Automatic Recognition of Facial Actions in Spontaneous Expressions

Marian Stewart Bartlett[1], Gwen C. Littlewort[1], Mark G. Frank[2], Claudia Lainscsek[1], Ian R. Fasel[1], Javier R. Movellan[1]

[1]Institute for Neural Computation, University of California, San Diego.
mbartlet@ucsd.edu, gwen@mplab.ucsd.edu, clainscsek@ucsd.edu, ianfasel@cogsci.ucsd.edu, movellan@mplab.ucsd.edu

[2]Department of Communication, University at Buffalo, State University of New York.
mfrank83@buffalo.edu

***Abstract*— Spontaneous facial expressions differ from posed expressions in both which muscles are moved, and in the dynamics of the movement. Advances in the field of automatic facial expression measurement will require development and assessment on spontaneous behavior. Here we present preliminary results on a task of facial action detection in spontaneous facial expressions. We employ a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). The system automatically detects frontal faces in the video stream and coded each frame with respect to 20 Action units. The approach applies machine learning methods such as support vector machines and AdaBoost, to texture-based image representations. The output margin for the learned classifiers predicts action unit intensity. Frame-by-frame intensity measurements will enable investigations into facial expression dynamics which were previously intractable by human coding.**

I. Introduction

A. The facial action coding system

In order to objectively capture the richness and complexity of facial expressions, behavioral scientists have found it necessary to develop objective coding standards. The facial action coding system (FACS) [17] is the most widely used expression coding system in the behavioral sciences. A human coder decomposes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscles. An example is shown in Figure 1.

FACS provides an objective and comprehensive language for describing facial expressions and relating them back to what is known about their meaning from the behavioral science literature. Because it is comprehensive, FACS also allows for the discovery of new patterns related to emotional or situational states. For example, what are the facial behaviors associated with driver fatigue? What are the facial behaviors associated with states that are critical for automated tutoring systems, such as interest, boredom, confusion, or comprehension? Without an objective facial measurement system, we have a chicken-and-egg problem. How do we build systems to detect comprehension, for example, when we don't know for certain what faces do when students are comprehending? Having subjects pose states such as comprehension and confusion is of limited use since there is a great deal of evidence that people do different things with their faces when posing versus during a spontaneous experience (e.g. [8], [14]). Likewise, subjective labeling of expressions has also been shown to be less reliable than objective coding for finding relationships between facial expression and other state variables. Some examples of this are discussed below, namely the failure of subjective labels to show associations between smiling and other measures of happiness, as well as failure of naive subjects to differentiate deception and intoxication from facial expression, whereas reliable differences were shown with FACS.

Objective coding with FACS is one approach to the problem of developing detectors for state variables such as comprehension and confusion, although not the only one. Machine learning of classifiers from a database of spontaneous examples of subjects in these states is another viable approach, although this carries with it issues of eliciting the state, and assessment of whether and to what degree the subject is experiencing the desired state. Experiments using FACS face the same challenge, although computer scientists can take advantage of a large body of literature in which this has already been done by behavioral scientists. Once a database exists, however, in which a state has been elicited, machine learning can be applied either directly to image primitives, or to facial action codes. It is an open question whether intermediate representations such as FACS are the best approach to recognition, and such questions can begin to be addressed with databases such as the one described in this paper. Regardless of which approach is more effective, FACS provides a general purpose representation that can be useful for many applications. It would be time consuming to collect a new database and train application-specific detectors directly from image primitives for each new appli-

© 2006 ACADEMY PUBLISHER

cation. The speech recognition community has converged on a strategy that combines intermediate representations from phoneme detectors plus context-dependent features trained directly from the signal primitives, and perhaps a similar strategy will be effective for automatic facial expression recognition.

There are numerous examples in the behavioral science literature where FACS enabled discovery of new relationships between facial movement and internal state. For example, early studies of smiling focused on subjective judgments of happiness, or on just the mouth movement (zygomatic major). These studies were unable to show a reliable relationship between expression and other measures of enjoyment, and it was not until experiments with FACS measured facial expressions more comprehensively, that a strong relationship was found: Namely that smiles which featured both orbicularis oculi (AU6), as well as zygomatic major action (AU12), were correlated with self-reports of enjoyment, as well as different patterns of brain activity, whereas smiles that featured only zygomatic major (AU12) were not (e.g. [16]). Research based upon FACS has also shown that facial actions can show differences between genuine and faked pain [8], and between those telling the truth and lying at a much higher accuracy level than naive subjects making subjective judgments of the same faces [26]. Facial Actions can predict the onset and remission of depression, schizophrenia, and other psychopathology [20], can discriminate suicidally from non-suicidally depressed patients [27], and can predict transient myocardial ischemia in coronary patients [42]. FACS has also been able to identify patterns of facial activity involved in alcohol intoxication that observers not trained in FACS failed to note [44].

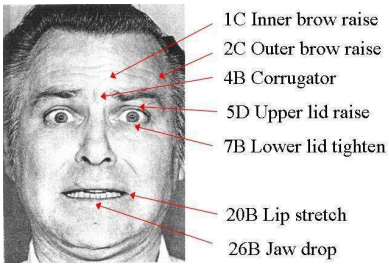

Figure 1. Example FACS codes for a prototypical expression of fear. Spontaneous expressions may contain only a subset of these Action Units.

