Use of Automated Facial Image Analysis for Measurement of Emotion Expression

Jeffrey F. Cohn
Department of Psychology
University of Pittsburgh

Takeo Kanade
Robotics Institute
Carnegie Mellon University

Facial expressions are a key index of emotion.   They have consistent correlation with self-reported emotion (Keltner, 1995; Rosenberg & Ekman, 1994; Ekman & Rosenberg, in press) and emotion-related central and peripheral physiology (Davidson, Ekman, Saron, Senulis, & Friesen, 1990; Fox & Davidson, 1988; Levenson, Ekman, & Friesen, 1990).   They putatively share similar underlying dimensions with self-reported emotion (e.g., positive and negative affect) (Bullock & Russell, 1984; Gross & John, 1997; Watson & Tellegen, 1985).   Facial expressions serve interpersonal functions of emotion by conveying communicative intent, signaling affective information in social referencing (Campos, Bertenthal, & Kermoian, 1992), and more generally contributing to the regulation of social interaction (Cohn & Elmore, 1988; Fridlund, 1994; Schmidt & Cohn, 2001).   As a measure of trait affect, stability in facial expression emerges early in life (Cohn & Campbell, 1992; Malatesta, Culver, Tesman, & Shephard, 1989). By adulthood, stability is moderately strong, comparable to that for self-reported emotion (Cohn, Schmidt, Gross, & Ekman, 2002), and predictive of favorable outcomes in emotion-related domains including marriage and personal well-being over periods as long as 30 years (Harker & Keltner, 2001).   Expressive changes in the face are a rich source of cues about intra- and interpersonal functions of emotion (cf. Keltner & Haitd, 1999).

To make use of the information afforded by facial expression for emotion science and clinical practice, reliable, valid, and efficient methods of measurement are critical. Until recently, selecting a measurement method meant choosing among one or another human-observer-based coding system (e.g., Ekman & Friesen, 1978 and Izard, 1983) or facial electromyography (EMG). While each of these approaches has advantages, they are not without costs. Human-observer-based methods are time consuming to learn and use, and they are difficult to standardize, especially across laboratories and over time (Bakeman & Gottman, 1986; Martin & Bateson, 1986). Facial EMG requires placement of sensors on the face, which may inhibit facial action and which rules out its use for naturalistic observation.

An emerging alternative to these methods is automated facial image analysis using computer vision. Computer vision is the science of extracting and representing meaningful information from digitized video and recognizing perceptually meaningful patterns.   An early focus in automated face image analysis by computer vision was face recognition (Kanade, 1973,

1977).   That area has sufficiently advanced that commercially viable applications have become available (Phillips, Grother, Micheals, Blackburn, Tabassi, & Bone, 2003).   Computer vision research in facial image processing has turned increasingly toward automated facial expression recognition.   In 1992, the National Science Foundation convened a seminal interdisciplinary workshop on this topic (Ekman, Huang, Sejnowski, & Hager, 1992), which brought together psychologists with expertise in facial expression and computer vision scientists with interest in facial image analysis. Since then, there has been considerable research activity, as represented by a series of six international meetings beginning in 1995 (http://image.korea.ac.kr/FG2004).

Several automated facial image analysis systems have been developed (Cootes, Edwards, & Taylor, 2001; Essa & Pentland, 1997; Lyons, Akamasku, Kamachi, & Gyoba, 1998; Padgett, Cottrell, & Adolphs, 1996; Wen & Huang, 2003; Yacoob & Davis, 1996; Zhang, 1999; Zhu, De Silva, & Ko, 2002).   They can classify a small set of emotion-specified expressions, such as joy and anger. Others (Bartlett, Hager, Ekman, & Sejnowski, 1999; Fasel & Luttin, 2000; Cohn, Zlochower, Lien, & Kanade, 1999; Pantic & Rothkrantz, 2000a; Tian, Kanade, & Cohn, 2001) have achieved some success in the more difficult task of recognizing facial action units of the Facial Action Coding System (FACS: Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002). Actions units (AU) are the smallest visibly discriminable changes in facial expression. Comprehensive reviews of the literature in automated facial expression analysis can be found in Pantic and Rothkrantz (2000b, 2003) and in Tian, Kanade, and Cohn (in press).   While many basic research issues remain (Bartlett, Movellan, Littlewort, Braathen, Frank, & Sejnowski, in press; Kanade, Cohn, & Tian, 2000; Matthews, Ishikawa, & Baker, 2004; Pantic & Rothkrantz, 2003; Smith, Bartlett, & Movellan, 2001; Tian, Kanade, & Cohn, in press), applications of automated facial image analysis to emotion science have begun (e.g., Schmidt, Cohn, & Tian, 2003), with broader adoption likely to follow as methods continue to evolve.

In this chapter, we present work, development, and progress of the CMU/Pitt Automated Facial Image Analysis (AFA) System, a leading approach to automatic recognition of facial action units and quantitative analysis of their timing. We describe how we have used it to assess emotion processes and discuss prospects for its broader use in emotion science and clinical practice.   AFA has progressed through 3 versions – I, II, and III.   In the remainder of this chapter, we distinguish between them when referring to features that are specific to one or the other version.

Automatic Facial Expression Analysis (AFA)

Figure 1 depicts the overall structure of the AFA system for recognition of facial action units and analysis of their dynamics. A digitized image sequence is input to the system. The region of the face and location of individual face features are delineated in the initial frame, either manually using a computer mouse or other pointing device or automatically using a module for head and feature detection. Head motion is recovered automatically and used to warp (or stabilize) the face image to a standard (i.e., canonical) view. Changes in both permanent (e.g., brows, eyes, lips) and transient (lines and furrows) facial features are automatically detected and tracked throughout the image sequence. Informed by FACS, we group the facial features into separate collections of feature parameters. Facial actions in the upper and lower face are relatively independent (Ekman & Friesen, 1978).   Parameters describe shape, motion, eye state,

lip state, motion of brow and cheek, and presence/absence and change in appearance of furrows and wrinkles. The extracted facial feature parameters are fed to two neural network-based classifiers. One is for upper face action units, the other for lower face action units. In addition to action unit recognition, the parameters quantify the timing of facial actions and head motion for studies of the timing of facial actions.

Insert Figure 1 about here

Face detection and facial feature localization

To locate the face and obtain the positions of facial features, AFA-I used hand initialization in the first video frame.   AFA-II used a combination of automatic face detector (Rowley, Baluja, & Kanade, 1998) and manual adjustment.   Recently, AFA-III uses more automatic approaches: one is a method developed by Zhou, Gu, and Zhang (2003) and the other by Matthews and Baker (2005).   The method of Zhou et al. is limited to mostly frontal images (See Figure 2); otherwise, manual adjustment remains necessary. The Matthews and Baker method performs well for moderate out-of-plane head rotation, but requires more extensive algorithm training.

Insert Figure 2 about here

Automatic recovery of 3-D head motion and image stabilization

Expressive changes in the face often co-occur with head movement.   People raise their head in surprise (Camras, Lambrecht, & Michel, 1996) and turn toward a friend while beginning to smile (Kraut & Johnson, 1979).   In a video sequence, both types of motion are likely to be present.   The effects of rigid (head) motion must be measured and removed prior to extracting information about non-rigid motion (expression) so that these two types of motion are not confounded.

AFA-III uses a cylindrical head model to estimate the 6 degrees of freedom 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 template of head appearance. For any given subsequent frame, the template is projected onto the image plane assuming the pose has remained unchanged from the previous frame.   We then compute the difference between the projected image and the current frame, and the difference provides the correction on the estimate of pose.   We iterate this process to further refine the estimate by using a model-based optical-flow algorithm. As new parts of the head become visible, their appearance is added to the cylinder surface for a more complete template of the head appearance (Xiao, Moriyama, Kanade, & Cohn, 2003).

Insert Figure 3 about Here

An example of system output is shown in Figure 3.   The image data are of spontaneous facial behavior from Frank and Ekman (1997). From the input image sequence (Figure 3A), the

head is tracked and its pose recovered (Figure 3B). The system stabilizes the face region by transforming the image to a common orientation (Figure 3C) and then localizes a region of interest. Figure 3D shows the localized eye region.   Note that even though the head pitches forward in the image sequence, size and orientation of the "stabilized" eye region remain the same.

