Please do not quote. In press, *Handbook of affective computing*. New York, NY: Oxford

Automated Face Analysis for Affective Computing

Jeffrey F. Cohn & Fernando De la Torre

Abstract

Facial expression communicates emotion, intention, and physical state, and regulates interpersonal behavior. Automated Face Analysis (AFA) for detection, synthesis, and understanding of facial expression is a vital focus of basic research. While open research questions remain, the field has become sufficiently mature to support initial applications in a variety of areas. We review 1) human-observer based approaches to measurement that inform AFA; 2) advances in face detection and tracking, feature extraction, registration, and supervised learning; and 3) applications in action unit and intensity detection, physical pain, psychological distress and depression, detection of deception, interpersonal coordination, expression transfer, and other applications. We consider user-in-the-loop as well as fully automated systems and discuss open questions in basic and applied research.

Keywords

Automated Face Analysis and Synthesis, Facial Action Coding System (FACS), Continuous Measurement, Emotion, Nonverbal Communication, Synchrony

1. Introduction

The face conveys information about a person's age, sex, background, and identity, what they are feeling, or thinking (Darwin, 1872/1998; Ekman & Rosenberg, 2005). Facial expression regulates face-to-face interactions, indicates reciprocity and interpersonal attraction or repulsion, and enables inter-subjectivity between members of different cultures (Bråten, 2006; Fridlund, 1994; Tronick, 1989). Facial expression reveals comparative evolution, social and emotional development, neurological and psychiatric functioning, and personality processes (Burrows & Cohn, In press; Campos, Barrett, Lamb, Goldsmith, & Stenberg, 1983; Girard, Cohn, Mahoor, Mavadati, & Rosenwald, 2013; Schmidt & Cohn, 2001). Not surprisingly, the face has been of keen interest to behavioral scientists.

Beginning in the 1970s, computer scientists became interested in the face as a potential biometric (Kanade, 1973). Later, in the 1990s, they became interested in use of computer vision and graphics to automatically analyze and synthesize facial expression (Ekman, Huang, & Sejnowski, 1992; Parke & Waters, 1996). This effort was made possible in part by the development in behavioral science of detailed annotation schemes for use in studying human emotion, cognition, and related processes. The most detailed of these systems, the Facial Action Coding System (Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002), informed the development of the MPEG-4 facial animation parameters (Pandzic & Forchheimer, 2002) for video transmission and enabled progress toward automated measurement and synthesis of facial actions for research in affective computing, social signal processing, and behavioral science.

Early work focused on expression recognition between mutually exclusive posed facial actions. More recently, investigators have focused on the twin challenges of expression detection in naturalistic settings, in which low base rates, partial occlusion, pose variation, rigid head motion, and lip movements associated with speech complicate detection, and real-time synthesis of photorealistic avatars that are accepted as live video by naïve participants.

With advances, automated face analysis (AFA) is beginning to realize the goal of advancing human understanding (Ekman et al., 1992). AFA is leading to discoveries in areas that include detection of pain, frustration, emotion intensity, depression and psychological distress, and reciprocity. New applications are emerging in instructional technology, marketing, mental health, and entertainment. This chapter reviews methodological advances that have made these developments possible, surveys their scope, and addresses outstanding issues.

2. Human-Observer Based Approaches to Measurement

Supervised learning of facial expression requires well-coded video. What are the major approaches to manually coding behavior? At least three can be distinguished: message-based, sign-based, and dimensional.

2.1 Approaches

2.1.1 Message-based measurement. In *message-based* measurment (Cohn & Ekman, 2005), observers make inferences about emotion or affective state. Darwin (1872/1998) described facial expressions for more than 30 emotions. Ekman and others (Ekman & Friesen, 1975; Izard, 1977; Keltner & Ekman, 2000; Plutchik, 1979) narrowed the list to a smaller number that they refer to as "basic" (See Figure 1) (Ekman, 1992; Keltner & Ekman, 2000). Ekman's criteria for "basic emotions" include evidence of univeral signals across all human groups, physiological specificity, homologous expressions in other primates, and unbidden occurrence (Ekman, 1992; Keltner & Ekman, 2000). Baron-Cohen and his colleagues proposed a much larger set of cognitive-emotional states that are less tied to an evolutionary perspective. Examples include concentration, worry, playfulness, and kindness (Baron-Cohen, 2003).

[INSERT FIGURE 1 HERE].

An appealing assumption of message-based approaches is that the face provides a direct "read out" of emotion (Buck, 1984). This assumption is problematic. The meaning of an expression is context dependent. The same expression can connote anger or triumph depending on where, with what, and how it occurs. The exhaltation of winning a hard faught match and the rage of losing can be difficult to distinguish without knowing context (Feldman Barrett, Mesquita, & Gendron, 2011). Similarly, smiles accompanied by cheek raising convey enjoyment; the same smiles accompanied by head lowering and turning to the side convey embarrassment (Cohn & Schmidt, 2004; Keltner & Buswell, 1997). Smiles of short duration and with a single peak are more likely to be perceived as polite (Ambadar, Cohn, & Reed, 2009). Too, expressions may be posed or faked. In the latter case, there is a dissociation between the assumed and the actual subjective emotion. For these reasons and others, there is reason to be dubious of one-to-one correspondences between expression and emotion (Cacioppo & Tassinary, 1990).

2.1.2 Sign-based measurement. An alternative to message-based measurement is to use a purely descriptive, *sign-based* approach and then use experimental or observational methods to discover the relation between such signs and emotion. The most widely-used method is the Facial Action Coding System (FACS) (Cohn, Ambadar, & Ekman, 2007; Ekman et al., 2002). FACS describes facial activity in terms of anatomically based action units (AUs) (Figure 2). The FACS taxonomy was developed by manually observing gray level variation between expressions in images, recording the electrical activity of facial muscles, and observing the effects of electrically stimulating facial muscles (Cohn & Ekman, 2005). Depending on the version of FACS, there are 33 to 44 AUs and a large number of additional "action descriptors" and other movements. AU may be coded using either binary (presence vs. absence) or ordinal (intensity) labels. Figure 2 and 3 show examples of each.

While FACS itself includes no emotion labels, empirically-based guidelines for emotion interpretation have been proposed. The FACS investigator's guide and other sources hypothesize mappings between AU and emotion (Ambadar et al., 2009; Ekman & Rosenberg, 2005; Knapp & Hall, 2010). Sign-based approaches in addition to FACS are reviewed in Cohn and Ekman (2005).

[INSERT FIGURES 2 AND 3 HERE].

2.1.3 Dimensional measurement. Both message- and sign-based approaches emphasize differences between emotions. An alternative emphasizes their similarities.

Schlosberg (1952, 1954) proposed that the range of facial expressions conforms to a circular surface with pleasantness-unpleastness (i.e. valence) and attention-rejection as the principal axes (Activity was proposed as a possible third). Russell and Bullock (1985), like Schlosberg, proposed that emotion conforms to a circumplex structure with pleasantness-unpleasantness (valence) as one axis, but replaced attention-rejection with arousal-sleepiness. Watson & Tellegen (1985) proposed an orthogonal rotation of the axes to yield positive and negative affect (PA and NA, respectively, each ranging in intensity from low to high). More complex structures have been proposed. Mehrabian (1998) proposed that dominance-submissiveness be included as a third dimension. Tellegen, Watson, and Clark (1999) proposed hierarchical dimensions.

Dimensional approaches have several advantages. They are well studied as indices of emotion (Fox, 2008). They are parsimonious, representing any given emotion in terms of two or three underlying dimensions. They lend themselves to continuous representations of intensity. Positive and negative affect (PA and NA), for instance can be measured over intensity ranges of hundreds of points. Last, they often require relatively little expertise. As long as multiple independent and unbiased ratings are obtained, scores may be aggregated across multiple raters to yield highly reliable measures. This is the case even when pairwise ratings of individual raters are noisy (Rosenthal, 2005). Such is the power of aggregating.

Some disadvantages may be noted. One, because they are parsiomonious, they are not well suited to representing discrete emotions. Pride and joy, for instance could be difficult to distinguish. Two, like the message-based approach, dimensional representations implicitly assume that emotion may be inferred directly from facial expression, which as noted above is problematic. And three, the actual signals involved in communicating emotion are unspecified.

2.2 Reliability

Reliabilility concerns the extent to which measurement is repeatable and consistent; that is, free from random error (Martin & Bateson, 2007). Whether facial expression is measured using a message, sign, or dimensional approach, we wish to know to what extent variability in the measurements represents true variation in facial expression rather than error. In general reliability between observers can be considered in at least two ways (Tinsley & Weiss, 1975). One is whether coders make exactly the same judgments (i.e. Do they agree?). The other is whether their judgments are consistent. When judgments are made on a nomimal scale, *agreement* means that each coder assigns the same score. When judgments are made on an ordinal or interval scale, *consistency* refers to the degree to which ratings from different sources are proportional when expressed as deviations from their means. Accordingly, agreement and consistency may show disassociations. If two coders always differ by *x* points in the same direction on an ordinal or interval scale, they have low agreement but high consistency. Depending on the application, consistency between observers may be sufficient. Using a dimensional approach to assess intensity of positve affect, for instance, it is unlikely that coders will agree exactly. What matters is that they are consistent relative to each other.

In general, message- and sign-based approaches are evaluated in terms of agreement; and dimensional approaches are evaluated in terms of consistency. Because base rates can bias uncorrected measures of agreement, statistics such as kappa and *F1* (Fleiss, 1981) afford some protection against this source of bias. When measuring consistency, intraclass correlation (Shrout & Fleiss, 1979) is preferable to Pearson correlation when mean differences in level are a concern. The choice of reliability type (agreement or consistency) and metric should depend on how measurements are obtained and how they will be used.

3 Automated Face Analysis

Automated Face Analysis (AFA) seeks to detect one or more of the measurement types discussed in Section 2. This goal requires multiple steps that include face detection and tracking, feature extraction, registration, and learning. Regardless of approach, there are numerous challenges. These include: (1) non-frontal pose and moderate to large head motion make facial image registration difficult; (2) many facial actions are inherently subtle making them difficult to model; (3) the temporal dynamics of actions can be highly variable; (4) discrete AUs can modify each other's appearance (i.e. non-additive combinations); (5) individual differences in face shape and appearance undermine generalization across subjects; and (6) classifiers can suffer from over-fitting when trained with insufficient examples.

