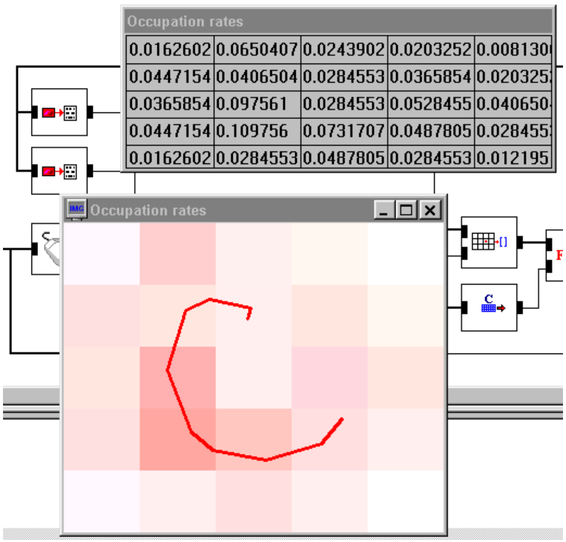

Fig. 2. Occupation rates

2.3. The EyesWeb Trajectory Analysis Library

The EyesWeb Trajectory Analysis Library contains a collection of blocks and patches for extraction of features from trajectories in 2D (real or virtual) spaces. It complements the EyesWeb Space Analysis Library and it can be used in conjunction with the EyesWeb Motion Analysis Library.

Blocks can deal with lot of trajectories at the same time, for example the trajectories of the body joints (e.g., head, hands, feet) or the trajectories of the points tracked using the Lucas-Kanade feature tracker available in the Motion Analysis sub-library. Features that can be extracted include geometric and kinematics measures.

Examples of geometric features are the length of a trajectory, its direction and its *Directness Index*. The Directness Index (DI) is a measure of how much a trajectory is direct or flexible. In the Laban's Theory of Effort it is related to the Space dimension. In the actual implementation the DI is computed as the ratio between the length of the straight line connecting the first and last point of a given trajectory and the sum of the lengths of each segment constituting the given trajectory. Therefore, the more it is near to one, the more direct is the trajectory (i.e., the trajectory is "near" to the straight line).

The available kinematics measures are velocity, acceleration, and curvature. Their instantaneous values are calculated on each input trajectory. Numeric derivatives can be computed using both the symmetric and the asymmetric backward methods (the user can select the one he prefers). Acceleration is available both in the usual x and y components and in the normal-tangent components.

Descriptive statistic measures can also be computed:

(i) *Along time*: for example, average and peak values calculated either on running windows or on all the samples between two subsequent commands (e.g., the average velocity of the hand of a dancer during a given motion phase)

(ii) *Among trajectories*: for example, average velocity of groups of trajectories available at the same time (e.g., the average instantaneous velocity of all the tracked points located on the arm of a dancer).

As in the case of the EyesWeb Space Analysis Library, trajectories can be real trajectories coming from tracking algorithms in the real world (e.g., the trajectory of the head of a dancer tracked using a tracker included in the EyesWeb Motion Analysis Library) or trajectories in virtual, semantic spaces (e.g., a trajectory representing a gesture in a semantic, expressive space).

The extracted measures can be used as input for clustering algorithms in order to group trajectories having similar features. In the real space this approach can be used to identify points moving in a similar way (e.g., points associated to the same limb in the case of the Lucas-Kanade feature tracker). In a semantic space, it could allow grouping similar gestures, e.g., gestures communicating the same expressive intention.

The EyesWeb Expressive Gesture Processing Library has been employed in a number of artistic events (list and description of performances available at the MEGA project website www.megaproject.org). It consists of a distinct and separate add-on with respect to the EyesWeb software platform and includes the research and development work carried out during the last year.

Novel blocks for the EyesWeb Gesture Processing Library are currently under development, including for example refined motion tracking, extraction of new cues, machine learning techniques for high-level gesture analysis.

Acknowledgements

We thank Matteo Ricchetti and Riccardo Trocca for discussions and their concrete contributes to this project. We also thank the other members of the EyesWeb staff.

This work has been partially supported by the EU – IST Project MEGA (Multisensory Expressive Gesture Applications) and by the National CNR Project CNRG0024AF "Metodi di analisi dell'espressività nel movimento umano per applicazioni in Virtual Environment".

References

[1] Camurri, A., De Poli G., Leman M. “MEGASE - A Multisensory Expressive Gesture Applications System Environment for Artistic Performances”, Proc. Intl. Conf. CAST01, GMD, St Augustin-Bonn, pp.59-62, 2001.

[2] Laban, R., Lawrence F.C., “Effort”, Macdonald & Evans Ltd. London, 1947.

[3] Laban, R., “Modern Educational Dance” Macdonald & Evans Ltd. London, 1963.

[4] Bobick, A.F., Davis J., “The Recognition of Human Movement Using Temporal Templates”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(3): 257-267, 2001.

[5] Lucas B., Kanade T., “An iterative image registration technique with an application to stereo vision” in Proceedings of the International Joint Conference on Artificial Intelligence, 1981.

[6] Camurri A., Lagerlöf I., Volpe G., “Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques”, International Journal of Human-Computer Studies, Elsevier Science, in press.

[7] Zhao, L., “Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures”, Ph.D Dissertation, University of Pennsylvania, 2001

[8] Kilian J., “Simple Image Analysis By Moments” OpenCV library documentation, 2001.

[9] Boone, R. T., Cunningham, J. G., “Children's decoding of emotion in expressive body movement: The development of cue attunement” Developmental Psychology, 34, 1007-1016, 1998.

[10] Camurri A., Mazzarino B., Trocca R., Volpe G. “Real-Time Analysis of Expressive Cues in Human Movement.” Proc. Intl. Conf. CAST01, GMD, St Augustin-Bonn, pp. 63-68, 2001.