Although FACS has a proven record for the scientific analysis of facial behavior, the process of applying FACS to videotaped behavior is currently done by hand and has been identified as one of the main obstacles to doing research on emotion [15], [25]. FACS coding is currently performed by trained experts who make perceptual judgments of video sequences, often frame by frame. It requires approximately 100 hours to train a person to make these judgments reliably and pass a standardized test for reliability. It then typically takes over two hours to code comprehensively one minute of video. Furthermore, although humans can be trained to code reliably the morphology of facial expressions (which muscles are active) it is very difficult for them to code the dynamics of the expression (the activation and movement patterns of the muscles as a function of time). There is good evidence suggesting that such expression dynamics, not just morphology, may provide important information [18]. For example, spontaneous expressions have a fast and smooth onset, with distinct facial actions peaking simultaneously, whereas posed expressions tend to have slow and jerky onsets, and the actions typically do not peak simultaneously [24].

Significant advances in computer vision open up the possibility of automatic coding of facial expressions at the level of detail required for such behavioral studies. Automated systems would have a tremendous impact on basic research by making facial expression measurement more accessible as a behavioral measure, and by providing data on the dynamics of facial behavior at a resolution that was previously unavailable. Such systems would also lay the foundations for computers that can understand this critical aspect of human communication. Computer systems with this capability have a wide range of applications in basic and applied research areas, including man-machine communication, security, law enforcement, psychiatry, education, and telecommunications [39].

B. Spontaneous Facial Expression

The importance of spontaneous behavior for developing and testing computer vision systems becomes apparent when we examine the neurological substrate for facial expression. There are two distinct neural pathways that mediate facial expressions, each one originating in a different area of the brain. Volitional facial movements originate in the cortical motor strip, whereas the more involuntary, emotional facial actions, originate in the subcortical areas of the brain (e.g. [33]. Research documenting these differences was sufficiently reliable to become the primary diagnostic criteria for certain brain lesions prior to modern imaging methods (e.g. [6].) The facial expressions mediated by these two pathways have differences both in which facial muscles are moved and in their dynamics. The two neural pathways innervate different facial muscles [41], and there are related differences in which muscles are moved when subjects are asked to pose an expression such as fear versus when it is displayed spontaneously [14]. Subcortically initiated facial expressions (the involuntary group) are characterized by synchronized, smooth, symmetrical, consistent, and reflex-like facial muscle movements whereas cortically initiated facial expressions are subject to volitional real-time control and tend to be less smooth, with more variable dynamics (see review by Rinn [40].) However, precise characterization of spontaneous expression dynamics has been slowed down by the need to use non-invasive technologies (e.g. video), and the difficulty of manually coding expression intensity frame-by-frame. Thus the importance of video based automatic coding systems.

© 2006 ACADEMY PUBLISHER

These two pathways appear to correspond to the distinction between biologically driven versus socially learned facial behavior. Researchers agree, for the most part, that most types of facial expressions are learned like language, displayed under conscious control, and have culturally specific meanings that rely on context for proper interpretation (e.g. [13]). Thus, the same lowered eyebrow expression that would convey "uncertainty" in North America might convey "no" in Borneo [9]. On the other hand, there are a limited number of distinct facial expressions of emotion that appear to be biologically wired, produced involuntarily, and whose meanings are similar across all cultures; for example, anger, contempt, disgust, fear, happiness, sadness, and surprise [13]. A number of studies have documented the relationship between these facial expressions of emotion and the physiology of the emotional response (e.g. [19], [20].) There are also spontaneous facial movements that accompany speech. These movements are smooth and ballistic, and are more typical of the subcortical system associated with spontaneous expressions (e.g. [40]). There is some evidence that arm-reaching movements transfer from one motor system when they require planning to another when they become automatic, with different dynamic characteristics between the two [12]. It is unknown whether the same thing happens with learned facial expressions. An automated system would enable exploration of such research questions.

C. The need for spontaneous facial expression databases

The machine perception community is in critical need of standard video databases to train and evaluate systems for automatic recognition of facial expressions. An important lesson learned from speech recognition research is the need for large, shared databases for training, testing, and evaluation, without which it is extremely difficult to compare different systems and to evaluate progress. Moreover, these databases need to be typical of real world environments in order to train data-driven approaches and for evaluating robustness of algorithms. An important step forward was the release of the Cohn-Kanade database of FACS coded facial expressions [28], which enabled development and comparison of numerous algorithms. Two more recent databases also make a major contribution to the field: The MMI database which enables greater temporal analysis as well as profile views [38], as well as the Lin database which contains 3D range data for prototypical expressions at a variety of intensities [30]. However, all of these databases consist of posed facial expressions. It is essential for the progress of the field to be able to evaluate systems on databases of spontaneous expressions. As described above, spontaneous expressions differ from posed expressions in both which muscles are moved, and in the dynamics of those movements. Development of these databases is a priority that requires joint effort from the computer vision, machine learning, and psychology communities. A database of spontaneous facial expressions collected at UT Dallas [34] was a significant contribution in this regard. The UT Dallas database elicited facial expressions using film clips, and there needs to be some concurrent measure of expression content beyond the stimulus category since subjects often do not experience the intended emotion and sometimes experience another one (e.g. disgust or annoyance instead of humor). FACS coding of this database would be extremely useful for the computer vision community. We present here a database of spontaneous facial expressions that has been FACS coded using the Facial Action Coding System.

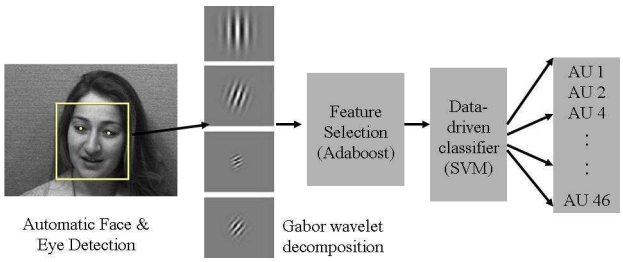

Figure 2. Overview of fully automated facial action coding system.