We have tested the head tracker 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, Moriyama, Kanade, & Cohn, 2003). We compared the recovered motion with ground truth obtained by a position and orientation measurement device that used markers attached to the head (Optotrak$^®$ 3020 Position Sensor). The AFA head tracker was highly consistent with ground truth measurements; for example, for 75° yaw, absolute error was 3.86^o (Xiao et al., 2003).

While a head shape is not actually a cylinder, a cylinder model is adequate for many facial actions and contributes to system stability and robustness. A cylinder model, however, does not take into account the depth variation on the face surface.   This is a problem for recognizing some facial action units such as lip pursing (AU 18). An alternative is to use an anatomically-based complete face model in which the exact proportions of facial features are represented (De Carlo & Mataxas, 1996; Essa & Pentland, 1997).   While powerful, such a person-specific anatomic model requires a large number of parameters that are dependent on the exact shape of the subject's individual face, which typically is unknown.   Until recently, therefore, short of laser-scanning individual faces (Wen, 2004) or making anthropometric measurements in advance of facial image analysis, use of anatomically based 3D face models was not feasible. A recently developed algorithm that is capable of extracting 3D shape and appearance parameters from a single video (Xiao et al., 2004) may change the situation. The subject specific shape information of the face can be obtained from the input video data to be analyzed itself (See Figure 4).

Insert Figure 4 about here

Feature extraction and representation

Contraction of the facial muscles produces changes in the appearance and shape of facial landmarks, such as the eyes and lips, and in the direction and magnitude of the motion on the skin surface resulting in the appearance of transient facial features. Transient features include facial lines and furrows that are not present at rest but appear with facial expressions.   Some of the transient facial features, such as crows-feet wrinkles, may become permanent with age.

Permanent facial features

To track permanent facial features, AFA uses several different approaches. These include optical flow, Gabor wavelets, multi-state models, and generative model fitting. Using multiple approaches increases the accuracy of action unit recognition (Tian, Kanade, & Cohn, 2001, 2002).

i.   Optical flow.   In FACS, each action unit is anatomically related to contraction of a specific facial muscle.   AU 12 (oblique raising of the lip corners), for instance, results from contraction of the *Zygomatic major* muscle, AU 20 (lip stretch) from the *Risorius* muscle, and AU 15

(oblique lowering of the lip corners) from the *Depressor anguli* muscle (See Appendix 1). Muscle contractions produce movement in the overlaying tissue.   Optical flow quantifies the magnitude and direction of such movement.   Figure 5 shows an example of dense flow extracted over the entire face. As the jaw drops (AU 26/27), the eyes widen (AU 5), and the brows are raised (AU 1+2), the flow captures these facial actions.

Insert Figures 5 and 6 about here

Obtaining smooth dense flow for the whole face image reliably requires incorporating a global model of motion (Wu, Kanade, Li, & Cohn, 2000), which is computationally intensive. It is more efficient to compute feature motion for localized facial regions.   Tracking specific "feature points" (Figure 6) in these regions yields motion that is still consistent with that obtained from dense flow.   Lien, Kanade, Cohn, & Li (2000) found that the two approaches to optical flow computation achieved similar high accuracy for action unit recognition.

Insert Figures 7 and 8 about here

ii. Gabor wavelets.   Gabor wavelets are filters of varying orientations (e.g., vertical, oblique, or horizontal image gradients) and resolution. An example of Gabor filters is shown in Figure 7. Various orientations of the filter are shown across rows, while resolutions are shown across columns; each image in Figure 7 is referred to as a Gabor kernel. Gabor coefficients for a given image are the correlation images between the image and a set of these Gabor kernels. Figure 8 shows an example of Gabor coefficients in the eye region. Note that the output of the Gabor filter differs for each eye state. It was found that Gabor coefficients in the eye region could discriminate between three action units (AU 41, AU 42, and AU 45) with accuracy comparable to that of manual FACS coding (Figure 8) (Tian, Kanade, and Cohn, 2002).

Insert Figures 7 & 8 about here

iii. Multi-state models. Facial features such as the mouth can exhibit both quantitative and qualitative change in appearance. An example of quantitative change is the amount of displacement of the lip corner as smile intensity increases. Optical flow works well for this type of change. Qualitative change in appearance is disappearance of features and appearance of totally new features such as the one that occurs when the lips tightly compress. Detecting this type of change is not easy with a technique like optical flow, as it involves more than detecting movement of features.   Multistate models of facial components address these issues.

Insert Figure 9 about here

Figure 9 shows an example of a three-state model of the lips. The model represents open, closed, and tightly closed lips. Different lip contour templates are prepared for different lip states. The open and closed lip contours are modeled by two parabolic arcs, which are described by six parameters: the lip center position (xc, yc), the lip shape (h1, h2, and w), and the lip orientation.   For tightly closed lips, the dark mouth line connecting the lip corners represents the position, orientation, and shape.

   iv. Generative model fitting approach for eye state analysis. AFA-II used an eye model similar to the lip model in which parabolic curves represent contours. The appearance of the eye region, however, is more complex than such a simple model. As Figure 10 illustrates, appearance of the eye varies within individuals (e.g., eye state, illumination, and orientation) and across individuals (e.g., race and gender).   Asiatic and European faces, for example, differ in having single or double upper eyelids, respectively.   To represent such variation, a more sophisticated model is needed. Figure 11a shows the structure generative eye model AFA-III uses. Structural individuality is represented by size and color of the iris, width and boldness of the eyelid, width of the bulge below the eye, and width of the proximal illumination reflection on the bulge and furrow.   Motion is represented by up-down positions of the upper and lower eyelids and 2D position of the iris.   By matching this model with the eye region of an input image by means of an extended Lucas-Kanade algorithm and other techniques, we obtain detailed measurement of eye region appearance and eye motion (Moriyama, Xiao, Cohn, & Kanade, In press) (Figure 11b).

Insert Figure 10 and 11a and 11b about here

Transient facial features

        Transient features provide crucial information for recognition of certain AUs. Wrinkles and furrows appear perpendicular to the direction of the motion of the activated muscles. Contraction of the *corrugator* muscle, for instance, produces vertical furrows between the brows, which is coded as AU 4 in FACS. Contraction of the medial portion of the *frontalis* muscle causes horizontal wrinkling in the center of the forehead (AU 1).   Some of these transient features may become permanent with age. Permanent crow's-feet wrinkles around the outside corners of the eyes, which are characteristic of AU 6, are common in adults but not in children. When wrinkles and furrows become permanent, contraction of the corresponding muscles accentuates their appearance, such as deepening or lengthening.

Insert Figure 12 about here

        AFA detects wrinkles and furrows in the nasolabial region, the nasal root, and the areas lateral to the outer corners of the eyes (Figure 12). These areas are located using the tracked locations of the corresponding permanent features. Presence or absence of wrinkles and furrows in these regions is determined by the strength and orientation of edge-like features using Gabor wavelet or edge detection technique.   The wrinkle/furrow state is classified as present if edge features increase from the neutral frame. For nasolabial furrows, the existence of vertical to diagonal connected edges is used for classification. If the length of connected edge pixels is longer than a threshold, the nasolabial furrow is determined to be present and is modeled as a line. The orientation of the furrow is represented as the angle between the furrow line and a line connecting the medial canthi (inner eye corners). This angle determines different action units. For example, the nasolabial furrow angle of AU 9 or AU 10 is larger than that of AU 12.

Facial feature representation and action unit recognition by pattern recognition

        Extracted features are transformed into a set of parameters for AU recognition.   Upper and lower facial features are divided into two groups of parameters: the upper and lower faces. With

a few exceptions (e.g., AU 9 effects on brow motion), facial actions in the upper and lower face have little interaction with each other. All parameters are either normalized for variation in face orientation and size (AFA-I and AFA-II) or computed from stabilized face images (AFA-III). Facial feature parameters are unaffected by variation in head position, rotation, and scale. Figure 13 shows the coordinate scheme and parameters used in AFA-II for the upper and lower face. We define a face coordinate system in AFA-II by using the inner corners of the eyes: the x-axis as the line connecting the two inner corners of the eyes and the y-axis as perpendicular to it pointing upward.   The positions of the two inner corners of the eyes are least affected by facial muscle contraction and can be most reliably detected.

Insert Figure 13 about here

We have experimented with various approaches of classifying feature parameters into action units. These include hidden Markov models (HMM) (Lien et al., 2000), discriminant analysis (Cohn et al., 1999), rule-based recognition (Moriyama et al., 2002; Cohn et al., 2003), and neural networks (Tian, Kanade, & Cohn, 2001, 2002).

