Decoding Emotions from Facial Animations

Shazia Afzal[*]
Computer Laboratory
University of Cambridge

Tevfik Metin Sezgin[†]
College of Engineering
Koç University

Peter Robinson[‡]
Computer Laboratory
University of Cambridge

Keywords: Facial expression analysis, animation

1 Introduction

Facial feature point tracking is used as a primary input in several systems that perform affect recognition using facial expressions. To determine the efficacy of automatically extracted facial feature points in encoding emotional content, we conducted an experiment that compared human raters judgements of emotional expressions between actual video clips and three automatically generated representations of them (point-light displays, stick-figure models and 3D animations, Fig. 1). Although our main objective was to assess the utility of automatically extracted facial feature points in conveying emotions, our results give interesting insights into optimal representation of facial displays in emotion judgments as well as in analysing the perceptual quality and realism of the animations. Here we discuss results from our experiment that we consider are of relevance and interest to this symposium. Preliminary findings and details of the experiment are discussed in [1], while a more detailed description and analysis is in progress.

2 The Experiment

14 participants in the age-group of 20 to 34 were asked to identify emotions in video stimuli presented to them in a randomised order. For each video a primary and an optional secondary emotion label was recorded. The primary label was considered as the true response emotional label and used to compute the accuracy of judgement. The secondary label, when present, was used as an indicator of ambiguity. Finally, replay counts and decision times were used as indicators of difficulty.

The stimulus material was compiled using samples taken from four different databases. These were selected to represent a range of posed and naturalistic experimental control conditions. Five examples for interest, confusion, boredom, happiness and surprise were taken from each database based on perfect agreement by three coders. State of art feature point tracking technology was then used to generate the three representations corresponding to each video sample. The point-based representation was created directly from the output of the face-tracker while minimal detail was added to these landmarks to produce the stick-figure animations. Finally, the automatically tracked feature points were directly converted into a set of MPEG-4 defined FAPs for driving the 3D animations.

3 Findings

Our study was designed to investigate how the accuracy of emotion recognition is affected by the nature and representation format of stimuli generated from automatically tracked facial feature points. The type of representation and database appeared consistently as the main influencing factors for accuracy, difficulty as well as ambiguity in classification performance. Moreover, type of emotion was found to be related to the source database in determining the accuracy. Original videos showed higher recognition rates consistently across representations and databases. Surprisingly however, the stick-figure models showed relatively higher levels of recognition accuracy compared to both the point-light and 3D animations.

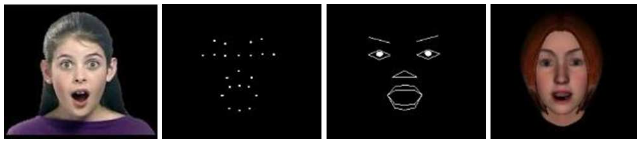

Figure 1: *Example of the three illustrations generated from an original video of surprised using automatically tracked facial feature points.*

Inter-rater agreements and recognition rates for emotions like happy and surprised were found to be consistently higher irrespective of the representation scheme used or the database sampled. This implies that the facial feature points commonly employed for emotion recognition using facial expression analysis may be suitable for inferring only some emotions and may not be sufficient to discriminate patterns for all emotions. States like confused and bored for instance, are accompanied by subtle changes in the face which are not adequately captured by the set of facial feature points.

Overall, these results provide new insights into perception of emotion from automatically generated facial displays. While the results indicate that automatic facial feature point tracking does in fact retain the underlying emotion dynamics, the efficiency of this largely depends on the type of data source as well as the emotion type. This becomes challenging specifically when handling naturalistic data.

The results suggest that an intermediate-level of representation, where only an outline of facial expressions is provided, affords better perception of emotion in automatically generated displays. This has implications for synthesis of emotions using computer animations. It is possible that the abstraction level of a stick-figure model allows rendering flaws to be ignored and to focus attention on emotionally salient movements. In contrast, complex models like the 3D animations may enhance flaws in renderings thereby diverting attention to non-significant areas or artefacts. The low recognition accuracy obtained for the 3D animation videos could in effect be attributed to the quality of the animations, or to the difficulty of automatic generation of well coordinated eye-gaze, facial displays and head gestures. There is some evidence supporting the Uncanny Valley Theory and users discomfort with highly realistic portrayals of embodied agent behaviour. Whether or not such a negative preference would explain the increase in difficulty and reduced accuracy in recognising 3D animations is a subject requiring further exploration.

4 References

[1] S. Afzal, T.M. Sezgin, Y. Gao and P.Robinson, Perception of Emotional Expressions in Different representations Using Facial Feature Points, International Conference on Affective Computing & Intelligent Interaction, 2009.

[*] e-mail: Shazia.Afzal@cl.cam.ac.uk
[†] e-mail: mtsezgin@ku.edu.tr
[‡] e-mail: Peter.Robinson@cl.cam.ac.uk