To address these and other issues, a large number of facial expression and AU recognition/detection systems have been proposed. The pipeline depicted in Figure 4 is common to many. Key differences among them include types of input images (2D or 3D), face detection and tracking, types of features, registration, dimensionality reduction, classifiers, and databases. The number of possible combinations that have been considered is exponential and beyond the bounds of what can be considered here. With this in mind, we review essential aspects. We then review recent advances in expression transfer (also referred to as automated face synthesis, or AFS) and applications made possible by advances in AFA.

[INSERT FIGURE 4 HERE].

3.1 Face and facial feature detection and tracking

AFA begins with face detection. In the case of relatively frontal pose, the Viola and Jones (2004) face detector may be the most widely used. This and others are reviewed in Zhang and Zhang (2010). Following face detection, either a sparse (e.g. eyes or eye corners) or dense set of facial features (e.g., the contours of the eyes and other permanent facial features) is detected and tracked in the video. An advantage of the latter is that it affords information from which to infer 3D pose (especially yaw, pitch, and roll) and viewpoint registered representations (e.g., warp face image to a frontal view).

To track a dense set of facial features, active appearance models (Cootes, Edwards, & Taylor, 2001) are often used. AAMs decouple the shape and appearance of a face image. Given a pre-defined linear shape model with linear appearance variation, AAMs align the shape model to an unseen image containing the face and facial expression of interest. The shape of an AAM is described by a 2D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape (Ashraf et al., 2009). The vertex locations correspond to a source appearance image, from which the shape is aligned. Since AAMs allow linear shape variation, the shape can be expressed as a base shape $\mathbf{s}_0$ plus a linear combination of m shape vectors $\mathbf{s}_i$. Because AAMs are invertible, they can be used both for analysis and for synthesizing new images and video. Theobald and Matthews (Boker et al., 2011; Theobald, Matthews, Cohn, & Boker, 2007) used this approach to generate real-time near video-realistic avatars, which we discuss below.

The precision of AAMs comes at a price. Prior to use, they must be trained for each person. That is, they are "person-dependent" (as well as camera- and illumination-dependent). To overcome this limitation, Saragih, Lucey, and Cohn (2011a) extended the work of Cristinacce and Cootes (2006) and others to develop what is referred to as a constrained local model (CLM). Compared with AAMs, CLMs generalize well to unseen appearance variation and offer greater invariance to global illumination variation and occlusion (S. Lucey et al., 2009; 2010). They are sufficiently fast to support real-time tracking and synthesis (S. Lucey, Wang, Saragih, & Cohn, 2010). A disadvantage of CLMs relative to AAMs is that they detect shape less precisely. For this reason, there has been much effort to identify ways to compensate for their reduced precision (Chew et al., 2012).

3.2 Registration

To remove the effects of spatial variation in face position, rotation, and facial proportions, images must be registered to a canonical size and orientation. 3D rotation is especially challenging because the face looks different from different orientations. 3D transformations can be estimated from monocular (up to a scale factor) or multiple cameras using structure from motion algorithms (Matthews, Xiao, & Baker, 2007; Xiao, Baker, Matthews, & Kanade, 2004) or head trackers (Morency, 2008; Xiao, Kanade, & Cohn, 2003). For small to moderate out-of-plane rotation a moderate distance from the camera (assume orthographic projection), the 2D projected motion field of a 3D planar surface can be recovered with an affine model of six parameters.

3.3 Feature extraction

Several types of features have been used. These include geometric (also referred to as shape), appearance, and motion.

3.3.1 Geometric features. Geometric features refer to facial landmarks such as the eyes or brows. They can be represented as fiducial points, a connected face mesh, active shape model, or face component shape parameterization (Tian, Cohn, & Kanade, 2005). To detect actions such as brow raise (AU 1+2) changes in displacement between points around the eyes and those on the brows can be discriminative. While most approaches model shape as 2D features, a more powerful approach is to use structure from motion to model them as 3D features (Saragih et al., 2011a) (Xiao et al., 2004). Jeni (2012) found that this approach improves AU detection.

Shape or geometric features alone are insufficient for some AUs. Both AU 6 and AU 7 narrow the eye aperture. The addition of appearance or texture information aids in discriminating between them. AU 6 but not AU 7, for instance, causes wrinkles lateral to the eye corners. Other AUs, such as AU 11 (nasolabial furrow deepener) and AU 14 (mouth corner dimpler) may be undetectable without reference to appearance because they occasion minimal changes in shape. AU 11 causes a deepening of the middle portion of the nasolabial furrow. AU 14 and AU 15 each cause distinctive pouching around the lip corners.

3.3.2 Appearance features. Appearance features represent changes in skin texture such as wrinkling and deepening of facial furrows and pouching of the skin. Many techniques for describing local image texture have been proposed. The simplest is a vector of raw pixel-intensity values. However, if an unknown error in registration occurs, there is an inherent variability associated with the true (i.e., correctly registered) local image appearance. Another problem is that lightning conditions affect texture in gray scale representations. Biologically inspired appearance features, such as Gabor wavelets or magnitudes (Jones & Palmer, 1987), (Movellan, Undated), HOG (Dalal & Triggs, 2005), and SIFT (Mikolajczyk & Schmid, 2005) have proven more robust than pixel intensity to registration error (Chew et al., 2012). These and other appearance features are reviewed in (De la Torre & Cohn, 2011) and (Mikolajczyk & Schmid, 2005).

3.3.3 Motion features. For humans, motion is an important cue to expression recognition, especially for subtle expressions (Ambadar, Schooler, & Cohn, 2005). No less is true for AFA. Motion features include optical flow (Mase, 1991) and dynamic textures or motion history images (MHI) (Chetverikov & Peteri, 2005). In early work, Mase (1991) used optical flow to estimate activity in a subset of the facial muscles. Essa and Pentland (1997) extended this approach, using optic flow to estimate activity in a detailed anatomical and physical model of the face. Yacoob and Davis (1997) bypassed the physical model and constructed a mid-level representation of facial motion directly from the optic flow. Cohen and colleagues (2003) implicitly recovered motion representations by building features such that each feature motion corresponds to a simple deformation on the face. Motion history images (MHIs) were first proposed by Bobick and Davis (2001). MHIs compress into one frame the motion over a number of consecutive ones. Valstar, Pantic, and Patras (2004) encoded face motion into Motion History Images. Zhao and Pietikainen (2007) used volume local binary patterns (LBP), a temporal extension of local binary patterns often used in 2D texture analysis. These methods all encode motion in a video sequence.

3.3.4 Data reduction/selection. Features typically have high dimensionality, especially so for appearance. To reduce dimensionality, several approaches have been proposed. Widely used linear techniques are principal components analysis (PCA) (Hotelling, 1933), Kernel PCA (Schokopf, Smola, & Muller, 1997), and Independent Components Analysis (Comon, 1994). Nonlinear techniques include Laplacian Eigenmaps (Belkin & Niyogi, 2001), local linear embedding (LLE) (Roweis & Saul, 2000), and locality preserving projections (LPP) (Cai, He, Zhou, Han, & Bao, 2007; Y. Chang, Hu, Feris, & Turk, 2006)). Supervised methods include linear discriminant analysis, AdaBoost, kernel LDA, and locally sensitive LDA.

3.4 Learning

Most approaches use supervised learning. In supervised learning, event categories (e.g., emotion labels or AU) or dimensions are defined in advance in labeled training data. In unsupervised learning, labeled training data are not used. Here, we consider supervised approaches. For a review of unsupervised approaches, see De la Torre and Cohn (2011).

Two approaches to supervised learning are: (1) static modeling—typically posed as a discriminative classification problem in which each video frame is evaluated independently; (2) temporal modeling—frames are segmented into sequences and typically modeled with a variant of dynamic Bayesian networks (e.g., Hidden Markov Models, Conditional Random Fields).

In static modeling, early work used neural networks (Tian, Kanade, & Cohn, 2001). More recently, support vector machine classifiers (SVM) have predominated. Boosting has been used to a lesser extent both for classification as well as for feature selection (Littlewort, Bartlett, Fasel, Susskind, & Movellan, 2006; Y. Zhu, De la Torre, Cohn, & Zhang, 2011). Others have explored rule-based systems (Pantic & Rothkrantz, 2000).

In temporal modeling, recent work has focused on incorporating motion features to improve performance. A popular strategy uses HMMs to temporally segment actions by establishing a correspondence between the action's onset, peak, and offset and an underlying latent state. Valstar and Pantic (Valstar & Pantic, 2007) used a combination of SVM and HMM to temporally segment and recognize AUs. Koelstra and Pantic (Koelstra & Pantic, 2008) used Gentle-Boost classifiers on motion from a non-rigid registration combined with an HMM. Similar approaches include a nonparametric discriminant HMM (Shang & Chan, 2009) and partially-observed Hidden Conditional Random Fields (K. Y. Chang, Liu, & Lai, 2009). In related work, Cohen and his colleagues (2003) used Bayesian networks to classify the six universal expressions from video. Naive-Bayes classifiers and Gaussian Tree-Augmented Naive Bayes (TAN) classifiers learned dependencies among different facial motion features. In a series of papers, Qiang and his colleagues (Li, Chen, Zhao, & Ji, 2013; Tong, Chen, & Ji, 2010; Tong, Liao, & Ji, 2007) used Dynamic Bayesian Networks to detect facial action units.

3.5 Databases

Data drives research. Development and validation of supervised and unsupervised algorithms requires access to large video databases that span the range of variation expected in target applications. Relevant variation in video includes pose, illumination, resolution, occlusion, facial expression, actions, and their intensity and timing, and individual differences in subjects. An algorithm that performs well for frontal, high-resolution, well-lit video with few occlusions may perform rather differently when such factors vary (Cohn & Sayette, 2010).

Most face expression databases have used directed facial action tasks; subjects are asked to pose discrete facial actions or holistic expressions. Posed expressions, however, often differ in appearance and timing from those that occur spontaneously. Two reliable signals of sadness, AU 15 (lip corners pulled down) and AU 1+4 (raising and narrowing the inner corners of the brow) are difficult for most people to perform on command. Even when such actions can be performed deliberately, they may differ markedly in timing from what occurs spontaneously (Cohn & Schmidt, 2004). Differences in the timing of spontaneous and deliberate facial actions are particularly important in that many pattern recognition approaches, such as hidden Markov Models (HMMs), are highly dependent on the timing of the appearance change. Unless a database includes both deliberate and spontaneous facial actions, it will likely prove inadequate for developing face expression methods that are robust to these differences.

