The FaceReader: Online facial expression recognition

M.J. den Uyl, H. van Kuilenburg
VicarVision, Sentient Systems B.V., Amsterdam

Abstract
This paper describes our FaceReader system, which we presented at Measuring Behavior 2005. This system is able to describe facial expressions and other facial features online with a remarkable accuracy. In this paper we will describe the possibilities of such a system and will discuss the technology used to make the system work. Currently, emotional expressions can be recognized with an accuracy of 89% and it can also classify a number of other facial features.

Keywords
Facial expressions, classification, model-based, Active Appearance Model, real-time.

1 Introduction

1.1 What a face can tell
Apart from being the means to identify other members of the species, the human face provides a number of signals essential for interpersonal communication in our social life. The face houses the speech production apparatus and is used to regulate the conversation by gazing or nodding, and to interpret what has been said by lip reading. It is our direct and naturally preeminent means of communicating and understanding somebody's affective state and intentions on the basis of the shown facial expression [4]. Personality, attractiveness, age and gender can also be seen from someone's face. Thus the face is a multi-signal sender/receiver capable of tremendous flexibility and specificity. In turn, automating the analysis of facial signals would be highly beneficial for fields as diverse as security, behavioral science, medicine, communication, education, and human-machine interaction.

1.2 How can FaceReading help?
In security contexts, apart from their relevance for person spotting and identification, facial signals play a crucial role in establishing or detracting from credibility. In medicine, facial signals are the direct means to identify when specific mental processes are occurring. In education, pupils' facial expressions inform the teacher of the need to adjust the instructional message. As far as natural interfaces between humans and machines (computers, robots, cars, etc.) are concerned, facial signals provide a way to communicate basic information about needs and demands to the machine. Where the user is looking (i.e., gaze tracking) can be effectively used to free computer users from the classic keyboard and mouse. Also, certain facial signals (e.g., a wink) can be associated with certain commands (e.g., a mouse click) offering an alternative to traditional keyboard and mouse commands. The human ability to read emotions from someone's facial expressions is the basis of facial affect processing that can lead to expanding interfaces with emotional communication and, in turn, to obtaining a more flexible, adaptable, and natural interaction between humans and machines.

Figure 1. *FaceReader demonstrator as shown at Measuring Behavior 2005.*

2 Face reading technology

The core problem of face analysis is how to simultaneously account for the three major source of variance in face images: pose/orientation, expression and lighting. To counter the problems caused by these sources of variation, the FaceReader classifies a face in three consecutive steps:

2.1 Face finding
Firstly, an accurate position of a face is found using a method called the Active Template Method (similar to the implementation described in [7]). The Active Template Method displaces a deformable face template over an image, returning the most likely face position or multiple positions if we allow more than one face to be analyzed.

2.2 Face modeling
Next, we use a model-based method called the Active Appearance Model (AAM) [2] to synthesize an artificial face model, which describes both the locations of key points, as well as the texture of the face in a very low dimensionality. The AAM uses a set of annotated images to calculate the main sources of variation found in face images and uses PCA compression to reduce the model dimensionality. New face models can then be described as deviations from the mean face, using a compact vector called the "appearance vector". As an example, Figure 2 shows the effects of varying the principle component of the appearance vector between -3 and +3 standard deviation. The AAM manages to compactly model individual facial variations in addition to variations related to pose/orientation, lighting and facial expression.

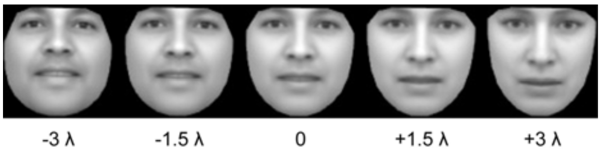

Figure 2. *Varying the first element of the appearance vector.*

Below an example is given of a synthesized face or AAM "fit" that can be automatically obtained from a face image. The AAM fit closely resembles the original face and only very little information is lost despite a very large reduction in dimensionality.

Figure 3. Example of a generated face model.

2.3 Face classification

The final stage in the FaceReader architecture is the actual classification of the expression or facial features we are interested in. This can be done in a very straightforward way by training an artificial neural network [1], which takes the AAM appearance vector as input. Given enough training data, we can train the network to learn to classify any facial feature as long as this feature is well modeled in the synthesized faces.

We have trained a network to classify the emotional expression shown on a face in one of the categories: happy, angry, sad, surprised, scared, disgust or neutral. These emotional categories are also known as the "basic emotions" or "universal emotions" [3]. As training material we have used the 'Karolinska Directed Emotional Faces' set [6] containing 980 high quality facial images.
Table 1 shows the performance results for this classifier. Horizontally, the actual expression shown on the presented images is shown and vertically you will see the emotion as predicted by the classifier. The total accuracy of the classification on the chosen set of emotional expressions is around 89% correct, which is among the highest performance rates reported.
Naturally, this method is not limited to the mentioned set of emotional expressions. To illustrate this, we have trained a classifier to detect 15 minimal facial actions, called "Action Units" described in the Facial Action Coding System [4]. This classifier had an average performance of 85% on the selected set of Action Units.
Besides expressions, we have also successfully trained classifiers on properties such as gender, ethnicity, age, facial hair (beard/moustache) or whether a person is wearing glasses or not. We are confident that any feature which a human observer can detect by observing a synthesized face can be learned by a classification network as well.

	happy	**angry**	**Sad**	**surprised**	**scared**	**disgust**	**neutral**	**recall**
happy	**138**	0	1	0	0	0	1	0.99
angry	1	**116**	2	1	3	11	0	0.87
sad	3	4	**109**	19	2	1	1	0.78
surprised	0	1	6	**128**	0	0	0	0.95
scared	0	8	5	2	**115**	5	3	0.83
disgust	1	5	3	0	3	**125**	0	0.91
neutral	0	11	2	1	1	0	**125**	0.89
precision	0.97	0.80	0.85	0.85	0.93	0.88	0.96	**0.89**

Table 1. *Confusion table for emotional expression classification.*

3 Conclusions and future work

We have managed to create a fully automatic facial classification system which is robust under varying conditions of pose, orientation and lighting using an implementation of the Active Appearance Model as core technology. Our FaceReader system can classify emotional expression with a very high accuracy and can be trained to classify almost any other facial feature.
Currently, we are working on further improving the accuracy of our system and extending the possibilities of classification, so that it will be possible in the near future to classify features which are located outside the modeled area of the face (for example the hair) or features which are poorly modeled by the AAM such as wrinkles, tattoos, piercings and birthmarks. We are also currently in the process of adding person identification to the system.

References:

1. Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.
2. Cootes, T.; Taylor, C. (2000). *Statistical models of appearance for computer vision.* Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering.
3. Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, **8**, 151–158.
4. Ekman, P.; Friesen, W.V.; Hager, J.C. (2002). *The Facial Action Coding System*. Weidenfeld & Nicolson, London.
5. Keltner, D.; Ekman, P. (2000). *Facial expression of emotion*. In: M. Lewis and J. Haviland-Jones (Eds.) Handbook of emotions, pp. 236-249, New York: Guilford Publications, Inc.
6. Lundqvist, D.; Flykt, A.; Öhman, A. (1998). The Karolinska Directed Emotional Faces – KDEF, Department of Clinical Neuroscience, Psychology section, Karolinska Institute.
7. Sung, K.K.; Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(1), 39–51.