D. System overview

Here we describe progress on a system for fully automated facial action coding, and show preliminary results when applied to spontaneous expressions. This was the first system for fully automated expression coding, presented initially in [3], and it extends a line of research developed in collaboration with Paul Ekman and Terry Sejnowski [11]. It is a user independent fully automatic system for real time recognition of facial actions from the Facial Action Coding System (FACS). The system automatically detects frontal faces in the video stream and codes each frame with respect to 20 Action units. In previous work, we conducted empirical investigations of machine learning methods applied to the related problem of classifying expressions of basic emotions [31]. We compared AdaBoost, support vector machines, and linear discriminant analysis, as well as feature selection techniques. Best results were obtained by selecting a subset of Gabor filters using AdaBoost and then training Support Vector Machines on the outputs of the filters selected by AdaBoost. An overview of the system is shown in Figure 2. Here we apply this system to the problem of detecting facial actions in spontaneous expressions.

E. Relation to other work

There have been major advances in the computer vision literature for facial expression recognition over the past 15 years. See [22], [36] for reviews. Much of the early work on computer vision applied to facial expressions focused on recognizing a few prototypical expressions of emotion produced on command (e.g., "smile"). Some systems describe facial expressions in terms of component movements, most notably coding standard developed for MPEG4 which focuses on automatic coding of a set of

© 2006 ACADEMY PUBLISHER

facial feature points [10]. While coding standards like MPEG4 are useful for animating facial avatars, behavioral research may require more comprehensive information. For example, MPEG4 does not encode some behaviorally relevant movements such as the contraction of the orbicularis oculi, which differentiates spontaneous from posed smiles. It also does not encode changes in surface texture such as wrinkles, bulges, and shape changes that are critical for the definition of action units in the FACS system. For example, a characteristic pattern of wrinkles and bulges between the brows, as well as the shape of the brow (arched vs. flat), are important for distinguishing a basic brow raise (AU 1+2) from a fear brow (AU 1+2+4 shown in Figure 1, both of which entail upward movement of the brows, but the brow raise is a common behavior that emphasizes speech, shows engagement in conversation, or indicates a question, whereas the fear brow occurs in situations of danger and sometimes deception [26].

Several research groups have recognized the importance of automatically coding expressions in terms of FACS [11], [29], [37], [45], [46]. Our approach differs from others in that instead of designing special purpose image features for each facial action, we employ machine learning techniques for data-driven facial expression classification. These machine learning algorithms are applied to image-based representations. Image-based machine learning methods have been shown to be highly effective for machine recognition tasks such as face detection [47], and do not suffer from drift which is a major obstacle to tracking methods. In this paper we show that such systems capture information about action unit intensity that can be employed for analyzing facial expression dynamics. The image-based representation employed here is the output of a bank of Gabor filters, although in previous work we have applied machine learning to the image features as well, and found that the Gabor features are related to those developed by machine learning [2]. Learned classifiers taking such representations as input are sensitive not only to changes in position of facial features, but also to changes in image texture such as those created by wrinkles, bulges, and changes in feature shapes. We found in practice that image-based representations contain more information for facial expression than representations based on the relative positions of a finite set of facial features. For example, our basic emotion recognizer [31] gave a higher performance on the Cohn-Kanade dataset than an upper-bound on feature tracking computed by another group based on manual feature point labeling. We distinguish here feature-point tracking from the general category of motion-based representations. One may describe motion with spatio-temporal Gabor filters, for example, resulting in a representation related to the one presented here. At issue is whether reducing the image to a finite set of feature positions is a good representation. Ultimately, combining image-based and motion based representations may be the most powerful.y

Tian et al. [45] employ traditional computer vision techniques for state-of-the-art feature tracking. In this approach, specialized image features such as contour parameters are defined for each desired facial action. Pantic and Rothcrantz [37] use robust facial feature detection followed by an expert system to infer facial actions from the geometry of the facial features. More recent work from this group presented approaches to measuring facial actions in profile views, and recognizing expression dynamics from temporal rules [35]. Their approach is more heuristic than the data-driven system presented here. A strength of data-driven systems is that they learn the variations in appearance of a facial action due to differences in physiognomy, age, and elasticity, and also when an action occurs in combination with other facial actions. Nonlinear support vector machines have the added advantage of being able to handle multimodal data distributions which can arise with action combinations, provided that the class of kernel is well matched to the problem. A group at MIT presented a prototype system for fully automated FACS coding that used infrared eye tracking to register face images [29]. The recognition component is similar in spirit to the one presented here, employing machine learning techniques on image-based representations. Kapoor et al. used PCA (eigenfeatures) as the feature vector, whereas we previously found that PCA was a much less effective representation than Gabor wavelets for facial action recognition (see [11], [31]. More recently, [46], applied a dynamical Bayesian model to the output of a front-end FACS recognition system based on the one developed in our laboratory [3], [4]. While [46] showed that AU recognition benefits from learning causal relations between AU's in the training database, the analysis was developed and tested on a posed expression database. It will be important to extend such work to spontaneous expressions for the reasons described above.

II. Automated System

A. Real-time Face Detection

We developed a real-time face detection system that employs boosting techniques in a generative framework [23] and extends work by [47]. Enhancements to [47] include employing Gentleboost instead of AdaBoost, smart feature search, and a novel cascade training procedure, combined in a generative framework. Source code for the face detector is freely available at http://kolmogorov.sourceforge.net. Accuracy on the CMU-MIT dataset, a standard public data set for benchmarking frontal face detection systems, is 90% detections and 1/million false alarms, which is state-of-the-art accuracy. The CMU test set has unconstrained lighting and background. With controlled lighting and background, such as the facial expression data employed here, detection accuracy is much higher. All faces in the training datasets, for example, were successfully detected. The system presently operates at 24 frames/second on a 3 GHz Pentium IV for 320x240 images.