HMM encodes extracted features into a sequence of a set of symbols. Sequences of symbols representing target action units and action unit combinations are modeled separately. These HMM models represent the most likely action units and action unit combinations and are used to evaluate encoded feature data for automatic action unit recognition (Lien et al., 2000). Discriminant analysis computes dimensions along which phenomena differ and obtains classification functions that predict class membership. Discrimination among action units is done by computing and comparing the *a posteriori probabilities* of action units (Lien et al., 2000). Neural networks can learn nonlinear as well as linear discriminants.   A single network models multiple action units.   We have used neural networks with three layers (with one layer hidden), and have used a standard back-propagation method for training.   When action units occur in combination, multiple output nodes are excited (Tian et al., 2001). When we have compared the classification results by various classifiers, we have found they perform similarly (Lien et al., 2000).   More important than the choice of classifier is that the selected features be measured precisely and have high specificity for the target action units.

Use of AFA in Studies of Facial Expression of Emotion

It is known that both the configuration of facial features and the timing of facial actions are important in emotion expression and recognition (Cohn, in press).   The configuration of facial action units in relation to emotion, communicative intent, and action tendencies has been a major research topic. Less is known about the timing of facial actions because with manual methods timing measurement is only coarse and time consuming. We know, however, that people are highly sensitive to the timing of facial actions (Edwards, 1998) in social setting. Slower facial actions, for instance, appear more genuine (Krumhuber & Kappas, 2003; Schmidt, Ambadar, & Cohn, submitted), as do those that are more synchronous in their movement (Frank & Ekman, 1997).   Facial electromyography (EMG) effectively quantifies the timing of covert muscle action (e.g., Dimberg & Thunberg, 1998), but the timing of observable facial action had been measured only coarsely by manual coding. AFA makes possible quantitative measurement of the timing of observable facial actions.

We have used AFA to recognize action units, make comparisons with criterion measures of facial dynamics, and investigate the timing of spontaneous smiles, multi-modal coordination, and infant expressions of joy and distress.   The first two versions of AFA (AFA-I and AFA-II) assume that head motion is mostly parallel to the image plane of the camera and is minimal in out-of-plane motion.   We therefore applied AFA-I and AFA-II to directed facial action tasks, in which subjects are asked to perform deliberate facial actions with small head movement (Kanade, Cohn, & Tian, 2000) and to spontaneous facial behavior in which out-of-plane motion was small (Cohn & Schmidt, in press; Schmidt, Cohn, & Tian, 2003).   Spontaneous facial behavior, however, often includes moderate to large out-of-plane head motion, for which AFA-I and AFA-II are not appropriate.   A major breakthrough in AFA-III was the capability to accommodate such motion, which enabled us to expand our research in spontaneous facial behavior.

Automatic recognition of FACS action units

Motivated by our interest in emotion expression and social interaction, we have focused on the action units that are most common in these contexts (e.g., Sayette et al., 2001).   In directed facial action tasks, AFA has shown high agreement with manual FACS coding for approximately 20 action units (AFA-I: Cohn et al., 1999, Lien et al., 2000; AFA-II: Tian et al., 2001, 2002; AFA-III: Cohn et al., 2003, Moriyama et al., 2002).   In the upper face, AFA recognizes AU 1, AU 2, AU 4, AU 5, AU 6, AU 7, AU 41, AU 42, AU 43/45, and neutral (AU 0). In the lower face, it recognizes AU 9, AU 10, AU 12, AU 15, AU 17, AU 20, AU 25, AU 26, AU 27, AU 23/ 24, and neutral. (See Appendix 1 for definitions of action units).   These action units include most of those that have been a focus in the literature on facial expression and emotion (Ekman & Rosenberg, in press).

We made extensive comparisons to evaluate the AFA system's ability to generalize to new subjects by training and testing in independent data sets collected and FACS coded in different laboratories.   Average recognition accuracy exceeded 93% regardless of what data set was used for training or testing.   Accuracy was high for all action units with the exception of AU 26 (jaw drop). This AU 26 action unit is one that manual FACS coders have found troubling, as well. The recently revised FACS manual (Ekman et al., 2002) addressed this difficulty by altering the criteria for AU 25 and AU 26.

Action units can occur singly or in combination (Kanade, Cohn, & Tian, 2000; Smith et al., 2001).   Recognizing action units when they occur in combination is difficult because action units may modify each other's appearance when proximal to each other, analogous to co-articulation effects in speech.   Recognizing an individual action unit even when it appears in combination is important because there are thousands of possible combinations. Had each combination be recognized separately, the task of training would become impractical. The AFA is capable of recognizing action units AU 1, AU 2, AU 4, AU 5, AU 6, AU 7, AU 9, AU 10, AU 15, AU 17, AU 20, AU 25, AU 26, AU 27, and AU 23/24 whether they occur alone or in combination (Tian et al., 2001).   In spontaneous facial behavior, AFA-III was tested for recognition of action unit 45 (blink) and flutter. Flutter is defined as multiple partial blinks in rapid succession.   Image data used were those from a study of deception by Frank and Ekman (1997). Ethnically diverse young men, some of whom wore glasses, were video recorded while

telling the truth or lying in a high-stakes situation. The video contained moderate head motion. The system achieved 98% agreement with manual FACS coding of blink and flutter (Cohn et al., 2003).

The FACS manual labels head orientation and gaze as "action descriptors" rather than "action units."   Codes for action descriptors are late entries to the FACS manual, and unlike action units they lack the thorough description and differentiation. For this reason, we have not compared the AFA to FACS for action descriptors.   We have instead compared the AFA results with motion-capture devices, which produce precise quantitative measurement of head motion and are considered the gold standard.   AFA-III demonstrated high concurrent validity with motion capture device for pitch and yaw as large as 40^0 and 75^0, respectively.   Average recovery accuracy was within 3 ° (Xiao et al., 2003). Preliminary work with the eye-state analyzer in AFA-III indicates similar high concurrent validity for gaze (Moriyama, Xiao, Cohn, & Kanade, In press).   Together, these findings suggest that AFA produces far better measures of head motion and at least comparable measure of gaze to that of manual FACS coding.

Comparison with criterion measures of facial dynamics

We evaluated the temporal precision of AFA by comparing it with manual feature tracking in digitized video and with facial EMG.   AFA was highly consistent with both.

Wachtman ,Cohn, VanSwearingen, and Manders (2001) compared facial feature tracking by AFA and manual feature tracking in directed facial action tasks in digitized video of individuals with facial neuromuscular disorder. The two methods were found highly consistent, with Pearson's $r = .96$ or higher, $p<.001$ for each of the facial actions.   Differences between the methods were small on the order of less than 1 pixel on average and comparable to the inter-observer reliability of the manual method.

Another useful comparison is between AFA results and facial EMG.   Facial EMG is a gold standard for measurement of facial muscle activity.   AFA output and *Zygomaticus major* EMG were compared for lip corner motion (AU 12) in Cohn and Schmidt (in press).   Lip corner motion was quantified by the total displacement, $\Delta d = \sqrt{\Delta x^2 + \Delta y^2}$.   These two methods were in agreement for lip corner motion in 72% of cases with distinct EMG onset. Because EMG can detect occult changes in muscle activation below the threshold of visible change, this percentage agreement represents a conservative comparison of AFA's sensitivity to AU 12. In smiles the two methods were highly correlated ($r = 0.95$, $p<.01$). Visible onset occurred an average of .23 seconds after the EMG onset (See Figure 14 for an example).   This kind of relation between physiological measurement (i.e., EMG) and visible behavior (lip motion) at this level of precision became possible only by AFA.

Insert Figure 14 about here

Timing of spontaneous smiles

Smiles, as one of the most important facial expressions, emerge early in development and occur throughout the lifespan with high frequency to express emotion and communicative

intention.   While the configuration of smiles is well studied (e.g., Ekman, 1993; Frijda & Tcherkassof, 1997; Fridlund, 1994; Izard, 1983; Malatesta et al., 1989; Matias, Cohn, & Ross, 1989; Pantic & Rothkrantz, 2000b, 2003; Tian et al., in press), with few exceptions little is about their timing (e.g., Frank, Ekman, & Friesen, 1993; Hess & Kleck, 1990).   We used AFA to investigate the timing of the onset phase of spontaneous smiles.   The onset phase provides the initial and most conspicuous change in appearance in smiling as perceived by human observers (Leonard, Voeller, & Kuldau, 1991).   Viewers respond in kind either overtly or covertly as early as 0.30-0.40 sec after viewing an image of a smile (Dimberg & Thunberg, 1998). Because this duration is well within the average duration of smile onsets (Bugental, 1986; Cohn & Schmidt, in press), it is likely that this phase of smiles functions as the initial social signal.