Variability within and among coders is an important source of error that too often is overlooked by database users. Human performance is inherently variable. An individual coder may assign different AU to the same segment on different occasions ("test-retest" unreliability); and different coders may assign different AU ("alternate-form" unreliability). Although FACS coders are (or should be) certified in its use, they can vary markedly in their expertise and in how they operationalize FACS criteria. An additional source of error relates to manual data entry. Software for computer-assisted behavioral coding can lessen but not eliminate this error source. All of these types of error in "ground truth" can adversely affect classifier training and performance. Differences in manual coding between databases may and do occur as well and can contribute to impaired generalizability of classifiers from one database to another.

Section 4 of this handbook and earlier reviews (Zeng, Pantic, Roisman, & Huang, 2009) detail relevant databases. Several very recent databases merit mention. DISFA (Mavadati, Mahoor, Bartlett, Trinh, & Cohn, 2013) consists of FACS-coded high-resolution facial behavior in response to emotion-inducing videos. AU are coded on a 6-point intensity scale (0 to 5). The Binghamton-Pittsburgh 4D database (BP4D) is a high-resolution, 4D (3D * time) AU-coded database of facial behavior in response to varied emotion inductions (X. Zhang et al., 2013). Several databases include participants with depression or related disorders (Girard et al., 2013; Scherer et al., 2013; Valstar et al., 2013; Wang et al., 2008). Human use restrictions limit access to some of these. Two other large AU-coded databases not yet publically are the Sayette Group Formation Task (GFT) (Sayette et al., 2012) and the AMFED facial expression database (McDuff, Kaliouby, Senechal, et al., 2013). GFT includes manually FACS-coded video of 720 participants in 240 3-person groups (approximately 30 minutes each). AMFED includes manually FACS-coded video of thousands of participants recorded via webcam while viewing commercials for television.

4. Applications

AU detection and, to a lesser extent, detection of emotion expressions, has been a major focus of research. Action units of interest have been those strongly related to emotion expression and that occur sufficiently often in naturalistic settings. As automated face analysis and synthesis has matured, many additional applications have emerged.

4.1 AU detection

There is a large, vigorous literature on AU detection (De la Torre & Cohn, 2011; Tian et al., 2005; Zeng et al., 2009). Many algorithms and systems have been bench-marked on posed facial databases, such as Cohn-Kanade (Kanade, Cohn, & Tian, 2000; P. Lucey et al., 2010), MMI (Pantic, Valstar, Rademaker, & Maat, 2005), and the UNBC Pain Archive (P. Lucey, Cohn, Prkachin, Solomon, & Matthews, 2011) . Benchmarking on spontaneous facial behavior has occurred more recently. The FERA 2011 Facial Expression Recognition Challenge enrolled 20 teams to compete in AU and emotion detection (Valstar, Mehu, Jiang, Pantic, & Scherer, 2012). Of the 20, 15 participated in the challenge and submitted papers. Eleven papers were accepted for publication in a double-blind review. On the AU detection sub-challenge, the winning group achieved an $F1$ score of 0.63 across 12 AUs at the frame level. On the less difficult emotion detection sub-challenge, the top alogrithm classified 84% correctly at the sequence level.

The FERA organizers noted that the scores for AU were well-above baseline but still far from perfect. Without knowing the $F1$ score for interobserver agreement (see Section 2.2, above), it is difficult to know to what extent this score may have been attenuated by measurement error in the ground truth AU. An additional caveat is that results were for a single database of rather modest size (10 trained actors portraying emotions). Further opportunities for comparative testing on spontaneous behavior are planned for the 3rd International Audio/Visual Emotion Challenge (http://sspnet.eu/avec2013/) and the Emotion

Recognition In The Wild Challenge and Workshop (EmotiW 2013) (http://cs.anu.edu.au/few/emotiw.html*)*. Because database sizes in these two tests will be larger than in FERA, more informed comparisons between alternative approaches will be possible.

4.2 Intensity

Message-based and dimensional measurement may be performed on both ordinal and continuous scales. Sign-based measurement, such as FACS, conventionally use an ordinal scale (0 to 3 points in the 1978 edition of FACS; 0 to 5 in the 2002 edition). Action unit intensity has been of particular interest. AU unfold over time. Initial efforts focused on estimating their maximum, or "peak," intensity (Bartlett et al., 2006). More recent work has sought to measure intensity for each video frame (Girard, 2013; Mavadati et al., 2013; Messinger, Mahoor, Chow, & Cohn, 2009).

Early work suggested that AU intensity could be estimated by computing distance from the hyperplane of a binary classifier. For posed action units in Cohn-Kanade, distance from the hyperplane and (manually coded) AU intensity were moderately correlated for maximum AU intensity ($r = .60$) (Bartlett et al., 2006b). Theory and some data, however, suggest that distance from the hyperplane may be a poor proxy for intensity in spontaneous facial behavior. In RU-FACS, in which facial expression is unposed (also referred to as spontaneous), the correlation between distance from the hyperplane and AU intensity for maximum intensity was $r = .35$ or less (Bartlett et al., 2006a). Yang, Liu, and Metaxas (2009) proposed that supervised training from intensity-labeled training data is a better option than training from distance from the hyperplane of a binary classifier.

Recent findings in AU-coded spontaneous facial expression support this hypothesis. All estimated intensity on a frame-by-frame basis, which is more challenging than measuring AU intensity only at its maximum. In the DISFA database, intra-class correlation (ICC) between manual and automatic coding of intensity (0 to 5 ordinal scale) was 0.77 for Gabor features (Mavadati et al., 2013). Using support vector regression in the UNBC Pain Archive, Kaltwang and colleagues (Kaltwang, Rudovic, & Pantic, 2012) achieved a correlation of about 0.5. In the BP4D database, a multiclass SVM achieved an ICC of 0.92 for AU 12 intensity (Girard, 2013), far greater than what was achieved using distance from the hyperplane of a binary SVM. These findings suggest that for spontaneous facial expression at the frame level, it is essential to train on intensity-coded AU and a classifier that directly measures intensity (e.g., multiclass SVM or support vector regression).

4.3 Physical pain

Pain assessment and management are important across a wide range of disorders and treatment interventions. Pain measurement is fundamentally subjective and is typically measured by patient self-report, which has notable limitations. Self-report is idiosyncratic; susceptible to suggestion, impression management, and deception; and lacks utility with young children, individuals with certain types of neurological impairment, many patients in postoperative care or transient states of consciousness, and those with severe disorders requiring assisted breathing, among other conditions.

Using behavioral measures, pain researchers have made significant progress toward identifying reliable and valid facial indicators of pain. In these studies pain is widely characterized by brow lowering (AU 4), orbital tightening (AU 6 and 7), eye closure (AU 43), nose wrinkling, and lip raise (AU 9 and 10) (Prkachin & Solomon, 2008). This development led investigators from the affective computing community to ask whether pain and pain intensity could be detected automatically. Several groups working on different datasets have found the answer to be yes. Littlewort and colleagues (Littlewort, Bartlett, & Lee) discriminated between actual and feigned pain. Hammal and Kunz (2012) discriminating pain from the six basic facial expressions and neutral. We and others detected occurrence and intensity of shoulder pain in a clinical sample (Ashraf et al., 2009; Hammal & Cohn, 2012; Kaltwang et al., 2012; P. Lucey, Cohn, Howlett, Lucey, & Sridharan, 2011).

From these studies, two findings that have more general implications emerged. One, pain could be detected with comparable accuracy whether features were fed directly to a classifier or by a two-step classification in which action units were first detected and AU then were input to a classifier to detect pain. The comparability of results suggests that the AU recognition step may be unnecessary when detecting holistic expressions, such as pain. Two, good results could be achieved even when training and testing on coarse (sequence level) ground truth in place of frame-by-frame behavioral coding (Ashraf et al., 2009). Future research will be needed to test these suggestions.

4.4 Depression and psychological distress

Diagnosis and assessment of symptom severity in psychopathology are almost entirely informed by what patients, their families, or caregivers report. Standardized procedures for incorporating facial and related nonverbal expression are lacking. This is especially salient for depression, for which there are strong indications that facial expression and other nonverbal communication may be powerful indicators of disorder severity and response to treatment. In comparison with non-depressed individuals, depressed individuals have been observed to look less at conversation partners, gesture less, show fewer Duchenne smiles, more smile suppressor movements, and less facial animation. Human-observer based findings such as these have now been replicated using automated analyses of facial and multimodal expression (Joshi, Dhall, Goecke, Breakspear, & Parker, 2012; Scherer et al., 2013). An exciting implication is that facial expression could prove useful for screening efforts in mental health.

To investigate possible functions of depression, we (Girard et al., 2013) recorded serial interviews over multiple weeks in a clinical sample that was undergoing treatment for Major Depressive Disorder. We found high congruence between automated and manual measurement of facial expression in testing hypotheses about change over time in depression severity.

The results provided theoretical support for the hypothesis that depression functions to reduce social risk. When symptoms were highest, subjects showed fewer displays intended to seek interpersonal engagement (i.e. less smiling as well as fewer sadness displays) and more displays that communicate rejection of others (i.e. disgust and contempt). These findings underscore the importance of accounting for individual differences (All subjects were compared to themselves over the course of depressive disorder); provide further evidence in support of AFA's readiness for hypothesis testing about psychological mechanisms; and suggest that automated measurement may be useful in detecting recovery and relapse, as well as in contributing to public health efforts to screen for depression and psychological distress.

4.5 Deception detection

Theory and some data suggest that deception and hostile intent can be inferred in part from facial expression (Ekman, 2009). The RU-FACS database (Bartlett et al., 2006a), which has been extensively used for AU detection, was originally collected for the purpose of learning to detect deception. While no deception results to our knowledge have yet been reported, others using different databases have realized some success in detecting deception from facial expression and other modalities. Metaxas, Burgoon, and their colleagues (Michael, Dilsizian, Metaxas, & Burgoon, 2010; Yu et al., 2013) proposed an automated approach that uses head motion, facial expression, and body motion to detect deception. Tsiamyrtzis (2006) and others achieved close to 90% accuracy using thermal cameras to image the face (Tsiamyrtzis et al.). Further progress in this area will require ecologically valid training and testing data. Too often, laboratory studies of deception have lacked verisimilatude or failed to include the kinds of people most likely to attempt deception or hostile actions. While the need for good data is well recognized, barriers to its use have been difficult to overcome. Recent work in deception detection will be presented at *FG 2013*: *Visions on Deception and Non-cooperation* (http://hmi.ewi.utwente.nl/vdnc-workshop/).

4.6 Interpersonal coordination

Facial expression of emotion most often occurs in an interpersonal context. Breakthroughs in automated facial expression analysis make possible to model patterns of interpersonal coordination in this context. With Messinger and colleagues (Hammal, Cohn, & Messinger, 2013; Messinger et al., 2009), we modeled mother and infant synchrony in action unit intensity and head motion. For both action unit intensity and head motion we found strong evidence of synchrony with frequent changes in phase, or direction of influence, between mother and infant. Figure 5 shows an example for mother and infant head nod amplitude. A related example for mother-infant action unit intensity is presented in Chapter 42 of this volume.