The automatically located faces were rescaled to 96x96 pixels. The typical distance between the centers of the eyes was roughly 48 pixels. Automatic eye detection

© 2006 ACADEMY PUBLISHER

[23] was employed to align the eyes in each image. The images were then passed through a bank of Gabor filters 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps) (See [31]). Output magnitudes were then passed to the classifiers. No feature selection was performed for the results presented here, although it is ongoing work that will be presented in another paper.

B. Facial Action Classification

Facial action classification was assessed for two classifiers: Support vector machines (SVM's) and AdaBoost.

a) SVM's.: SVM's are well suited to this task because the high dimensionality of the Gabor representation O(10^5) does not affect training time, which depends only on the number of training examples O(10^2). In our previous work, linear, polynomial, and radial basis function (RBF) kernels with Laplacian, and Gaussian basis functions were explored [31]. Linear and RBF kernels employing a unit-width Gaussian performed best on that task. Linear SVMs are evaluated here on the task of facial action recognition.

b) AdaBoost.: The features employed for the AdaBoost AU classifier were the individual Gabor filters. This gave 9x8x48x48= 165,888 possible features. A subset of these features was chosen using AdaBoost. On each training round, the Gabor feature with the best expression classification performance for the current boosting distribution was chosen. The performance measure was a weighted sum of errors on a binary classification task, where the weighting distribution (boosting) was updated at every step to reflect how well each training vector was classified. AdaBoost training continued until 200 features were selected per action unit classifier. The union of all features selected for each of the 20 action unit detectors resulted in a total of 4000 features.

III. Facial expression data

A. The RU-FACS Spontaneous Expression Database

Mark Frank, in collaboration with Javier Movellan and Marian Bartlett, has collected a dataset of spontaneous facial behavior with rigorous FACS coding. The dataset consists of 100 subjects participating in a 'false opinion' paradigm. In this paradigm, subjects first fill out a questionnaire regarding their opinions about a social or political issue. Subjects are then asked to either tell the truth or take the opposite opinion on an issue where they rated strong feelings, and convince an interviewer they are telling the truth. Interviewers were retired police and FBI agents. A high-stakes paradigm was created by giving the subjects $50 if they succeeded in fooling the interviewer, whereas if they were caught they were told they would receive no cash, and would have to fill out a long and boring questionnaire. In practice, everyone received a minimum of $10 for participating, and no one had to fill out the questionnaire. This paradigm has been shown to elicit a wide range of emotional expressions as well as speech-related facial expressions [26]. This dataset is particularly challenging both because of speech-related mouth movements, and also because of out-of-plane head rotations which tend to be present during discourse.

Subjects faces were digitized by four synchronized Dragonfly cameras from Point Grey. (See Figure 3). The analysis in this paper was conducted using the video stream from the frontal view camera. Two minutes of each subject's behavior is being FACS coded by two certified FACS coders. FACS codes include the apex frame as well as the onset and offset frame for each action unit (AU). To date, 33 subjects have been FACS-coded. Here we present preliminary results for a system trained on two large datasets of FACS-coded posed expressions, and tested on the spontaneous expression database. Future work will include spontaneous expressions in training as well. Here we explore how well a system trained on posed expressions under controlled conditions performs when applied to real behavior.

Figure 3. Sample synchronized camera views from the RU-FACS spontaneous expression database.

B. Posed expression databases

Because the spontaneous expression database did not yet contain sufficient labeled examples to train a data-driven system, we trained the system on a larger set of labeled examples from two FACS-coded datasets of posed images. The first dataset was Cohn and Kanade's DFAT-504 dataset [28]. This dataset consists of 100 university students ranging in age from 18 to 30 years. 65% were female, 15% were African-American, and 3% were Asian or Latino. Videos were recoded in analog S-video using a camera located directly in front of the subject. Subjects were instructed by an experimenter to perform a series of 23 facial displays. Subjects began each display with a neutral face. Before performing each display, an experimenter described and modeled the desired display. Image sequences from neutral to target display were digitized into 640 by 480 pixel arrays with 8-bit precision for grayscale values. The facial expressions in this dataset were FACS coded by two certified FACS coders.

© 2006 ACADEMY PUBLISHER

The second dataset consisted of directed facial actions from 24 subjects collected by Ekman and Hager. (See [11].) Subjects were instructed by a FACS expert on the display of individual facial actions and action combinations, and they practiced with a mirror. The resulting video was verified for AU content by two certified FACS coders.

IV. Training

The combined dataset contained 2568 training examples from 119 subjects. Separate binary classifiers, one for each AU, were trained to detect the presence of the AU regardless of the co-occurring AU's. We refer to this as context-independent recognition. Positive examples consisted of the last frame of each sequence which contained the expression apex. Negative examples consisted of all apex frames that did not contain the target AU plus neutral images obtained from the first frame of each sequence, for a total of 2568-N negative examples for each AU.

V. Generalization Performance *Within* Dataset

We first report performance for generalization to novel subjects *within* the Cohn-Kanade and Ekman-Hager databases. Generalization to new subjects was tested using leave-one-subject-out cross-validation in which all images of the test subject were excluded from training. Results for the AdaBoost classifier are shown in Table I. System outputs were the output of the AdaBoost discriminant function for each AU. All system outputs above threshold were treated as detections.

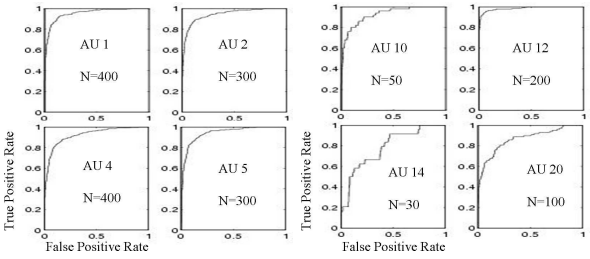

Figure 4. ROC curves for 8 AU detectors, tested on posed expressions.

