A Comparative Study of Alternative FACS Coding Algorithms

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

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

Technical Report CMU-RI-TR-02-06

OVERVIEW

Two groups were contracted to experiment with coding of FACS (Ekman & Friesen, 1978) action units on a common database. One group is ours at CMU and the University of Pittsburgh, and the other is at UCSD. The database is from Frank and Ekman (1997) who video-recorded an interrogation in which subjects lied or told the truth about a mock crime. 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. This report describes the results of automated facial expression analysis by the CMU/Pittsburgh group. An interdisciplinary team of consultants, who have combined expertise in computer vision and in facial analysis, will compare the results of this report with those in a separate report submitted by the UCSD group.

BACKGROUND

People communicate not only by speech and written language but also by their tone of voice, the way they stand or move and their patterns of gaze. These modes of nonverbal behavior communicate emotion and often are referred to as paralinguistic because they modify, substitute for, and improve the understanding of spoken communication. Of the various modes of nonverbal communication, the human face is especially important. Facial expressions can indicate emotion and pain, regulate social behavior, and reveal brain function. A large literature in psychology (Ekman & Rosenberg, 1997), comparative biology (Darwin, 1872/1998; Fridlund, 1994; Schmidt & Cohn, in press), rehabilitative medicine (VanSwearingen & Cohn et al., 1998, 1999), and neuroscience (Rinn, 1984) informs the interpretation of facial expression. Available methods for coding facial expression, however, are human-observer dependent, labor intensive, and difficult to standardize. These problems tend to limit the use of facial expression analysis in clinical and forensic settings and in human-computer interaction where they are needed. To make optimal use of the information afforded by facial expression, reliable, valid and efficient methods of measurement are critical.

Within the past decade, there has been significant effort toward analysis of human facial expression using computer vision. Several such systems (e.g., Essa & Pentland, 1994, 1997; Padgett, Cottrell, & Adolphs, 1996; Yacoob & Davis, 1994) have recognized under controlled conditions a small set of emotion-specified expressions, such as joy and anger. This focus on emotion-specified expressions follows from the work of Darwin (1872) and more recently Ekman (1993) and Izard (Izard et al.,1983) who proposed that basic emotions have corresponding prototypic facial expressions. In everyday life, however, such prototypic expressions occur relatively infrequently. Instead, emotion more often is communicated by changes in one or two discrete features, such as tightening the lips in anger or obliquely lowering the lip corners in sadness (Gosselin et al., 1995). Change in isolated features, especially in the area of the brows or eyelids, is typical of paralinguistic displays (e.g., Eibl-Eibesfeldt, 1989). Subtle changes in facial expression are associated with negative emotion and intention in high-stakes contexts (Ekman, 2001). To capture the full range of facial expression, detection, tracking, and classification of fine-grained changes in facial features are needed.

The anatomically based Facial Action Coding System (FACS: Ekman & Friesen, 1978a) currently is the most comprehensive manual method of analyzing facial displays. FACS consists of 44 action units. Thirty are anatomically related to contraction of specific facial muscles while the anatomic basis of another 14 is unspecified. Using FACS and viewing videotaped facial behavior in slow motion, coders can manually code all possible facial displays. Although Ekman and Friesen (1978) proposed that specific combinations of FACS action units represent prototypic expressions of emotion, it should be noted that emotion expressions are not part of FACS; they are coded in separate systems, such as EMFACS (Friesen & Ekman, 1983) or MAX (Izard et al., 1983). FACS itself is purely descriptive, uses no emotion or other inferential labels, and provides the necessary ground truth with which to describe facial expression.

Previous work by the UCSD and CMU/Pittsburgh groups. In previous work, the CMU/Pittsburgh and UCSD groups have achieved some success in recognizing FACS action units under controlled conditions. The CMU group has demonstrated automatic recognition of 18 action units using a feature-based approach whether they occur alone or in as many as 30 combinations (Cohn, Zlochower, Lien, & Kanade, 1999; Tian, Kanade, & Cohn, 2000, 2001). The UCSD group using a computational neuroscience approach has recognized 12 action units (Bartlett, Hager, Ekman, and Sejnowski, 1999; Donato et al., 1999). A limitation of this work, and indeed of almost all research to date in automated facial expression analysis (cf. Schmidt & Cohn, 2001), is that it is limited to deliberate facial expressions 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 (Kanade, Cohn, & Tian, 2000). These challenges are well represented in the database used in the research reported here. This is the first research effort to attempt automated action unit recognition in naturally occurring (spontaneous) facial behavior.

To accomplish this goal, the CMU/Pittsburgh group has developed a third version of its Automated Face Analysis. The CMU/Pittsburgh system automatically recognizes action units in the context of 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 action units in spontaneous facial behavior. Manual processing is limited to marking seven feature points in the initial image of the stabilized image sequence. All other processing is automatic. In initial tests, reported below, the system recognized blinks (AU 45) with 100% accuracy (kappa = 1). Action units in the brow region were recognized with 57% accuracy (kappa = .33). Excluding brow-down, for which training data were limited to only 13 sequences, recognition accuracy for brow motion increased to 80% (kappa = .58, $p < .0001$).

DATABASE

Image data were from a study of deception by Frank & Ekman (1997). The subjects were 20 young adult men. Seven were Euro-American, 2 African-American, and 1 Asian. Two subjects 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. Twelve subjects stole $50 and 8 told the truth. 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; brow motion was coded in an additional 7 subjects. Certified FACS coders from the CMU/Pittsburgh group confirmed all coding.