[INSERT FIGURE 5 HERE].

The pattern of association we observed for head motion and action units between mothers and infants was non-stationary with frequent changes in which partner is leading the other. Hammal and Cohn (2013) found similar nonstationarity in the head pose coordination of distressed intimate adults. Head amplitude and velocity for pitch (nod) and yaw (turn) was strongly correlated between them, with alternating periods of instability (low correlation) followed by brief stability in which one or the other partner led the other. Until recently, most research in affective computing has focused on individuals. Attention to temporal coordination expands the scope of affective computing and has implications for robot-human communication as well. To achieve more human like capabilities and make robot-human interaction feel more natural, designers might broaden their attention to consider the dynamics of communicative behavior.

4.7 Expression transfer

Many approaches to automated face analysis are invertible. That is, their parameters can be used to synthesize images that closely resemble or are nearly identical to the originals. This capability makes possible expression transfer from an image of one person's face to that of another (Theobald & Cohn, 2009). Theobald, Matthews, and their colleagues developed an early prototype for expression transfer using AAM (Theobald, Bangham, Matthews, & Cawley, 2004). This was followed by a real-time system implemented over an audio-visual link in which naïve participants interacted with realistic avatars animated by an actual person (Theobald et al., 2009) (Figure 6). Similar though less realistic approaches have been developed using CLM (Saragih, Lucey, & Cohn, 2011b). Expression transfer has been applied in computational behavioral science and media arts.

[INSERT FIGURE 6 HERE].

4.7.1 Expression transfer in computational behavioral science. In conversation, expectations about another person's identity are closely involved with their actions. Even over the telephone, when visual information is unavailable, we make inferences from the sound of the voice about the other person's gender, age, and background. To what extent do we respond to whom we think we are talking to rather than to the dynamics of their behavior? This question had been unanswered because it is difficult to separately manipulate expectations about a person's identity from their actions. An individual has a characteristic and unified appearance, head motions, facial expressions, and vocal inflection. For this reason, most studies of person perception and social expectation are naturalistic or manipulations in which behavior is artificially scripted and acted. But scripted and natural conversations have different dynamics. AFA provides a way out of this dilemma. For the first time, static and dynamic cues become separable (Boker et al., 2011).

Pairs of participants had conversations in a video-conference paradigm (Figure 6). One was a confederate for whom an AAM had previously been trained. Unbeknownst to the other participant, a resynthetized avatar was substituted for the live video of the confederate (Figure 7). The avatar had the face of the confederate or another person of same or opposite sex. All were animated by the actual motion parameters of the confederate.

[INSERT FIGURE 7 HERE].

The apparent identity and sex of a confederate was randomly assigned and the confederate was blind to the identity and sex that they appeared to have in any particular conversation. The manipulation was believable in that, when given an opportunity to guess the manipulation at the end of experiment, none of the naive participants was able to do so. Significantly, the amplitude and velocity of head movements were influenced by the dynamics (head and facial movement and vocal timing) but not the perceived sex of the partner.

These findings suggest that gender-based social expectations are unlikely to be the source of reported gender differences in head nodding between partners. Although men and women adapt to each other's head movement amplitudes it appears that adaptation may simply be a case of people (independent of sex) adapting to each other's head movement amplitude. A shared equilibrium is formed when two people interact.

4.7.2 Expression transfer in media arts

Expression transfer has been widely used in the entertainment industry where there is an increasing synnergy between computer vision and computer graphics. Well-known examples in film include *Avatar* and the *Hobbit* (http://www.iainm.com/iainm/Home.html). Emotion transfer has made significant inroads in gaming and other applications as well. Sony's Everquest II, as but one example, enables users to animate avatars in multi-person games (Hutchings, 2012).

4.8 Other applications

4.8.1 Discriminating between subtle differences in related expressions

Most efforts to detect emotion expressions have focused on the basic emotions defined by Ekman. Others have discriminated between posed and unposed smiles (Cohn & Schmidt, 2004; Valstar, Gunes, & Pantic, 2007) and between smiles of delight and actual and feigned frustration (Hoque & Picard, 2011). Ambadar and colleagues (2009) found that smiles perceived as polite, embarrassed, or amused varied in both the occurrence of specific facial actions and in their timing. Whitehill and colleagues (Whitehill, Littlewort, Fasel, Bartlett, & Movellan) developed an automatic smile detector based on appearance features. Gratch (Gratch, 2013) used automated analysis of smiles and smile controls in testing the hypothesis of Hess that smiling is determined by both social context and appraisal. Together, these studies highlight the potential of automated measurement to make fine-grained discrimination among emotion signals.

4.8.2 Marketing. Until a few years ago, self-report and focus groups were the primary means of gauging reaction to new products. With the advent of AFA, more revealing approaches have become possible. Using web-cam technology, companies are able to record thousands of viewers in dozens of countries and proces their facial expression to infer liking or disliking of commercials and products (McDuff, Kaliouby, & Picard, 2013; Szirtes, Szolgay, Utasi, Takacs, Petras, & Fodor, 2013). The methodology is well suited to the current state of the art. Participants are seated in front of a monitor, which limits out-of-plane head motion and facial expression is detected in part by knowledge of context (i.e., strong priors).

4.8.3 Drowsy-driver detection. Falling asleep while driving contributes to as many as 15% of fatal crashes. A number of systems to detect drowsy driving and take preventive actions have been proposed and are in various stages of development. Using either normal or infrared cameras, some monitor eyeblink patterns (Danisman, Bilasco, Djeraba, & Ihaddadene, 2010), while others incorporate additional behaviors, such as yawning and face touching (Matsuo & Khiat, 2012; Vural et al., 2010), head movements (Lee, Oh, Heo, & Hahn, 2008), and pupil detection (Deng, Xiong, Zhou, Gan, & Deng, 2010).

4.8.4 Instructional technology. Interest, confusion, rapport, frustration, and other emotion and cognitive-emotional states are important process variables in the classroom and in tutoring (Craig, D'Mello, Witherspoon, & Graesser, 2007). Until recently, they could be measured reliably only offline, which limited their usefulness. Recent work by Whitehill and Littlewort (Whitehill et al., 2011) evaluates the feasibility of realtime recognition. Initial results are promising. In the course of demonstrating feasibility, they found that in some contexts smiles are indicative of frustration or embarrassment rather than achievement. This finding suggests that automated methods have sufficient precision to distinguish in realtime between closely related facial actions that signal student cognitive-emotional states.

4.9 User-in-the-loop

While fully automated systems are desirable, significant advantages exist in systems that integrate user and machine input. With respect to tracking, person-specific AAM and manually initialized head tracking are two examples. Person-specific AAMs that have been trained using manually labeled video achieve higher precision than fully automatic generic AAM or CLM. Some head trackers (Jang & Kanade, 2008) achieve higher precision when users first manually initialize them on one or more frames. User-in-the-loop approaches have been applied in several studies to reveal the dynamics of different types smiles. In an early application, (Cohn & Schmidt, 2004; Schmidt, Ambadar, Cohn, & Reed, 2006) and also (Valstar, Pantic, Ambadar, & Cohn, 2006) found that manually-coded spontaneous and deliberate smiles systematically differed in their timing as measured using AFA. Extending this approach, (Ambadar et al., 2009) used a combination of manual FACS coding and automated measurement to discover variation between smiles perceived as embarrassed, amused, and polite. FACS coders first detected the onset and offset of smiles (AU 12 along with AU 6 and smile controls, e.g., AU 14). Amplitude and velocity then were measured using AFA. They found that the three types of smiles systematically varied in both shape and timing. These findings would not have been possible with only manual measurement.

Manual FACS coding is highly labor intensive. Several groups have explored the potential of AFA to reduce that burden (Simon, De la Torre, Ambadar, & Cohn, 2011; L. Zhang, Tong, & Ji, 2008). In one, referred to as Fast-FACS, manual FACS coders first detect AU peaks. An algorithm then automatically detects their onsets and offsets. Simon, De la Torre, & Cohn (2011) found that Fast-FACS achieved more than 50% reduction in the time required for manual FACS coding. Zhang, Tong, and Ji (L. Zhang et al., 2008) developed an alternative approach that uses active learning. The system performs initial labeling automatically; a FACS coder manualy makes any corrections that are needed; and the result is fed back to the system to further train the classifier. In this way, system performance is iteratively improved with a manual FACS coder in the loop. In other work, Hammal (Hammal, 2011) proposed an automatic method for successive detection of onsets, apexes, and offsets of consecutive facial expressions. All of these efforts combine manual and automated methods with the aim of achieving synergistic increases in efficiency.

5. Discussion

Automated facial analysis and synthesis is progressing rapidly with numerous initial applications in affective computing. Its vitality is evident in the breadth of approaches (in types of features, dimensionality reduction, and classifiers) and emerging uses (e.g., AU, valence, pain intensity, depression or stress, marketing, and expression transfer). Even as new applications come online, open research questions remain.

Challenges include more robust real-time systems for face acquisition, facial data extraction and representation, and facial expression recognition. Most systems perform within a range of only 15 to 20 degrees of frontal pose. Other challenges include illumination, occlusion, subtle facial expressions, and individual differences in subjects. Current systems are limited to indoors. Systems that would work in outdoor environments or with dynamic changes in illumination would greatly expand the range of possible applications. Occlusion is a problem in any context. Self-occlusion from head-turns or face touching and occlusion by other persons passing in front of the camera are common. In a 3-person social interaction in which participants have drinks, occlusion occurred about 10% of the time (Cohn & Sayette, 2010). Occlusion can spoil tracking, especially for holistic methods such as AAM, and accuracy of AU detection. Approaches to recovery of tracking following occlusion and estimation of facial actions in presence of occlusion are research topics.

Zhu, Ramanan, and their colleagues (X. Zhu, Vondrick, Ramanan, & Fowlkes, 2012) in object recognition raised the critical question: Do we need better features and classifiers or more data? The question applies as well to expression detection. Because most datasets to date are relatively small, the answer so far is unknown. The FERA GEMEP corpus (Valstar, Mehu, Jiang, Maja Pantic, & Scherer, 2012) consisted of emotion portrayals from only 10 actors. The widely-used Cohn-Kanade (Kanade et al., 2000; P. Lucey et al., 2010) and MMI (Pantic et al., 2005) corpuses have more subjects but relatively brief behavioral samples from each. To what extent is classifier performance attenuated by the relative paucity of training data? Humans are pre-adapted to perceive faces and facial expressions (i.e. strong priors) and have thousands of hours or more of experience in that task. To achieve human-like accuracy, both access to big data and learning approaches that can scale to it may be necessary. Initial evidence from object recognition (X. Zhu et al., 2012), gesture recognition (Sutton, 2011), and smile detection (Whitehill et al., 2009) suggest that datasets orders of magnitude larger than those available to date will be needed to achieve optimal AFA.