The system obtained a mean of 91% agreement with human FACS labels. Overall percent correct can be an unreliable measure of performance, however, since it depends on the proportion of targets to non-targets, and also on the decision threshold. In this test, there was a far greater number of non-targets than targets, since targets were images containing the desired AU (N in Table I, and non-targets were all images not containing the desired AU (2568-N). A more reliable performance measure is area under the ROC (receiver-operator characteristic curve.) This curve is obtained by plotting hit rate (true positives) against false alarm rate (false positives) as the decision threshold varies. See Figure 4. The area under this curve is denoted A′. A′ is equivalent to percent correct in a 2-alternative forced choice task, in which the system must choose which of two options contains the target on each trial. Mean A′ for the posed expressions was 92.6.

TABLE I.
PERFORMANCE FOR POSED EXPRESSIONS.

Shown is fully automatic recognition of 20 facial actions, generalization to novel subjects in the Cohn-Kanade and Ekman-Hager databases. N: Total number of positive examples. P: Percent agreement with Human FACS codes (positive and negative examples classed correctly). Hit, FA: Hit and false alarm rates. A′: Area under the ROC. The classifier was AdaBoost.

AU	Name	N	P	Hit	FA	A′
1	Inn. brow raise	409	92	86	7	95
2	Out. brow raise	315	88	85	12	92
4	Brow lower	412	89	76	9	91
5	Upper lid raise	286	92	88	7	96
6	Cheek raise	278	93	86	6	96
7	Lower lid tight	403	88	89	12	95
9	Nose wrinkle	68	100	88	0	100
10	Lip Raise	50	97	29	2	90
11	Nasolabial	39	94	33	4	74
12	Lip crnr. pull	196	95	93	5	98
14	Dimpler	32	99	20	0	85
15	Lip crnr. depr.	100	85	85	14	91
16	Lower Lip depr.	47	98	29	1	92
17	Chin raise	203	89	86	10	93
20	Lip stretch	99	92	57	6	84
23	Lip tighten	57	91	42	8	70
24	Lip press	49	92	64	7	88
25	Lips part	376	89	83	9	93
26	Jaw drop	86	93	58	5	85
27	Mouth stretch	81	99	100	1	100
	Mean		90.9	80.1	8.2	92.6

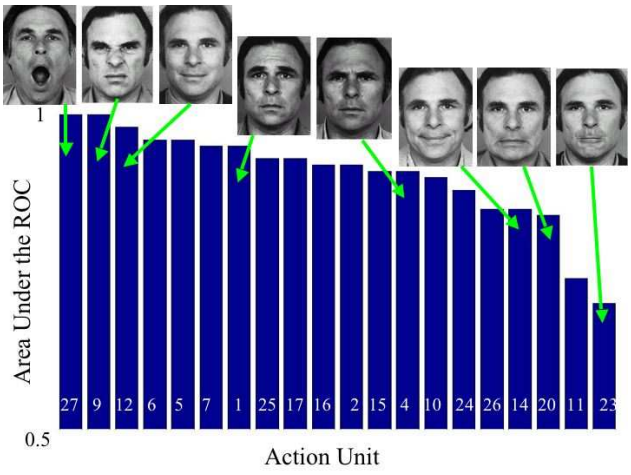

Figure 5. System performance (area under the ROC) for posed facial actions. Actions were sorted in order of detection performance. High, middle, and low-performing AU's are illustrated.

Figure 5 contains a graphical depiction of performance sorted in order of A'. High, middle, and low performing action units are illustrated. We note that both the highest and lowest performance was obtained with lower-face AU's (AU's 9-27), while mid-range performance was obtained for brow and eye region actions (AU's 1-7).

A. Effect of training set size

We next investigated to what degree the performance variation related to training set size. Inspection of the ROC curves in Figure 4 suggests a dependence of sys-

© 2006 ACADEMY PUBLISHER

tem performance on the number of training examples. Figure /reffig:numex plots area under the roc against training set size for all 20 AU's tested. The scatter plot is elbow-shaped, where action units with the fewest training examples also had the lowest performance, and then the plot flattens for training set sizes substantially greater than 100. Action units 27 and 9 are reliably detected despite relatively small training set sizes, suggesting these two may be easier to detect.

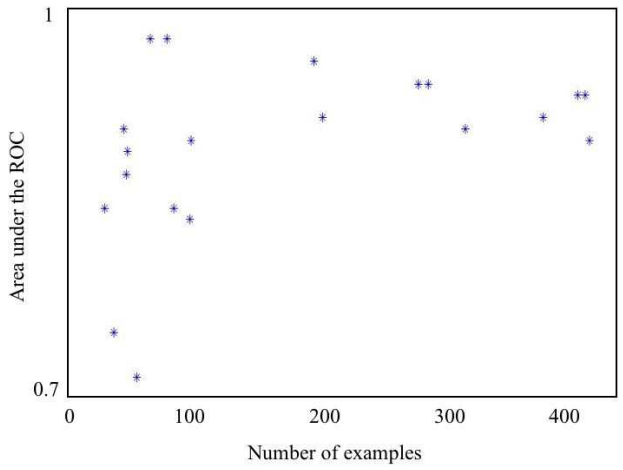

Figure 6. Scatter plot of system performance (area under the ROC) against training set size for the 20 AU's in Table I.

VI. Generalization to Spontaneous Expressions

We then tested the system described in Section IV on a new dataset of spontaneous expressions, the RU-FACS dataset. The dataset included speech related mouth and face movements, and significant amounts of in-plane and in-depth rotations.

