# Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library

Antonio Camurri, Barbara Mazzarino, and Gualtiero Volpe

InfoMus Lab, DIST – University of Genova
Viale Causa 13, I-16145 Genova, Italy
{Antonio.Camurri,Barbara.Mazzarino,Gualtiero.Volpe}@unige.it
http://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 gesture in full-body human movement. As a main concrete result of our research work, we present a collection of algorithms and related software modules for the EyesWeb open architecture (freely available from 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 fields of performing arts, therapy and rehabilitation, museum interactive installations, and other immersive augmented reality and cooperative virtual environment applications. The work has been carried out at DIST - InfoMus Lab in the framework of the EU IST Project MEGA (Multisensory Expressive Gesture Applications, www.megaproject.org).

## 1 Introduction

Our research is focused on the design of multimodal interfaces involving full-body human interaction, explicitly considering and enabling the communication of nonverbal expressive, emotional content. The general objective is to improve state of the art in immersive, experience-centric mixed reality and virtual environment applications. Computational models of expressiveness in human gestures can contribute to new paradigms for the design of interactive systems, improved presence and physicality in the interaction [1]. Main research directions include (i) multimodal analysis and classification of expressive gestures in musical signals and human movement, (ii) real-time generation and post-processing of audio and visual content depending on the output of the analysis, (iii) study of the interaction mechanisms and mapping strategies enabling the results of the (multimodal) analysis to be employed (transformed) in the process of automatic generation of audio and visual content, and possibly of behavior of mobile robots (e.g. a moving scenery on stage, a robot for museums) [2,3,4].

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

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

A. Camurri and G. Volpe (Eds.): GW 2003, LNAI 2915, pp. 460–467, 2004.

The generation of a particular output (e.g., sound, color, images) 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), of expressive qualities (e.g. fluent/rigid, light/heavy), and ideally in term of the "tension" in the artistic performance. A layered approach [2] has been proposed to model these aspects that we named "expressive gesture". This approach models expressive gesture from low-level physical measures (e.g., in the case of human movement: position, speed, acceleration of body parts) up to descriptors of overall (motion) features (e.g., fluency, directness, impulsiveness).

Models and algorithms are here presented with reference to a concrete output of the research: the EyesWeb Expressive Gesture Processing Library, a collection of software modules for the EyesWeb open software platform (distributed for free at www.eyesweb.org). This library has been developed in the three-year EU IST MEGA project. MEGA 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 (www.megaproject.org).

## 2 The EyesWeb Expressive Gesture Processing Library

The *EyesWeb Expressive Gesture Processing Library* includes a collection of software modules and patches (interconnections of modules) in 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. It is based on one or more videocameras and other sensor systems.
- *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 of gesture in semantic, abstract 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 spaces may again be either physical spaces or 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 from one or more videocameras. Background subtraction techniques can be used to segment the body silhouette. Given a 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.

For example, an algorithm has been developed to segment the body silhouette in sub-regions using spatio-temporal projection patterns (see [6] for an example of how such patterns are employed for gait analysis). This algorithm also provides a way to compute more robustly the position of the body center of gravity and sub-regions in the silhouette (see Figure 1a). Software modules for extracting silhouette's contour and computing its convex hull are also available (see Figure 1b).

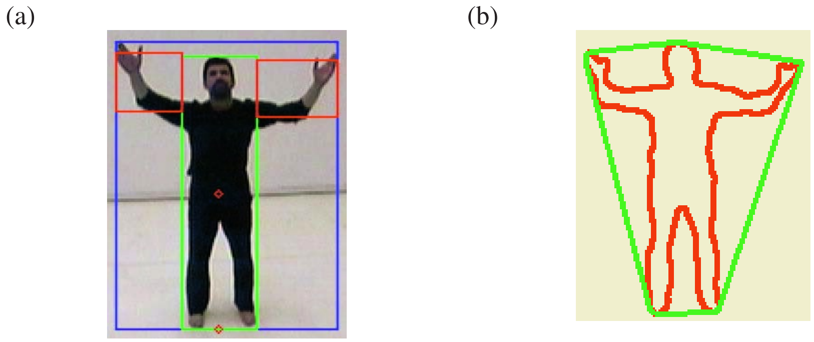


**Fig. 1.** (a) Sub-regions of the body silhouette and body center of gravity; (b) contour and convex hull of the body silhouette.

Starting from body 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) [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. Algorithms are provided to compute both the overall QoM and the QoM internal to the body silhouette.

- *Silhouette shape/orientation of body parts*. It is based on an analogy between 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 2 shows the ellipse calculated on a reference dance fragment. 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* (CI), 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 [10][11]). 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).

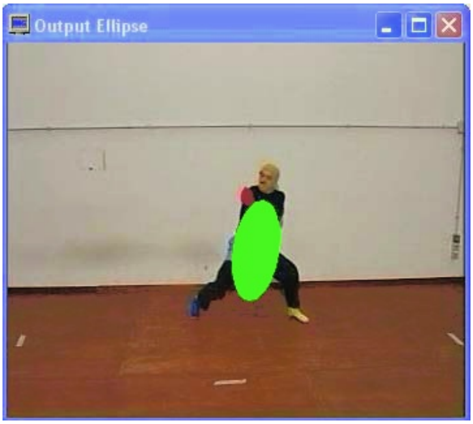


**Fig. 2.** Silhouette shape and orientation. The ellipse approximates the silhouette; its axes give an approximation of the silhouette's orientation.

The EyesWeb Motion Analysis Library also includes blocks and patches extracting 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 [12] 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 was considered to be moving if the area of the motion image (i.e., the QoM) was 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 [13]. The space is divided into active cells forming a grid. A point moving in the space is considered and tracked. Three main kinds 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 (e.g., 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 [11]) 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.