Blink. Measurement of blink (AU 45 in FACS) is important in several fields, including neurology, physiology, and psychology. Control of blinking is distributed among cranial nerves 3 and 7, higher motor pathways, and facial muscles *levator palpebrae superioris, orbicularis oculi,* and *pars palpebralis*. Blink rate varies with physiological and emotional arousal, dopaminergic activity and personality (Blin et al., 1990; Depue et al., 1994), cognitive effort (Holland & Tarlow, 1972; Karson, 1988), and incentive motivation (Meyer et al., 1953). Blink rate is decreased in Parkinson's disease (Karson et al., 1984) and increased in schizophrenia (e.g., Karson, 1988). Increased blink rate is an indicator of deception (Ekman, 2001). We included for analysis all instances of blink (AU 45) for which the 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.

We classified separately the few instances of multiple blinks that were present in the database. Multiple blink (eyelid "flutter") is defined as two or more rapidly repeating blinks (AU 45), which may be separated by AU 42 (eyelids appear as a 'slit,' or nearly closed) rather than full eyelid opening.

Brow motion. Brow motion is important in emotion and paralinguistic communication. The combination AU 1+2 (*frontalis*), which raises both the inner and outer portions of the brow, is a component of the prototypic expressions of surprise (Ekman, 1984, 1993), and is common in paralinguistic communication (e.g., brow flash: Eibl-Eibesfeldt, 1989). AU 4 (*corrugator supercilii),* which draws the brows medially and down, is an index of negative affect (Cacioppo et al., 1986; Ekman, 1984) and concentration (Scherer, 1992). AU 9 (*levator labii superioris alaeque nasi*) is a component of disgust expressions. AU 9 wrinkles the nasal root and lowers the medial portion of the brows when moderate to strong. Action units were aggregated into brow-up (AU 1+2) and brow-down (AU 4 or AU 9) because there were too few sequences in which the component action units occurred singly. The Pittsburgh and Rutgers coders agreed on 77% of the brow-up but only 52% of the brow-down. While analysis was limited to sequences on which coders were in agreement, the low reliability and small number and heterogeneity of exemplars of brow-down were limiting factors.

CMU/PITTSBURGH AUTOMATED FACE ANALYSIS V.3

Figure 1 depicts an overview of the CMU/PITTSBURGH face analysis system. 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 (Rowley, Baluja, & Kanade, 1998). Head motion (6 *DOF*) is recovered automatically. Using the recovered motion parameters, the face region is stabilized. In the only necessary manual step, a few feature points are marked around the subject's right eye (image left) and brow in the first image of the stabilized image sequence. Eye and brow features are extracted in the image sequence, and action units and action unit combinations then are recognized. Figure 2 shows the graphical user interface we developed to implement this system.

Automated Recovery of 3D Head Motion and Stabilization

Expressive changes in the face often occur together with head movement. Raised brows, for instance, often occur as the head pitches back (Camras et al., 1996), although the opposite action or head turns to the side also can be observed. Expression may vary too as a result of individual differences in facial proportions (Farkas & Munro, 1987). Completely removing the effects of head movement from the input image sequence would be very difficult. It may even require a complicated transformation that is dependent on the knowledge of the exact shape of the individual face. When out-of-plane rotation of the head is negligible or absent (Bartlett et al., 1999; Cohn et al., 1999), an affine or perspective transformation is adequate to align images so that face position, size, and orientation are kept relatively constant across subjects (Cohn et al., 1999; Lien et al., 2000). For significant out-of-plane motion, which is common in naturally occurring facial behavior, modeling and tracking of the head in 3D becomes necessary. We seek a model that is computationally fast, automatic, and contains the smallest number of parameters necessary for robust 3D head tracking, motion recovery, and image warping of the face region. Our goal is to stabilize the face image so that the effects of rigid motion do not interfere with feature extraction or action unit recognition.

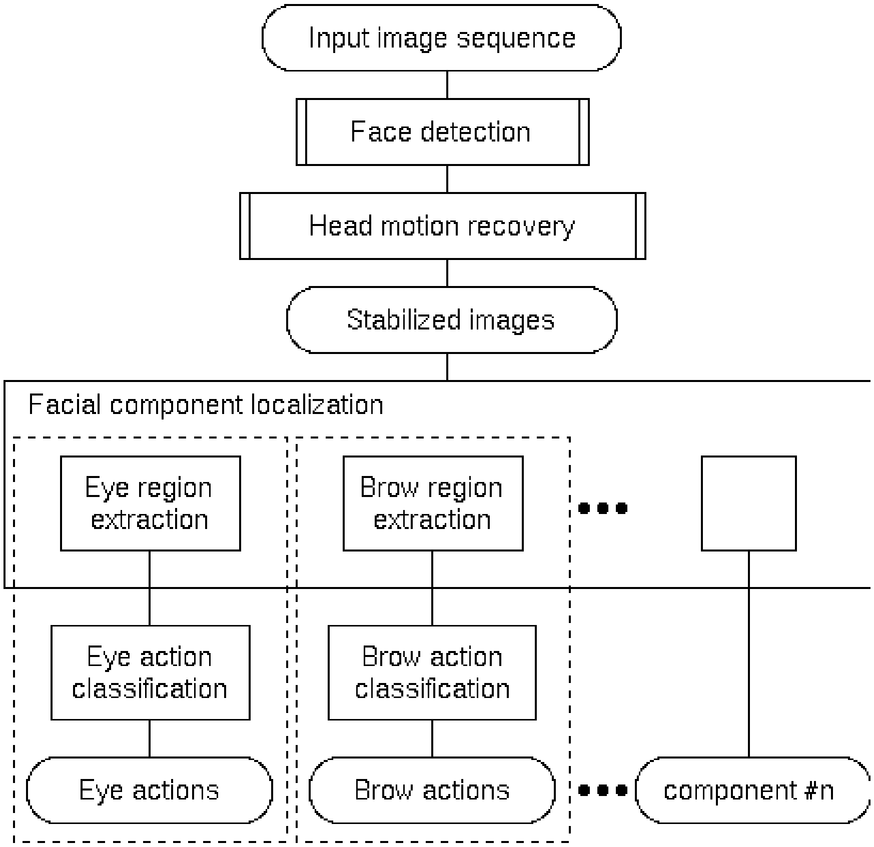


Figure 1. Overview, CMU/Pittsburgh Automated Face Analysis v.3.

