Gestures Saliency: a Context-based Analysis

Matei Mancas[1], Donald Glowinski[2], Pierre Bretéché[3], Jonathan Demeyer[1], Thierry Ravet[1], Gualtiero Volpe[2], Antonio Camurri[2]

[1] Engineering Faculy of Mons (FPMs) – TCTS Lab, Bd. Dolez, 31, 7000 Mons, Belgium
{Matei.Mancas, Jonathan.Demeyer, Thierry.Ravet}@fpms.ac.be

[2] University of Genoa - Casa Paganini/InfoMus Lab, Viale Causa, 13, 16124 Genoa, Italy
{Donald.Glowinski, Gualtiero.Volpe, Antonio.Camurri}@unige.it

[3] University of Franche-Comté – Laseldi Lab, Cours L. Leprince Ringuet, 25211 Montbéliard, France
Pierre.Breteche@univ-fcomte.fr

Abstract. Gestures' expressivity, as perceived by humans, may be related to the amount of attention they attract. In this paper, we present two experiments that quantify behavior saliency by the rarity of selected motion and gestural features in a given context. The first one deals with the current quantity of motion of a person's silhouette compared to a brief history of his quantity of motion values. The second one focuses on the motion speed of a person compared to the motion speed of other persons around him. Considering both features (speed and quantity of motion) and contexts (space and time), we compute a rarity index providing cues on the behavior novelty. This can be considered as a preliminary step to an expressive gesture analysis based on behavioral changes.

Keywords: computational attention, saliency, rarity, data fusion, tracking.

1 Introduction: towards a context-based gestural analysis

A lot of research effort has been devoted to robustly track humans in a scene and to analyze their gestures in order to individuate and characterize their behaviors. Gestural analysis, often applies in situations where either the human on which the analysis is carried on is preliminary selected or the same kind of analysis is performed to all the subjects that can be distinguished in the scene. A recent field of research aims at investigating collective behaviors [1]. Still, the object of the analysis is already defined and the work mainly focuses on characterizing collective displacements. The possibility of dynamically selecting the person to carry analysis on or to adapt analysis to the current behavior of a person in a context-dependent way would open new directions for gesture research. Human beings naturally show the capacity to dedicate limited perceptual resources to what is of particular interest. However, computers capacity to exhibit a behavior worthy of attention remains very limited. In this paper, we present a computational attention-based [2] setup that deals with (i) fusing data to obtain a robust tracking of elements in the scene (ii) implementing an index of rarity that highlights a salient behavior of a subject on which further gesture analysis can be carried out.

2 Implementation

Robust Tracking

We designed a non-invasive and reliable tracking technique that can adapt to flexible environments with regard to illumination changes. This technique, implemented in real-time on the EyesWeb XMI software platform [3], fuses information on the participants' head positioning (X and Y position) from two different video tracking approaches applied to two different video streams (color and infra-red). The data fusion is achieved by using a weighted mean between both sets of coordinates. The weight (confidence level) changes according to the participants' visibility with respect to the camera's field of view (FOV) and obstruction possibility from 0 when they are not visible by the camera and 1 when they are. For the color modality, the confidence level also gradually changes between 0 and 1 depending on blob area variations: abrupt variations dues for example to sudden illumination changes decrease the confidence level whereas stable blob areas over time increase it. Our data fusion approach enables the tracking system to obtain reliable positioning data over time.

Salient Gestures: an attention filter

Following our previous work, mainly achieved on static images [2], we found that a factor leading to saliency is event rarity which induces surprise. Two real-world scenarios were elaborated to test our motion saliency in the time and spatial contexts. The first scenario deals with the current quantity of motion for a participant's silhouette compared to a 4-second history of his quantity of motion values. Feature saliency is here inversely proportional to its rarity. For this scenario, the computer exhibits human-like reactions to the participant's gestures: if the person performs a high quantity of motion after a period of rest, this is interesting. But when he keeps on with a high quantity of motion, the saliency exponentially decreases. Moreover, if, after more than 4 seconds, the person decides to slow down or to stop moving, its saliency will be high again. A second scenario was conducted to test a spatial context of attention with a different feature: motion speed. Two participants were instructed to move simultaneously (e.g., swinging together) whereas a third one was performing a contrasted action compared with the two others (e.g., staying still). Higher feature contrast is related here with a higher saliency: if two participants have a low speed and the third one is faster, the computer is attracted by the third participant, but if two participants move with a similar speed while a third one stopped, the computer will pay more attention to the stopped participant than to the moving ones. Both cases show that a feature is not salient by itself but only in relation with a given context: the lack of motion can be more salient than motion in a specific context.

References

1. Hongeng, S. and Nevatia, R., 2001, "Multi-agent event recognition" in ICCV, pp. II: 84-91.
2. Mancas M., Gosselin B., Macq B., 2007, "A Three-Level Computational Attention Model", Proc. of ICVS Workshop on Computational Attention & Applications, Germany.
3. Camurri, A., De Poli, G., Leman, M., and Volpe, G., 2005. "Toward Communicating Expressiveness and Affect in Multimodal Interactive Systems for Performing Art and Cultural Applications", IEEE Multimedia, 12,1 (Jan. 2005), 43-53.