Engineering Applications of Artificial Intelligence 51 (2016) 124–135

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Associating gesture expressivity with affective representations

Lori Malatesta [a], Stylianos Asteriadis [b], George Caridakis [a,d], Asimina Vasalou [c], Kostas Karpouzis [a]

[a] Image, Video and Multimedia Lab, National Technical University of Athens, Greece
[b] Department of Knowledge Engineering, University of Maastricht, Netherlands
[c] London Knowledge Lab, UCL, United Kingdom
[d] Department of Cultural Technology, University of the Aegean, Greece

ARTICLE INFO

Available online 15 February 2016

Keywords:
Affective computing
Gesture expressivity parameters
Association of representations
Modelling expressivity
Expressivity judgment study
Neuro-fuzzy network

ABSTRACT

Affective computing researchers adopt a variety of methods in analysing or synthesizing aspects of human behaviour. The choice of method depends on which behavioural cues are considered salient or straightforward to capture and comprehend, as well as the overall context of the interaction. Thus, each approach focuses on modelling certain information and results to dedicated representations. However, analysis or synthesis is usually done by following label-based representations, which usually have a direct mapping to a feature vector. The goal of the presented work is to introduce an interim representational mechanism that associates low-level gesture expressivity parameters with a high-level dimensional representation of affect. More specifically, it introduces a novel methodology for associating easily extracted, low-level gesture data to the affective dimensions of activation and evaluation. For this purpose, a user perception test was carried out in order to properly annotate a dataset, by asking participants to assess each gesture in terms of the perceived activation (active/passive) and evaluation (positive/negative) levels. In affective behaviour modelling, the contribution of the proposed association methodology is twofold: On one hand, when analysing affective behaviour, it can enable the fusion of expressivity parameters alongside with any other modalities coded in higher-level affective representations, leading, in this way, to scalable multimodal analysis. On the other hand, it can enforce the process of synthesizing composite human behaviour (e.g. facial expression, gestures and body posture) since it allows for the translation of dimensional values of affect into synthesized expressive gestures.

1. Introduction

Human behaviour has been often studied both for the purposes of inferring internal affective aspects from observations, and for fostering the quality and believability of synthesized actions of virtual characters. Whether studying low level behavioural cues – and, thus, focusing on aspects like gestures, body, posture and facial expressions – or emphasizing high level phenomena such as affective states and dispositions, often the case is that researchers will adopt different computational models or representations in order to model emotions and related phenomena. Such variation is not necessarily problematic since it enables the investigation of subtle differences amongst approaches. Nevertheless, when it comes to building an affect aware system (either targeting analysis, or synthesis components) such variations raise interesting questions regarding the correspondence between entities in each representation scheme.

Affective cues can be broadly categorised across two broad families, verbal and non-verbal communication channels. In the presented work, we focus on the analysis of non-verbal behavioural channels. A substantive body of research has focused on non-verbal behaviour, within the fields of psychology, cognitive science and human computer interaction, stressing the importance of *qualitative* expressive characteristics of body motion, posture, gestures, facial expressions (Ioannou et al., 2007), eye gaze (with eye gaze still necessitating specialized hardware (Bengoechea et al., 2013; Jennett et al., 2008) and overall human action recognition during an interaction session (Wallbott, 1998; Pelachaud, 2008; Knapp and Hall, 2013). Qualitative affective cues contain significant information about the user's non verbal behaviour and communication. In this context computationally formulated qualitative expressivity features (e.g. fluidity of a gesture performance) correspond to the intermediate layer between extracted quantitative features (e.g. coordinates of the hand position) and the

E-mail addresses: lori@image.ntua.gr (L. Malatesta), stelios.asteriadis@maastrichtuniversity.nl (S. Asteriadis), gcari@image.ntua.gr, gcari@aegean.gr (G. Caridakis), A.Vasalou@ucl.ac.uk (A. Vasalou), kkarpou@image.ntua.gr (K. Karpouzis).

http://dx.doi.org/10.1016/j.engappai.2016.01.010
0952-1976/

conveyed emotion in the form of the adopted affective representation approach.

Behaviour expressiveness is an integral part of the communication process since it can provide information about the current emotional state and the profile of the interlocutor, as well as his performance.

Although the research field of human behaviour analysis has primarily focused on human to human interaction, there has been growing recognition of the importance of accounting for Human-Computer Interaction (HCI). Most examples of studies incorporating gesture expressivity in the HCI context (Vinciarelli et al., 2012; Caramiaux et al., 2015), however, have tended to focus on the expressively-enhanced *synthesis* of gestures by virtual agents and ECAs (Caridakis et al., 2007; Cassell et al., 2004; Martin et al., 2006; Kipp et al., 2007; Pelachaud, 2008; Hartmann et al., 2005) which, many times, follow animation patterns that depend on low-level features (e.g. tracking) and only partly depend on semantic interpretation of human's emotional or cognitive state.

The present work focuses on expressivity in gesturing. The approach adopted is holistic in the sense that the gestures studied are not recognized or broken down to their components. Emphasis is given to the expressive content of a closed set of singular gestures with clear semantic meaning (such as waving goodbye or clapping). The contribution of the research work presented in this article lies in the association of existing results on automatically calculated expressivity parameters (Caridakis et al., 2006), with dimensional representations of affect. This is done by incorporating a properly, data-driven trained neuro-fuzzy network. The proposed association allows for the inclusion of expressivity parameters in the fusion process with other modalities that commonly use dimensional representations of affect.