As AFA is increasingly applied to real-world problems, the ability to apply trackers and classifiers across different contexts will become increasingly important. Success will require solutions to multiple sources of database specific biases. For one, approaches that appeal to domain-specific knowledge may transfer poorly to domains in which that knowledge fails to apply. Consider the HMM approach of Li and colleagues (Li et al., 2013). They improved upon detection of AU 12 (oblique lip-corner raise) and AU 15 (lip corners pulled down) by incorporating a constraint that these AU are mutually inhibiting. While this constraint may apply in the posed and enacted portrayals of amusement that they considered, in other contexts this dependency may be troublesome. In situations in which embarrassment (Keltner & Buswell, 1997) or depressed mood (Girard et al., 2013) are likely, AU 12 and AU 15 have been found to be positively correlated. AU 15 is a “smile control,” defined as an action that counteracts the upward pull of AU 12. In both embarrassment and depression, occurrence of AU 12 increases the likelihood of AU 15. Use of HMM to encode spatial and temporal dependencies requires thoughtful application. Context (watching amusing videos versus clinical interview with depressed patients) may be especially important for HMM approaches.

Individual differences among persons affect both feature extraction and learning. Facial geometry and appearance change markedly over the course of development (Bruce & Young, 1998). Infants have larger eyes, greater fatty tissue in their cheeks, larger heads relative to their bodies, and smoother skin than adults. In adulthood, permanent lines and wrinkles become more common, and changes in fatty tissue and cartilage alter appearance. Large differences exist both between and within males and females and different ethnic groups. One of the most challenging factors may be skin color. Experience suggests that face tracking more often fails in persons that have very dark skin. Use of depth cameras, such as the Leap (Leap Motion) and Microsoft Kinect (Sutton, 2011), or infrared cameras (Buddharaju et al., 2005), may sidestep this problem. Other individual differences include characteristic patterns of emotion expression. Facial expression encodes person identity (Cohn, Schmidt, Gross, & Ekman, 2002; Peleg et al., 2006).

Individual differences affect learning, as well. Person-specific classifiers perform better than ones that are generic. Recent work by Chu and colleagues (Chu, Torre, & Cohn, 2013) proposed a method to narrow the distance between person-specific and generic classifiers. Their approach, referred to as a Selective Transfer Machine (STM), simultaneously learns the parameters of a classifier and selectively minimizes the mismatch between training and test distributions. By attenuating the influence of inherent biases in appearance, STM achieved results that surpass non-personalized generic classifiers and approach the performance of classifiers that have been trained for individual persons (i.e. person-dependent classifiers).

At present, taxonomies of facial expression are based on observer-based schemes, such as FACS. Consequently, approaches to automatic facial expression recognition are dependent on access to corpuses of well-labeled video. An open question in facial analysis is whether facial actions can be learned directly from video in an unsupervised manner. That is, can the taxonomy be learned directly from video? And unlike FACS and similar systems that were initially developed to label static expressions, can we learn dynamic trajectories of facial actions? In our preliminary findings on unsupervised learning using RU-FACS database (Zhou, De la Torre, & Cohn, 2010), moderate agreement between facial actions identified by unsupervised analysis of face dynamics and FACS approached the level of agreement that has been found between independent FACS coders. These findings suggest that unsupervised learning of facial expression is a promising alternative to supervised learning of FACS-based actions.

Because unsupervised learning is fully empirical, it potentially can identify regularities in video that have not been anticipated by the top-down approaches such as FACS. New discoveries become possible. Recent efforts by Guerra-Filho and Aloimonos (2007) to develop vocabularies and grammars of human actions suggest that this may be a fruitful approach.

Facial expression is one of several modes of nonverbal communication. The contribution of different modalities may well vary with context. In mother-infant interaction, touch appears to be especially important and tightly integrated with facial expression and head motion (Messinger et al., 2009). In depression, vocal prosody is highly related to severity of symptoms. We found that over 60% of the variance in depression severity could be accounted for by vocal prosody. Multimodal approaches that combine face, body language, and vocal prosody represent upcoming areas of research. Interdisciplinary efforts will be needed to progress in this direction.

While much basic research still is needed, AFA is becoming sufficiently mature to address real-world problems in behavioral science, biomedicine, affective computing, and entertainment. The range and depth of applications is just beginning.

Figures

Figure 1. Basic emotions. from left to right: amusement, sadness, anger, fear, surprise, disgust, contempt, and embarrassment.

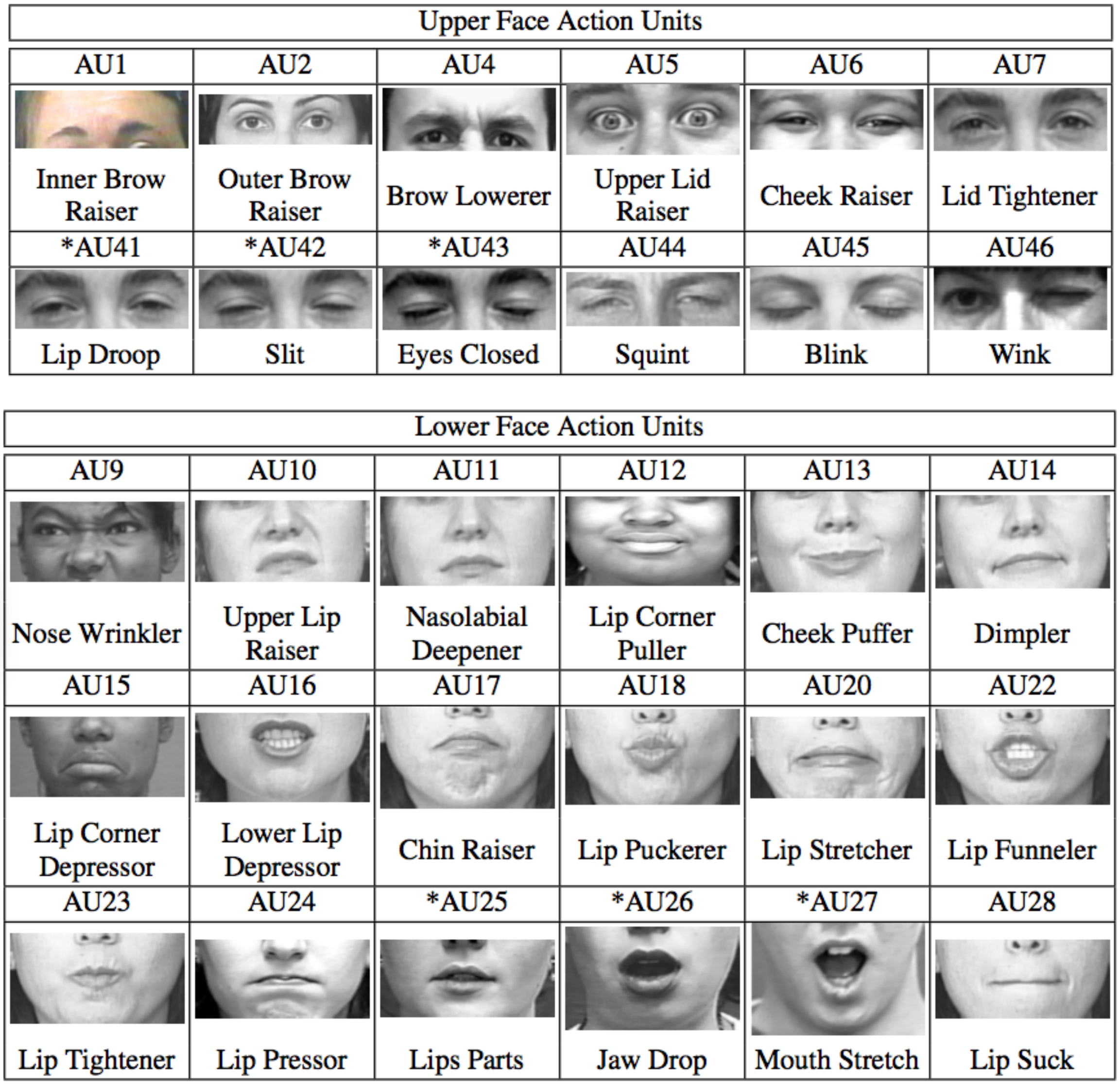

Figure 2. Action units (AUs), facial action coding system.

Sources: Ekman & Friesen (1978); Ekman et al., (2002). Images from C-K database, Kanade et al. (2000).

Figure 3. Intensity variation in AU 12.

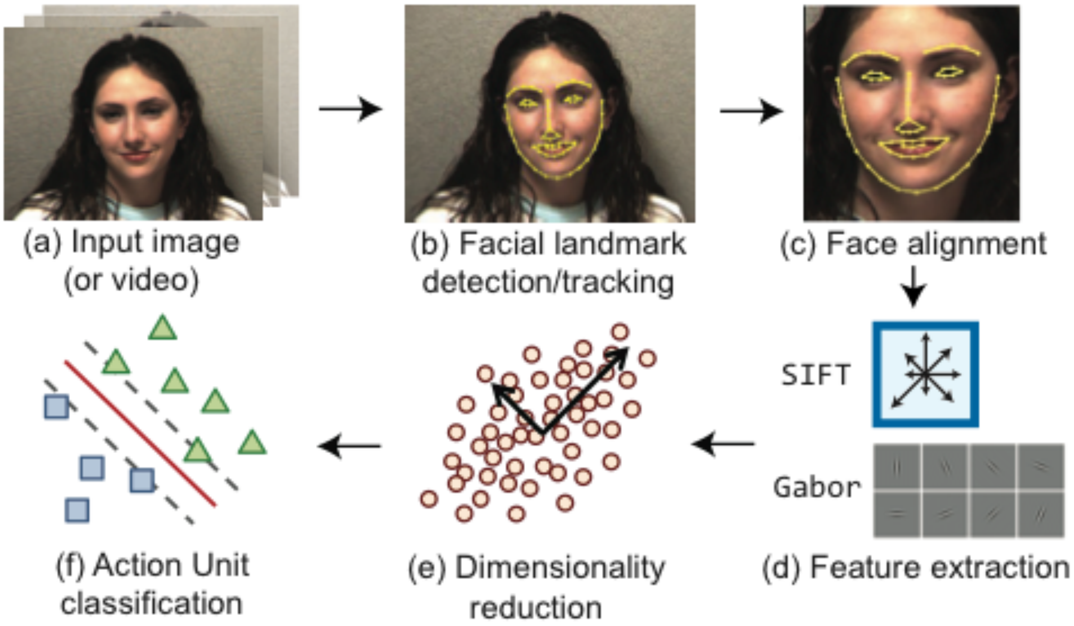

Figure 4. Example of the facial action unit recognition system.

