Automatic Recognition of Eye Blinking in Spontaneously Occurring Behavior

Tsuyoshi Moriyama[1] Takeo Kanade[1] Jeffrey F. Cohn [1,2]
Jing Xiao[1] Zara Ambadar[2] Jiang Gao[1] Hiroki Imamura[1]

[1]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213
[2]Department of Psychology, University of Pittsburgh, Pittsburgh 15260

Abstract

Previous research in automatic facial expression recognition has been limited to recognition of gross expression categories (e.g., joy or anger) in posed facial behavior under well-controlled conditions (e.g., frontal pose and minimal out-of-plane head motion). We have developed a system that detects discrete and important facial actions, (e.g., eye blinking), in spontaneously occurring facial behavior with non-frontal pose, moderate out-of-plane head motion, and occlusion. The system recovers 3D motion parameters, stabilizes facial regions, extracts motion and appearance information, and recognizes discrete facial actions in spontaneous facial behavior. We tested the system in video data from a 2-person interview. Subjects were ethnically diverse, action units occurred during speech, and out-of-plane motion and occlusion from head motion and glasses were common. The video data were originally collected to answer substantive questions in psychology, and represent a substantial challenge to automated AU recognition. In analysis of 335 single and multiple blinks and non-blinks, the system achieved 98% accuracy.

1. Introduction

Within the past decade, there has been significant effort toward automatic recognition of human facial expression using computer vision. Several such systems [4, 8, 11) have recognized under controlled conditions a small set of gross emotion categories, such as joy and anger. Others have achieved some success in the more difficult task of recognizing FACS [3] action units [1, 7, 10]. Action units represent the smallest visible change in facial expression. A limitation of almost all research to date in automatic facial expression recognition is that it is limited to deliberate facial expression recorded under controlled conditions that omit significant head motion and other factors that complicate analysis.

Automatic recognition of facial action units in spontaneously occurring facial behavior presents several technical challenges. These include rigid head motion, non-frontal pose, occlusion from head motion, glasses, and gestures, talking, low intensity action units, and rapid facial motion [6]. This paper reports one of the first attempts to automatically recognize action units, in particular eye blinking, in spontaneous facial behavior during social interaction with non-frontal pose, moderate out-of-plane head motion, and moderate occlusion. The face analysis system recovers 3D motion parameters, stabilizes facial regions, extracts motion and appearance information, and recognizes action units in spontaneous facial behavior. Manual processing is limited to marking several feature points in the initial image of the stabilized image sequence. All other processing is automatic. In an initial test, reported below, the system recognized single and multiple blinks and non-blinks with 98% accuracy.

2. Database

We used video data from a study of deception by [5]. Subjects were 20 young adult men. Data from 10 were available for analysis. Seven of the 10 were Euro-American, 2 African-American, and 1 Asian. Two wore glasses. Subjects either lied or told the truth about whether they had stolen a large sum of money. Prior to stealing or not stealing the money, they were informed that they could earn as much as $50 if successful in perpetuating the deception and could anticipate relatively severe punishment if they failed. By providing strong rewards and punishments, the manipulation afforded ecological validity for deception and for truth-telling conditions.

Subjects were video recorded using a single S-Video camera. Head orientation to the camera was oblique and out-of-plane head motion was common. The tapes were digitized into 640x480 pixel arrays with 16-bit color resolution. A certified FACS coder at Rutgers University under the supervision of Dr. Frank manually FACS-coded start

1051-4651/02 $17.00 (c) 2002 IEEE

and stop times for all action units in 1 minute of facial behavior in the first 10 subjects. Certified FACS coders from the University of Pittsburgh confirmed all coding.

In this report we focus on automatic analysis of blinks (AU 45 in FACS). Measurement of blink is important in several fields, including neurology, physiology, and psychology. Blink rate varies with physiological and emotional arousal, cognitive effort, and deception [e.g., 2]. We included for analysis all instances of blink (AU 45) for which two independent teams of certified FACS coders agreed; 95% of blinks (AU 45) met this criterion and were included in the analyses. We also included an equal number of non-blink sequences of equal duration for comparison. The database contains a few instances of multiple blinks in a short period of time, or eyelid "flutter," defined as two or more rapidly repeating blinks (AU 45), which may be separated by AU 42.