We found the onset phase of spontaneous smiles has highly consistent temporal characteristics regardless of context and the occurrence of other action units, including AU 6 and masking movements. The larger the intensity of the onset phase, the faster is the peak velocity, with an average $R^2 = 0.82$.   This finding that intensity and velocity of smile onsets have a strong relationship is consistent with ballistic motion. Previous attempts to examine this issue were limited to relatively gross measures, such as the duration of manually coded action units (Frank et al., 1993).   The fact that AFA produces quantitative measures of rate of change – velocity in particular – allowed for more rigorous kinematic analyses to test hypotheses about the timing of spontaneous smiles.

Multimodal coordination of facial action, head motion, and gaze

We investigated coordination among head motion, facial action, and gaze that occurs in spontaneous smiles. We focused on spontaneous smiles that occurred following directed facial action tasks. Keltner (1995) found that smiles in this context were frequently associated with embarrassment. Following Keltner, we hypothesized a pattern of coordination of head motion, gaze, and facial expression; smiles associated with embarrassment involve motion of looking down and away while beginning to smile. We found strong support for this hypothesis.   Facial action, as indicated by lip-corner displacement during spontaneous smiles, was moderately correlated with all 6 *df* of head motion and with eye motion, as suggested by neuroscience literature (King, Lisberger, & Fuchs, 1976; Klier, Hongying, & Crawford, 2003).   Further, the patterns of correlation we found appeared to be specific to embarrassment and part of a coordinated motor routine (Michel & Camras, 1992). Smile intensity increases as the face and gaze pitch down and move away from the experimenter; followed by decreasing intensity as the orientation of the face comes back toward the experimenter (See figure 15). (For details, see Cohn, Reed, Moriyama, Xiao, Schmidt, & Ambadar, 2004).   As we did not collect self-report measures, we cannot say with certainty that the smiles we observed were related to feelings of embarrassment or relief at the task's completion. The findings, however, provide strong quantitative support for existence of dynamic coordination of multi-modal actions, and suggest that the detection of such a coordination can disambiguate smiles that are otherwise morphologically similar (e.g., smiles of embarrassment versus those of enjoyment).

Insert Figure 15 about here

Infant expressions of joy and distress

Previous literature proposes that cheek raising (AU 6) increases observers' perceptions of smile intensity in infants (Messinger, Fogel, & Dickson, 1999, 2001). This hypothesis has been difficult to test in perceptual judgment studies because infant head orientation typically is confounded with smile and distress intensity, and because manual FACS coding of intensity is relatively coarse.   The head tracking and face stabilization features of AFA allow us to overcome these difficulties.   After recovering 3D head motion, AFA-III warps the face images to a common orientation and precisely measures smile intensity as the lip corner displacement (described above).   We (Bolzani-Dinehart, Messinger, Acosta, Cassel, Ambadar, & Cohn, 2003) then used the measurement to create experimental stimuli for use in a judgment study. We found that smiles with mouth opening, cheek raising, and greater lip corner displacement were perceived by raters as more emotionally positive than equivalent smiles without these features. In related work, we have begun to use AFA to track changes in facial expression of infants during mother-infant face-to-face interaction. An example of facial feature tracking is shown in Figure 16.   The infant's head motion and facial expression are quantitatively measured in a way manual coding could only approximate.   Precise measurement of infant and parent behavior during face-to-face interaction enables us to more rigorously test parent-infant bidirectional influence than possible previously (e.g., Cohn & Tronick, 1988) and provides new capability to investigate the dynamic processes in emotion and emotion regulation.

Insert Figure 16 about here

Discussion

Automated facial image analysis exemplified by AFA is an emergent option for assessing facial expression of emotion.   AFA has shown good agreement in action unit recognition with manual FACS coding in deliberate facial action tasks, and in head motion with ground-truth measures in the more challenging case of spontaneous facial behavior.   Automated face image analysis has proven especially effective in revealing the dynamics of facial action, head motion, and gaze. It affords quantitative power similar to EMG, and yet is specific to observable facial actions and quantifies head motion and eye position as well.   The study of emotion dynamics in facial behavior is an especially exciting domain since until now it could only be studied in a coarse way other than using facial EMG sensors.

In addition to dynamic aspects of emotion expression, a major application of automated facial image analysis will be recognition of FACS action units and emotion-specified expressions.   Initial efforts with AFA have been encouraging.   Nearly all of the action units prevalent in emotion expression are recognized by AFA, and similar results also were reported by several other facial image analysis systems.   An important qualification, however, is that high level of performance has been demonstrated only in deliberate facial actions.   Automated action unit recognition in spontaneous facial behavior is more difficult and needs more research before it becomes broadly useful. Toward that end, AFA has made a small step by demonstrating recognition of a few action units in spontaneous behavior.   In video of spontaneous facial behavior from a study of deception by Frank and Ekman (1997), AFA achieved 98% agreement with manual FACS coding for blinks (AU 45) and flutter (Cohn et al., 2003). In another study

(Cohn & Schmidt, in press) , the system demonstrated strong concurrent validity with facial EMG for continuous measurement of *Zygomatic major* intensity, which is the primary measure of positive affect in facial EMG studies (Cacioppo Martzke, Petty, & Tassinary, 1988).

To accomplish this goal will require not only algorithm development, but also use of rigorously FACS coded image data for training and testing algorithms.   For deliberate facial action tasks, we created a large representative database, the Cohn-Kanade AU-Coded Facial Expression Database (Kanade et al., 2000). The database consists of FACS coded directed facial action tasks in over 200 adult men and women of varying ethnicity. The database has been widely distributed for research in automated facial image analysis and is serving as a testbed and benchmark for algorithm development and testing.   Comparable FACS coded data sets of spontaneous facial behavior will be required for fast progress. The emotion science community can be of invaluable help in this regard by making available to researchers in this area facial expression image data with associated manual codes.

A number of technical challenges exist for AFA.   Among these, the most important are how to parse the stream of behavior, prevent error accumulation, and increase automation.   AFA and other facial image analysis approaches have assumed that expressions involve a single facial action or expression, and that they begin and end from a neutral position.   In actuality, facial expression is more complex.   Action units occur in combinations or show serial dependence. Transitions among action units may involve no intervening neutral state. Parsing the stream of facial action units under these circumstances is a challenge.   Human coders meet this task in part by having a mental representation of a neutral face. However, even for human coders, defining events and transitions is not a solved problem. For automated facial image analysis, parsing will likely involve higher order pattern recognition than has been considered to date.

Many of the methods used in automated facial image analysis so far involve dynamic templates for which estimates are continually updated.   With dynamic templates, error tends to propagate and accumulate across an image sequence. So far, most of AFA applications have involved relatively short image sequences up to 10 seconds or so, for which error accumulation was not a significant problem. As we begin to process much longer sequences, an appropriate measure is required.   The head tracking module in AFA overcomes this problem through a combined use of robust regression and reference images.   Robust regression identifies and discounts the effects of outliers, and reference images provide a way to reinitialize estimates so as to attenuate error accumulation. For head tracking, this approach has been highly successful. The cylinder model head tracker has performed well for image sequences as long as 20 minutes. Similar capability will be needed for action unit recognition.

Current methods involve some degree of initialization, such as delimiting face regions to process, adjusting templates of facial features, or personalizing active appearance models.   For example, current active appearance models require a fair amount of manual input during the training phase.   While a fully automated system is not always necessary for all applications, increased automation will accelerate the adoption of AFA in emotion science and clinical practice.

In summary, automated facial image analysis for measurement of facial expression is advanced, its application to the study of emotion has started to inform our understanding of emotion processes, and new types of findings, such as the timing of multi-modal behavior in spontaneous smiles, have begun to emerge.

Author Notes

References

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

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., Movellan, J.R., Littlewort, G.C., Braathen, B., Frank, M.G., & Sejnowski, T.J. (in press). Towards automatic recognition of spontaneous facial actions. Afterword by J.R. Movellan and M.S. Bartlett: The next generation of automatic facial expression measurement. In P. Ekman (Ed.), *What the Face Reveals, 2nd Edition*, NY, NY: Oxford University Press.

