Advances in face and gesture analysis

H. van Kuilenburg, M.J. den Uyl, M.L. Israël, and P. Ivan
VicarVision, Amsterdam, The Netherlands, info@vicarvision.nl

Recent technological developments are opening new possibilities for consumer behaviour analysis and consumer interaction. Faces are one of the primary channels for humans to transmit emotional signals, such as the level of appreciation or interest. These are most often uncontrolled spontaneous signals which are more honest and instantaneous and thus more useful for consumer behaviour research than self-reported values.

This article describes a number of new developments in face and gesture analysis which allow a more accurate, more extensive and more reliable analysis of faces, plus the ability to determine the point of interest and basic head gestures.

Facial expression analysis

In 2007 VicarVision and Noldus Information Technology launched FaceReader™, a system for fully automatic facial expression analysis. Although the FaceReader system opens many new possibilities for behavioural research, the scope of settings under which FaceReader can be used has its limitations. Future releases will include technological improvements that will further increase the range of use of the system.

What makes automatic facial expression analysis so difficult? The technical answer to this question is: the high dimensionality of the underlying mathematical problem space. Clearly, an image or video of a face is not suitable for facial analysis in its raw form (an array of pixels), so steps must be taken to reduce the raw data into a graspable set of features describing a face before it can be accurately analyzed. Sources of variation that need to be accounted for in order to reduce the dimensionality include: location of the face in the image, size and orientation of the face, lighting of the face, personal variations in for example gender / age / ethnicity and finally the facial expressions we are interested in.

Face detection is the first step in obtaining the important information from a face image or video frame. Several well known methods exist for this task, but each method has some tradeoffs in terms of speed, number of detections and/or framing accuracy. The latest face detection system that will be used in the next FaceReader release uses a unique combination of two face detection algorithms to find faces under a very wide range of variations while still creating a accurate face framing (and in a reasonable amount of time). The popular Viola-Jones algorithm [5] is used to roughly detect the presence of a face after which a deformable template method [4] creates a more accurate framing containing information about the likely in-plane rotation of a face.

The next phase in the FaceReader system is an accurate (photorealistic) modelling of the face using an algorithmic approach based on the Active Appearance method described by Cootes and Taylor [1]. A trained appearance model has limits on the amount of variation that it is able to model, which manifests itself for example as a lower modelling accuracy (or failure) for people of certain ethnicities and age and difficulties modelling faces under certain lighting and orientation. Currently the FaceReader system works best for Caucasian middle-aged people with little facial hair and a frontal camera position and lighting. Dedicated models are being developed for people from Asian origin and for children, which should greatly increase the worldwide usability of the FaceReader system.

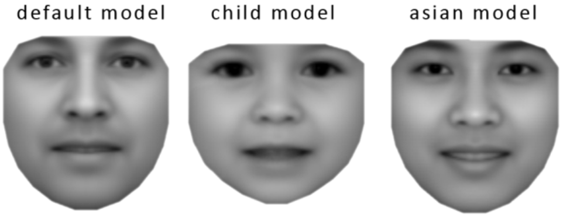
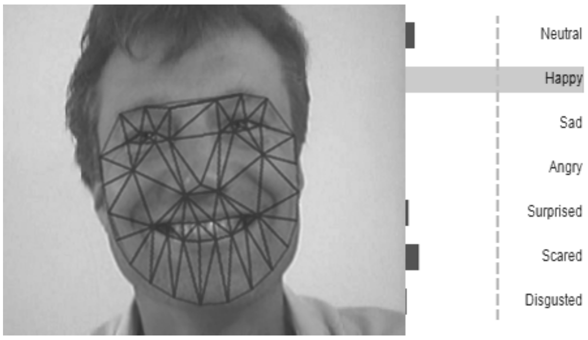


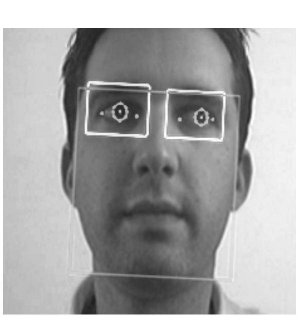
Figure 1. *Mean faces of FaceReader appearance models*

In addition to creating more of these dedicated models, a new architecture is being considered which should expand the range of acceptable faces even further. If instead of having a few models, which the user can choose from, we have a (large) number of partly overlapping models, we can create an automatic selection mechanism that would choose the most appropriate model for the current person/settings. The training examples used to create these models could be automatically selected by a mechanism designed to include a certain amount of variation in each model which is known to work well. Increased processing speed and parallel architectures in modern PCs are making it a feasible solution to run multiple models parallel in real-time.

