Automatic Recognition of Eye Blinking in Spontaneously Occurring Behavior

Jeffrey F. Cohn,[1] Jing Xiao, Tsuyoshi Moriyama, [2] Zara Ambada[1], & Takeo Kanade [2]

University of Pittsburgh[1]

Carnegie Mellon University[2]

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 a discrete and important facial action (e.g., eye blinking) in spontaneously occurring facial behavior that has been measured 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. The 10 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 blinks, the system achieved 98% accuracy.

Introduction

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. Facial displays indicate emotion (Ekman, 1993) and pain (Craig, Hyde, & Patrick, 1991), regulate social behavior (Cohn & Elmore, 1988; Eibl-Eibesfeldt, 1989; Fridlund, 1994), reveal brain function (Ekman, Davidson, & Friesen, 1990; Fox & Davidson, 1988) and pathology (Katsikitis & Pilowsky, 1988; Rinn, 1984), and signal developmental transitions in infants (Campos, Bertenthal, & Kermoian, 1992; Emde, Gaensbauer, & Harmon, 1976). To make use of the information afforded by facial expression, reliable, valid, and efficient methods of measurement are critical.

Human-observer-based (i.e., manual) methods of coding facial expression are labor intensive, semi-quantitative, and difficult to standardize across laboratories or over time. Training is time consuming (approximately 100 hours with the most descriptive methods), and coding criteria may drift with time (Bakeman & Gottman, 1986; Martin & Bateson, 1986). Implementing comprehensive systems is reported to take up to 10 hours of coding time per minute of behavior depending upon the comprehensiveness of the system and the density of behavior changes (Ekman, 1982). Such extensive effort discourages standardized measurement and may encourage the use of less specific coding systems with unknown convergent validity (Matias, Cohn, & Ross, 1989). These problems tend to promote the use of smaller sample sizes (of subjects and behavior samples), prolong study completion times, and thus limit the generalizability of study findings.

Within the past decade, there has been significant effort toward automatic recognition of human facial expression using computer vision. Several such systems (Essa & Pentland, 1997; Padgett, Cottrell, & Adolphs, 1996; Yacoob & Davis, 1997) have recognized under controlled conditions a small set of emotion-specified expressions, such as joy and anger. Others (Bartlett, Hager, Ekman, & Sejnowski, 1999; Lien, Kanade, Cohn, & Li, 2000; Tian, Kanade, & Cohn, 2000, 2001, 2002) have achieved some success in the more difficult task of recognizing facial action units of the Facial Action Coding System (FACS: Ekman & Friesen, 1978). Actions units (AU) are the smallest visibly discriminable changes in facial expression. In a series of studies (Cohn et al., 1999; Lien et al., 2000; Tian et al., 2001, 2002), our group has developed a system that recognizes 20 of approximately 30 action units that have a known anatomic basis and occur most frequently in emotion expression and paralinguistic communication (Kanade, Cohn, & Tian, 2000; Sayette, Cohn, Wertz, & Perrott, 2001). Action units are recognized whether they occur alone or in combinations. The ability to recognize action units whether they occur alone or in combinations is an important feature because action units may occur in thousands of combinations (Ekman, 1982), and the appearance and timing of action units can vary depending on the action units with which they co-occur (analogous to co-articulation effects in speech).

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 multiple challenges. These include camera orientation, head motion, and occlusion, differences between deliberate and spontaneous expression, and individual differences among subjects. We first review each of these factors. We then report one of the first attempts to

automatically recognize action units in spontaneous facial behavior during social interaction with non-frontal pose, moderate out-of-plane head motion, and moderate occlusion.

Camera orientation and rigid head motion

Most research in facial expression recognition has been limited to image sequences in which the face is oriented toward the camera. While frontal pose is a reasonable assumption when subjects are asked to model facial expressions or talk in front of a camera (as in 'talking head' or TV news applications), in most social interaction research cameras are oriented about 15-30 degrees to the side in order not to interfere with face-to-face interaction. With variation in camera orientation, face appearance changes qualitatively. As one example, the relative symmetry of facial features is lost (Liu, Schmidt, Cohn, & Weaver, 2002) and facial features become occluded when camera orientation is no longer frontal. Systems that have been trained on relatively symmetric frontal views may generalize poorly to face images acquired from oblique camera angles.

A related problem is head motion. Moderate to large head motion is common in naturally occurring behavior and often accompanies change in expression. Kraut and Johnson (1979) found that smiling typically occurs while turning toward another person. Camras, Lambrecht, and Michel, (1996) found that infant surprise expressions often occur as the infant pitches her head back. To control for head motion, previous research has used head mounted cameras (Pantic, 2002) or selected image sequences in which head motion is absent (Bartlett et al., 1999; Donato, Bartlett, Hager, Ekman, & Sejnowski, 1999). When head motion is allowed, it typically is limited to motion that is parallel to the image plane of the camera (i.e., planar head motion) (Tian et al., 2001; Yacoob & Davis, 1997). For head motion that is planar or nearly so, an affine or perspective transformation of images (Lien, Kanade, Cohn, & Li, 2000) can align images so that face position, size, and orientation are kept relatively constant across subjects, and these factors do not interfere significantly with feature extraction. The problem occurs with out-of-plane motion of more than about 5 degrees, which is typical in spontaneous facial behavior. Out-of-plane head motion is a particular challenge for computer vision systems because the face looks different when the head pitches up or down or turns from side to side (i.e., yaw), which occludes facial features. A computer vision system must be robust to variation in camera orientation and out-of-plane head motion.