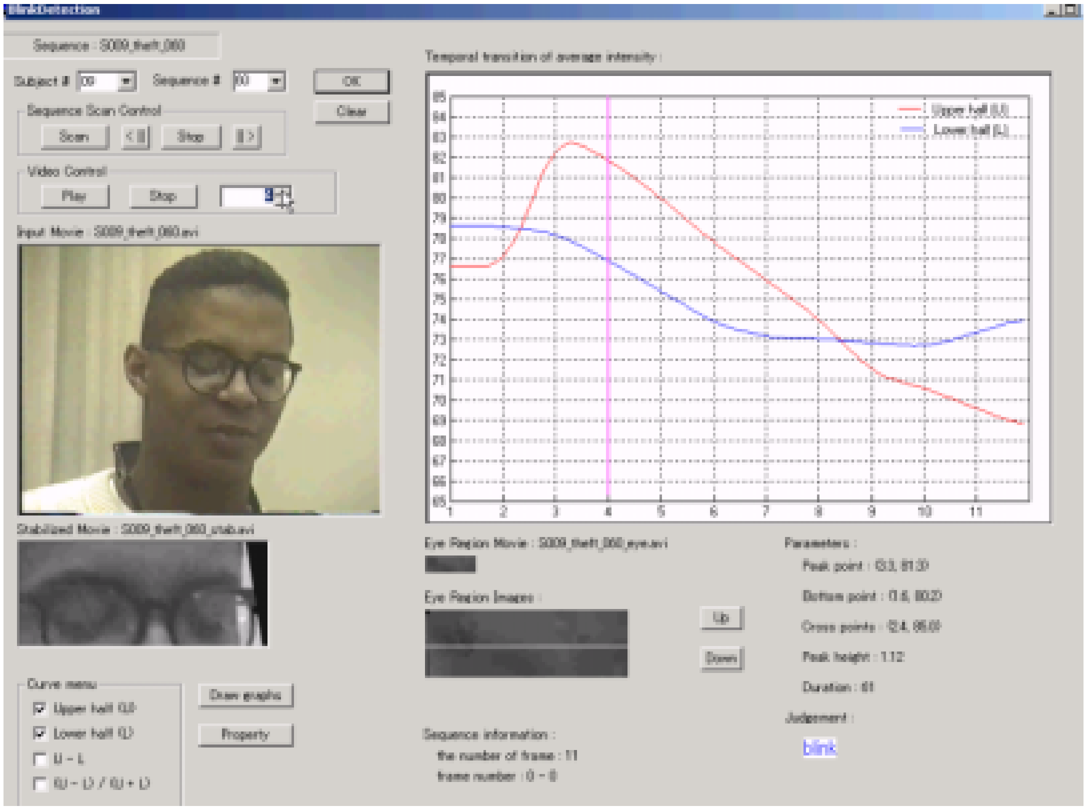
Initial processing includes face detection, which could be performed automatically (Rowley et al., 1998), histogram matching, for global lighting changes, and 2D color-blob face tracking. The latter provides preliminary estimates of 2D head location in the new image; from these initial estimates, the horizontal and vertical translations can be computed prior to 3D motion recovery. 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. 3D head pose is estimated automatically. Experimental tests suggest that the system is insensitive to small variations in the initial fit of the head model (Xiao, Kanade, & Cohn, Submitted).

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 outliers from the templates. A pixel (x, y) within this region will be removed from the new template as an outlier if,

$$| I(f(x, y; \mu), t+1) - \hat{I}(f(x, y; \mu), t) | > c\sigma_R$$

where *c* is a constant that represents the strictness of judgment on outliers. This procedure contributes to system robust to occlusion and non-rigid motion.

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 (For further detail, please see Xiao, Kanade, & Cohn, submitted) (Appendix 6). An example of system output using the Frank & Ekman (1997) data can be seen in Figure 2.



Sequence : S009_theft_060

Subject # 09 Sequence # 60

Input Movie : S009_theft_060.avi

Stabilized Movie : S009_theft_060_stab.avi

Eye Region Movie : S009_theft_060_eye.avi

Eye Region Images :

Parameters :
Peak point : (3.3, 81.3)
Bottom point : (1.6, 80.2)
Cross points : (2.4, 65.0)
Peak height : 1.12
Duration : 61

Judgement :
blink

Sequence information :
the number of frame : 11
frame number : 0 - 0

☑ Upper half (U)
☑ Lower half (L)
☐ U - L
☐ (U - L) / (U + L)

Figure 2. Graphical user interface for CMU/Pittsburgh Automated Face Analysis system. For demonstration, please see www-2.cs.cmu.edu/~tmoriyam/blink/.