Bolzani-Dinehart, L., Messinger, D. S., & Acosta, S., Cassel, T., Ambadar, Z. & Cohn, J.F. (2003, April). A dimensional approach to infant facial expressions. *Society for Research in Child Development.* Tampa, Florida.

Bullock, M., & Russell, J.A. (1984). Preschool children's interpretation of facial expressions of emotion. *International Journal of Behavioral Development, 7*, 193-214.

Bugental, D. (1986). Unmasking the "polite smile": situational and personal determinants of managed affect in adult-child interaction. *Personality and Social Psychology Bulletin, 12(1)*, 7-16.

Cacioppo, J. T., Martzke, J. S., Petty, R. E., & Tassinary, L. G. 1988. Specific forms of facial EMG response index emotions during an interview: From Darwin to the continuous flow hypothesis of affect laden information processing. *Journal of Personality and Social Psychology*, *54*, 592–604.

Campos, J.J., Bertenthal, B.I.., & Kermoian, R. (1992). Early experience and emotional development: The emergence of wariness of heights. *Psychological Science, 3*, 61-64.

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

Cohn, J. F. (in press). Automated analysis of the configuration and timing of facial expression. In P. Ekman & E. Rosenberg, *What the face reveals* (2nd edition): *Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press Series in Affective Science.* New York: Oxford.

Cohn, J. F. and Campbell, S. B. (1992). Influence of maternal depression on infant affect regulation. In D. Cicchetti and S. Toth (Eds.), *Rochester Symposium on Developmental*

Psychopathology, A developmental approach to affective disorders (Vol. 4, pp. 105-130). Hillsdale, NJ: Erlbaum.

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., Reed, L., Moriyama, T., Xiao, J., Schmidt, K., & Ambadar, Z. (2004). Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG'04)*, Seoul, Korea, xxx-xxx.

Cohn, J. F. & Schmidt, K. L. (2004, In press). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing, 2*, 1-12.

Cohn, J. F., Schmidt, K., Gross, R., & Ekman, P. (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 (ICMI 2002)*, Pittsburgh, PA, 491-496.

Cohn, J. F. & Tronick, E. Z. (1988). Mother-infant interaction: Influence is bidirectional and unrelated to periodic cycles in either partner's behavior. *Developmental Psychology*, *24*, 386-392.

Cohn, J. F., Xiao, J., Moriyama, T., Ambadar, Z., & Kanade, T. (2003). Automatic recognition of eye blinking in spontaneously occurring behavior. *Behavior Research Methods, Instruments, and Computers, 35,* 420-428.

Cohn, J. F., Zlochower, A., Lien, J. J., Hua, W., & Kanade, T. (2000). Automated face analysis. In C. Rovee-Collier & L. Lipsitt (Eds.), *Progress in infancy research*, *1*, 155-182. Hillsdale, NJ: Erlbaum.

Cootes, T.F., Edwards, G.J., & Taylor, C.J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 681-685.

Davidson, R. J., Ekman, P., Saron, C.D., Senulis, J.A., & Friesen, W. (1990). Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I . *Journal of Personality & Social Psychology. 58,* 330-341.

DeCarlo, 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*, San Fancisco, CA, 231-238.

Dimberg, U., Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scandinavian Journal of Psychology, 39*, 39-45.

Edwards, K. (1998). The face of time: Temporal cues in facial expressions of emotion. *Psychological Science, 9(4)*, 270-276.

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

Ekman, P. Friesen, W., & Hager, J. (2002). *Facial Action Coding System*. Salt Lake City, Utah: Research Nexus.

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

Ekman, P., Huang, T.S., Sejnowski, T.J., & Hager, J.C. (July 30 to August 1, 1992). *Final report to NSF of the planning workshop on facial expression understanding,* Washington, DC: NSF.

Ekman, P. & Rosenberg, E. (In press). *What the face reveals: Basic and applied studies of spontaneous facial expression using the Facial Action Coding System (FACS). 2ⁿᵈ Edition.* NY, NY: Oxford.

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

Fasel, B. & Luttin, J. Recognition of asymmetric facial action unit activities and intensities. *Proceedings of the International Conference on Pattern Recognition (ICPR 2000),* Barcelona, Spain.

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. (1993). Not all smiles are created equal: The differences between enjoyment and non-enjoyment smiles. *Humor: International Journal of Humor Research, 6,* 9-26.

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 and Social Psychology, 64,* 83-93.

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

Frijda, N.H. & Tcherkassof, A. (1997). Facial expressions as modes of action readiness. In Russell & & Fernandez-Dols, *The psychology of facial expression,* pp. 78-102.

Gross, J.J., & John, O.P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology, 72,* 435-448.

Harker, L.A. & Keltner, D. (2001). Expressions of positive emotions in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality and Social Psychology, 80,* 112-124.

Hess, U. & Kleck, R. (1990). Differentiating emotion elicited and deliberate emotional expressional facial expressions. *European Journal of Social Psychology, 20,* 369-385.

Izard, C.E. (1983). *The Maximally Discriminative Facial Movement Coding System.* Unpublished Manuscript, University of Delaware.

Kanade, T. (1973). *Picture processing system by computer complex and recognition of human faces.*   Doctoral dissertation, Kyoto University.

Kanade, T. (1977). *Computer recognition of human faces.* Stuttgart and Busel: Birkhauser Verlag.

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

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality & Social Psychology, 68,* 441-454.

Keltner, D. & Haidt, J. (1999). Social functions of emotions at four levels of analysis. Cognition and Emotion, 13, 505-521.

King, W.M., Lisberger, S.G., & Fuchs, A.F. (1976). Response of fibers in medial longitudinal fasciculus (MLF) of alert monkeys during horizontal and vertical conjugate eye movements evoked by vestibular or visual stimuli. *Neurophysiology, 39,* 1135-49.

Klier, E.M., Hongying, W., & Crawford, J.D. (2003). Three-dimensional eye-head coordination is implemented downstream from the superior colliculus. *Journal of Neurophysiology, 89,* 2839-2853.

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

Krumhuber, E. & Kappas, A. (2003, September). *Moving smiles: The influence of the dynamic components on the perception of smile-genuineness*. Paper presented at the 10th European Conference Facial Expressions: Measurement and Meaning, Rimini, Italy.

Leonard, C. M., Voeller, K. K. S., & Kuldau, J. M. (1991). When's a smile a smile? or how to detect a message by digitizing the signal. *Psychological Science, 2(3),* 166-172.

Levenson, R.W., Ekman, P., & Friesen, W.V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27*, 363-384.

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.

Lucas, B. & Kanade, T. (1981). An interactive image registration technique with an application in stereo vision. *International Joint Conference on Artificial Intelligence, pp. 674-679.*

Lyons, M., Akamasku, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *Proceedings of the International Conference on Face and Gesture Recognition, Nara, Japan.*

Malatesta, C.Z., Culver, C., Tesman, J.R., & Shepard, B. (1989). The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development, 54*, Serial No. 219.

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

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

Matthews, I. & Baker, S. (2005, in press). Active appearance models revisited. *International Journal of Computer Vision.*

Matthews, I., Ishikawa, T., & Baker, S. (2004). The template update problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26,* 810 - 815.

Messinger, D., Fogel, A., & Dickson, K. (1999). What's in a smile? *Developmental Psychology, 35(3),* 701-708.

Messinger, D., Fogel, A., & Dickson, K. L. (2001). *All smiles are positive, but some smiles are more positive than others. Developmental Psychology, 37*, 642-653.

Michael, G.F. & Camras, L.A. (1992). Infant interest expressions as coordinative motor structures. *Infant Behavior and Development, 15*, 347-358.

Moriyama, T., Kanade, T., Cohn, J. F., Xiao, J., Ambadar, Z., Gao, J., et al. (2002).   Automatic recognition of eye blinking in spontaneously occurring behavior. *Proceedings of the International Conference on Pattern Recognition (ICPR 2002)*, Quebec, Canada, 78-81.

Moriyama, T., Xiao, J., Cohn, J.F., & Kanade, T. (In press). Detailed eye model and its application to analysis of facial image. *Proceedings of the IEEE Conference on Society, Man, and Cybernetics,* The Hague, The Netherlands.

NexGen Ergonomics, *Optotrak® 3020 Position Sensor.*

Padgett, C. & Cottrell, G.W. (1996). Representing face images for emotion classification. *Proceedings Advances in Neural Information Processing Systems, 894-900.*

Pantic, M. & Rothkrantz, M. (2000a). Expert system for automatic analysis of facial expression. *Image and Vision Computing, 18,* 881-905.