Deliberate versus spontaneous facial expression

Most facial expression data have been collected by asking subjects to perform a series of expressions. These directed facial action tasks typically are more intense and include more appearance changes than those that occur spontaneously (Zlochower, 2001). They also are likely to differ in appearance and timing from spontaneously occurring behavior (Cohn & Schmidt, In press). Deliberate and spontaneous facial behavior are mediated by separate motor pathways, the pyramidal and extra-pyramidal motor tracks, respectively (Rinn, 1984). As a consequence, fine-motor control of deliberate facial actions often is inferior to and less symmetric than that which occurs spontaneously. Many people, for instance, are able to raise their outer brows spontaneously while leaving their inner brows at rest; few can perform this action voluntarily. Spontaneous depression of the lip corners (AU 15) and raising and narrowing the inner corners of the brow (AU 1+4) are common signs of sadness. Without training, few people can perform these actions deliberately, which incidentally is an aid in lie detection (Frank & Ekman, 1997). Spontaneous smiles show a highly consistent relation between duration and amplitude that is characteristic of ballistic motion; for deliberate smiles, these parameters are uncorrelated, to which observers may be particularly sensitive (Cohn & Schmidt, In press). Differences in the temporal organization of spontaneous and deliberate facial actions are particularly important to automatic facial expression analysis in that many pattern recognition approaches, such as Hidden Markov Modeling, are highly dependent on the timing of appearance change.

Individual differences among subjects

Face shape, texture, color, and facial and scalp hair vary with sex, ethnic background, and age (Bruce & Young, 1998). Infants, for instance, have smoother, less textured skin and often lack facial hair in the brows or scalp. The eye opening and contrast between iris and sclera differ markedly between Asians and Northern Europeans, which may affect the robustness of eye tracking and facial feature analysis more generally. Beards, eyeglasses, or jewelry may obscure facial features. Such individual differences in appearance can have important consequence for face analysis. Few attempts to study their influence exist. In our experience (Cohn, Tian, & Forbes, 2000), algorithms for optical flow and high-gradient component detection that have been optimized for young adults perform less well when used in infants. The reduced texture of infants' skin, their increased fatty tissue, juvenile facial conformation, and lack of transient furrows may all have contributed to the differences observed in face analysis between infants and adults.

In addition to individual differences in appearance, there are individual differences in expressiveness, which refers to the degree of facial plasticity, relative intensity of facial expression, bias for particular facial expressions, and overall rate of expression. Individual differences in these characteristics are well established and are important aspects of individual identity (Cohn, Schmidt, Gross, & Ekman, 2002; Moore, Cohn, & Campbell, 1997; Schmidt & Cohn, 2001). An extreme example of variability in expressiveness occurs in individuals who have incurred damage to the facial nerve or central nervous system (Van Swearingen, Cohn, & Bajaj-Luthra, 1999). To develop algorithms that are robust to individual differences in facial features and expressiveness, it is essential to include subjects of varying ethnic background, age, and sex, and to allow for facial hair, jewelry or eyeglasses that may obscure or hide facial features.

We report 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 in diverse subjects. The image data were collected in order to investigate a socially and psychologically important topic, deception, rather than to test facial expression algorithms. We focus here on eye blinking. Measurement of blinking is important in several fields, including neurology, physiology, and psychology. Blink rate varies with physiological and emotional arousal, cognitive effort, and deception (Holland. & Tarlow, 1972; Ekman, 2001; Karson, 1988).

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 blinks with 98% accuracy.

Method

Database

We used video data from a study of deception by Frank and Ekman (1997). 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 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). We included for analysis all instances of blink (AU 45) for which two independent teams of certified FACS coders agreed; 95% of those examined (167 blinks) met this criterion and were included in the analyses. The average number of blinks per subject was 16.70 with standard deviation of 10.86. We also include the few instances of eyelid "flutter" that occurred, defined as two or more rapidly repeating blinks (AU 45) with only partial eye opening (AU 42 or AU 41) between them. In flutter, the modal interval between blinks was 1 frame, with a maximum of 2 frames. These instances of flutter occurred in 3 subjects, with 1 subject accounting for 12 of 14 that occurred. For each subject, we included an equal number of non-blink intervals of equal duration for comparison.

Overview of face analysis system

Figure 1 depicts an overview of the face analysis system (Automated Face Analysis version 3) used for automatic recognition of blinks (FACS AU 45) and flutter. A digitized image sequence is input to the system. The face region is delineated in the initial frame either manually or by using a face detector (Rowley, Baluja, & Kanade, 1998). The image in which the head is most upright is chosen as the reference image. Head motion (6 *df*) is recovered automatically. Using the recovered motion parameters, the face region is stabilized; that is, warped to a view of common orientation. Facial features are extracted in the image sequence and action units are recognized. The current study is limited to analysis of the eye region. The modules for feature extraction and analysis of other face regions (e.g., forehead and brows) are not considered.

Insert Figure 1 About Here

Automatic recovery of 3d head motion and stabilization of eye region

To estimate 3d head motion, one option is to use an anatomically based face model in which the exact proportions of facial features are represented (De Carlo & Mataxas, 1996; Essa & Pentland, 1997). To work well, such anatomic models require fitting a large number of parameters that are dependent on the exact shape of the individual face, which typically is unknown. These parameters must be precisely estimated throughout an image sequence. When precise initialization and tracking are not available, better recovery may be achieved by using a simpler 3d geometric model. In our approach, we use a cylindrical head model to estimate the 6 degrees of freedom (*df*) of head motion, whose parameters are horizontal and vertical position, distance to the camera (i.e. scale), pitch, yaw, and roll.

A cylindrical model is fit to the initial face region, and the face image is cropped and "painted" onto the cylinder as the initial appearance template. For any given frame, the template is the head image in the previous frame that is projected onto the cylindrical model. To estimate head motion, we first warp the head image to the cylindrical model with pose that of the template; that is, we assume that pose has remained unchanged from the template to the current frame. We then compute the difference between the warped image and the template to provide the initial estimate of pose. To obtain the final estimate of pose, we iterate this process to further refine the estimate by using a model-based optical flow algorithm (Lucas & Kanade, 1981; Xiao, Kanade, & Cohn, 2002). This algorithm is implemented with an iterative hierarchical image pyramid (Poelman, 1995) so that rapid and large head motion may be tracked. While the algorithm assumes that lighting remains unchanged from one frame to the next, we have found that the algorithm is relatively robust to the typical variations in lighting found in indoor settings.

