The emotional computer

Peter Robinson, Tadas Baltrušaitis, Ian Davies, Tomas Pfister, Laurel Riek and Kevin Hull

University of Cambridge Computer Laboratory

Abstract. When people talk to each other, they express their feelings through facial expressions, tone of voice, body postures and gestures. They even do this when they are interacting with machines. These hidden signals are an important part of human communication, but most computer systems ignore them. A major challenge for pervasive computing is to appreciate emotions as part of the context in which man-machine communications are conducted. This paper reports on work by a team in the Computer Laboratory at the University of Cambridge who are exploring the role of emotions in human-computer interaction. The research is also shown in the accompanying video, *The Emotional Computer*.

1 Introduction

Charles Darwin published *The expression of the emotions in man and animals* in 1872, exploring the role of emotional expression in communication between humans [5]. Over a century later, Rosalind Picard observed that effective communication between people and computers also requires emotional intelligence [10]; computers must have the ability to recognize and express emotions.

The study of affective computing has blossomed subsequently. This paper presents a summary of some of the challenges involved in affective computing, and illustrates them with examples drawn from work in the Computer Laboratory at the University of Cambridge[1]; it is not intended as a comprehensive survey. The accompanying video, *The Emotional Computer*[2], gives a light-hearted account of some of the projects.

2 Recognising emotions

Although Darwin concentrated on facial features to convey emotions in Expression, he also mentions vocal sounds, other sounds, body posture and gesture, and physiological responses as further indications of emotion. All of these channels have been considered as ways of automatic monitoring emotion in humans, although these sensors used for some are more invasive than for others. Signals that can be monitored non-invasively are of particular interest for pervasive computing.

Pervasive 2011, June 12 - 15, San Francisco, CA, USA

[1] http://www.cl.cam.ac.uk/emotions/

[2] http://www.youtube.com/watch?v=whCJ4NLUSB8

2.1 Facial expressions

People routinely express their mental states through their facial expressions and this is one of the clearest channels for communication. Inference from facial expressions has been studied using a variety of techniques but mostly restricted to six basic emotions [7]. Recognising complex, cognitive mental states is more difficult, but probably more useful as part of general interaction with computer systems. We have developed a fully automatic system requiring no human intervention which operates in real-time [8,9]. Our Facial Affect Inference System uses a multi-level representation of the video input, combined in a Bayesian inference framework operating at four levels: facial feature points, FACS action units (AUs) [6], gestures composed of several AUs, and mental states.

Videos from the Mind Reading DVD [2] were used to train statistical classifiers in the inference system. The evaluation considered six conditions including 29 of the 412 underlying mental state concepts chosen to be particularly relevant for human-computer interaction. For a mean false positive rate of 4.7%, the overall accuracy of the system is 77%. The system also generalises well to faces not included in the training data.

2.2 Non-verbal aspects of speech

The voice provides another significant channel for the expression of emotions. Features such as the pitch, energy and tempo can reveal a lot about the mood of the speaker. However, it seems that it is not possible to identify features that indicate particular mental states directly. It may be possible to distinguish between using two emotions using one or two particular features, but a different set of features may be required to distinguish those emotions from others.

Our approach [14] has been to calculate a large collection of about 170 features for each utterance. A training phase uses data mining to identify the features that separate each pair of emotional conditions. The operational phase then uses these pair-wise comparisons as preferences in a voting scheme to give an overall ranking. Two voting schemes have been considered: Condorcet attempts to find a single winner and a threshold system allows multiple winners.

Selecting a single winner gives an overall recognition accuracy of 70% between nine conditions. The threshold method includes the correct result in 83% of the trials, compared with a random rate of 14%. The approach also generalises well to speakers other than those use for training and even to other languages.

2.3 Body posture and gesture

The third natural channel for expression of emotions includes body posture and gesture. However, we need to discount elements that are governed by the movement being considered and who is doing it before we can analyse how it is being done. Movement involves a strong individual bias, so the analysis is harder than for facial expressions or voice [3,4].

Our approach breaks complex motions down into a system of isolated elements whose dynamic cues can be used to distinguish affects. An average feature vector is then calculated over all the motions by an individual and this is used to factor out the individual's motion idiosyncrasies. Finally, support vector machines with a polynomial kernel are used to classify the emotion. As with speech, pair-wise comparisons are used on individual motion segments, and then the segment is classified using a majority vote. A complete motion is then classified by a majority vote of the classifications of its component segments.