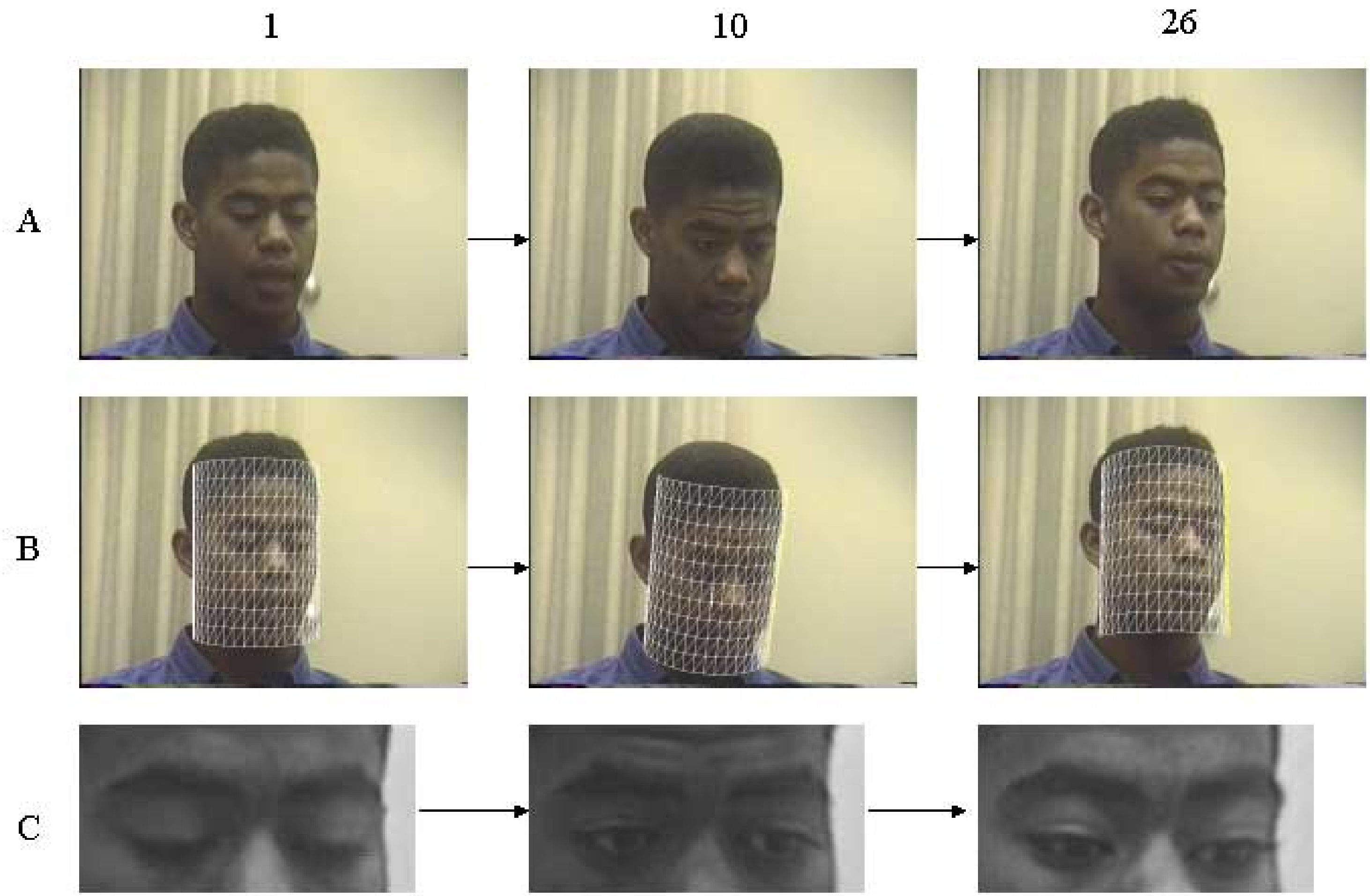


Figure 3. Automated recovery of 3D head motion and image stabilization. A) Frames 1, 10, and 26 from original image sequence. B) Face tracking in corresponding frames. C) Stabilized eye and brow regions.

ANALYSIS OF BLINK

The eye region (Figure 4) consists of the iris, sclera, upper and lower eyelids and the eyelashes, which can be regarded as vertically symmetric with the exception of vertical motion of the iris and the difference between upper and lower eyelashes. The visible part of the iris is approximately rectangular in the absence of moderate to strong AU 5.

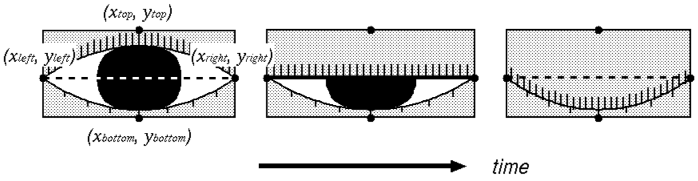


Figure 4. 2D eye model.

If we divide an eye region into upper and lower portions, the difference between the upper and lower portion would be reflected in the difference of the statistical feature of illumination associated with the eyelashes. In blinking, illumination changes occur when the eyelashes descend into the lower portion of the eye region.

Automated Feature Extraction in the Eye Region

The input face image sequence (Figure 3A) has been automatically preprocessed to obtain the stabilized image sequence (Figure 3C and as described above). The face region is tracked in the image sequence and change from the first frame (head position angle = **0**) in terms of head motion is recovered. Then, by manually giving the feature points (Figure 5) $\{x_i, y_i; i = left, top, right, bottom\}$ 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, which is the target sequence of eye action classification here (Figure 6).

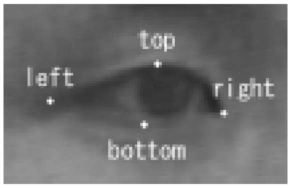


Figure 5. Feature points used to define size of eye region in initial frame.

Now we treat only the right eye (image left) here because it is assumed for now that the target eye actions are symmetric between left and right. The target eye actions are blink, multiple blink (eyelid 'flutter') and non-blink.

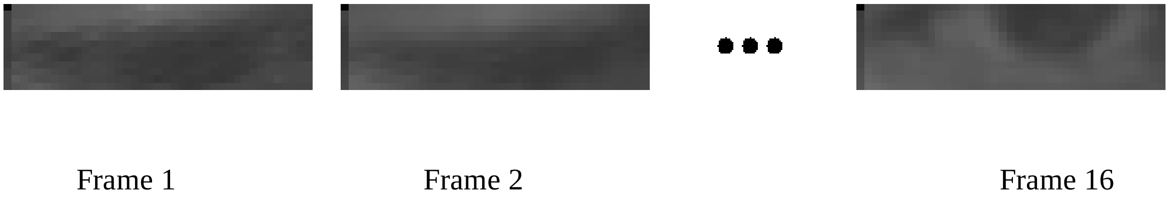


Figure 6. Automatically stabilized eye images.