Major potential sources of error in estimating head motion include non-rigid motion (e.g., facial expression) and occlusion (e.g., a hand moving in front of the face). When they occur, some pixels in the template may change in the processed image or even disappear. We do not want these pixels to contribute to motion estimation. When estimating optical flow from one frame to the next, these pixels will have relatively large error. To minimize their influence, pixels with large error are given less weight in estimating motion between frames so that they contribute less than other pixels. The specific technique that we use is iteratively re-weighted least squares (Black, 1992). In this way, the potential influence of non-rigid motion and occlusion are minimized in estimating rigid head motion.

Head templates change while tracking. Once head pose is estimated in a new frame, the region facing the camera is extracted as the new template. Because head poses are recovered using templates that are constantly updated 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 when the face reappears after 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 was tested successfully in image sequences that include maximum pitch and yaw as large as 40° and 75°, respectively, and time duration of up to 20 minutes (Xiao, Kanade, & Cohn, 2002). The precision of recovered motion was evaluated with respect to the 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^{o}). For details, see Xiao, Kanade, & Cohn (2002). While a head shape is not actually a cylinder, a cylinder model is found to be adequate and indeed contributes to system stability and robustness.

Insert Figure 2 About Here

An example of system output can be seen in Figure 2. From the input image sequence (Figure 2A), the head is tracked and its pose recovered as described above (Figure 2B). Once the head pose is recovered, we can stabilize the face region by transforming the image to a common orientation (Figure 2C) and then localize a region of interest, which in the current study is the eye region (Figure 2D), as described below.

Eye action classification

The eye region consists of the iris, sclera, upper and lower eyelids, and the eyelashes. If we divide the eye region into upper and lower portions (Figure 3), the grayscale intensity distribution of the upper and lower portion would change as the eyelid closes and opens during blinking. Grayscale intensity can range from 0 (black) to 255 (white). When the eye is open, the relatively light grayscale intensity of sclera is averaged with the darker grayscale values of the upper eyelashes, pupil, and iris. As the eyelid closes, grayscale intensity in the upper eye region increases as the eyelashes move into the lower region and the pupil and

iris become occluded by the relatively brighter (i.e., higher grayscale intensity) skin of the now exposed upper-eyelid. In pilot testing and in the present study, we found that even in dark-skinned subjects, grayscale intensity in the upper eye region is greater when the eyes are closed rather than open. We can use this knowledge to automatically track closing and opening of the eye and recognize blinks from non-blinks once the eye region is stabilized with respect to rigid motion.

Insert Figure 3 About Here

Automatic feature extraction in the eye region

The input face image sequence (Figure 2A) has been automatically processed to obtain the stabilized image sequence (Figure 2C as described above). We then define the eye region by manually marking four feature points in the first frame of the stabilized image sequence. These feature points in the first frame are the inner and outer eye corners, the center of the upper eyelid, and just below the center of the lower eyelid. The pixel coordinates of these feature points in the first frame then are used to define the upper and lower eye regions in each frame of the stabilized image sequence. Figure 3 shows an example from frames 1, 5, 6, 7, and 9 of a stabilized image sequence in which the eye changes from open to closed. Note that as the eye closes, the grayscale intensity of the upper eye region becomes lighter as the upper eyelid progressively covers the iris and pupil and the eyelashes on the upper eyelid move into the lower eye region.

Insert Figure 4 About Here

For now we treat only the right eye (image left). The classification categories of eye actions are blink, flutter, and non-blink. For this classification, the average grayscale intensity is calculated for the upper and for the lower eye regions. When the eye is open, mean grayscale intensity in the upper half is smaller than that in the lower half and reverses when closed. When mean intensities for the upper and lower regions are plotted over time, they cross when an eye closes and opens. The intensity curve for the upper eye peaks when the eye is completely closed (AU 45). By counting the number of crossings (Nc) and the number of peaks (Np) in the grayscale intensity curves, we can detect the timing, number, and duration of eye blinking:

$Np = 0$ → Non-blink
$Np = 1$ and $Nc \neq 0$ → Blink
$Np \geq 2$ and $Nc \neq 0$ → Flutter

Where Np = Number of peaks and Nc = Number of crossings in the grayscale intensity curves for the upper and lower eye regions (Figure 4).

Insert Figure 4 About Here

Figure 5 shows examples of the grayscale intensity curves for blink (Figure 5A), non-blink (Figure 5B), and flutter (Figure 5C).

Insert Figure 5 About Here

Results

The accuracy of blink detection was assessed in two ways. First, we compared the sensitivity of the algorithm to quantitative changes in eye closure to that of human judges. Second, we compared the accuracy of automatic blink detection with that of human FACS coders.

To evaluate the sensitivity of the blink detection algorithm to variation in eye opening and closing, we compared automatic and manual (human) processing. Stabilized eye image sequences in which AU 45 (blink) occurred were analyzed using the algorithm described above. Grayscale intensity of the upper eye region as determined automatically was normalized over the range of 0 to 1. Normalized intensity of 1.00 corresponded to the open eye, and 0.00 corresponded to the closed eye. The digitized images then were randomly sorted. Two researchers, blind to the results of automatic processing, then manually sorted each sequence from eye open to closed to open. They next estimated the degree of eye opening on a scale from 0 (eye closed) to 1 (eye open). In each of 10 sequences examined, the manual ratings of eye open by each of the two researchers were highly consistent with automatic measurement of grayscale intensity. Figure 6 shows graphical results for one of the 10 sequences examined. The manual ratings by each of the raters were highly consistent with each other and with the results of automatic processing of grayscale intensity. In the example shown, the consistency between Rater 1 and automatic processing (r = .95) was comparable to the consistency found in manual ratings between Rater 1 and Rater 2 (r = .96). These findings were typical of the close correspondence between manual ratings of eye state and automatic processing.