The method was tested on a corpus of about 1200 motion samples, representing roughly equal numbers of neutral, happy, angry and sad expressions of four different actions. The average recognition rate of 81% is comparable to the rates achieved by human observers of similar data. We are currently looking at multimodal inference which combines two or more channels for improved accuracy, and at possible applications.

3 Expressing emotions

In psychology it is well understood that humans and some non-human mammals can convey empathetic responses through involuntary facial mimicry. Might this enhance human-robot interaction? A simple preliminary experiment used a robot to mirror back some expressions a human makes to it in real-time [11]. Participants were invited to talk to a robot about their experiences arriving in the town and travelling to the laboratory. It reacted in two ways – either by mimicking the subject or by moving randomly– and we found that participants in the facial-mimicking group found the interaction more satisfying. Of course, this experiment is naïve, but it does indicate that expression of emotions by robots plays a part in their communication with humans.

This gives rise to questions concerning the degree of human-likeness required in the appearance of robots that interact with humans. A further experiment asked subjects to watch video clips of robots with a variety of forms [12]. Two clips were shown for each character, one showed the robot being treated cruelly and the other was emotionally neutral. The participants were asked how sorry they felt for the character, and the responses were directly correlated with human-likeness.

The technology is currently being tested using a high-fidelity robotic head which simulates movements disorders that might be encountered by trainee doctors [13]. The same system is being used to identify how well emotions are conveyed in different synthetic representations [1]. We are currently looking at its use as an intervention for autism therapy.

4 Acknowledgements

The work described here was undertaken by a team of gifted colleagues in the Computer Laboratory at the University of Cambridge – notably Andra Adams, Shazia Afzal, Tadas Baltrušaitis, Ntombi Banda, Daniel Bernhardt, Ian Davies, Yujian Gao, Rana el Kaliouby, Marwa Mahmoud, Tomas Pfister, Laurel Riek, Metin Sezgin and Tal Sobol Shikler – with valuable advice from Neil Dodgson and Alan Blackwell. Simon Baron-Cohen and his colleagues in the Autism Research Centre have provided numerous technical insights. Kevin Hull produced and directed the *Emotional Computer* video. Parts of the work have been undertaken with support from the EPSRC, the European Commission, the Cambridge-MIT Institute, the Gates Cambridge Trust, Hanson Robotics, the Qualcomm Studentship, Thales Research & Technology, Toyota ITC, Vicon UK, the Neil Wiseman Fund and the University of Cambridge.

References

1. Tadas Baltrušaitis, Laurel D. Riek, and Peter Robinson. Synthesising expressions using facial feature point tracking: How emotion is conveyed. In *ACM Workshop on Affective Interaction in Natural Environments*, Florence, Italy, October 2010.
2. Simon Baron-Cohen, Ofer Golan, Sally Wheelwright, and J. Hill. Mind reading: the interactive guide to emotions. DVD, 2004.
3. Daniel Bernhardt and Peter Robinson. Detecting affect from non-stylised body motions. In *Affective Computing and Intelligent Interaction*, pages 59–70, Lisbon, Portugal, 2007. Springer-Verlag.
4. Daniel Bernhardt and Peter Robinson. Detecting emotions from connected action sequences. In *International Visual Informatics Conference*, Kuala Lumpur, 2009.
5. Charles Darwin. *The expression of the emotions in man and animals*. John Murray, London, 1872.
6. Paul Ekman and Wallace V. Friesen. *Facial action coding system: a technique for the measurement of facial movement*. Consulting Psychologists Press, 1978.
7. Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the human face*. Pergamon Press, New York, 1972.
8. Rana el Kaliouby and Peter Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *Computer Society Conference on Computer Vision and Pattern Recognition*, page 154, Washington, DC, 2004.
9. Rana el Kaliouby and Peter Robinson. Generalization of a vision based computational model of mind-reading. In *Affective Computing and Intelligent Interaction*, pages 582–589, Beijing, China, 2005.
10. Rosalind W. Picard. *Affective computing*. MIT Press, Boston, 1997.
11. Laurel D. Riek, Philip Paul, and Peter Robinson. When my robot smiles at me: enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, 3(1-2):99–108, July 2010.
12. Laurel D. Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In *Affective Computing and Intelligent Interaction*, Amsterdam, The Netherlands, October 2009.
13. Laurel D. Riek and Peter Robinson. Synthesising head and facial movement disorders on android robots. In *International Symposium on Facial Analysis and Animation*, Edinburgh, Scotland, October 2010.
14. Tal Sobol Shikler and Peter Robinson. Classification of complex information: inference of co-occurring affective states from their expressions in speech. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(7):1284–1297, July 2010.