Using motion capture to recognize affective states in humans

Nadia Bianchi-Berthouze
UCLIC, University College London, London, UK, n.berthouze@ucl.ac.uk

Humans convey and recognize affective states from a broad spectrum of verbal and nonverbal modalities. An historical focus on face as a primary modality for conveying and recognizing emotions spurred the computer science community to research methods for computer systems to automatically recognize emotions from facial expressions. Today, the Facial Action Coding System (FACS) [1] is the most popular standard for systematically categorising (facial) expressions of emotions. Recent psychology studies, however, have revealed that another form of nonverbal communication, body posture, can prove a very good indicator for various categories of emotion. Whilst these studies have been used rather effectively to enable artificial systems to express affective behaviour through posture (e.g., Sony's AIBO), posture still has no equivalent to FACS and most existing studies use only coarse-grained posture descriptors (e.g. leaning forward, slumping back).

Over the last 5 years, my research has been focused on investigating the extent to which low-level features of body postures provide the information necessary to recognize not only basic emotional states but also more subtle nuances, cross-cultural differences, and affective dimensions. In this paper, I briefly review how I used motion capture to understand how humans recognize affective states from postures and how build automatic affective posture recognition models that could make technology more engaging.

Low-level description of posture

To establish the groundwork for a FACS-like formal model, we used motion capture to record 3D affective gestures from actors of different age, gender and race. Each actor was asked to perform an in-place gesture expressing happiness, sadness, fear or anger. The actors were not forced in their acting, but were allowed to express the emotions in their own natural way. The actors were not allowed to observe each other during performance. Each actor was dressed in a suit with 32 markers on the joints and body segments. 32 markers provide sufficient information to describe the posture. Each gesture was captured by 8 cameras and represented by consecutive frames describing the position of the 32 markers in the 3D space. A total of 109 gestures were collected. For each gesture, we selected the frame (i.e., a static posture) that the actor evaluated as being the most expressive instant of the gesture. Each frame was described using 24 posture features, chosen based on the concept of ''sphere of movement'' used in dance to convey emotion. Direction and volume of the body were described by projecting each marker on the 3 orthogonal planes and measuring the lateral, frontal and vertical extension of the body. Each posture was initially rotated to simulate a frontal view of the posture. The features computed include rotation angles describing head and torso configurations as well as a series of distances between key anatomical landmarks. Each feature was normalized according to the body structure of the actor, i.e., according to the maximal extension of his/her body. For example, the lateral opening of the right arm was computed by the ratio of the distance between the right hand and the left shoulder along the X-direction, and the maximum lateral extension of the arm. In [2], we showed how a trained associative neural network could successfully classify just over 70% of 102 postures extracted from natural human motion capture data. Adding a measure of the direction of the movement to the postural descriptor allowed for a significant improvement (+8%) of postures that showed the lowest inter-observer agreement [3]. In [4], we tested the informational content of the posture descriptors by applying mixed discriminant analysis (MDA) and looking at whether the features could account for different levels (high, low) of three affective dimensions: arousal, valence and action tendency. The results showed a 1% error on arousal, 20% on valence and 25% on action tendency. Finally, we showed that by using both supervised and unsupervised learning mechanisms, nuances of emotions could be recognized from these low-level features, with performance similar to that of human observers [5].

How do people interpret affective posture?

The above motion data enabled us to create faceless computer characters to build an understanding of how people interpret affective postures. Using the above descriptive system, our studies revealed significant effects of both culture [6] and gender [7] on the affective appraisal of body postures. Interestingly, our set of low-level feature descriptors did also provide a mechanistic explanation to recent findings in neuroscience suggesting that the face fusiform area (FFA) – the brain area responsible for facial processing – was involved in processing postural expressions of affect even when facial cues were removed [8]. Indeed, our statistical analysis showed that features related to head configuration (e.g., inclination and rotation of the head) were very important in discriminating between emotions [9] and between nuances of a given emotion in particular [5]. This body of work thus suggests that posture could be used, if not as an alternative to facial expressions, at least in conjunction with facial expressions to provide for finer grain appraisals and increased discriminatory power in the case of ambiguous or incongruent information. This is not the only contribution of posture to our study of emotion in human-machine interaction, however.