The algorithm to classify blink, multiple blink (eyelid flutter), and non-blink from feature vectors is:

```
Input : T frames, eye region images (size M x N)
Output : Eye action class (blink, multiple blink [eyelid flutter], non-blink)
Begin

1. for   $1 \le t \le T$
  1.1 $\mathrm{Im}_u(t) := \frac{1}{D} \sum_{x_{left} \le x \le x_{right}, y_{top} \le y \le y_{bottom}/2} I(x, y, t)$
  1.2 $\mathrm{Im}_l(t) := \frac{1}{D} \sum_{x_{left} \le x \le x_{right}, y_{top}/2 \le y \le y_{bottom}} I(x, y, t)$
  where, $D = \frac{(x_{right} - x_{left})(y_{top} - y_{bottom})}{2}$
2. BC := 0
3. BlinkFlag := 0
4. for $1 \le t \le T$
  4.1 if BlinkFlag=0 and ($\mathrm{Im}_u(t) > \mathrm{Im}_l(t)$) and ($\mathrm{Im}_u(t+1) < \mathrm{Im}_u(t)$),
      then, BC=BC+1、BlinkFlag=1
  4.2 if BlinkFlag=1 and ($\mathrm{Im}_u(t) < \mathrm{Im}_l(t)$), then BlinkFlag = 0
5. if
  5.1 BC = 0, then Non-Blink
  5.2 BC = 1, then Blink
  5.3 BC $\ge$ 2, then Multiple Blink (eyelid flutter)
end
```

In 1., the average illumination intensity in the upper and the lower half of the eye region image of frame #t is calculated, where 1.1 is correspondent to the upper half ($\mathrm{Im}_u(t)$), 1.2 corresponds to the lower half ($\mathrm{Im}_l(t)$).

2. initializes *BC* which denotes the number of blinks. 3. initializes the eye closure flag *BlinkFlag*. 4. calculates the number of blinks, where in 4.1, *BC* is incremented when the eye closure condition is satisfied, and *BlinkFlag* is set to *1*, in 4.2, *BlinkFlag* is reset to *0*. In 5., the eye action is classified based on the number of *BC*.

Figure 7 shows examples of luminance curves for blink, multiple blink (eyelid flutter), and non-blink. Additional examples may be found at http://www.cs.cmu.edu/~tmoriyam/blink.

Recognition Results for Blink

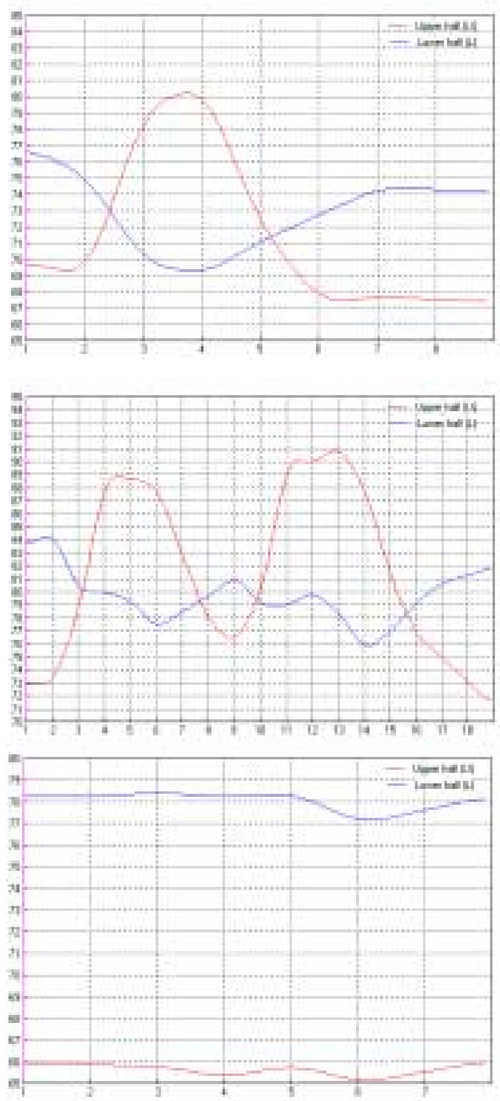
The algorithm achieved an overall accuracy of 98%, with 100% accuracy between blinks and non-blinks (Table 1). Six of 14 multiple blinks were incorrectly recognized as single blinks. Rapid transitions from AU 45 to AU 42 to AU 45, in which eye closure remains nearly complete, were occasionally recognized as a single blink. The symmetry metric in this case was not consistently sensitive to the slight change from eye closed to AU 42. Additional statistical measures may be needed to more consistently recognize instances of AU 42. In previous work, we have found that Gabor wavelets can perform well for this purpose (Tian, Kanade, & Cohn, 2000).

Table 1. Comparison of Manual FACS Coding and Automated Face Analysis Recognition of Blink.

		Automated Face Analysis v.3		
		Blink (AU 45)	Multiple Blink	Non-Blink
Manual FACS Coding	Blink (AU 45)	153	0	0
	Multiple Blink	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, agreement = 100% (kappa = 1).

The current findings are encouraging in light of the challenges presented in this image database. Ethnic background of the 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. Reflection from the eyeglasses was a further challenge. Overall accuracy was 98%. Combining single and multiple blinks (which is common practice among FACS coders), accuracy was 100%. In future work, we will refine and formulate the model of eye motion and will examine the ability of this approach quantitatively in terms of factors such as noise and image resolution. A major goal is to test the system on additional action units.



Blink

Multiple Blink

Non-Blink

Figure 7. Examples of luminance curves for blink, multiple blink, and non-blink.

ANALYSIS OF BROW MOTION

As noted above, we focused on brow-up (AU 1+2) and brow-down (AU 4 or AU 9 or AU 1+4 if AU 4 preceded the AU 1). We aggregated action units in this way because the number of individual action units of each type was too small for analysis. Even so, the number of brow-down was only 13. The target action units tended to occur at low intensity. In brow-down, for instance, only in two cases did intensity exceed the ‘b’ level on a 5-point ordinal scale from ‘a’ (barely perceptible), ‘b’ minimal intensity, to ‘e’ (maximal intensity).