3. Overview of Face Analysis System

Figure 1 depicts an overview of the face analysis system used for automatic recognition of blinks (FACS AU 45). A digitized image sequence is input to the system. The face region is delimited in the initial frame either manually or using a face detector [9]. The image in which the head is most upright is chosen as the reference image. Head motion (6 *DOF*) is recovered automatically. Using the recovered motion parameters, the face region is stabilized. Facial features are extracted in the image sequence, and action units are recognized.

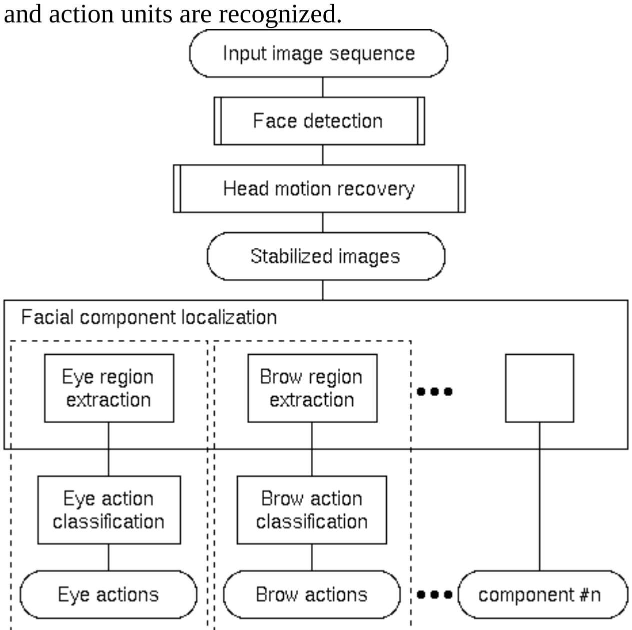


Figure 1. Overview, Face Analysis System

4. Automatic Recovery of 3D Head Motion and Stabilization of Eye Region

The 3D model that our head motion recovery uses is a cylinder. In the absence of knowing the physical size of the face or the distance between face and camera, the head model and its initial location will be up to a scale, but it does not matter for our purpose of cancelling the effect of pose. A cylindrical model is fit to the initial face region, and its image is cropped and "painted" onto the cylinder as the appearance template. Experimental tests suggest that the system is insensitive to small variations in the initial fit of the head model.

The head motion is tracked in terms of the change from the first frame. While tracking, the templates change dynamically. Once head pose is estimated in a new frame, the region facing the camera is extracted as the new template. Robust statistics are applied to remove outlier pixels, such as a hand or object placed in front of the face, from being included in the templates. This procedure contributes to system robustness to occlusion and non-rigid motion.

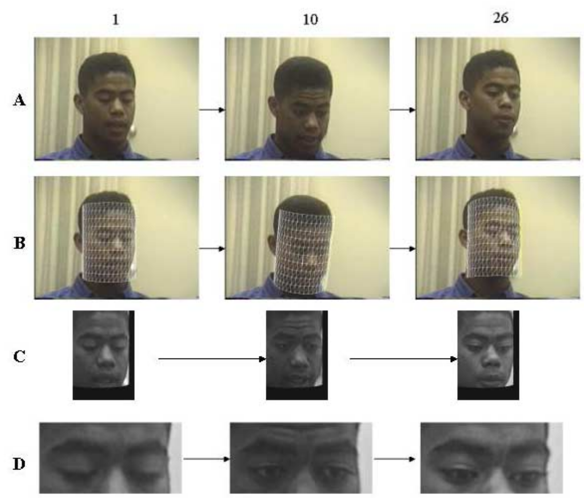


Figure 2. Automatic recovery of 3D head motion and image stabilization. A) Frames 1, 10, and 26 from original image sequence. B) Face tracking in corresponding frames. C) Stabilized face image. D) Localized face region.

Because head poses are recovered using dynamic templates and the pose estimated for the current frame is used in estimating the pose in the next frame, errors would accumulate unless otherwise prevented. To solve this problem, the first frame and the initial head pose are stored as a reference. When the estimated pose for the new frame is close to the initial one, the system rectifies the current pose estimate by registering this frame with a reference one. The re-registration prevents errors from accumulating and enables the system to recover head pose following occlusion, such as when the head moves momentarily out of the camera's view. By re-registering the face image, the system can run indefinitely.