Pantic, M. & Rothkrantz, M. (2000b). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1424-1445.

Pantic, M. & Rothkrantz, M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91,* 1371-1390.

Phillips, P.J., Grother, P., Michaels, R.J., Blackburn, D.M., Tabasi, E., & Bone, J.M. (2003). Face recognition vendor test 2002. http://www.frvt.org/FRVT2002/documents.htm

Rosenberg, E. & Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition & Emotion, 8,* 201-229.

Rowley, H.A. Baluja, S., & Kanade, T. (1998). Neural network-based face detection. Pattern Analysis and Machine Vision, x, xxx-xxx.

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., Ambadar, Z., & Cohn, JF. (Submitted). Timing of lip corner movement affects perceived genuineness of spontaneous smiles.

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., Cohn, J. F., & Tian, Y. L. (2003). Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles. *Biological Psychology, 65, 49-66.*

Smith, E., Bartlett, M.S., and Movellan, J.R. (2001). Computer recognition of facial actions: A study of co-articulation effects. *Proceedings of the 8th Annual Joint Symposium on Neural Computation.*

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. L., Kanade, T., & Cohn, J. F. (2002). Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FG'02)*, Washington, DC, 229-234.

Tian, Y.L., Kanade, T., & Cohn, J.F. (in press). Facial expression analysis. In S.Z. Li & A.K. Jain (Eds.), *Handbook of face recognition*.   NY: Springer.

Wachtman, G. S., Cohn, J. F., Van Swearingen, J. M., & Manders, E. K. (2001). Automated tracking of facial features in facial neuromotor disorders. *Plastic and Reconstructive Surgery*, *107*, 1124-1133.

Watson, D. & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin,* 219-235.

Wen, Z. (January 2004). Face processing research at the University of Illinois Champaign-Urbana. *Face Processing Meeting, Center for Multimodal Learning and Communication*, Carnegie Mellon University, Pittsburgh, PA.

Wen, Z. & Huang, T.S. (2003). Capturing subtle facial motions in 3D face tracking. *International Conference on Computer Vision,* Nice, France, xxx-xxx.

Wu, Y. T., Kanade, T., Li, C. C., & Cohn, J. F. (2000).   Image registration using wavelet-based motion model.   *International Journal of Computer Vision*, *38*, 129-152.

Xiao, J., Baker, S., Matthews, I., & Kanade, T. (2004). Real-time combined 2D+3D active appearance models. *IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, xxx-xxx.

Xiao, J. & Kanade, T. (2004). Non-rigid shape and motion recovery: Degenerate deformations. *IEEE Conference on Computer Vision and Pattern Recognition,* Washington, DC, xxx-xxx.

Xiao, J., Moriyama, T., Kanade, T., & Cohn, J. F. (2003). Robust full-motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology*, *13*, 85-94.

Yacoob, Y. & Davis, L. (1996). Recognizing human facial expression from long image sequences using optical flow. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 18,* 636-642.

Zhang, Z (1999). Feature-based facial expression recognition: Sensitivity analysis and experiments with multi-layer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence, 13,* 893-911.

Zhou, Y., Gu, L., & Zhang, H.J. (2003). Bayesian tangent shape model: Estimating shape and pose parameters via Bayesian inference. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR2003*, pp. xxx-xxx.

Zhu, Y., De Silva, L.C., & Ko, C.C. (2002). Using moment invariants and HMM in facial expression recognition. *Pattern Recognition Letters, 23,* 83-91.

Appendix 1

Action Units of the Facial Action Coding System (Ekman & Friesen, 1978)

FACS Action Units.

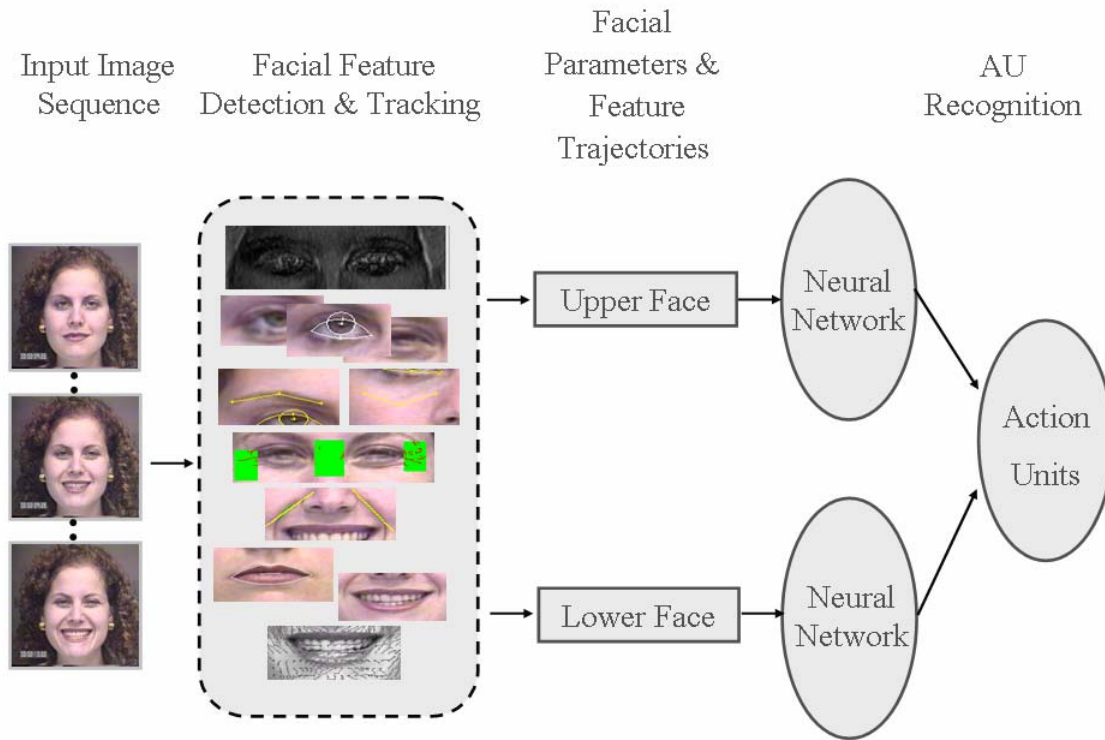| AU	Facial muscle	Description of muscle movement
1	*Frontalis, pars medialis*	Inner corner of eyebrow raised
2	*Frontalis, pars lateralis*	Outer corner of eyebrow raised
4	*Corrugator supercilii, Depressor supercilii*	Eyebrows drawn medially and down
5	*Levator palpebrae superioris*	Eyes widened
6	*Orbicularis oculi, pars orbitalis*	Cheeks raised; eyes narrowed
7	*Orbicularis oculi, pars palpebralis*	Lower eyelid raised and drawn medially
9	*Levator labii superioris alaeque nasi*	Upper lip raised and inverted; superior part of the nasolabial furrow deepened; nostril dilated by the medial slip of the muscle
10	*Levator labii superioris*	Upper lip raised; nasolabial furrow deepened producing square-like furrows around nostrils
11	*Levator anguli oris* (a.k.a. *Caninus)*	Lower to medial part of the nasolabial furrow deepened
12	*Zygomaticus major*	Lip corners pulled up and laterally
13	*Zygomaticus minor*	Angle of the mouth elevated; only muscle in the deep layer of muscles that opens the lips
14	*Buccinator*	Lip corners tightened. Cheeks compressed against teeth
15	*Depressor anguli oris (*a.k.a*. Triangu-*	Corner of the mouth pulled downward and