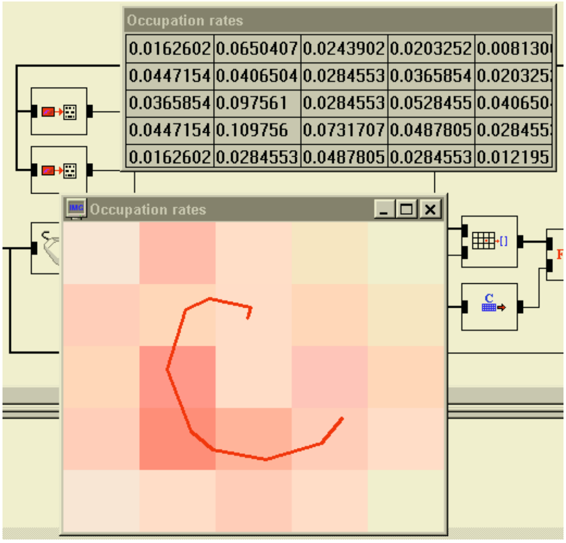
The EyesWeb Space Analysis Library implements the model and includes blocks allowing the definition of interacting discrete potentials on 2D spaces, the definition of regions, and the extraction of cues (such as, for example, the occupation rates of regions in the space). For example, Figure 3 shows 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.

### 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, and feet tracked by means of color tracking techniques – occlusions are not dealt with at this stage) 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.



Occupation rates

| | | | | |
|---|---|---|---|---|
| 0.0162602 | 0.0650407 | 0.0243902 | 0.0203252 | 0.008130 |
| 0.0447154 | 0.0406504 | 0.0284553 | 0.0365854 | 0.020325 |
| 0.0365854 | 0.097561 | 0.0284553 | 0.0528455 | 0.040650 |
| 0.0447154 | 0.109756 | 0.0731707 | 0.0487805 | 0.028455 |
| 0.0162602 | 0.0284553 | 0.0487805 | 0.0284553 | 0.012195 |

**Fig. 3.** Occupation rates calculated with the EyesWeb Space Analysis Library for a trajectory in a 2D space divided into 25 cells. The displayed trajectory refers to the last 25 frames (i.e., 1 s), but the occupations rates are calculated for the whole trajectory since the start of the gesture.

Examples of geometric features are the length of a trajectory, its direction and its *Directness Index* (DI). The Directness Index is a measure of how much a trajectory is direct or flexible. In the Laban's Theory of Effort [10] it is related to the Space dimension. In the current 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 kinematical 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/she 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 physical space this approach can be used to idntify 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 whose axes might be for example some of the expressive cues discussed above, the approach could allow grouping similar gestures, or gestures communicating the same expressive intention. Further developments and experiments in these direction are presented in a forthcoming paper.

## 3 Evaluation and Validation of Expressive Cues

Some of the algorithms for the extraction of expressive cues included in the EyesWeb Motion Analysis Library have been recently validated through a collection of perceptual experiments.

For example, in an experiment on segmentation of dance performances in pause and motion phases, spectators were asked to segment dance fragments in pause and motion phases (corresponding to a gesture or to a part of it). Two different dance fragments were used in the experiment, characterized by hard and smooth movement. Preliminary results evidenced the main effect of the expressive qualities (hard vs. smooth) in the perceived segmentation. Results were compared with automatic segmentation performed by the developed algorithms. The algorithms generally revealed to be more sensible than humans since they could identify motion and pause phase that humans were not able to distinguish.

In another experiment, subjects were asked to indicate through continuous measurement how much energy they perceive in a dance fragment and how much contraction they perceive in the dancer body. Results (available in a forthcoming paper) are compared with the automatically measured Quantity of Motion and Contraction Index cues.

## 4 Conclusion

The EyesWeb Expressive Gesture Processing Library has been employed in a number of artistic events (list and description of performances available at the EU IST MEGA project website www.megaproject.org), and in therapy and rehabilitation [14].
This library consists of a distinct and separate add-on with respect to the EyesWeb open software platform and includes some of the research and development carried out during the three-year EU IST Project MEGA.

Novel algorithms and related software modules for the EyesWeb Expressive Gesture Processing Library are currently under development, including for example refined motion tracking (e.g. with multiple cameras), extraction of new cues, machine-learning techniques for higher-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, and in particular Paolo Coletta, Massimiliano Peri, and Andrea Ricci.

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., Mazzarino B., Ricchetti M., Timmers R., Volpe G. (2004), "Multimodal analysis of expressive gesture in music and dance performances", in A. Camurri, G. Volpe (Eds.), "Gesture-based Communication in Human-Computer Interaction", LNAI 2915, Springer Verlag, 2004.
2. 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.
3. Camurri, A., Coglio, A. "An Architecture for Emotional Agents". IEEE MULTIMEDIA, 5(4):24-33, Oct-Dec 1998.
4. Camurri, A., Ferrentino, P. "Interactive Environments for Music and Multimedia. ACM MULTIMEDIA SYSTEMS, 7:32-47, Special issue on *Audio and Multimedia*, January 1999, ACM-Springer.
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. Liu Y., Collins R., Tsin Y., "Gait Sequence Analysis using Frieze Patterns" European Conference on Computer Vision, Copenhagen, May 2002, pp.657-671.
7. 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.
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. Laban, R., Lawrence F.C., "Effort", Macdonald & Evans Ltd. London, 1947.
11. Laban, R., "Modern Educational Dance" Macdonald & Evans Ltd. London, 1963.
12. 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.
13. 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.
14. Camurri A., Mazzarino B., Volpe G., Morasso P., Priano F., Re C., "Application of multimedia techniques in the physical rehabilitation of Parkinson's patients", Journal of Visualization and Computer Animation (In Press).