Brow motion may be detected by change in both the position of face components (i.e., brows) and in the appearance of transient wrinkles. The brows are raised in AU 1+2 and lowered and drawn medially in AU 4. AU 9 tends to lower the brows. Wrinkling perpendicular to the direction of muscle contraction provides additional cues. In AU 1+2, wide horizontal wrinkles appear across the forehead. In AU 4, vertical wrinkles form between the brows and oblique wrinkles may appear beginning near and above the inner corners. In AU 9, horizontal wrinkles may occur at the nasal root. Because the action units we observed were typically of low intensity, wrinkles often failed to form. In addition, because action units occurred at low intensity, the amount of brow motion often was often less than 2 pixels in magnitude. Under these conditions, small errors in image stabilization or in feature extraction could result in error.

Automated Feature Extraction in the Brow Region

Figure 8 indicates the general flow of our method. The input sequence refers to the brow region image sequence, which was obtained by the automated preprocessing described in Figure 3. The brow region image is processed by two parallel modules. To detect wrinkles, we use edge detectors in the rectangular region just above the brow. To quantify motion, we quantify the pixel displacement from the initial position. Based on these measurements, brow action classification is performed.

In this system, the input face image sequence (Figure 3A) is preprocessed to obtain the stabilized image sequence (3C) by detecting the face region through the sequence and recovering the changes from the first frame (head position angle = **0**) in terms of head motion. Then by giving the feature points manually (Fig. 9 $\{x_{li}, y_{li}, x_{ri}, y_{ri}; i = left, center, right\}$) on the first image, the brow regions for the rest of the image sequence

$$I(x,y) \quad \begin{array}{l} x: \; x_{lleft} \le x \le xl_{right} + (x_{rleft} - x_{lright})/2 \\ y: \; if \; y_{lbottom} \le y_{rbottom}, 0 \le y \le y_{rbottom}, else \, 0 \le y \le y_{lbottom} \} \end{array}$$

are obtained, which is the target sequence of brow action classification here. In this report we treat only the left brow here because it is assumed that the target brow actions here are symmetric between left and right.

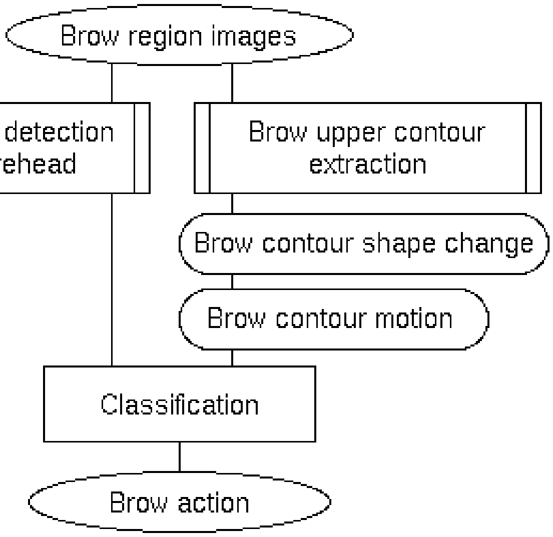

Brow region images
Wrinkle detection in forehead
Brow upper contour extraction
Brow contour shape change
Brow contour motion
Classification
Brow action

Figure 8. Flow of CMU/Pittsburgh Automated Analysis of Brow Motion

Now the algorithm to classify the target brow region image sequence into these categories is:

```
Input : T frames, brow region images (size M x N)
Output : Brow action class (Brow-up, Brow-down, non-brow motion)
begin
  1. for  $1 \le t \le T$
   if $D(t) \le -2.0 \, and \, C(t) \le 0.9$, then classified as Brow-down
   where, $D(t) = Displacement, C(t) = Correlation\ coefficients$

  2. for  $1 \le t \le T$
   $average\_displacement\ AD = \frac{\sum_t D(t)}{T}$, if $AD > 0$, then classified into Brow-up
  3. for  $1 \le t \le T$
   $edge\_variance\ EV = \frac{\sqrt{\sum_t (D(t) - \overline{D})^2}}{T}$, if $EV > 1.0$, then classified into Brow-up
   where, $\overline{D}$ is average edge power.
  4. otherwise  classified as Non-brow motion
end
```

In 1., Brow-down is sifted first based on the displacement and the decrease of the correlation coefficients. Then in 2., if the average displacement is positive, it is classified into brow-up, and in 3., if the variance of temporal edge curve is more than 1.0, it is classified into brow-up. Otherwise, it is classified as non-brow motion.

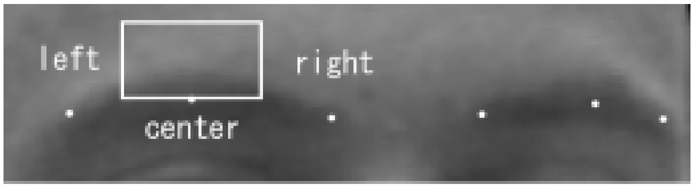


Figure 9. Brow region from automatically stabilized image sequence.