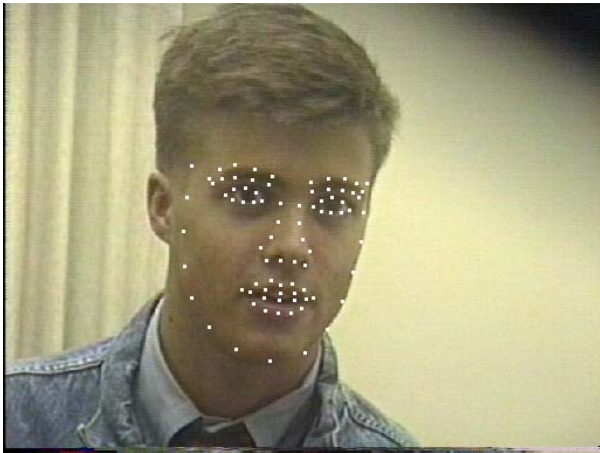|		
	laris)	inward
16	*Depressor labii inferioris*	Lower lip pulled down and laterally
17	*Mentalis*	Skin of chin elevated
18	*Incisivii labii superioris and Incisivii labii inferioris*	Lips pursed
20	*Risorius w/ platysma*	Lip corners pulled laterally
22	*Orbicularis oris*	Lips everted (funneled)
23	*Orbicularis oris*	Lips tightened
24	*Orbicularis oris*	Lips pressed together
25	*Depressor labii inferioris*, or relaxation of *mentalis,* or *orbicularis oris*	Lips parted
26	*Masseter;* relaxed temporal and internal *pterygoid*	Jaw dropped
27	*Pterygoids* and *digastric*	Mouth stretched open
28	*Orbicularis oris*	Lips sucked
41	Relaxation of *levator palpebrae superioris*	Upper eyelid droop
42	*Orbicularis oculi*	Eyelid slit
43	Relaxation of *Levator palpebrae* superioris; orbicularis oculi, pars palpebralis	Eyes closed
44	Orbicularis oculi, pars palpebralis	Eyes squinted
45	Relaxation of *Levator palpebrae superioris; Orbicularis oculi, pars palpebralis*	Blink
46	Relaxation of *Levator palpebrae superioris; orbicularis oculi, pars palpebralis*	Wink

Note. Entries are limited to action units that have a known anatomical basis. Action descriptors and codes for head and eye position are omitted.
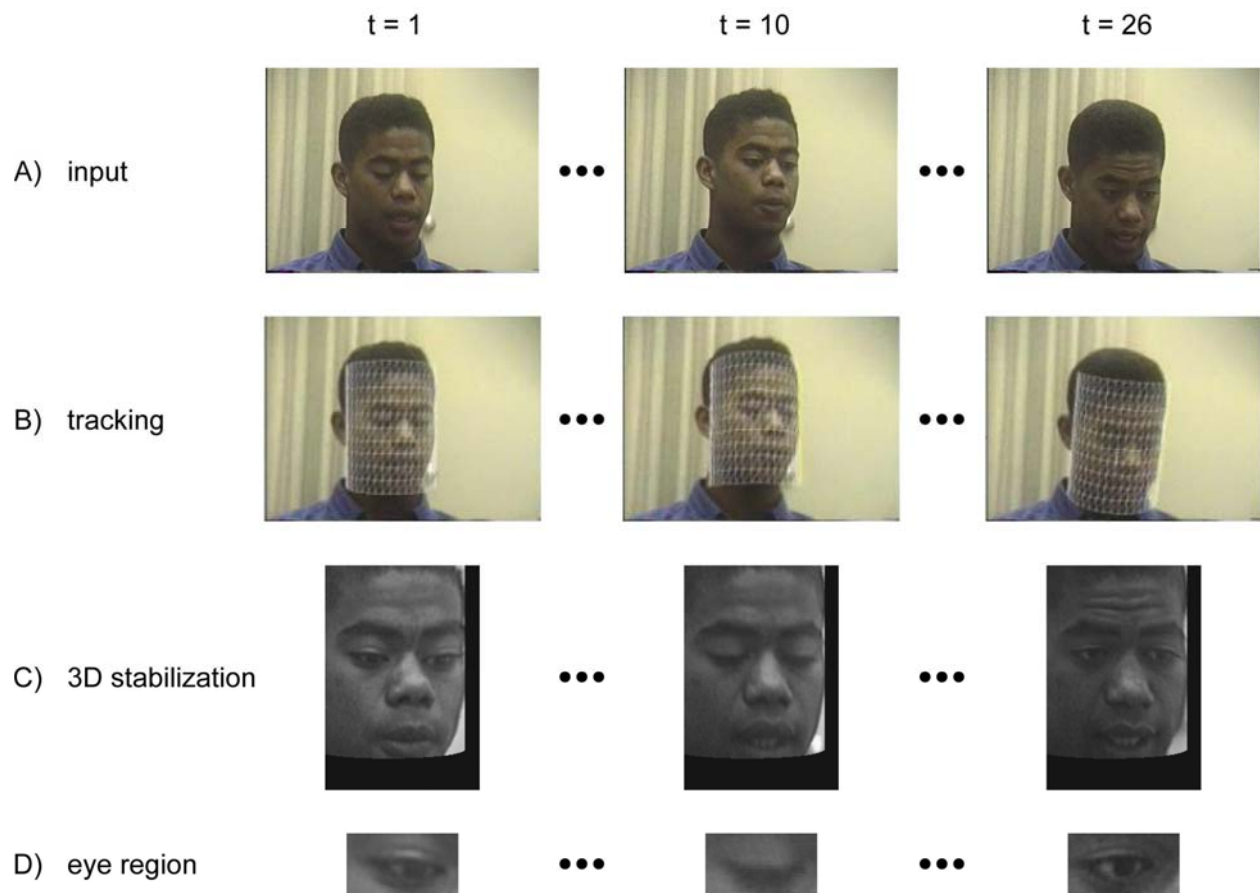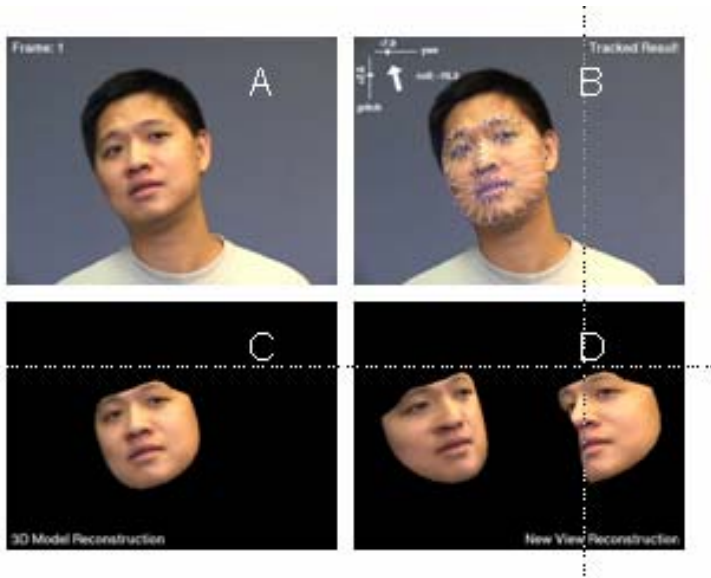
Figure Captions

Figure 1.   Overview of CMU/Pitt Automated Face Analysis (AFA) System.

Figure 2.   Example result of automatic face and facial feature detection.

Figure 3.   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. From Cohn et al., 2003.

Figure 4.   Sample results for 3D+2D AAM: A) the input image, B) the tracked result, C) the 3DMM reconstruction, and D) two new view reconstructions. From Xiao et al., 2004.
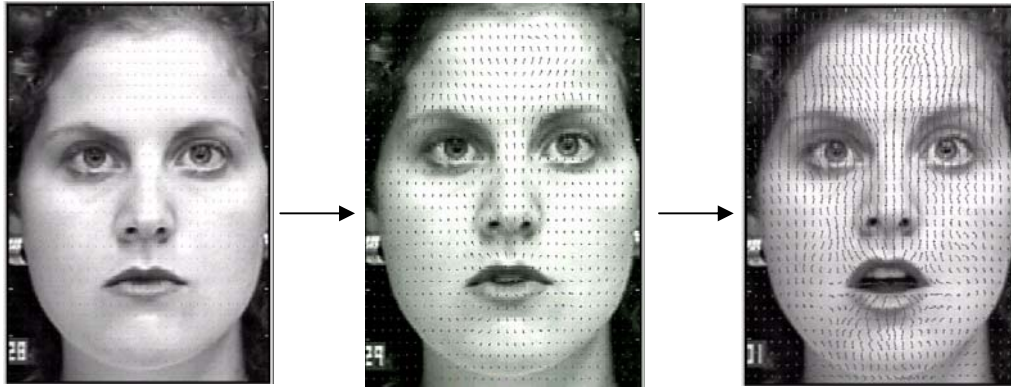
Figure 5.   Example of dense flow extraction. From Lien et al. (2000).

Figure 6.   Example of facial feature tracking. From Cohn et al. (1999).

Figure 7.   Gabor filters. Orientation of the filter systematically varies across rows, while resolution varies across columns.