Insert Figure 6 About Here

Table 1 shows recognition results for blink detection in all image data in comparison with the manual FACS coding. The algorithm achieved an overall accuracy of 98%. Six of 14 instances of flutter 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. Transitions from AU 45 to AU 41 (drooping upper eyelid) to AU 45 were more frequently detected. The measure we used (crossing of average intensities) was not consistently sensitive to the slight change between complete closure (AU45) and closure that was nearly complete (AU 42). If blink and flutter are combined into a single category (which is common practice among FACS coders), classification accuracy of eye closure and opening was 100%.

Insert Table 1 About Here

Discussion

This study is one of the first to attempt automatic action unit recognition in naturally occurring facial behavior. All other work in automated facial expression recognition has been limited to analysis of deliberate facial expressions that have been collected under controlled conditions for purposes of algorithm development and testing. We analyzed image data from Frank and Ekman (1997) who collected them under naturalistic conditions in the course of psychological research on deception and not with the intention of automated facial expression analysis. We analyzed spontaneously occurring behavior rather than posed expressions. The data presented significant challenges in terms of heterogeneity of subjects, brightness, occlusion, pose, out-of-plane head motion, and the low intensity of action units.

To meet these challenges, we developed Automated Face Analysis version 3 that automatically estimates 3D motion parameters, stabilizes face images for analysis, and recognizes facial actions using a face-component based approach to feature extraction. We emphasized the aspect of automated analysis of feature extraction, localization and tracking; manual processing was limited to feature marking in a single initial frame, and all other processing was fully automatic. This focus contrasts with recent efforts by Bartlett, Braathen, Littlewort, Sejnowski, & Movellan (2001) to analyze spontaneous facial behavior that requires manual labeling and registration of each image.

Automated Face Analysis v.3 successfully recognized blinks from non-blinks for all the examples in the database. It also was able to distinguish, with lower accuracy flutter. We found that automated analysis of facial images still

presents significant challenges. Many previously published algorithms, including our own, that worked well for frontal faces and good lighting conditions fail with images under non-frontal facial poses, full 6 *df* head motions, and ordinary lighting. Precise and reliable extraction and localization of features is the key to the success of automated FACS coding, facial expression and emotion analysis. The 3D-model based stabilization technique presented here for stabilizing the arbitrary and unknown head motion is one such example. In the next phase of our research, we expand the size and diversity of our database of FACS-coded spontaneous facial behavior and increase the number and complexity of action units that can be recognized automatically in this context. We also include explicit analysis of the timing of facial actions in relation to context and communicative intention (Schmidt & Cohn, Submitted; Cohn & Schmidt, In press). While challenges remain, these findings support the feasibility of developing and implementing comprehensive, automated facial expression analysis in research on emotion and non-verbal communication.

References

Bakeman, R. & Gottman, J.M. (1986). *Observing behavior: An introduction to sequential analysis*. Cambridge: Cambridge University Press.

Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J. (1999). Measuring facial expressions by computer image analysis. P*sychophysiology, 36*, 253-263.

Bartlett, M.S., Braathen, B., Littlewort, G., Sejnowski, T.J., & Movellan, J.R. (2001). *A comparative study of alternative FACS coding algorithms* (Tech Rep. MPLAB-TR-2001-06*)*. . San Diego: Institute for Neural Computation, University of California San Diego.

Black, M. (1992). *Robust incremental optical flow*. Unpublished doctoral dissertation, Yale University.

Bruce, V. & Young, A. (1998*). In the eye of the beholder: The science of face perception*. Oxford University Press.

Campos et al., (1992). Early experience and emotional development: The emergence of wariness of heights. *Psychological Science, 3*, 61-64.

Camras, L.A., Lambrecht, L., & Michel, G.F. (1996). Infant "surprise" expressions as coordinative motor structures. *Infant Behavior and Development, 20*, 183-195.

Cohn, J.F. & Elmore, M. (1988). Effect of contingent changes in mothers' affective expression on the organization of behavior in 3-month-old infants. *Infant Behavior and Development, 11*, 493-505.

Cohn, J.F. & Schmidt, K. (In press). The Timing of Facial Motion in Posed and Spontaneous Smiles. *Proceedings of the International Conference on Active Media Technology, 2*, xxx-xxx. Chongqing, China.

Cohn, J.F., Tian, Y.L., & Forbes, E. (July 2000). Detection, tracking, and classification of facial action units in infant and caregiver facial expression [Abstract]. *Proceedings of the International Conference on Infant Studies,* Brighton, England.

Cohn, J.F., Zlochower, A., Lien, J., & Kanade, T. (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology, 36*, 35-43.

Cohn, J.F., Schmidt, K., Gross, R., & Ekman, P. (October 2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. *Proceedings of the International Conference on Multimodal User Interfaces, 4*, 491-496.

Craig, K. D., Hyde, S. A., & Patrick, C. J. (1991). Genuine, suppressed and faked facial behavior during exacerbation of chronic low back pain. *Pain, 46*, 161-171.

De Carlo, D. & Metaxas, D. (1996). The integration of optical flow and deformable models with applications to human face shape and motion estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 15*, 231-238.

Donato, G.L., Bartlett, M.S., Hager, J.C., Ekman, P., & Sejnowski, T.J. (1999). Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 974-989.

Eibl-Eibesfeldt, I. (1989). *Human ethology*. Hawthorne, NY: Aldine de Gruyter.

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

Ekman, P. (1993). Facial expression and emotion, *American Psychologist 48*, 384-392.

Ekman, P. (1982). Methods for measuring facial action. In K.R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 45-90). Cambridge: Cambridge University Press.

Ekman, P., Davidson, R.J., & Friesen, W.V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology, 58*, 342-353.

Ekman, P. & Friesen, W. (1978). *Facial action coding system.* Palo Alto, CA: Consulting Psychologists Press.