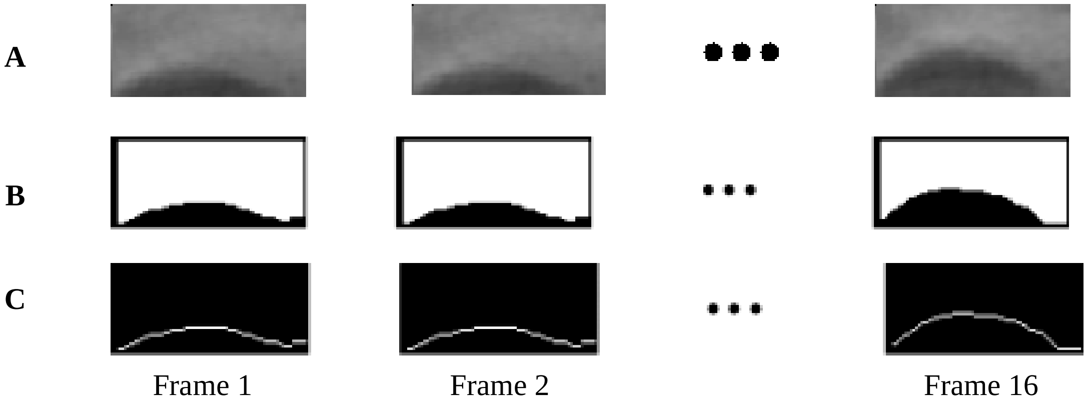


Figure 10. Feature extraction in brow region. A: Stabilized brow region. B: Binary images. C: Contour.

The input (brow region images) are shown in Figure 10A, and correspondent binary images are shown in Figure 10B. For the binary images, the boundary between skin and brow region (darker part) was searched from the lower edge in each frame, respectively (Figure 10C). Contours can be regarded as a 1D signal on the horizontal axis. We can store the contour of the first frame as the template to be compared with that of the following frames and calculate the correlation coefficients and the displacement from the initial position.

Results of Brow Classification

Accuracy in the brow region was higher for brow-up than for brow-down. Average accuracy across the three categories was 57 % (Table 2), which represents moderate agreement between manual and automated recognition.

Compared with the eye region, accuracy in the brows was lower. Several factors may have contributed to this difference. Action unit intensity was one factor. While blinks are defined by qualitative closing of the eyelid (except when combined with AU 42), brow motion can vary in degree. In the brow motions we analyzed, intensity typically was low. With only two exceptions, brow-up in particular was no higher than the ‘b’ level of a 5-point scale ranging from ‘a’ (barely perceptible) to ‘e’ (maximum intensity).

Table 2. Comparison of Manual FACS Coding and Automated Face Analysis Recognition of Brow Motion.

	Automated Face Analysis v.3		
Manual FACS Coding	Brow-Up (AU 1+2)	Brow-Down (AU 4 or AU 9)	Non-Brow Motion
Brow-Up (AU 1+2)	23	17	8
Brow-Down (AU 4 or AU 9)	1	6	6
Non-Brow motion	8	13	40

Note. Overall agreement = 57%, kappa = .33. Omitting brow-down, agreement = 80%, kappa = .58.

Brow-down is a heterogeneous category, which included both AU 4 and AU 9. While both action units lower the brows, AU 4 pulls them medially while AU 9 does not, and the occurrence and pattern of wrinkling differs both within and between these action units. Given the small number of brow-down (n = 13), this heterogeneity was particularly challenging. Occlusion from eyeglasses was another factor especially in the case of AU 4 and AU 9.

Human FACS coders had similar difficulty with brow-down, agreeing only about 50% in this dataset. The combination of occlusion from eyeglasses and correlation of forward head pitch with brow-down complicated FACS coding. Reliability between human FACS coders for brow-down was comparable to that between human FACS coders and the CMU/Pittsburgh analysis system. Were we to omit brow-down from analysis and only consider brow-up and non-brow motion, for which the amount of data was adequate and manual reliability higher, recognition accuracy would increase to 80% in the brow region. In future work, it will be important to increase substantially the amount and reliability of training data in the brow region.

GENERAL 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 Mark 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, luminance, occlusion, pose, out-of-plane head motion, and the low intensity of action units.

To meet these challenges, the CMU/Pittsburgh group developed Automated Face Analysis v.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 frame, and all other processing was fully automatic. This contrasts with the emphasis of the UCSD group, which requires manual labeling and registration of each image followed by manual localization of facial regions but emphasizes the various classification schemes of extracted features (Bartlett et al., unpublished manuscript, 2001). The two groups had complementary approaches and emphases that made this collaborative effort most productive.

The Automated Face Analysis v.3 of the CMU/Pittsburgh group successfully recognized blinks from non-blinks for all the examples in the database (which turned out to be a relatively easy task). It was also able to distinguish, with lower accuracy, however, multiple blinks separated by as few as one frame and partial eye closure (AU 42). Accuracy in automatically recognizing brow-up, brow-down, and non-brow motion was not as high; it was 57%. The small number (n=13) of samples, heterogeneity, and lower reliability of brow-down were limiting factors. Omitting brow-down, accuracy in the brow region increased to 80%, and intersystem agreement between human FACS coders and Automated Face Analysis v.3 approached or was comparable to that of human FACS coders, which is the current gold standard.

We found that automated analysis of facial images still present significant challenges. Many previously published algorithms, including our own, that worked well for frontal faces and good lighting condition images fail with images under non-frontal facial poses, full 6-DOF 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 will 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. While challenges remain, these findings support the feasibility of developing and implementing comprehensive, automated facial expression analysis in applied settings.

ACKNOWLEDGEMENTS

We thank Marian Bartlett, Javier Movellan, and Terry Sejnowski from the UCSD group for their productive and collaborative working relationships that we enjoyed and that gave insight for improving our automated face expression analysis system. We thank consultants Paul Ekman, Mark Frank, Pietro Perona, and Yaser Yacoob for invaluable comments, suggestions, and leadership, and Kathleen Miritello for her guidance in making possible this collaborative effort.

APPENDICES

1) PowerPoint Demo (Distributed in August).
2) On-line demo from www-2.cs.cmu.edu/~tmoriyam/blink/ presents detailed results, including comparisons with manual pattern recognition.
3) 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 and Autonomous Systems*, *31*, 131-146. www.cs.cmu.edu/~face.
4) Tian, Y., Kanade, T., and Cohn, J.F. (October 2000). Eye-state detection by local regional information. *Proceedings of the International Conference on Multimedia Interfaces*, pp. xxx-xxx. Beijing, China. www.cs.cmu.edu/~face.
5) 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. www.cs.cmu.edu/~face.
6) Xiao, J., Kanade, T., & Cohn, J.F. (Submitted). A real-time system of 3D head motion recovery. *Proceedings of the IEEE International Conference on Automated Face and Gesture Recognition.* This is a revised version with additional experiments of a manuscript distributed in August.

