7th Report, Innovative Interfaces

Robert Hoff

December 5, 2011

1 Context

As previously established, detecting affective states and understanding emotion in human behaviour may have many beneficial applications in computer applications and HCI [6]. This week's theme is about analysing and inferring emotions, by observing variables related to facial expressions [5], body motions [2] and speech [9]. Similar to each technique is observing certain metrics that can be reduced mathematically, to spacial points, or spectral frequencies in the case of speech analysis. All methods use data from external libraries compiled by external studies. Two of the papers are based on the Mind Reading library from the University of Cambridge [7]. The techniques differ in their mathematical approach and algorithms. There are a number of existing mathematical techniques such as support vector machines (SVM) [1], and dynamic bayesian networks (DBN) [8]. The two papers on speech and body motions are analogous to each other for inferring affective states from non-stylised behaviour, but use different variables as input and analytic approach. The papers on facial expressions and speech both recognises that affective states occur simultaneously, and their calculations are composed of attributions from different classes of emotions.

2 Summaries

2.1 Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures

Complex states include many emotions that fall outside of a basic set [10]. With respect to HCI it is some of the complex emotions that are considered the most important, in this paper - *agreeing, concentrating, disagreeing, interested, thinking and unsure.* Real-time interference is important to enable real-time responses in HCI. The data analysed was taken from the Mind Reading compilation [7]. A probabilistic approach (DBN) is used based on spacial points associated with facial features, the timetraces of points can be mapped to facial expressions, and thereby emotions. The method uses models of different layers, where increasingly higher levels of inferences are made.

A probability distribution is associated with the given six emotional states that were considered important. In tests there is a high agreement between the results and the actual results, but there are some false interpretations. However, compared to human inferences (humans also make mistakes), the system performs comparably.

2.2 Classification of Complex Information: Inference of Co-Occurring Affective States from their Expressions in Speech

The paper describes a method of analysing different emotions based on non-stylised motions, which are motions secondary to another performed action, such as walking or knocking a door. The basis for the technique are measurements of points on the body (taken from a pre-compiled library [11]) described by time-dependent spacial vectors (x,y,z). The experiment in the paper is concerned by analysing knocking, where the main points of interest are the right shoulder and arm. Motion-sequences and calculations based on the measurements may reveal something about the person's basic emotion, an energetic knock for example be related to excitement or anger, while the knock of a sad person may appear relatively slow and weak.

This particular approach requires knowledge of the intended action before analysis, but may be something that can later be solved by computer vision. The set of emotions were limited to only 4, neutral, happy, angry and sad. Given the complexity of human emotions, this is may be to coarse to have useful application. Data was taken from a library compiled from actors performing the actions (by enacting specific situations) that corresponding to one of the emotional states [11], so the sampling size was limited, and perhaps synthetic. This is in general a worry in all such datasets. The paper does illustrate that based on a small input, inferences of emotion can still be possible. Which is also a remarkable observation of human interpretation; the ability to match emotions based on very limited data (just light-points associated with joints on the body).

2.3 Detecting Affect from Non-stylised Body Motions

This paper concerns inferring affective states from non-verbal expressions in speech. In speech and voice, characteristics like intonation, rate and tempo are important. The system used data from the Mind Reading Database [7] which features a rich set of emotions. The technique works by extracting a large collection of metrics from the voice patterns, the metrics are related to frequency, energy and spectral content. The extracted metrics are compared pair-wise to pre-compiled results representing affective groups. Based on the comparison a tabulation of characteristics can be made for each voice pattern. The system is built for learning, that is it can add to it's compiled database to over time to make better predictions.

The paper emphasises that emotions should be considered as co-occurring, and that emotions can be considered as compositions. The paper shows that a selected of basic emotions (called affective-groups) may be used to determining more complex emotions. It is not that clear though how the basic set should be chosen, why the ones given should be considered more fundamental than others. The set of emotions chosen for analysis is similar to the paper on facial expressions and was made with applicability for HCI.

3 Evaluation

Inferring emotions can be divided into the observed variables, previous knowledge of emotions (based on databases), and the definition of emotions (based on psychology). Mathematical techniques vary between approaches, which may indicate further research is needed regarding underlying theory. One unified probabilistic framework has been proposed for the recognition of facial expressions [12]. A governing problem of determining emotion relates to their complexity, and in most cases lack of absolute answers. Verification is by surveying the participant, or checking if inferences match opinion of other people. Recent work on classification is on the extended Cohn-Kanade Database [4]. The techniques for facial expressions have been further refined, but still based Ekman's work on FACS [3]. Other databases that are important are the MMI-Facial Expression Database and the JAFFE database.

We may consider possible results of combining all the techniques described. Invariably, inferring emotions is inherently related to probabilities because of the inexact nature of their definitions. Therefore, we can expect systems to perform better if measurements are taken from across a range of observations.

References

[1] Kapur A. and et al. Gesture-based affective computing on motion capture data. *Affective Computing and Intelligent Interaction*, pages 1–7, 2005.

[2] Bernhardt D. and Robinson P. Detecting affect from non-stylised body motions. *Affective Computing and Intelligent Interaction*, 4738:59–70, 2007.

[3] Ekman P. and Friesen W. Facial action coding system. *Consulting Psychologists Press*, 1974.

[4] Lucey P. and et al. The extended cohn-kanade dataset. *IEEE Computer Vision and Pattern Recognition Workshop*, 2010.

[5] Kaliouby R. and Robinson P. Real-time inference of complex mental states from facial expressions and head gestures. *Real-Time Vision for Human-Computer Interaction*, pages 181–200, 2005.

[6] Picard R.W. *Affective Computing*. MIT Press, 1997.

[7] Baron-Cohen S., Golan O., Wheelwright S., and Hill J.J. Min-dreading: The interactive guide to emotions. *Jessica Kingsley Limited*, 2004.

[8] Park S. and Aggarwal J.K. Semantic-level understanding of human actions and interactions using event hierarchy. *IEEE Workshop on Articulated and Non Rigid Motion*, 2004.

[9] Sobol-Shikler T. and Robinson P. 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.

[10] Michel F. Valstar, Bihan Jiang, Marc Mehu, Maja Pantic, and Klaus Scherer. The first facial expression recognition and analysis challenge. *Automatic Face & Gesture Recognition and Workshops*, 2011.

[11] Ma Y., Paterson H.M., and Pollick F.E. A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods*, 38:134–141, 2006.

[12] Tong Y. A unified probabilistic framework for spontaneous facial action modeling and understanding. *Pattern Analysis and Machine Learning*, 2010.