Emde, R.N., Gaensbauer, T.J. & Harmon, R.J. (1976). Emotional expression in infancy: A biobehavioral study. *Psychological Issues, 10*, NY: International Universities.

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

Fox, N. & Davidson, R.J. (1988). Patterns of brain electrical activity during facial signs of emotion in ten-month-old infants. *Developmental Psychology, 24*, 230-236.

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.

Fridlund, A.J. (1994). *Human facial expression.* NY: Academic.

Holland, M.K. & Tarlow, G. (1972). Blinking and mental load. *Psychological Reports, 31*, 119-127.

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.

Karson, C.N. (1988). Physiology of normal and abnormal blinking. *Advances in Neurology, 49,* 25-37.

Katsikitis, M. & Pilowsky, I. (1988). A study of facial expression in Parkinson's disease using a novel microcomputer-based method. *Journal of Neurology, Neurosurgery, and Psychiatry, 51*, 362-366.

Kraut, R.E. & Johnson, R. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology, 37,* 1539-1553.

Lien, J.J.J., Kanade, T., Cohn, J.F., & Li, C.C. (2000). Detection, tracking, and classification of subtle changes in facial expression. *Journal of Robotics & Autonomous Systems***,** *31*, 131-146.

Liu, Y., Schmidt, K., Cohn, J.F., & Weaver, R.L. (May 2002). Human facial asymmetry for expression-invariant facial identification. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 5,* 208-214.

Lucas, B.D. & Kanade, T. (1981). An iterative image registration technique with an application in stereovision. *Proceedings of the International Joint Conference on Artificial Intelligence, 7,* 674-679.

Martin, P. & Bateson, P. (1986). *Measuring behavior: An introductory guide*. Cambridge: Cambridge University Press.

Matias, R., Cohn, J.F., & Ross, S. (1989). A comparison of two systems to code infants' affective expression. *Developmental Psychology, 25,* 483-489.

Moore, G.A., Cohn, J.F., & Campbell, S.B. (1997), Mothers' affective behavior with infant siblings: Stability and change. *Developmental Psychology, 33,* 856-860.

Padgett, C., Cottrell, G.W., & Adolphs, B. (1996). Categorical perception in facial emotion classification. *Proceedings of the Cognitive Science Conference, 18*, 249-253.

Pantic, M. (August 2002). Facial gesture recognition from static dual-view face images. *Proceedings of the International Conference on Measuring Behavior, 4,* 195-197.

Poelman, C.J. (1995). *The paraperspective and projective factorization methods for recovering shape and motion* (Tech. Rep. CMU-CS-95-173*).* Pittsburgh, PA: Carnegie Mellon University, Robotics Institute.

Rinn, W.E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95,* 52-77.

Rowley, H.A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. . *IEEE Transactions on Pattern Recognition and Machine Intelligence, 20,* 23-38.

Sayette, M.A., Cohn, J.F., Wertz, J.M., Perrott, M.A., & Parrott, D.J. (2001). A psychometric evaluation of the Facial Action Coding System for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25,* 167-186.

Schmidt, K.L. & Cohn, J.F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression. *Yearbook of Physical Anthropology, 44,* 3-24.

Schmidt, K.L. & Cohn, J.F. (Manuscript submitted for publication). Signal characteristics of spontaneous facial expression: Automatic movement in solitary and social smiles.

Tian, Y.L., Kanade, T., & Cohn, J.F. (May 2002). Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 5,* 229-234.

Tian, Y.L, Kanade, T., & Cohn, J.F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23,* 97-116.

Tian, Y., Kanade, T., and Cohn, J.F. (October 2000). Eye-state detection by local regional information. In T. Tan, Y. Shi, & W. Gao (Eds.). Advances in Multimodal Interfaces – ICMI 2000. Third International Conference, Beijing, China. *Lecture Notes in Computer Science, 1948,* pp. 143-150. NY: Springer.

VanSwearingen, J.M., Cohn, J.F., & Bajaj-Luthra, A. (1999). Specific impairment of smiling increases severity of depressive symptoms in patients with facial neuromuscular disorders. *Journal of Aesthetic Plastic Surgery, 23,* 416-423.

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.

Xiao, J., Kanade, T., & Cohn, J.F. (2002). Robust full motion recovery of head by dynamic templates and re-registration techniques. *Proceedings of the IEEE*

International Conference on Automatic Face and Gesture Recognition, 5, 163-169.

Zlochower, A. (2001). *Automated face analysis distinguishes deliberate from spontaneous smiles on the basis of quantitative differences in intensity and asymmetry*. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, Pennsylvania.

Author Note

This research was supported by grants R01 MH51435 from the National Institute of Mental Health and CIA Contract #2000-A128400-000 to Drs. Cohn and Kanade. We wish to thank Paul Ekman, Mark Frank, Marian Bartlett, Kathleen Miritello, Javier Movellan, Pietro Perona, Terry Sejnowski, and Yaser Yacoob for invaluable assistance, comments, and suggestions. Paul Ekman and Mark Frank generously provided image data.

Portions of the findings were presented at the International Conference on Methods and Techniques in Behavioral Research, Amsterdam, August 2002, and the International Conference on Pattern Recognition, Quebec, Canada, August 2002.

Correspondence should be addressed to Jeffrey F. Cohn, Department of Psychology, 4327 Sennott Square, the University of Pittsburgh, Pittsburgh, PA 15260. Tel: 412-624-8825. Fax: 412-624-2023. Email address: jeffcohn@pitt.edu.