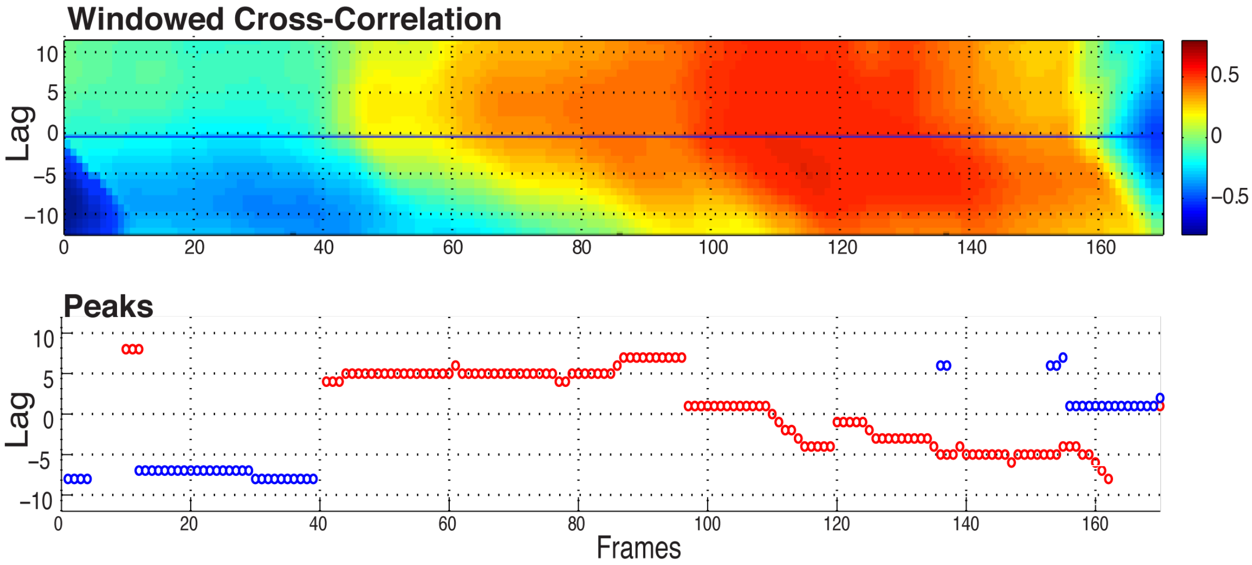

Figure 5. Top panel: Windowed cross-correlation within a 130-frame sliding window between mother and infant head-pitch amplitude. The area above the midline (Lag > 0) represents the relative magnitude of correlations for which the mother's head amplitude predicts her infant's; the corresponding area below the midline (Lag < 0) represents the converse. The midline (Lag = 0) indicates that both partners are changing their head amplitudes at the same time. Positive correlations (red) convey that the head amplitudes of both partners are changing in the same way (i.e., increasing together or decreasing together). Negative correlation (blue) conveys that the head amplitudes of both partners are changing in the opposite way (e.g., head amplitude of one partner increases as that of the other partner decreases). Note that the direction of the correlations changes dynamically over time. Bottom panel: Peaks (r >.40) in the windowed cross-correlations as found using an algorithm proposed by Boker (Boker, Rotondo, Xu, & King, 2002).

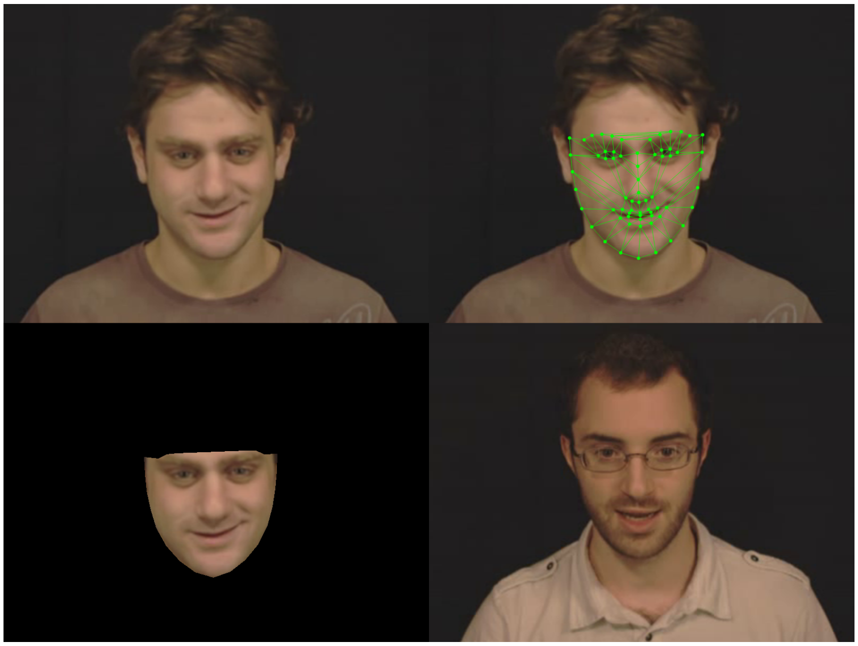

Figure 6. Illustration of video-conference paradigm. Clockwise from upper left: Video of the source person; AAM tracking of the source person; their partner; and the AAM reconstruction that is viewed by the partner.

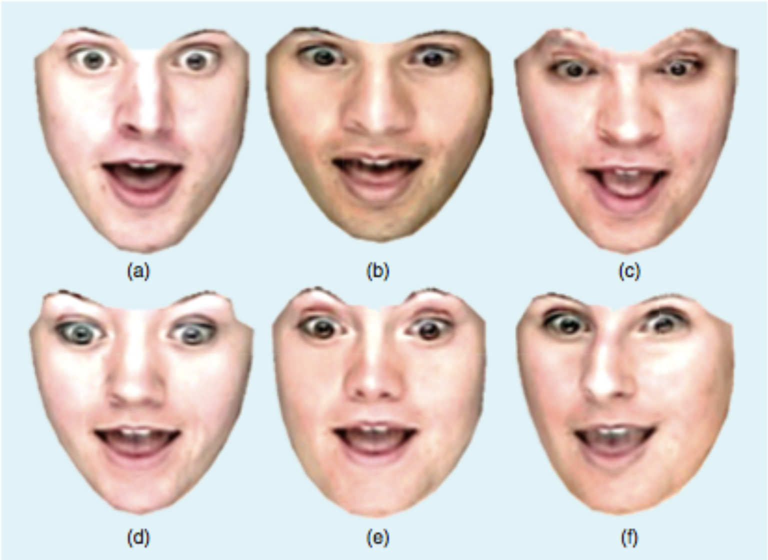

Figure 7. Applying expressions of a male to the appearances of other persons. In (a), the avatar has the appearance of the person whose motions were tracked. In (b) and (c), the avatars have the same-sex appearance. Parts (d) through (f) show avatars with opposite-sex appearances.

Source: Images courtesy of the American Psychological Association.

Acknowledgements

Research reported in this chapter was supported in part by the National Institutes of Health (NIH) under Award Number MHR01MH096951 and by the US Army Research Laboratory (ARL) under the Collaborative Technology Alliance Program, Cooperative Agreement W911NF-10-2-0016. We thank Nicole Siverling, Wen-Sheng Chu, and Zakia Hammal for their help. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or ARL.

References

Ambadar, Z., Cohn, J. F., & Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior, 33*(1), 17-34.

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Prkachin, K. M., & Solomon, P. E. (2009). The painful face: Pain expression recognition using active appearance models. *Image and Vision Computing, 27*(12), 1788–1796.

Baron-Cohen, S. (2003). *Mind reading: The interactive guide to emotion*

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006a). Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*. doi: http://mplab.ucsd.edu/~marni/pubs/Bartlett_JMM06.pdf

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., & Movellan, J. R. (2006b). *Fully automatic facial action recognition in spontaneous behavior*. Paper presented at the Seventh IEEE International Conference on Automatic Face and Gesture Recognition, Southampton, England.

Belkin, M., & Niyogi, P. (2001). Laplacian Eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems, 14*, 586–691.

Bobick, A. F., & Davis, J. W. (2001). The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(3), 257-267.

Boker, S. M., Cohn, J. F., Theobald, B. J., Matthews, I., Mangini, M., Spies, J. R., . . . Brick, T. R. (2011). Something in the way we move: Motion, not perceived sex, influences nods in conversation. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 874-891. doi: 10.1037/a0021928

Boker, S. M., Rotondo, J. L., Xu, M., & King, K. (2002). Windowed cross–correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods, 7*(1), 338-355.

Bråten, S. (2006). *Intersubjective communication and cmotion in early ontogeny* New York, NY: Cambridge University.

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

Buck, R. (1984). *The Communication of emotion*. New York: The Guilford Press.

Buddharaju, P., Dowdall, J., Tsiamyrtzis, P., Shastri, D., Pavlidis, I., & Frank, M. G. (2005, June). *Automatic thermal monitoring system (ATHEMOS) for deception detection*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Burrows, A., & Cohn, J. F. (In press). Comparative anatomy of the face. In S. Z. Li (Ed.), *Handbook of biometrics* (2nd ed.). Berlin Heidelberg: Springer.

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45*(1), 16-28.

Cai, D., He, X., Zhou, K., Han, J., & Bao, H. (2007). Locality sensitive discriminant analysis. In International Joint Conference on Artifical intelligence. *International Joint Conference on Artificial Intelligence*.

Campos, J. J., Barrett, K. C., Lamb, M. E., Goldsmith, H. H., & Stenberg, C. (1983). Socioemotional development. In M. M. Haith & J. J. Campos (Eds.), *Handbook of child psychology* (4th ed., Vol. II, pp. 783-916). New York: Wiley.

Chang, K. Y., Liu, T. L., & Lai, S. H. (2009). *Learning partially-observed hidden conditional random fields for facial expression recognition*. Paper presented at the IEEE International Conference on Computer Vision and Pattern Recognition.

Chang, Y., Hu, C., Feris, R., & Turk, M. (2006). Manifold based analysis of facial expression. *Image and Vision Computing, 24*, 605-614.