The system has been tested successfully in image sequences that include maximum pitch and yaw as large as 50° and 90°, respectively, and time duration of up to 20 minutes [12]. The precision of recovered motion was evaluated with ground truth obtained by a precise position and orientation measurement device with markers attached to the head, and found to be highly consistent (e.g., for 75° yaw, absolute error averaged 3.86°). For details, see [12]. While a head shape is not actually a cylinder, a cylinder model is adequate and indeed contributes to system stability and robustness.

An example of system output can be seen in Figure 2A and 2B. Once the head pose is recovered, we can stabilize the face region by transforming the image to its cylindrical canonical view through the recovered position and orientation of the cylindrical head model. (Figure 2C)

5. Eye action classification

The eye region consists of the iris, sclera, upper and lower eyelids and the eyelashes (Figure 3). If we divide an eye region into upper and lower portions, the intensity distribution of the upper and lower portion would change as the eyelid closes and opens during blinking.

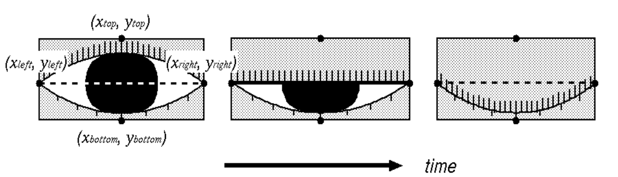
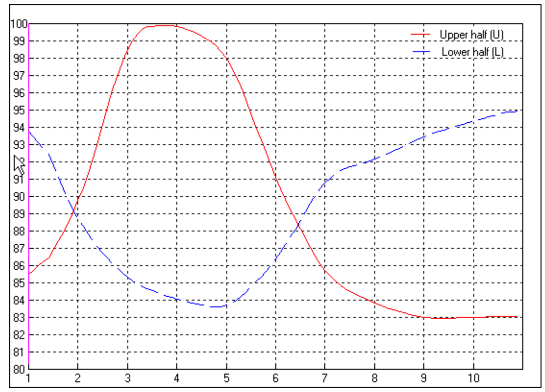
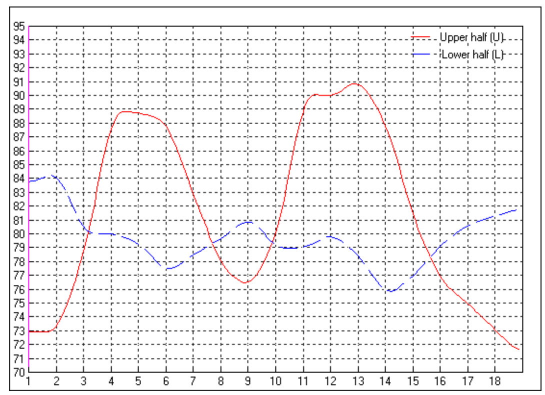


Figure 3. 2D eye model.

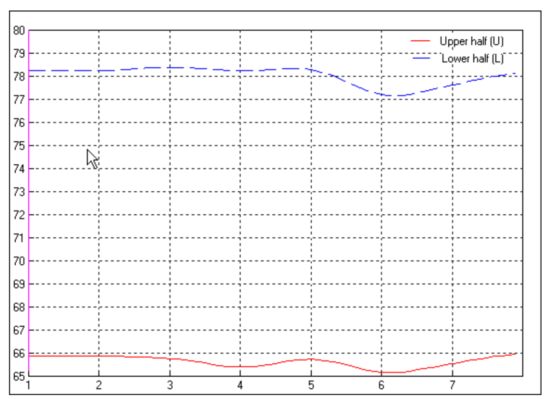
The input face image sequence (Figure 2A) has been automatically processed to obtain the stabilized image sequence (Figure 2C as described above). By manually giving the feature points $\{x_i, y_i; i = left, top, right, bottom\}$ in the first frame of the stabilized image sequence, the eye regions for the rest of the sequence $I(x,y) = \{I(x,y) \mid x_{left} \le x \le x_{right}, y_{top} \le y \le y_{bottom}\}$ are obtained (Figure 2D).