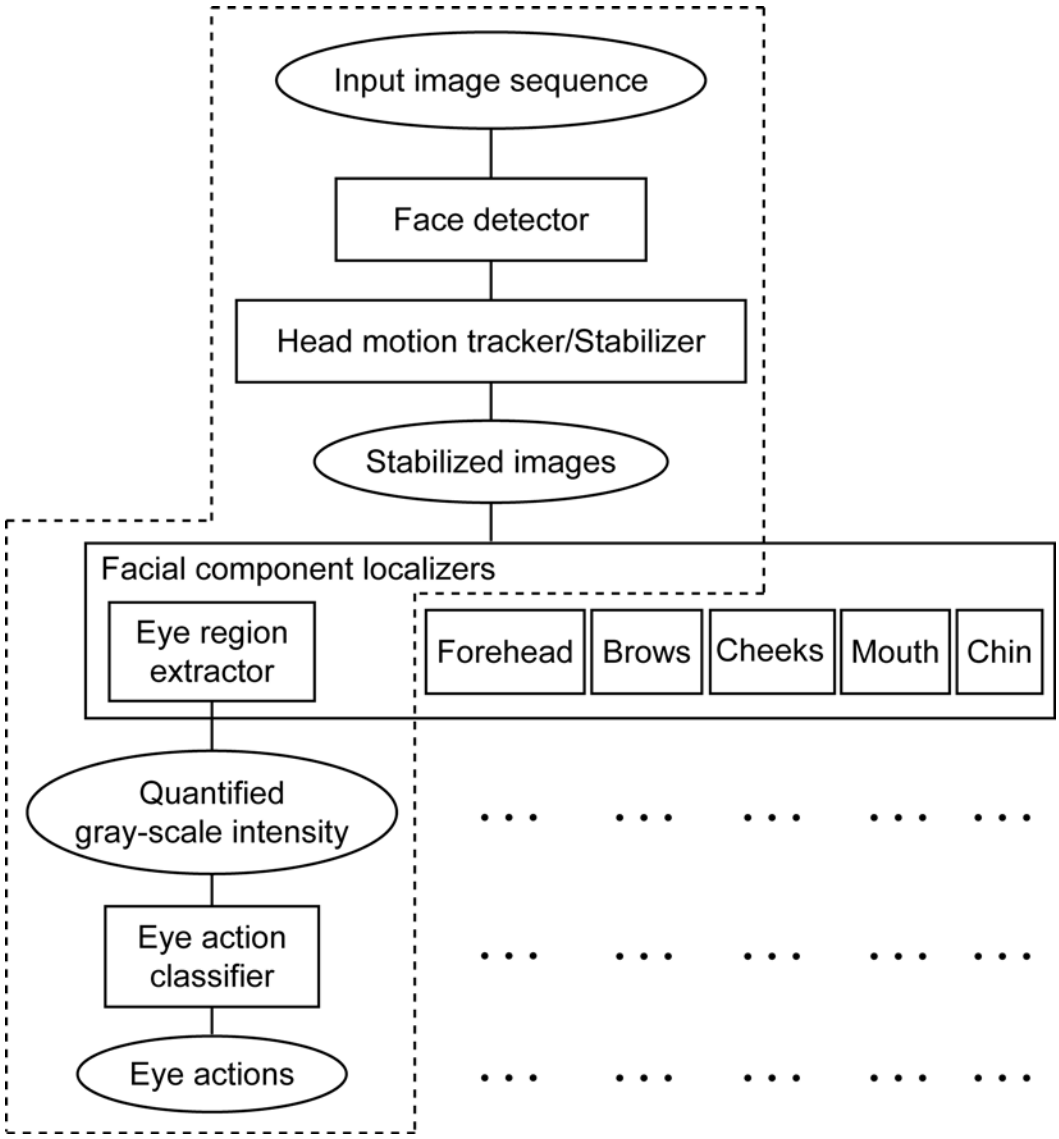
Table 1
Comparison of Manual FACS Coding and Automatic Recognition.

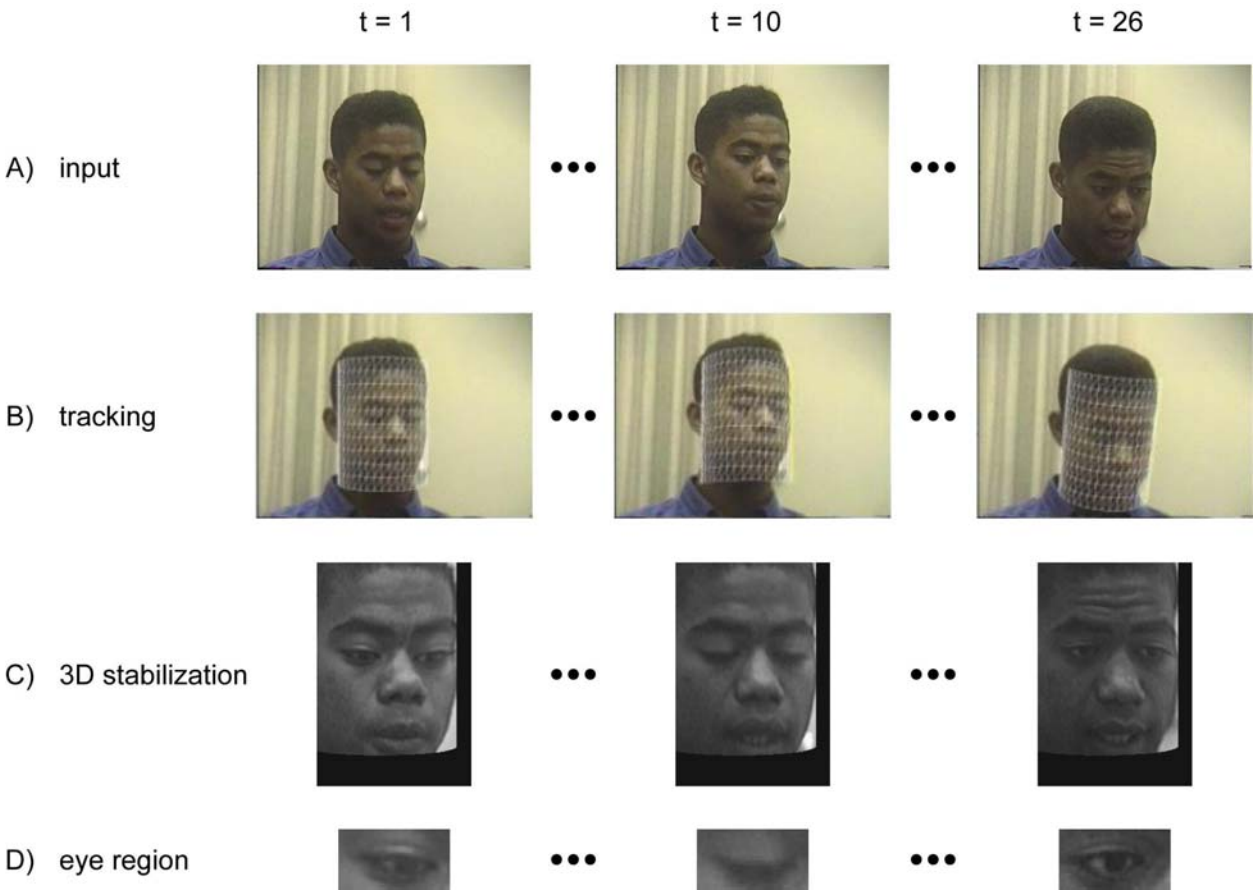
		Automatic Recognition		
		Blink (AU45)	Flutter	Non-Blink
Manual FACS Coding	Blink (AU45)	153	0	0
	Flutter	6	8	0
	Non-Blink	0	0	168