Chetverikov, D., & Peteri, R. (2005). A brief survey of dynamic texture description and recognition *Computer Recognition Systems: Advances in Soft Computing* (Vol. 30, pp. 17-26).

Chew, S. W., Lucey, P., Lucey, S., Saragih, J. M., Cohn, J. F., Matthews, I., & Sridharan, S. (2012). In the pursuit of effective affective computing: The relationship between features and registration. *IEEE Transactions on Systems, Man, and Cybernetics - Part B, 42*(4), 1-12. doi: 10.1109/TSMCB.2012.2194485

Chu, W.-S., Torre, F. D. l., & Cohn, J. F. (2013). *Selective transfer machine for personalized facial action unit detection*. Paper presented at the IEEE International Conference on Computer Vision and Pattern Recognition.

Cohen, I., Sebe, N., Garg, A., Lew, M. S., & Huang, T. S. (2003). Facial expression recognition from video sequences. *Computer Vision and Image Understanding, 91*(1-2), 160-187.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan & J. J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science* (pp. 203-221). New York, NY: Oxford University.

Cohn, J. F., & Ekman, P. (2005). Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences* (pp. 9-64). New York: Oxford.

Cohn, J. F., & Sayette, M. A. (2010). Spontaneous facial expression in a small group can be automatically measured: An initial demonstration. *Behavior Research Methods, 42*(4), 1079-1086.

Cohn, J. F., & Schmidt, K. L. (2004). 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. L., Gross, R., & Ekman, P. (2002). Individual differences in facial expression: Stability over time, relation to self-reported emotion, and ability to inform person identification. *International Conference on Multimodal User Interfaces*, 491-496.

Comon, P. (1994). Independent component analysis: A new concept? *Signal Processing, 36*(3), 287–314.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(6), 681-685.

Craig, S. D., D'Mello, S. K., Witherspoon, A., & Graesser, A. (2007). Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive-affective states during learning. *Cognition and Emotion, 22*, 777-788.

Cristinacce, D., & Cootes, T. F. (2006). *Feature detection and tracking with constrained local models*. Paper presented at the British Machine Vision Conference.

Dalal, N., & Triggs, B. (2005). *Histograms of oriented gradients for human detection*. Paper presented at the IEEE International Conference on Computer Vision and Pattern Recognition.

Danisman, T., Bilasco, I. M., Djeraba, C., & Ihaddadene, N. (2010). *Drowsy driver detection system using eye blink patterns*. Paper presented at the International Cofnerence on Machine and Web Intelligence (ICMWI). http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5628557

Darwin, C. (1872/1998). *The expression of the emotions in man and animals (3rd edition)*. New York: Oxford University.

De la Torre, F., & Cohn, J. F. (2011). Visual analysis of humans: Facial expression analysis. In T. B. Moeslund, A. Hilton, A. U. Volker Krüger & L. Sigal (Eds.), *Visual analysis of humans: Looking at people* (pp. 377-410): Springer.

Deng, L., Xiong, X., Zhou, J., Gan, P., & Deng, S. (2010). *Fatigue detection based on infrared video puillography*. Paper presented at the Bioinformatics and Biomedical Engineering (iCBBE).

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3/4), 169-200.

Ekman, P. (2009). *Telling lies*. New York, NY: Norton.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to emotions from facial cues*. Englewood Cliffs, NJ: Prentice-Hall Inc.

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

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system*: Research Nexus, Network Research Information, Salt Lake City, UT.

Ekman, P., Huang, T. S., & Sejnowski, T. J. (1992). Final report to NSF of the planning workshop on facial expression understanding. Washington, D.C.: National Science Foundation.

Ekman, P., & Rosenberg, E. (2005). *What the face reveals* (2nd ed.). New York: 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.

Feldman Barrett, L., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*.

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

Fox, E. (2008). *Emotion science: Cognitive and neuroscientific approaches to understanding human emotions*. NY: Palgrave Macmillan.

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

Girard, J. M. (2013). *Automatic detection and intensity estimation of spontaneous smiles.* (M.S.), University of Pittsburgh, Pittsburgh.

Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., & Rosenwald, D. (2013). *Social risk and depression: Evidence from manual and automatic facial expression analysis*. Paper presented at the Tenth International Conference on Automatic Face and Gesture Recognition, Shanghai, China.

Gratch, J. (2013). *Felt emotion and social context determine the intensity of smiles in a competitive video game*. Paper presented at the IEEE International Conference on Automatic Face and Gesture Recognition, Shanghai, China.

Guerra-Filho, G., & Aloimonos, Y. (2007). A language for human action. *Computer, 40*(5), 42-51.

Hammal, Z. (Ed.). (2011). *Efficient detection of consecutive facial expression apices using biologically based log-normal filters*. Berlin: Springer.

Hammal, Z., & Cohn, J. F. (2012). *Automatic detection of pain intensity*. Paper presented at the International Conference on Multimodal Interaction, Santa Monica, CA.

Hammal, Z., Cohn, J. F., Baiile, T., George, D. T., Saragih, J. M., Nuevo-Chiquero, J., & Lucey, S. (2013). Temporal coordination of head motion in couples with history of interpersonal violence. *IEEE International Conference on Automatic Face and Gesture Recognition*.

Hammal, Z., Cohn, J. F., & Messinger, D. S. (2013). *Head movement dynamics during normal and perturbed parent-infant interaction*. Paper presented at the Affective Computing and Intelligent Interaction, Geneva, Switzerland.

Hammal, Z., & Kunz, M. (2012). Pain monitoring: A dynamic and context-sensitive system. *Pattern Recognition, 45*, 1265–1280.

Hoque, M. E., & Picard, R. W. (2011). Acted vs. natural frustration and delight: Many people smile in natural frustration. *IEEE International Conference on Automatic Face and Gesture Recognition (FG2011).*

Hotelling, H. (1933). Analysis of complex statistical variables into principal components. *Journal of Educational Psychology, 24*(6), 417-441.

Hutchings, E. (2012). Sony technology gives gaming avatars same facial expressions as players. Retrieved March 24, 2013, from http://www.psfk.com/2012/08/avatars-human-facial-expressions.html

Izard, C. E. (1977). *Human emotions*. New York: Plenum.

Jang, J.-S., & Kanade, T. (2008, September). *Robust 3D head tracking by online feature registration*. Paper presented at the IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands.

Jeni, L. A., Lorincz, A., Nagy, T., Palotai, Z., Sebok, J., Szabo, Z., & Taka, D. (2012). 3D shape estimation in video sequences provides high precision evaluation of facial expressions. *Image and Vision Computing Journal.*

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1233-1258.

Joshi, J., Dhall, A., Goecke, R., Breakspear, M., & Parker, G. (2012). *Neural-net classification for spatio-temporal descriptor based depression analysis*. Paper presented at the IEEE International Conference on Pattern Recognition, Tsukuba, Japan.

Kaltwang, S., Rudovic, O., & Pantic, M. (2012). Continuous pain intensity estimation from facial expressions. *Lecture Notes in Comptuer Science, 7432*, 368-377.

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

Kanade, T., Cohn, J. F., & Tian, Y. (2000). *Comprehensive database for facial expression analysis*. Paper presented at the Fourth International Conference on Automatic Face and Gesture Recognition, Grenoble. http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*(3), 250-270.

Keltner, D., & Ekman, P. (2000). Facial expression of emotion. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (2nd ed., pp. 236-249). New York: Guilford.

Knapp, M. L., & Hall, J. A. (2010). *Nonverbal behavior in human communication* (7th ed.). Boston: Wadsworth/Cengage.

Koelstra, S., & Pantic, M. (2008). *Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics* Paper presented at the FG, Amsterdam.

Leap Motion. (2013). Leap. from http://thecomputervision.blogspot.com/2012/05/leap-motion-new-reliable-low-cost-depth.html

Lee, D., Oh, S., Heo, S., & Hahn, M.-S. (2008). *Drowsy driving detection based on the driver's head movement using infrared sensors*. Paper presented at the Second International Symposium on Universal Communication.

Li, Y., Chen, J., Zhao, Y., & Ji, Q. (2013). Data-free prior model for facial action unit recognition. *Transactions on Affective Computing*.

Littlewort, G. C., Bartlett, M. S., Fasel, I. R., Susskind, J., & Movellan, J. R. (2006). Dynamics of facial expression extracted automatically from video. *Journal of Image & Vision Computing, 24*(6), 615-625.

Littlewort, G. C., Bartlett, M. S., & Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing, 27*(12), 1797-1803.

Lucey, P., Cohn, J. F., Howlett, J., Lucey, S., & Sridharan, S. (2011). Recognizing emotion with head pose variation: Identifying pain segments in video. *IEEE Transactions on Systems, Man, and Cybernetics - Part B, 41*(3), 664-674.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J. M., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kande Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression. *Third IEEE Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*, 1-8.

Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., & Matthews, I. (2011). Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database. *IEEE International Conference on Automatic Face and Gesture Recognition (FG2011)*.

Lucey, S., Wang, Y., Saragih, J. M., & Cohn, J. F. (2010). Non-rigid face tracking with enforced convexity and local appearance consistency constraint. *Image and Vision Computing, 28*(5), 781-789.

Martin, P., & Bateson, P. (2007). *Measuring behavior: An introductory guide* (3rd ed.). Cambridge, UK: Cambridge University Press.

Mase, K. (1991). Recognition of facial expression from optical flow. *IEICE Transactions on*

Information and Systems, E74-D(10), 3474-3483.

Matsuo, H., & Khiat, A. (2012). *Prediction of drowsy driving by monitoring driver's behavior*. Paper presented at the International Conference on Pattern Recognition (ICPR).

Matthews, I., Xiao, J., & Baker, S. (2007). 2D vs. 3D deformable face models: Representational power, construction, and real-time fitting. *International Journal of Computer Vision, 75*(1), 93–113.

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). DISFA: A non-posed facial expression video database with FACS-AU intensity coding. *IEEE Transactions on Affective Computing*.

McDuff, D., Kaliouby, R. E., & Picard, R. (2013). *Predicting online media effectiveness based on smile responses gathered over the internet*. Paper presented at the Tenth International Conference on Automatic Face and Gesture Recognition, Shanghai, China.

McDuff, D., Kaliouby, R. E., Senechal, T., Amr, M., Cohn, J. F., & Picard, R. (2013). *AMFED facial expression dataset: Naturalistic and spontaneous facial expressions collected "in-the-wild"*. Paper presented at the Fifth IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Portland, OR.