(a) Blink



(b) Flutter



(c) Non-blink

Figure 4. Examples of luminance curves for blink, multiple blink, and non-blink

For now we treat only the right eye (image left). The classification categories of eye actions are blink, multiple blink (eyelid 'flutter'), and non-blink. For this classification, the average illumination intensity is calculated for the upper and for the lower half of the eye region. When mean intensities for the upper and lower halves are plotted over time (e.g., the curves in Figure 4), we notice that they

Table 1. Comparison of Manual FACS Coding and Automatic Recognition.				
		Automatic Recognition		
		Blink AU45	Flutter	Non-Blink
Manual FACS Coding	Blink (AU45)	153	0	0
	Flutter	6	8	0
	Non-Blink	0	0	168

Note. Multiple blinks are 3 triple blinks and 2 doubles, separated by 1-2 frames of AU 42 or open eye. Overall agreement = 98% (kappa = .97). Combining blink and multiple blink (flutter), agreement = 100%.

cross when an eye closes and opens. When the eye is open, mean intensity in the upper half is smaller than that in that in the lower half, and reverses when closed. Therefore by counting the number of crossings we can detect the timing, number, and duration of eye blinking.

6. Recognition Results for Blink

Table 1 shows the recognition results and comparison with the manual coding. The algorithm achieved an overall accuracy of 98%. If we combine blink and flutter into a single category (which is common practice among FACS coders), then classification accuracy of eye closure and opening was 100% Six of 14 multiple blinks were incorrectly recognized as single blinks. Rapid transitions from AU 45 to AU 42 to AU 45, in which eye closure remains nearly complete, were occasionally recognized as a single blink. The measure we used (crossing of average intensities) was not consistently sensitive to the slight change between complete closure (AU45) and partial closure (AU 42)

7. Summary

We have demonstrated automated recognition of eye blinking, one of the psychologically important, discrete, and common facial actions, in video from Frank's and Ekman's deception database, which was collected in the real world environment to answer a substantive question in psychology, rather than video taken for the sake of testing vision algorithms. Ethnic background of the psychology subjects was varied, several wore glasses, which occluded the brows, orientation to the camera was typically non-frontal, behavior was spontaneous, and out-of-plane motion was common. We found that reliable and precise compensation of head motion is critical, reflection from the eyeglasses poses serious difficulty, and appropriate discriminating measures need to be developed for discrete action unit detection and recognition.

8. References

[1] Bartlett, M.S. et al. (1999). Measuring facial expressions by computer image analysis. *Psychophysiology, 36*, 253-263.

[2] Ekman, P. (2001). *Telling lies.* NY: Norton.

[3] Ekman, P. & Friesen, W. (1978). *Facial Action Coding System.* Consulting Psychologists Press, Palo Alto, CA.

[4] Essa. I., and Pentland, A. (1997). "Coding, analysis, interpretation and recognition of facial expressions." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 7, 757-763.

[5] Frank, M. & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies**.** *Journal of Personality & Social Psychology, 72,* 1429-1439.

[6] Kanade, T., Cohn, J.F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, 4,* 46-53.

[7] Lien, J.J.J. et al. (2000). Detection, tracking, and classification of subtle changes in facial expression. *Journal of Robotics & Autonomous Systems***,** *31*, 131-146.

[8] Padgett, C, et al. (1996). Categorical perception in facial emotion classification. *Cognitive Science*.

[9] Rowley, H.A., et al. (1998). Neural network-based face detection. . *IEEE Transactions on Pattern Recognition and Machine Intelligence, 20,* 23-38.

[10] Tian, Y.L., et al., J.F. (2001). Recognizing action units for facial expression analysis*. IEEE Transactions on Pattern Recognition and Machine Intelligence, 23,* 97-116.

[11] Yacoob, Y. & Davis, L. (1997). Recognizing human facial expression from long image sequence using optical flow. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 18,* 636-642.

[12] Xiao, J., Kanade, T., & Cohn, J.F. (May 2002). Robust full motion recovery of head by dynamic templates and re-registration techniques. *Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, 5, xxx-xxx.* Washington, DC.