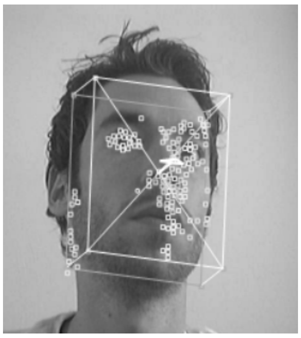


Gaze tracking

In addition to classifying emotional states, determining the object or source of the affect is often of equal importance for consumer behaviour research. In laboratory settings the number of stimuli can be limited, but 'in the wild' it is nearly impossible to determine the causes of all emotional expressions. For example consider a supermarket where a FaceReader system registers a "pleasantly surprised" emotion. Without knowing exactly which product or advertisement the person was looking at when the emotion was registered, this information is of little use.



VicarVision is developing an eye-tracking tool designed to be used with low cost off-the-shelf equipment for viewers that may move around freely over some distance. In contrast to special purpose eye-tracking systems, our system uses standard webcams without infrared capabilities and infrared lighting. Infrared lighting has clear advantages for eye-tracking as the pupil of the eye shows distinct and precise corneal reflections when lit with infrared light. However, standard webcams are easily available everywhere and do not require special lighting that may not work well over larger viewer distances. In a study we showed how the Active Appearance Model is a suitable candidate for creating accurate eye models containing both texture information and positions of key landmark points within the eyes such as the pupil centre and eye corners [2]. This information combined with head position and orientation information allows us to determine the gaze direction with reasonable accuracy.



Robust head orientation tracking

In uncontrolled settings, people tend to look around a lot and make fast head movements which cannot be handled by either the face detection or the face modelling algorithm. The result of this is that the face can only be tracked and analyzed a fraction of the time and all other behaviour of the subject is lost.

To overcome these problems, a module based on the “KLT point tracking algorithm” described by Shi and Tomasi [3], combined with a cylindrical head model is being developed. The point tracking algorithm finds features, within the bounding box of the face, which can be tracked both in the x- and y-direction, such as corner points and checkered textures. In an initial frame these key points are projected on one half of a cylinder (which forms a rough approximation of the shape of a face) with an initial orientation. In following frames the transformation of the cylinder that best fits the translation of the feature points in this frame is estimated. From this change in orientation of the cylinder (and thus change in orientation of the face) can be derived. This technique enables reliable orientation estimation for up to 90° away from the camera and a continuous head tracking in all orientations. Also, head gestures like nodding “yes” or shaking “no” can be recognized.

In addition to more robust face tracking, the face orientation information is also crucial for the gaze estimation module discussed before. In a project with a leading producer of skin-care products, this method also proved to be most valuable for analyzing consumers using products in front of a mirror while not being restricted in head orientation.

Conclusions

Several algorithmic improvements in the FaceReader and the development of new tools will lead to an increased robustness and usability of face analysis technologies . On the long term, significant advances are made to being able to analyze faces in completely uncontrolled settings and extract more information than the emotional expression only. All this is meant to open many new possibilities for consumer behaviour research.

References

1. T. Cootes and C. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, 2000.
2. P. Ivan. Active appearance models for gaze estimation. Master thesis, BWI, Free University Amsterdam, The Netherlands, 2007.
3. J. Shi and C. Tomasi. Good features to track. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'94), pages 593-600, IEEE Computer Society, Seattle, Washington, June 1994.
4. K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20(1)**, 39–51, 1998.
5. P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *In proceedings CVPR*, 2001.