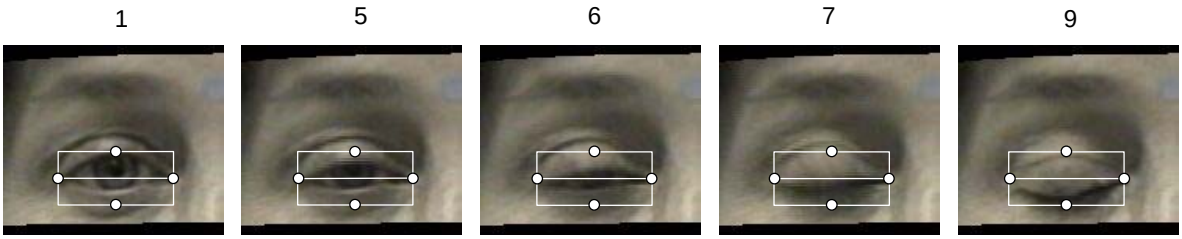
Note. Overall agreement = 98% (kappa = .97). Combining blink and flutter agreement = 100%.

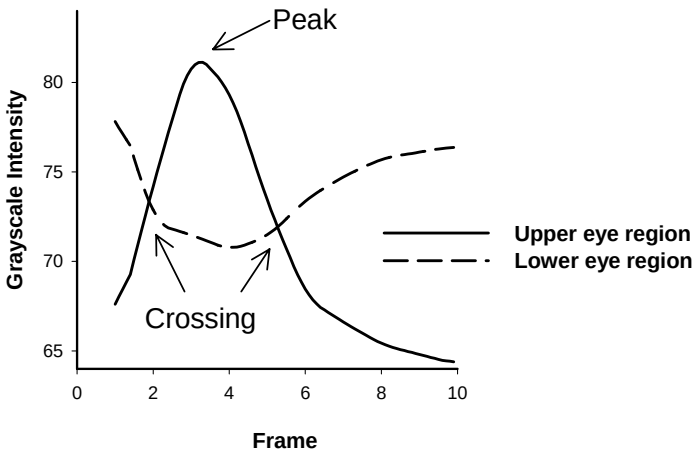
Figure Caption

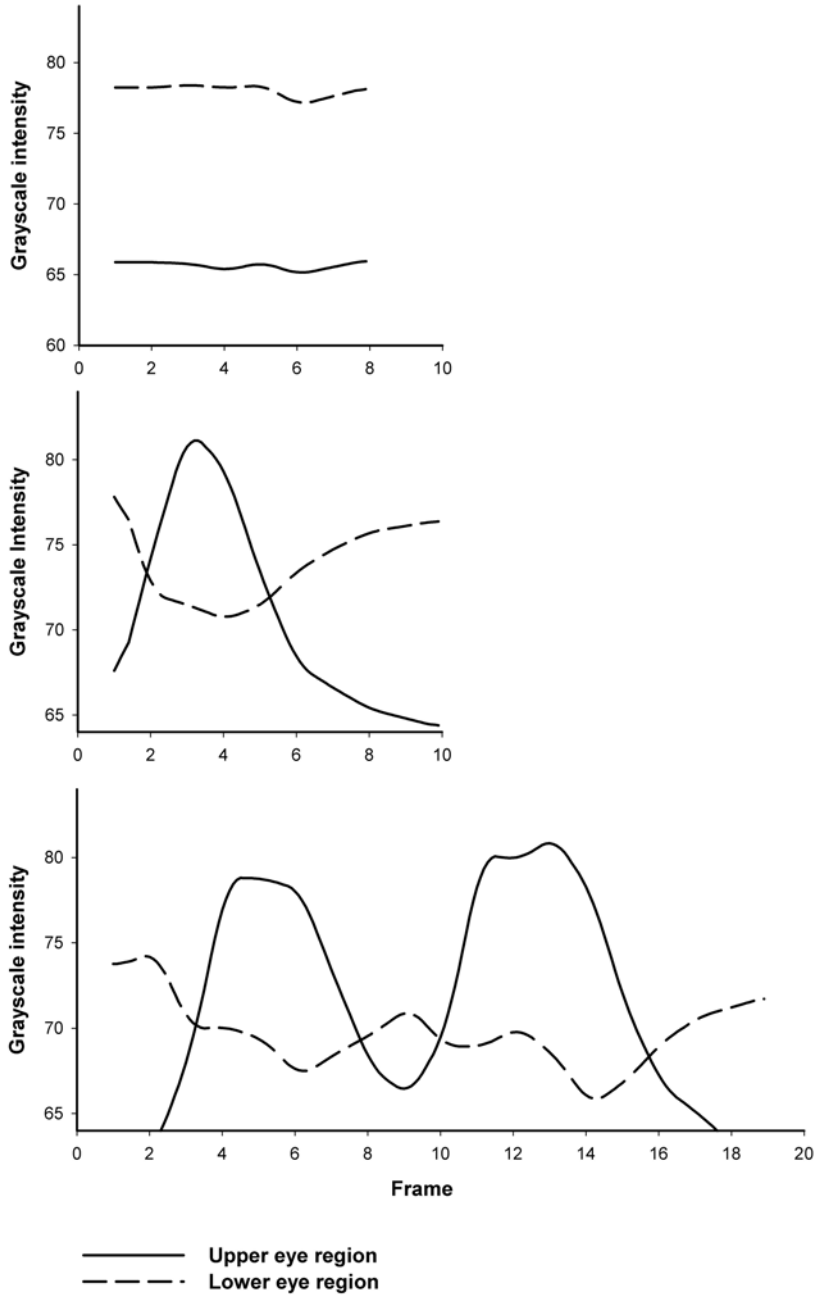
Figure 1. Overview Automated Face Analysis version 3.
Figure 2. Automatic recovery of 3D head motion and image stabilization. A) Frames 1, 10, and 26 from the original image sequence. B) Automatic face tracking in the corresponding frames. C) Stabilized face images. D) Localized eye regions.
Figure 3. Upper and lower portions of the eye region in frames 1, 5, 6, 7, and 9 from a stabilized image sequence in which the eye changes from open to closed.
Figure 4. Blink detection algorithm. The number and duration of non-blinks, blinks and flutter are detected by counting the number of peaks (Np) and crossings (Nc) in the grayscale intensity curves. If $Np = 0 \rightarrow$ non-blink; if $Np = 1$ and $Nc \neq 0 \rightarrow$ blink; if $Np \geq 2$ and $Nc \neq 0 \rightarrow$ flutter.