2. Related work

In computational behaviour analysis, according to a survey paper by Kleinsmith and Bianchi-Berthouze (2013), research on non-verbal affect recognition has mostly focused on facial expressions starting with the FACS coding system developed by Ekman and Friesen (1977) and moving to more recent computational approaches (Zhao et al., 2003; Pantic and Rothkrantz, 2000). The research shift towards bodily expressions has only started recently. According to the same survey, specific features of bodily expressivity have been identified to contribute to the recognition of specific affective states. In the case of upper body expressivity and gestures, there exist several manual annotation approaches on gesture analysis (Foster, 2004; Ferré et al., 2007; Kipp and Martin, 2009), while research on the *automatic analysis* of gesture expressivity is ongoing (Varni et al., 2010; Caridakis et al., 2006; Sanghvi et al., 2011; Pantic et al., 2007; Griffin et al., 2013; Glowinski et al., 2011, Kleinsmith and Bianchi-Berthouze (2007), Kleinsmith et al. (2011)), rendering human action analysis asymmetrically less studied with regards to its synthesis counterpart. The main reason behind this is that robust software or dedicated motion capture hardware (Pfeiffer et al., 2013) are needed in order to support analytic methods such as hand trajectory extraction which returns an abundance of data of high detail and richness, especially when it comes to these observations taking place in spontaneous, natural interaction contexts (Cowie et al., 2008). The attribution of affective labels on such data is not a straight forward task. Gesture expressivity – similarly to other types of bodily expressivity – can be interpreted in various ways, leaving a lot of room for subjective assessment. In order to establish ground truth for emotion expression, it is common practice to rely on the judgment of observer coders (Kleinsmith and Bianchi-Berthouze, 2013).

Another trend that has attracted attention in non-verbal behaviour studies is the role of multimodality (Caridakis et al., 2010): signals coming from different emotional channels (Zeng et al., 2009) inform a system's computational intelligence module regarding the emotional or cognitive state of the user. Synergy of multiple modalities (Kapoor et al., 2007) is expected to overcome problems related to reliability, noise and personalization. Other typical examples are reported in Castellano et al. (2009) and Sanghvi et al. (2011) where a Bayesian network uses information coming from posture and gaze, in order to detect engagement with a robot companion (Van Breemen et al., 2005) that is able to pose various expressions. Salem et al. (2012) also investigated the gesture and posture expressivity aspects from a Human Robot Interaction perspective.

Expressivity of body movement (Laban and Lawrence, 1974) is a qualitative cue that is, or at least should be, incorporated in the design process of such applications. In the words of Alex Pentland (1996): "The problem, in my opinion, is that our current computers are both deaf and blind: they experience the world only by way of a keyboard and a mouse… I believe computers must be able to see and hear what we do before they can prove truly helpful". Moving a step further, we might add, that they should also interpret appropriately what they see and hear.

Behaviour expressiveness is an integral part of the communication process since it can provide information about the person's current emotional state, the profile of the interlocutor and metrics of his/her performance. Many researchers have studied characteristics of human movement and coded them in binary categories such as slow/fast, restricted/wide, weak/strong, small/big, unpleasant/pleasant in order to properly model expressivity. *Expressivity dimensions* are considered as the most complete approach to body expressivity modelling, since they cover the entire spectrum of expressivity parameters related to emotion and affect (Karpouzis et al., 2007; Sykes, 2003).

3. Motivation

Non-verbal behaviour has been frequently broken down to its communicative functions (start/end conversation, emphasize, depict object etc) and the behaviours that manifest these functions (nod, body posture, gaze aversion etc.) (Kopp et al., 2006; Vilhjálmsson et al., 2007). One communicative function can be expressed through one or more behaviours and, vice versa, one single behaviour can express one or more functions. In our case we have chosen as behaviours a closed set of expressive gestures with a non-ambiguous semantic meaning. Our goal is to investigate and attempt to quantify how the same gestures, with the same functions, can convey different affective messages through their expressivity features.

An important aspect when studying gesture expressivity is that of subjectivity, mostly in terms of perceiving the conveyed emotion when a gesture is performed. In the case of facial expressions, several sets of universally recognisable emotions exist, with Ekman's being the most prominent (Ekman and Friesen, 1977). However, in the case of gesturing, the cultural background of the interactants plays an important role both when performing a gesture, as well as interpreting it in the receiving end. Kita (2009) (Schroder, 2004) elaborates on the culture-specific conventions for form-meaning associations in emblem gestures (e.g., the thumbs-up sign), and on how cognitive and cultural differences shape iconic and deictic gestures expressing spatial or temporal concepts. As a result, a perceiver-based annotation scheme is needed so as to obtain labels and ratings which can be used as 'ground truth' for any machine learning approach.

This work extends research by Caridakis et al. (2006) on gesture expressivity parameters. These parameters are the result of a qualitative approach to modelling non-verbal upper body expressivity based on computer vision algorithms. Castellano et al. (2009) and, later, Glowinski (Glowinski et al., 2011) have also studied abstract representations of gesture expressivity and their relation to emotion