A. characterization of head pose during spontaneous expression

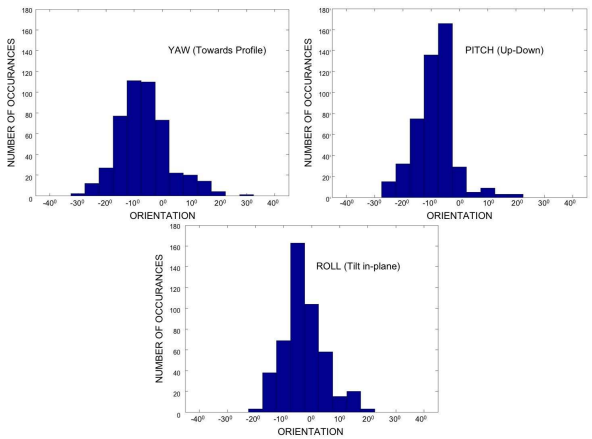

Figure 7. Distribution of head poses in frontal view camera during expression apex.

TABLE II.
HEAD POSE DISTRIBUTION DURING EXPRESSION APEX.

	Mean	sd	67% Interval	96% Interval
Yaw	-6.5^0	8.9^0	$[-15^0, 2^0]$	$[-24^0, 11^0]$
Pitch	-8.5^0	7.1^0	$[-16^0, 2^0]$	$[-23^0, 5^0]$
Roll	-2.8^0	7.4^0	$[-10^0, 4^0]$	$[-18^0, 2^0]$

In contrast to the highly controlled conditions of posed expression databases, in spontaneous behavior the head pose of subjects varies from frontal. In order to characterize the distribution of head pose during spontaneous facial expressions, we manually labeled the head pose for the action unit apex frames for 473 images from 21 subjects. Head pose was labeled by rotating a 3D head model with the arrow keys until it matched the head pose in the image. To assist alignment, an initial estimate of head pose was calculated from eye, nose, and mouth positions. In addition, the face image was projected onto the 3D head model and then back into the image plane from the estimated pose, in order to match the projected face with the face in the image.

Figure 7 shows the distribution of yaw, pitch, and roll during expression apex. The histograms show that each of these three head orientation measures ranges from approximately -30^0 to 20^0. Table II gives the mean and standard deviations of yaw, pitch, and roll. The mean head pose in this dataset is 8^0 down and 6^0 to the left, which is likely a result of the constraints of camera placement, in which the camera was placed over the right shoulder of the interviewer. Roll (or in-plane rotation) is the most straightforward to correct in computer vision systems. In both yaw and pitch, approximately 30% of all apex frames were between 15^0 and 25^0 from frontal, and no images in our sample were rotated beyond 30^0

B. AU recognition: Spontaneous Expressions

Preliminary recognition results are presented for 12 subjects. This data contained a total of 1689 labeled events, consisting of 33 distinct action units, 19 of which were AU's for which we had trained classifiers. Face detections were accepted if the face box was greater than 150 pixels width, both eyes were detected with positive position, and the distance between the eyes was $>$ 40 pixels. This resulted in faces found for 95% of the video frames. Most non-detects occurred when there was head rotations beyond$\pm 10^0$ or partial occlusion. All detected faces were passed to the AU recognition system.

Here we present benchmark performance of the basic frame-by-frame system on the video data. Figure 8 shows sample system outputs for one subject, and performance is shown in Figure 5 and Table III. Performance was assessed several ways. First, we assessed overall percent correct for each action unit on a frame-by-frame basis, where system outputs that were above threshold inside the onset and offset interval indicated by the human FACS codes, and below threshold outside that interval were

© 2006 ACADEMY PUBLISHER

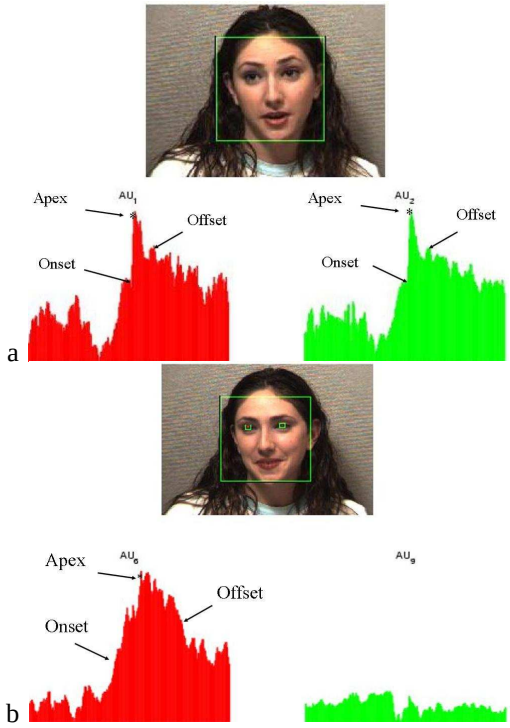

Figure 8. Sample system outputs for a 10-second segment containing a brow-raise (FACS code 1+2). System output is shown for AU 1 (left) and AU 2 (right). Human codes are overlayed for comparison (onset, apex, offset).

TABLE III.
RECOGNITION OF SPONTANEOUS FACIAL ACTIONS.

AU: Action unit number. N: Total number of testing examples. P: Percent correct over all frames. Hit, FA: Hit and false alarm rates. A′: Area under the ROC. A'_{Δ}: Area under the ROC for interval analysis (see text). The classifier was AdaBoost.