Mehrabian, A. (1998). Correlations of the PAD Emotion Scales with self-reported satisfaction in marriage and work. *Genetic, Social, and General Psychology Monographs 124(3):311-334]* (3), 311-334.

Messinger, D. S., Mahoor, M. H., Chow, S. M., & Cohn, J. F. (2009). Automated measurement of facial expression in infant-mother interaction: A pilot study. *Infancy, 14*(3), 285-305.

Michael, N., Dilsizian, M., Metaxas, D., & Burgoon, J. K. (2010). *Motion profiles for deception detection using visual cues*. Paper presented at the European Conference on Computer Vision.

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(10), 1615-1630.

Morency, L.-P. (2008). Watson user guide (Version 2.6A).

Movellan, J. R. (Undated). Tutorial on Gabor Filters. University of California San Diego.

Pandzic, I. S., & Forchheimer, R. (Eds.). (2002). *MPEG-4 facial animation: The standard, implementation and applications*. New York: John Wiley.

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

Pantic, M., Valstar, M. F., Rademaker, R., & Maat, L. (2005). *Web-based database for facial expression analysis*. Paper presented at the IEEE International Conference on Multimodal Interfaces, Amsterdam.

Parke, F. I., & Waters, K. (1996). *Computer facial animation*. Wellesley, MA: A.K. Peters.

Peleg, G., Katzir, G., Peleg, O., Kamara, M., Brodsky, L., Hel-Or, H., . . . Nevo, E. (2006, October 24, 2006). From the cover: Hereditary family signature of facial expression. Retrieved 43, 103, from http://www.pnas.org/cgi/content/abstract/103/43/15921

Plutchik, R. (1979). *Emotion: A psychoevolutionary synthesis*. NY: Harper & Row.

Prkachin, K. M., & Solomon, P. E. (2008). The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain, 139*, 267-274.

Rosenthal, R. (2005). Conducting judgment studies. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), *Handbook of nonverbal behavior research methods in the affective sciences* (pp. 199-236). NY: Oxford.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*(5500), 2323–2326.

Russell, J. A., & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology, 48*(5), 1290-1298.

Saragih, J. M., Lucey, S., & Cohn, J. F. (2011a). Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision, 91*(2), 200-215. doi: 10.1007/s11263-010-0335-9

Saragih, J. M., Lucey, S., & Cohn, J. F. (2011b). *Real-time avatar animation from a single image*. Paper presented at the 9th IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA.

Sayette, M. A., Creswell, K. G., Dimoff, J. D., Fairbairn, C. E., Cohn, J. F., Heckman, B. W., . . . Moreland, R. L. (2012). Alcohol and group formation: A multimodal investigation of the effects of alcohol on emotion and social bonding. *Psychological Science, 23*(8), 869-878 doi: 10.1177/0956797611435134

Scherer, S., Stratou, G., Gratch, J., Boberg, J., Mahmoud, M., Rizzo, A. S., & Morency, L.-P. (2013). *Automatic behavior descriptors for psychological disorder analysis*.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology, 44*, 229-237.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review, 61*, 81-88.

Schmidt, K. L., Ambadar, Z., Cohn, J. F., & Reed, L. I. (2006). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior, 30*, 37-52.

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary perspectives in facial expression research. *Yearbook of Physical Anthropology, 116*, 8-24.

Schokopf, B., Smola, A., & Muller, K. (1997). Kernel principal component analysis. *Artificial Neural Networks*, 583–588.

Shang, C. F., & Chan, K. P. (2009). *Nonparametric discriminant HMM and application to facial expression recognition*. Paper presented at the IEEE International Conference on Computer Vision and Pattern Recognition.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Simon, T. K., De la Torre, F., Ambadar, Z., & Cohn, J. F. (2011, October 9 to 12). *Fast-FACS: A computer vision assisted system to increase the speed and reliability of manual FACS coding*. Paper presented at the HUMAINE Association Conference on Affective Computing and Intelligent Interaction, Memphis, Tennesse.

Sutton, J. (2011). Body part recognition: Making Kinect robust. *IEEE International Conference on Automatic Face and Gesture Recognition (FG2011).*

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*(4), 297-303.

Theobald, B. J., Bangham, J. A., Matthews, I., & Cawley, G. C. (2004). Near-videorealistic synthetic talking faces: Implementation and evaluation. *Speech Communication, 44*, 127-140.

Theobald, B. J., & Cohn, J. F. (2009). Facial image synthesis. In D. Sander & K. R. Scherer (Eds.), *Oxford companion to emotion and the affective sciences* (pp. 176-179). NY: Oxford University Press.

Theobald, B. J., Matthews, I., Cohn, J. F., & Boker, S. M. (2007, November). *Real-time expression cloning using appearance models*. Paper presented at the Proceedings of the ACM International Conference on Multimodal Interfaces, Nagoya, Japan.

Theobald, B. J., Matthews, I., Mangini, M., Spies, J. R., Brick, T., Cohn, J. F., & Boker, S. M. (2009). Mapping and manipulating facial expression. *Language and Speech, 52*(2 & 3), 369-386.

Tian, Y., Cohn, J. F., & Kanade, T. (2005). Facial expression analysis. In S. Z. Li & A. K. Jain (Eds.), *Handbook of face recognition* (pp. 247-276). New York: Springer.

Tian, Y., Kanade, T., & Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*(2), 97-115.

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgements. *Journal of Counseling Psychology, 22,* 358-376.

Tong, Y., Chen, J., & Ji, Q. (2010). A unified probabilistic framework for spontaneous facial action modeling and understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(2), 258-273.

Tong, Y., Liao, W., & Ji, Q. (2007). Facial action unit recognition by exploiting their dynamic and semantic relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(10), 1683 - 1699.

Tronick, E. Z. (1989). Emotions and emotional communication in infants. *American Psychologist, 44*(2), 112-119.

Tsiamyrtzis, P., J. Dowdall, Shastri, D., Pavlidis, I. T., Frank, M. G., & Ekman, P. (2006). Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision, 71*, 197-214.

Valstar, M. F., Gunes, H., & Pantic, M. (2007). How to distinguish posed from spontaneous smiles using geometric features. *ACM International Conference on Multimodal Interfaces*, 38-45.

Valstar, M. F., Mehu, M., Jiang, B., Maja Pantic, F., & Scherer, K. (2012). Meta-analysis of the first facial expression recognition challenge. *IEEE Transactions on Systems, Man, and Cybernetics - Part B, 42*(4). doi: 10.1109/TSMCB.2012.2200675

Valstar, M. F., Mehu, M., Jiang, B., Pantic, M., & Scherer, K. (2012). Meta-analyis of the first facial expression recognition challenge. *IEEE Transactions of Systems, Man and Cybernetics -- Part B, 42*(4), 966 - 979

Valstar, M. F., & Pantic, M. (2007). *Combined support vector machines and hidden Markov models for modeling facial action temporal dynamics*. Paper presented at the IEEE Conference on Computer Vision (ICCV'07) Rio de Janiero, Brazil.

Valstar, M. F., Pantic, M., Ambadar, Z., & Cohn, J. F. (2006). *Spontaneous vs. posed facial behavior: Automatic analysis of brow actions*. Paper presented at the ACM International Conference on Multimodal Interfaces, Banff, Canada.

Valstar, M. F., Pantic, M., & Patras, I. (2004). *Motion history for facial action detection in video*. Paper presented at the Proceedings of the IEEE Conference on Systems, Man, and Cybernetics, the Hague, the Netherlands.

Valstar, M. F., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., . . . Pantic, M. (2013).

AVEC 2013 – The continuous audio/visual emotion and depression recognition challenge. Paper presented at the Third International Audio/Video Challenge Workship, Barcelona, Spain.

Viola, P. A., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision, 57*(2), 137-154.

Vural, E., Bartlett, M., Littlewort, G., Cetin, M., Ercil, A., & Movellan, J. (2010). *Discrimination of moderate and acute drowsiness based on spontaneous facial expressions*. Paper presented at the IEEE International Conferene on Machine Learning, Istanbul.

Wang, P., Barrett, F., Martin, E., Milonova, M., Gurd, R. E., Gur, R. C., . . . Verma, R. (2008). Automated video-based facial expression analysis of neuropsychiatric disorders. *Journal of Neuroscience Methods, 168*, 224–238.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219-235.

Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M. S., & Movellan, J. R. (2009). Towards practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(11), 2106 - 2111.

Whitehill, J., Serpell, Z., Foster, A., Lin, Y.-C., Pearson, B., Bartlett, M., & Movellan, J. (2011). *Towards an optimal affect-sensitive instructional system of cognitive skills*. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition Workshop on Human Communicative Behavior.

Xiao, J., Baker, S., Matthews, I., & Kanade, T. (2004). Real-time combined 2D+3D active appearance models. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 535-542.

Xiao, J., 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. (1997). Recognizing human facial expression from long image sequence using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 18*, 636-642.

Yang, P., Liu, Q., & Metaxas, D. N. (2009). Boosting encoded dynamic features for facial expression recognition *Pattern Recognition Letters, 30*, 132-139.

Yu, X., Zhang, S., Yan, Z., Yang, F., Huang, J., Dunbar, N., . . . Metaxas, D. (2013). *Interactional dissynchrony a clue to deception: insights from automated analysis of nonverbal visual cues*. Paper presented at the Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, Hawaii.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, 31*(1), 31-58. doi: 10.1109/TPAMI.2008.52

Zhang, C., & Zhang, Z. (2010). A survey of recent advances in face detection *Microsoft Research Technical Report*.

Zhang, L., Tong, Y., & Ji, Q. (2008). Active image labeling and its application to facial action labeling. In D. Forsyth, P. Torr & A. Zisserman (Eds.), *Lecture notes in computer science: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II* (Vol. 5303/2008, pp. 706-719). Berlin / Heidelberg: Springer.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., & Liu, P. (2013). *A 3D spontaneous dynamic facial expression database*. Paper presented at the Tenth

International Conference on Automatic Face and Gesture Recognition, Shanghai, China.

Zhao, G., & Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhou, F., De la Torre, F., & Cohn, J. F. (2010). Unsupervised discovery of facial events. *IEEE International Conference on Computer Vision and Pattern Recognition*.

Zhu, X., Vondrick, C., Ramanan, D., & Fowlkes, C. C. (2012). *Do we need more training data or better models for object detection*. Paper presented at the British Machine Vision Conference, Surrey, England.

Zhu, Y., De la Torre, F., Cohn, J. F., & Zhang, Y.-J. (2011). Dynamic cascades with bidirectional bootstrapping for action unit detection in spontaneous facial behavior. *IEEE Transactions on Affective Computing, 2*(2), 1-13.