Figure 8.   Results of Gabor filtering of the eye state (AU 41, AU 42, and AU 45). From Tian et al. (2002).

Figure 9.   Multi-state lip model. From Tian et al. (2001).

Figure 10   Diversity in the appearance of eye images.   Moriyama et al. (In press).

Figure 11a.   Generative eye model.

Figure 11b.   Eye state analysis including contour fitting of the upper and lower eyelids and detection of the center of the iris. Cropped images from 3 subjects of varied ethnic background are shown.

Figure 12.   Results of Canny edge filtering of facial lines and furrows in areas of the nasal root and lateral to the outer eye corners. From Tian et al. (2001).

Figure 13.   AFA-II parameters for (A) upper and (B) lower face. From Tian et al. (2001).

Figure 14.   Relation between *Zygomaticus major* EMG and lip-corner displacement in a spontaneous smile.   From Cohn & Schmidt (in press).

Figure 15.   Top panel: Selected frames from image sequence showing the relation between lip-corner displacement, head rotation, and eye motion. Bottom panel: Corresponding time series for these parameters. From Cohn et al. (2004).

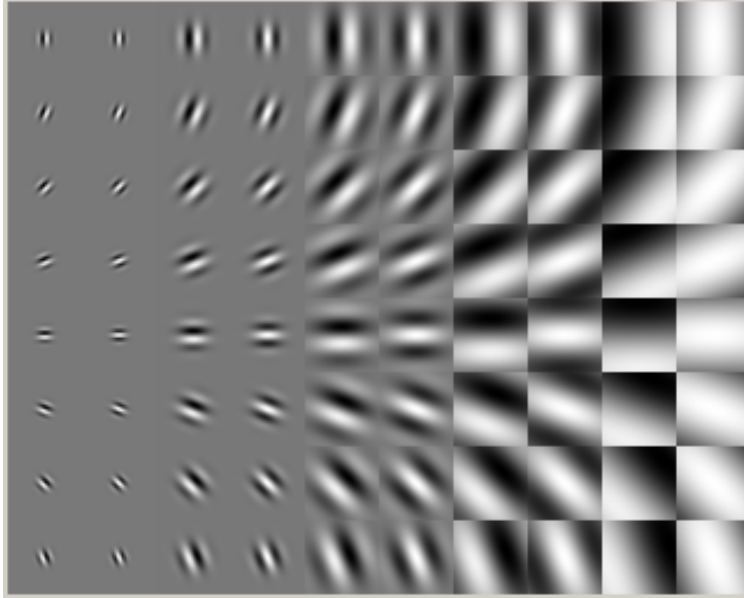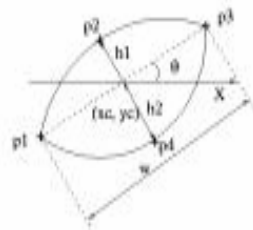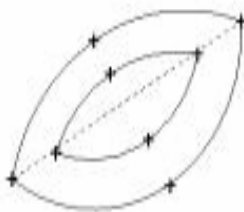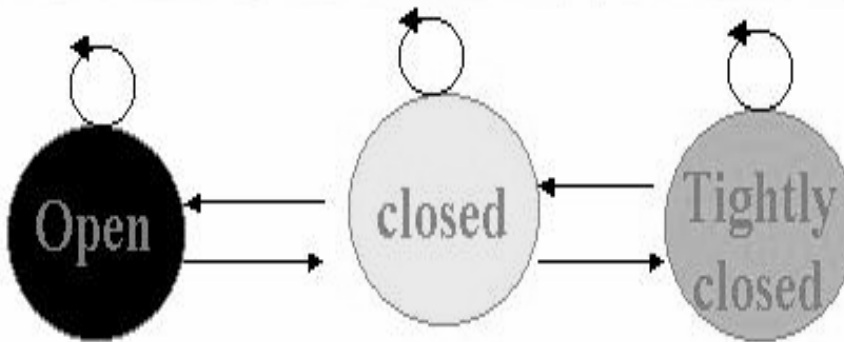Figure 16.   Head and feature tracking using active appearance model.

Input Image
Sequence

Facial Feature
Detection & Tracking

Facial
Parameters &
Feature
Trajectories

AU
Recognition

Upper Face

Neural
Network

Lower Face

Neural
Network

Action

Units

A) input

B) tracking

C) 3D stabilization
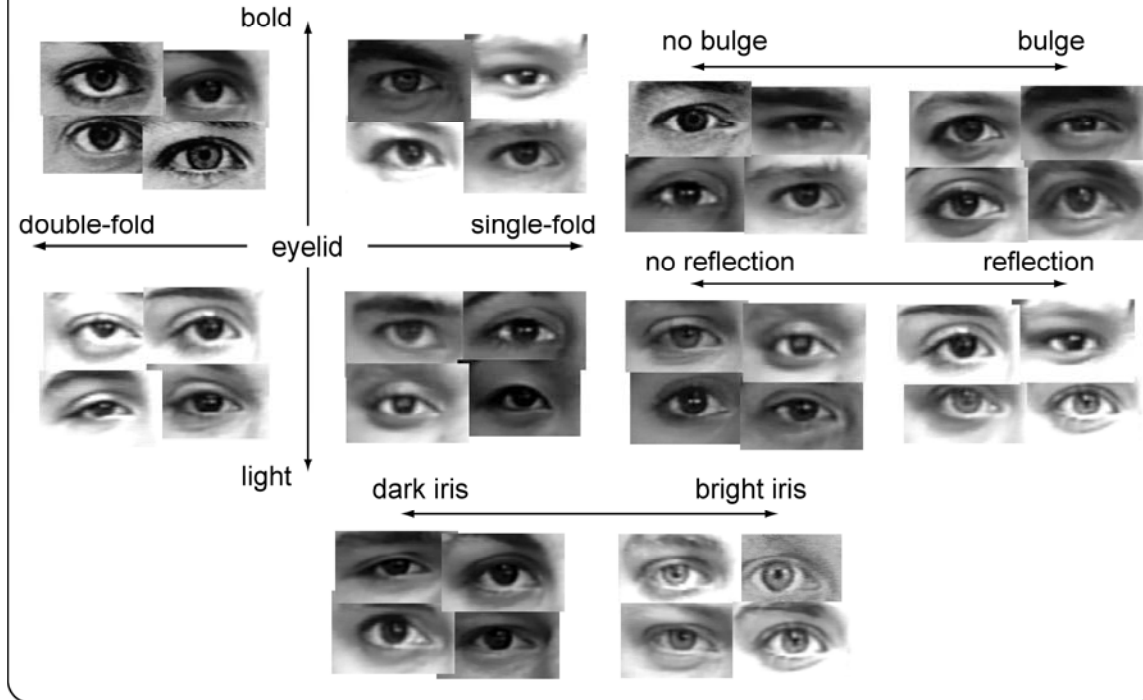
D) eye region

Original Images   Gabor Images

AU41

AU42

AU43

Variance from structural individuality

bold

double-fold ←————— eyelid —————→ single-fold

light

no bulge   bulge

no reflection   reflection

dark iris ←————→ bright iris

Variance from motion of a particular eye

motion A
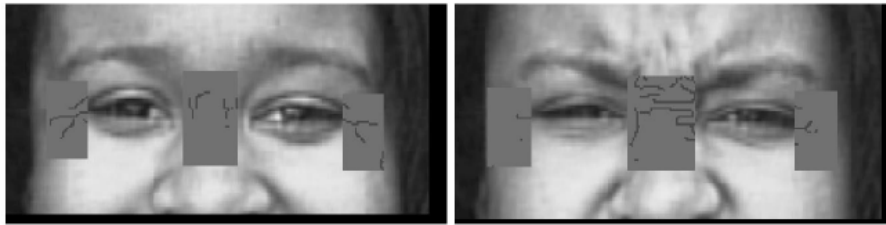
motion B

A

double-fold eyelid $\{\mathbf{d}\ (d_{width},\ d_{boldness}),\ R_d\}$

upper eyelid $\{\mathbf{u}\ (\upsilon_{height},\ \upsilon_{skew})\}$

sclera
$: R_s$ corners
$: R_c$ iris
$: \{i_{size},\ i_{color},\ i_x,\ i_y\}$

lower eyelid
$: \{\mathbf{l}\ (\lambda_{height})\}$ bulge
$: \{R_b\}$ reflection
$: \{R_r\}$ infraorbital furrow
$: \{\mathbf{f}\}$

B

(a)

(b)

A

B