Figure 5. Examples of grayscale intensity curves for non-blink, blink, and flutter.
Figure 6. Comparison of manual ordering of eye closure by Rater 1 and Rater 2 and automatic ordering using automatic facial expression analysis.

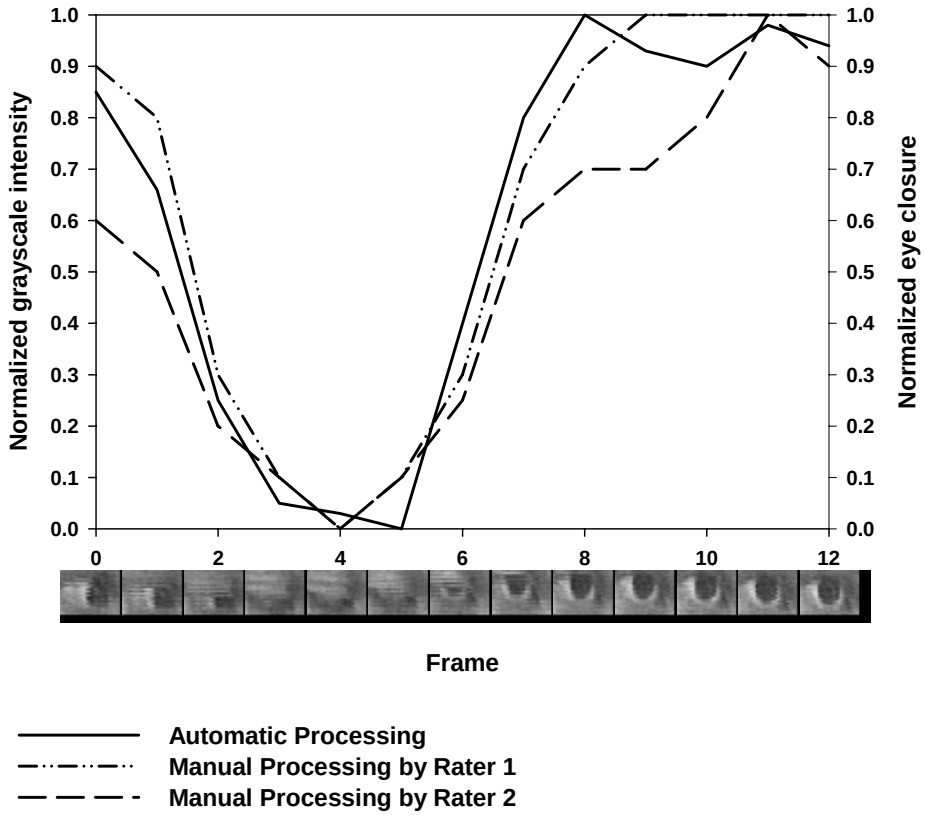

Input image sequence
Face detector
Head motion tracker/Stabilizer
Stabilized images
Facial component localizers
Eye region extractor
Forehead
Brows
Cheeks
Mouth
Chin
Quantified gray-scale intensity
Eye action classifier
Eye actions


t = 1
t = 10
t = 26
A) input
B) tracking
C) 3D stabilization
D) eye region


1
5
6
7
9


Peak
Crossing
Grayscale Intensity
80
75
70
65
0
2
4
6
8
10
Frame
Upper eye region
Lower eye region


Grayscale intensity
80
75
70
65
60
0
2
4
6
8
10
Grayscale Intensity
80
75
70
65
0
2
4
6
8
10
Grayscale intensity
80
75
70
65
0
2
4
6
8
10
12
14
16
18
20
Frame
Upper eye region
Lower eye region


Normalized grayscale intensity
Normalized eye closure
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
12
Frame
Automatic Processing
Manual Processing by Rater 1
Manual Processing by Rater 2