Beyond acted postures

Whereas the previous studies relied on acted postures, our current research is concerned with real-world scenarios in which the expressions can be more subtle and mixed. More specifically, the AffectME project (http://www.cs.ucl.ac.uk/staff/n.berthouze/AffectME.html) considers two scenarios: games involving full-body controllers (e.g., Wii), and rehabilitation of patients with chronic pain. The game industry has recently introduced full-body controllers, presumably to help create a more natural and engaging experience for the players. Yet, there have been few studies aimed to understand the relationship between body movement and engagement. In [10], we used using motion capture to quantify differences in movements between players using a traditional game pad controller and those using a full-body controller, and correlated them with measures of engagement [11]. In the clinical study, we are applying the above framework to automatically discriminate between different communicative roles of body movement in chronic pain patients. Although pain, as such, is not an emotion, it is associated with a set of negative emotions that will express at the postural level. Studies in non-verbal behavior and pain have shown that movement in patients convey three different types of information: the physical reaction to pain; the affective experience related to pain; and the search for empathy and attention of solicitous others (e.g., a partner or a practitioner). While we are still at very preliminary stage in this study, our aim is to create a computational model of body movement able to separate such components so as to enable the creation of technology to support patients in self-directed rehabilitation programs. Motion capture systems provide us with a unique source of accurate and rich data to inform the design and the validation of models with use in various areas of human machine interaction.

References

1. Ekman, P. and Friesen, W. (1978). Manual for the facial action coding system. Consulting Psychology Press, Palo Alto.
2. Bianchi-Berthouze, N., Kleinsmith, A. (2003). A categorical approach to affective gesture recognition. *Connection Science* **15(4)**, 259–269.
3. Kleinsmith, A., Fushimi, T., Berthouze, N. (2005). An incremental and interactive affective posture recognition system. *Int. Workshop on Adapting the Interaction Style to Affective Factors*, *Edinburgh, UK*.
4. Kleinsmith, A., de Silva, R., Bianchi-Berthouze, N. (2005). Grounding affective dimensions into posture descriptions. *Int. Conference on Affective Computing and Intelligent Interaction*, 263–270.
5. de Silva, R., Kleinsmith, A., Bianchi-Berthouze, N. (2005). Towards unsupervised detection of affective body nuances. *Int. Conference on Affective Computing and Intelligent Interaction*, 32–39.
6. Kleinsmith, A., De Silva, R., Bianchi-Berthouze, N.: Cross-cultural differences in recognizing affect from body posture. Interacting with Computers, 18 (6), 1371-1389, 2006
7. Kleinsmith, A., Bianchi-Berthouze, N., Berthouze, L. (2006). An effect of gender in the interpretation of affective cues in avatars. *Workshop on Gender and Interaction, Venice, Italy*.
8. de Gelder, B., Snyder, J., Greve, D., Gerard, G., Hadjikhani, N. (2003). Fear fosters fight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *PNAS* **101(47)**, 16701–16706.
9. de Silva, R., Bianchi-Berthouze, N. (2004). Modeling human affective postures: An information theoretic characterization of posture features. *J. Computational Agents and Virtual Worlds* **15**, 269–276.
10. Bianchi-Berthouze, N., Whan Woong, K., Patel, D. (2007). Does body movement engage you more in digital play? And why? *International Conference on Affective Computing and Intelligence Interaction, Lisboa (Portugal)*, LNCS **4738**, 102-113.
11. Cairns, P. Cox, A., Berthouze, N., Dhoparee, S. (2006). Quantifying the experience of immersion in games. *Workshop on Cognitive Science of Games and Gameplay, Vancouver, Canada*.