AU	N	P	Hit	FA	A′	A'_{Δ}
1	169	87	35	9	78	83
2	153	84	29	13	62	68
4	32	97	15	2	74	84
5	36	97	7	1	71	76
6	50	92	32	4	90	92
7	46	91	12	7	64	66
9	2	99	0	0	88	93
10	38	95	0	0	62	65
11	3	99	0	0	73	83
12	119	86	45	7	86	88
14	87	94	0	0	70	77
15	77	94	23	4	69	73
16	5	99	0	0	63	57
17	121	93	15	2	74	76
20	12	99	0	0	66	69
23	24	98	0	0	69	75
24	68	95	7	3	64	63
25	200	54	68	50	70	73
26	144	91	2	1	63	64
Mean		93	15	5	71	75

considered correct. This gave an overall accuracy of 93% correct across AU's for the AdaBoost classifier. Mean area under the ROC was .71.

Next an interval analysis was performed, which was intended to serve as a baseline for future analysis of output dynamics. The interval analysis measured detections on intervals of length I. Here we present performance for intervals of length 21 (10 on either side of the apex), but performance was stable for a range of choices of I. A target AU was treated as present if at least 6/21 frames were above threshold, where the threshold was set to 1 standard deviation above the mean. Negative examples consisted of the remaining 2 minute video stream for each subject, outside the FACS coded onset and offset intervals for the target AU, parsed into intervals of 21 frames. This simple interval analysis raised the area under the ROC to .75.

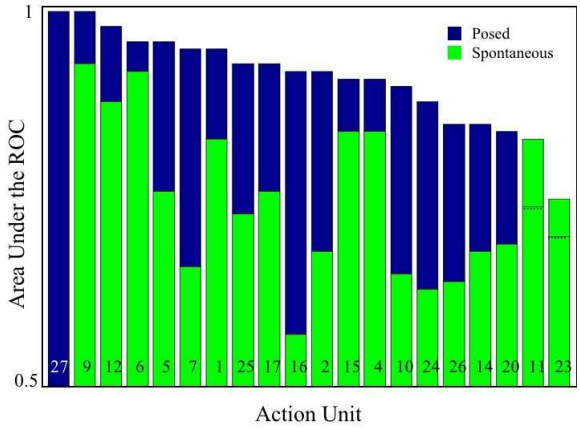

Figure 9. AU recognition performance (area under the ROC) for spontaneous facial actions. Performance is overlayed on the posed results of Figure 5.

TABLE IV.
COMPARISON OF ADABOOST TO LINEAR SVM'S.

The task is AU classification in the spontaneous expression database. A'_{Δ}: Area under the ROC for interval analysis (see text).

		AdaBoost		SVM	
AU	N	A′	A'_{Δ}	A′	A'_{Δ}
1	169	78	83	73	83
2	153	62	68	63	76
4	32	74	84	74	86
5	36	71	76	63	73
10	38	62	65	60	71
12	119	86	88	84	90
14	87	70	77	65	73
20	12	66	69	60	74
	Mean	71.1	76.3	67.8	78.3

C. AdaBoost v. SVM performance

Table IV compares AU recognition performance with AdaBoost to a linear SVM. In previous work with posed expressions of basic emotions, AdaBoost performed similarly to SVM's, conferring a marginal advantage over the linear SVM [31]. Here we support this finding for recognition of action units in spontaneous expressions. AdaBoost had a small advantage over the linear SVM which was statistically significant on a paired t-test (

© 2006 ACADEMY PUBLISHER

t(7)=3.1, p=.018). A substantial performance increase was incurred for both classifiers by employing the interval analysis. Here the output y was first converted to Z-scores for each subject $z = (y - \mu/\sigma)$, and then z was integrated over a window of 11 frames. The The temporal information in the classifier outputs contain considerable information that we intend to exploit in future work.

VII. The Margin Predicts Expression Intensity

Figure 10 shows a sample of system outputs for a 2 minute 20 second continuous video stream from the spontaneous expression database. Inspection of such output streams suggested that the system output, which was the distance to the separating hyperplane (the margin), contained information about expression intensity. A stronger relationship was observed for the outputs of the posed data, which had less noisy image conditions. System outputs for full image sequences of test subjects from the posed data are shown in Figure 11. Although each individual image is separately processed and classified, the outputs change smoothly as a function of expression magnitude in the successive frames of each sequence.

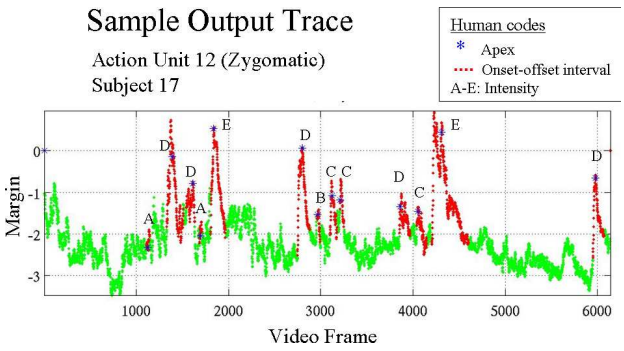

Figure 10. Output trajectory for a 2 minute 20 sec. video (6000 frames), for one subject and one action unit. Shown is the margin (the distance to the separating hyperplane). The human FACS labels are overlaid for comparison. Blue stars indicate the frame at which the AU apex was coded. The frames within the onset and offset of the AU are shown in red. Letters A-E indicate AU intensity, with E highest.

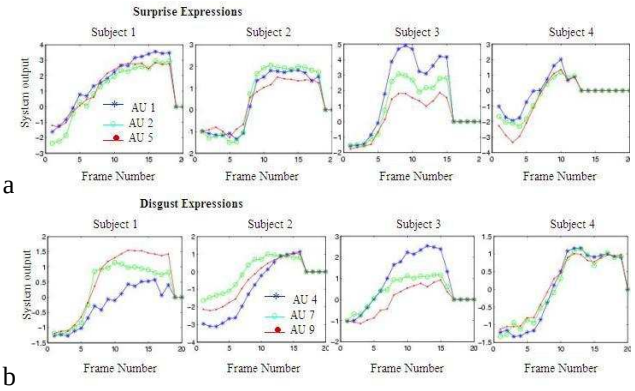

Figure 11. Automated FACS measurements for full image sequences. a. Surprise expression sequences from 4 subjects containing AU's 1,2 and 5. b. Disgust expression sequences from 4 subjects containing AU's 4,7 and 9.

