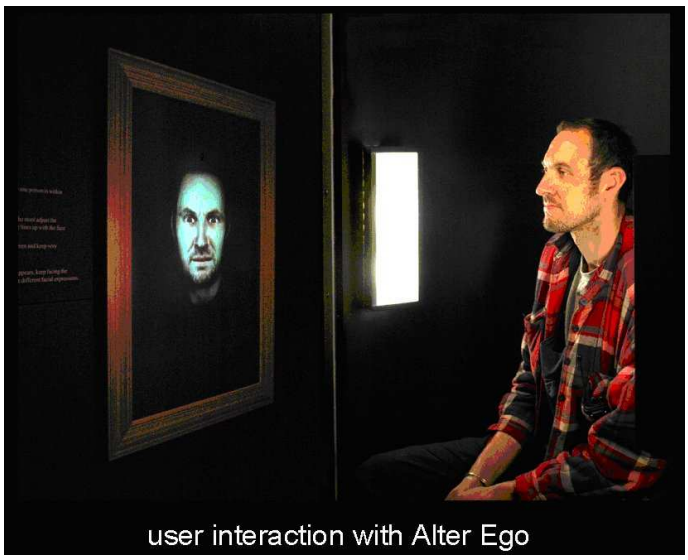
Alter Ego — live video classification of facial expression

Alf Linney*[1], Darren McDonald[1] & Alexa Wright[2]

[1] Ear Institute, University College London
[2] Ear Institute, University College London and CARTE University of Westminster

1 Introduction

The methodology and results described in this paper relate to research for the Alter Ego project (Wright *et al.*, 2005), an art/science collaboration that led to a publicly accessible interactive installation which has been shown in museums and galleries nationally. In essence, a 3D image of the autonomous "alter ego" of the user is created as a mirror reflection in real time. Initially, this "alter ego" behaves like a mirror, reflecting the subject's face on the screen and mimicking his or her expressions. After a short period, the image no longer behaves like a mirror, but responds to expressions made by the user in semi-deterministic way.



user interaction with Alter Ego

Because the context for this project is that of a publicly accessible artwork, the system must be robust and fully automatic. This poses some challenging scientific and technological problems, the most complex of these being to create the appearance of an automatic, real-time visual and emotional response from a computing machine. We divided this task into several distinct processes including the detection and tracking of facial landmarks at video rates; the analysis of these measurements in relation to particular facial expressions; the derivation of a decision tree able to reliably classify more than a dozen facial expressions; and the creation of a series of morph targets representing the end point of each of these expressions. In reality, humans may well interpret facial expression as a continuum, but we do commonly recognise end points or discrete states which we describe linguistically: for example, a smile, a frown and so forth.

One of the most basic tasks required from the machine in any pattern recognition or human computer interaction application is that of automatic classification. Over the past decade there has been a growing interest in the automatic detection of facial expression in both still images and live video. There are a variety of reasons for this. The significance of emotional cues in man/machine interaction; the availability of increasingly powerful computer resources; advances in face detection and tracking. As well as the more obvious applications for synthetic face animation in communications and the media, automatic facial expression analysis and classification in live images is also used in psychological studies (Ekman and Davidson, 1993), facial nerve studies in medicine (Dulguerov *et al.*, 1999) and lie detection (Ekman, 2001). Because of this, a number of research groups worldwide have attempted to automatically make sets of salient measurements which can be transformed by analysis into a single facial expression from both dynamic and static facial images, although this goal has not yet been fully achieved (Pantic and Rothkrantz, 2000; Fasel and Luettin, 2003).

2 Methods

In the Alter Ego installation images of the subject's face are captured continuously from a Webcam with an attached telephoto lens, allowing the face to occupy a large part of the frame. Using video in this way it is possible to make dynamic measurements on the moving face, and at the same time to look at "snapshots" represented by single frames. During research towards the project almost one hundred people were observed making either "enacted" or spontaneous facial expressions. Spontaneous expressions were gathered by videoing responses to carefully edited emotive video clips. Other people were asked to make a series of expressions in a specific order. A sample is presented below. This process allowed decisions to be made about what might be the most important features to measure. In the final analysis twenty-two facial features or "landmarks" were used. The positions of these features were tracked using software provided by a commercial company (Eyematic Interfaces Inc., Inglewood, California, USA).

When the expression database was created it was noted that the difference between a spontaneous expression and one "made to order" is easily distinguished by a human observer, but is extremely difficult to measure. Several research groups have attempted to create an automatic system for making this distinction. They are based, most notably, on a method proposed by Paul Ekman and Wallace Friesen for reducing facial expression to a series of specific facial movements related to particular muscle actions (Ekman and Friesen, 1978), which they called the Facial Action Coding System (FACS). This method enabled Ekman and Friesen to develop techniques for reading facial expression in terms of emotion and for deciding whether a subject's expression was a true reflection of emotion or a fake expression consciously made to deceive. Although this work was carried out more than quarter of a century ago, the earlier roots of this idea can be traced further back to Duchenne de Boulogne who, in the mid 1800's,