REFERENCES

Bartlett, M.S., Braathen, B., Littlewort-Ford, G., Hershey, J., Fasel, I., Marks, T., Smith, E., Sejnowski, T.J., & Movellan, J.R. (unpublished manuscript, 2001). Automatic analysis of spontaneous facial behavior: A final project report. University of California at San Diego.

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

Black, M.J. & Yacoob, Y. (1994). Recognizing facial expressions under rigid and non-rigid facial motions. *International Workshop on Automatic Face- and Gesture Recognition, F&G94,* (Zurich), 12-17.

Black, M.J., Yacoob, Y., Jepson, A.D., and Fleet, D.J. (June, 1997). Learning Parameterized Models of Image Motion. *Proceedings of the International Conference on Computer Vision and Pattern Recognition CVPR97*, 561-567.

Blin, O., Masson, G., Azulay, J.P., Fondarai, J., & Serratrice, G. (1990). Apomorhine-induced blinking and yawning in healthy volunteers. *British Journal of Clinical Pharamacology, 30,* 769-773.

Cacioppo, J.T., Petty, R. E., Losch, M. E. & Kim, H.-S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50,* 260-268.

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

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.

Darwin, C.R. (1872/1998). *The expression of emotions in man and animals*. Third edition with an introduction, afterword, and commentaries by Paul Ekman. NY: Oxford University.

Depue, R.A., Luciana, M., Arbisi, P., Collins, P., & Loeon, A. (1994). Dopamine and the structure of personality: Relation of agonist-induced dopamine activity to positive emotionality. *Journal of Personality and Social Psychology, 67,* 485-498.

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

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

Ekman, P. (1984). *Expression and the nature of emotion*. In K.R. Scherer & P. Ekman (Eds.), Approaches to emotion. Hillsdale, NJ: LEA.

Ekman, P. (2001). *Telling lies: Clues to deception in the marketplace, politics, and marriage.* NY: Norton.

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

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

Ekman, P. and Rosenberg, E.L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS).* Oxford University Press, New York, 1997.

Essa, I.A. & Pentland, A. (1994). A vision system for observing and extracting facial action parameters. *IEEE CVPR 1994*.

Farkas, L.G. & Munro, I.R. (1987). *Anthropometric facial proportions in medicine*, Springfield, Illinois: Charles C. Thomas.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions*. NY: Wiley.

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.

Frank, M., Ekman, P., & Friesen, W. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality & Social Psychology*, *64*, 83-93.

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.

Frank, M., Ekman, P., & Friesen, W. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality & Social Psychology*, *64*, 83-93

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

W.V. Friesen, P. Ekman, *EMFACS-7: Emotional Facial Action Coding System,* Unpublished manuscript, University of California at San Francisco, 1983.

Gosselin, P., Kirouac, G., & Dore, F. Y. (1995). Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology, 68,* 83-96.

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

Izard, C.E., Dougherty, L.M., & Hembree, E.A. (1983). *A system for identifying affect expressions by holistic judgments.* Unpublished Manuscript, University of Delaware.

Kanade, T., Cohn, J.F., & Tian, Y. (March 2000). Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), pp. 46-53. Grenoble, France.

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

Karson, C.N., Burns, R.S., LeWitt, P.A., Foster, N.L., & Newman, R.P. (1984). Blink rates and disorders of movement. *Neurology, 34,* 677-678.

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 and Autonomous Systems**,** 31, 131-146.

Meyer, D.R., Bahrick, H.P., & Fitts, P.M. (1953). Incentive anxiety, and the human blink rate. *Journal of Experimental Psychology, 45,* 183-187.

Padgett, C., Cottrell, G.W., & Adolphs, B. (1996). Categorical perception in facial emotion classification. *Cognitive Science*, *x,* xxx-xxx.

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 Analysis and Machine Intelligence, 20,* 23-38.

Scherer, K.R. (1992). What does facial expression express? In K.T. Strongman (Ed.), *International Review of Studies on Emotion, 2,* pp. 138-165. NY: Wiley.

Schmidt, K.L., Monaco, V., Peters, B., VanSwearingen, J., Tian, Y., and Cohn, J.F. (October 2000). Relation between automatic tracking of lip-corner motion and facial surface EMG of the *zygomaticus major* muscle during spontaneous smiles. *Society for Psychophysiological Research,* San Diego, California.

Schmidt, K.L. and Cohn, J.F. (In press). Human facial expressions as adaptations: Evolutionary questions in facial expression. *Yearbook of Physical Anthropology*.

Schmidt, K. & Cohn, J.F. (August 2001). Dynamics of facial expression: Normative characteristics and individual differences. *IEEE International Conference on Multimedia and Expo, ICME2001*, pp. 728-731. Tokyo, Japan.

Tian, Y., Kanade, T., and Cohn, J.F. (October 2000). Eye-state detection by local regional information. *Proceedings of the International Conference on Multimedia Interfaces*, pp. xxx-xxx. Beijing, China.

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.

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.

VanSwearingen, J.M., Cohn, J.F., Turnbull, J., Mrzai, & Johnson, P. (1998). Psychological distress, impairment, and disability in facial neuro-motor disorders. *Otolaryngology, Head and Neck Surgery, 118*, 790-796.

Yacoob, Y. & Davis, L. (1994). Computing spatio-temporal representations of human faces. In *Proceedings in Computer Vision and Pattern Recognition, CVPR-94*, 70-75.

Xiao, J., Kanade, T., & Cohn, J.F. (Submitted). A real-time system of 3D head motion recovery. *IEEE International Conference on Automated Face and Gesture Recognition*.