noted differences between real and fake smiles (Duchenne de Boulogne, 1862). More recently, a number of research teams have developed computer imaging and video systems that can, to a certain degree, detect the facial actions defined by FACS, and hence characterize facial expression. However these are either time-consuming or have been reported to achieve limited accuracy (Pantic M. and Rothkrantz, 2000). In order to derive a set of rules that would enable us to relate our measurements to actual facial expressions, we used the See5 data mining tool (RuleQuest Research Pty Ltd, NSW, Australia) which is the upgraded commercial version of the C4.5 algorithm and is based on the methods developed in the ID3 algorithm by Ross Quinlan (Quinlan, 1993). The method is one of learning by induction and its representation by decision or classification trees (Quinlan, 1979; 1986). We create a set of records. Each record has the same structure, consisting of a number of attributes consisting of facial measurements. Another attribute represents the category, in this case the facial expression of the record. The problem is to determine a decision tree that, on the basis of answers to queries about the non-category attributes, predicts correctly the value of the category attribute. Given a set of classified examples a decision tree is induced, biased by an information gain measure, which heuristically leads to small trees. In such algorithms, the information gain is measured by the change in entropy introduced by each new node in the decision tree. This tool, which heuristically searches for patterns in any given set of data, enabled us to establish a set of classifiers that are expressed as a binary decision tree. This tree is used to identify the most likely expression from a set of measured facial landmark positions.

3 Results

To illustrate the success rate of this method of classifying expressions, a confusion (or truth) matrix was constructed in which human assessment of facial images was used as the gold standard. As expected, the concordance between human and machine classification is not perfect, but the results obtained compared well with other reported methods and were generated fast enough for real time use. A confusion matrix of results from the decision tree is shown below.

Human classification:											
Angry frown	big smile	big surprise	disgust	laughing	lip pucker	neutral	nose wrinkle	sad	small surprise	small smile	tongue poke	Computer
13			2	2		5		7				angry f
	13		7	8			1			2		b smile
		11		2					11		3	b surprise
7	2		**5**	2	3	1	2	6			1	disgust
	14	1	1	**12**			2		1			laughing
		2	2		**9**	4	1	5		1	6	lip p
1	1					**23**	1	5		1		neutral
4			14	3			7	1			1	nose wr
3		1	5	2	2	6		**8**	2	1		sad
		7			2	5		2	**13**		1	s surprise
	9		4	3		5		1		**9**		s smile
		1		1	2						**26**	tongue p

To create both a living “mirror” and an apparently autonomous “alter ego” image, a series of generic three dimensional (3D) polygonal facial models representing the end-point of fifteen facial expressions were constructed. (Figure 5) These include small and broad smiles, surprise, anger, laughter, disgust, sadness and fear as well as more self-conscious expressions such as winking and poking out the tongue. The 3D models are warped to fit key landmark distances on a two dimensional video image of the individual face of each user. The 2D image of the user’s

face is then mapped onto the 3D model, which is animated by interpolation between models representing different expressions. Standard computer graphics techniques are used to render the images of the resulting 3D models simulating the individual adopting different expressions. The decision tree represents a procedure for using differences in facial measurements between the neutral face and an unknown expression to decide what that expression is most likely to be. At each stage in the tree a binary decision is made. The end point is the classification of the unknown expression. Typically a branching decision in the tree will be based on whether a particular measurement such as, for example, the change in width of the mouth between the unknown expression and neutral, is greater than or equal to a specific value. We have seen no other reports using this technique for facial expressions. After mimicking the facial expressions of the user for some seconds, in the installation the avatar then begins to react to these expressions. This is termed the "alter ego" phase. The expressions made by the "alter ego" are not random, but are generated by each classified subject expression being linked to a subset of response expressions. There are three randomly chosen possibilities for each response. This ensures that the response made to a subject's expression is reasonable, but not the same every time.

References

Duchenne de Boulogne, G.B. (1862). The Mechanism of Human Facial Expression. R. Andrew Cuthbertson, (ed., trans.). Cambridge, Cambridge University Press, 1990.

Dulguerov, P., Marchal, F., Wang, D., Gysin, C., Gidley, P., Gantz, B., Rubinstein, J., Sei, S., Poon, L., Lun, K., Ng, Y. (1999). Review of objective topographic facial nerve evaluation methods. *Am. J. Otol.,* **20(5)**, 672-678.

Ekman, P. and Friesen, W. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Paulo Alto, California, Consulting Psychologists Press.

Ekman, P. (2001). Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. New York, W.W. Norton.

Ekman, P. and Davidson, R.J. (1993). Voluntary smiling changes regional brain activity. *Psychological Science*, **4**, 342-345.

Fasela, B. and Juergen, L. (2003). Automatic facial expression analysis: a survey. *Pattern Recognition*, **36**, 259-275.

Pantic, M. and Rothkrantz, L.J.M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22(12)**, 1424-1445.

Quinlan, J.R. (1979). Discovering rules by induction from large collections of examples, in Expert Systems in the Microelectronic Age. Michie, D., (ed.), Edinburgh, Edinburgh University Press.

Quinlan J.R. (1986). Induction of decision trees. *Machine Learning*, **1**, 81Ű106.

Quinlan J.R. (1993). C4.5: Programs for Machine Learning. San Mateo, Morgan Kaufman.

Wright, A., Shinkle, E. and Linney, A. (2005). Alter Ego: Computer Reflections of Human Emotions. *Proceedings of the 6th. DAC Conference*, 191-199, Copenhagen, December. ISBN: 87-7949-107-3. Documented at: `www.alteregoinstallation.co.uk`