Intensity correlation: Posed data: A correlation analysis was performed in order to explicitly measure the relationship between the output margin and expression intensity. In order to assess this relationship in low noise conditions, we first performed the analysis for posed expressions. The posed data contains no speech, negligible head rotation, and consequently less luminance variation than the spontaneous data. In addition, the posed database was the training database, which enabled us to measure the degree to which the SVM learned about expression intensity for validation subjects in the same database as the training set.

Ground truth for action unit intensity was measured as follows: Five certified FACS coders labeled the action intensity for 108 images from the Ekman-Hager database. The images were four upper-face actions (1, 2, 4, 5) and two lower-face actions (10, 20), displayed by 6 subjects. Three images from each sequence were displayed: one immediately after onset, one at apex, and one intermediate frame. Images were presented in random order, and the FACS experts were asked to label both the AU and the AU intensity. In keeping with FACS coding procedures, the experts scored intensity on an A through E scale, where A is lowest, and E is highest. The experts did not always agree with the FACS label in the database, particularly for the lowest intensity frames. Disagreement rate was 5.3% for the upper-face actions and 8.5% for the lower-face actions. Intensities were included in the subsequent analysis only for frames on which the experts agreed with the AU label in the database.

We first measured the degree to which expert FACS coders agree with each other on intensity. Correlations were computed separately for each action unit. Correlations were computed between intensity scores by each pair of experts, and the mean correlation was computed across all expert pairs. The results are given in Table V. Because individual faces differ in the amount of wrinkling and movement for a given facial action, the correlations were computed two ways: within-subject and between-subject, where 'subject' refers to the person displaying the action unit. Within-subject correlations effectively remove differences in mean and variance between individuals, and is similar to computing a Z-score prior to correlating. For the within-subject analysis, correlations were computed separately for each subject, and the mean was computed across subjects. Mean correlation between expert FACS coders within subject was 0.84.

Correlations of the automated system with the human expert intensity scores were next computed. The SVM's were retrained on the even-numbered subjects of the Cohn-Kanade and Ekman-Hager datasets, and then tested on the odd-numbered subjects of the Ekman-Hager set, and vice versa. Correlations were computed between the SVM margin and the intensity ratings of each of the five expert coders. The analysis was again performed *within* subject, and then means were computed for each AU by collapsing across subject. The results are shown in Table VA. Overall mean correlation between the SVM

© 2006 ACADEMY PUBLISHER

margin and the expert FACS coders was 0.83, which was very similar to the human-human correlation of .84.

The analysis was next repeated between-subjects, and the results are shown in Table VB. For the between-subject analysis, all subjects displaying the action unit were included in a single correlation for each AU. We see that the expert agreement dropped about 10% to .73 by doing the between-subject correlation. For the SVM, the agreement with expert humans dropped to .53. This shows that the system will benefit from online learning of scale and threshold for each subject, since the within-subject analysis effectively removed mean and scale differences between subjects.

TABLE V.
HUMAN-SVM INTENSITY CORRELATIONS.

Expert-Expert is the mean correlation (r) across 5 human FACS experts. SVM-Expert is the mean correlation of the margin with intensity rating from each of the 5 experts. A. Correlations were computed *within* subject displaying the facial action. B. Correlations were computed *across* subjects displaying the facial action.

A. Within Subject, posed

	Action Unit						
	1	2	4	5	10	20	Mean
Expert-Expert	.92	.77	.85	.72	.88	.88	.84
SVM-Expert	.90	.80	.84	.86	.79	.79	.83

B. Between Subject, posed

	Action Unit						
	1	2	4	5	10	20	Mean
Expert-Expert	.77	.64	.79	.63	.75	.79	.73
SVM-Expert	.47	.85	.45	.39	.36	.69	.53

TABLE VI.
HUMAN-SVM INTENSITY CORRELATIONS, SPONTANEOUS DATA.

Action Unit								
1	2	4	5	10	12	14	20	Mean
.31	.09	.38	.51	.29	.75	.16	.33	.35

Intensity correlation: Spontaneous data: The correlation analysis was then repeated for the spontaneous expression data. This effectively tests how well the relationship between the margin and the intensity generalizes to a new dataset, and also how well it holds up in the presence of noise from speech and head movements. The intensity codes in the RU-FACS dataset were used as ground truth for action intensity. Correlations were computed between the margin of the linear SVM and the AU intensity as coded by the human coders for each subject for the 8 AU's shown in Table IV. As above, correlations were computed within subject, and then collapsed across subject to provide a mean correlation for each AU. The overall mean correlation was r=0.35. There was much variability in the correlations across AU. AU 12, for example, had a correlation of 0.75 between the margin and FACS intensity score. The correlations for the spontaneous expression data were overall substantially smaller than for the posed data. Nevertheless, for some facial actions the system extracted a signal about expression intensity on this very challenging dataset.

VIII. PERFORMANCE FACTORS

A. Effect of Compression

For many applications of automatic facial expression analysis, image compression is desirable in order to make an inexpensive, flexible system. The image analysis methods employed in this system, such as Gabor filters, may be more robust to lossy compression compared to other image analysis methods such as optic flow. We therefore investigated the relationship between AU recognition performance and image compression. Detectors for three action units (AU 1, AU2, and AU4) were compared when tested at five levels of compression: No loss (original bmp images), and 4 levels of jpeg compression quality: 100%, 75in Figure 12. Performance remained consistent across substantial quantities of lossy compression. This finding is of practical importance for system design.

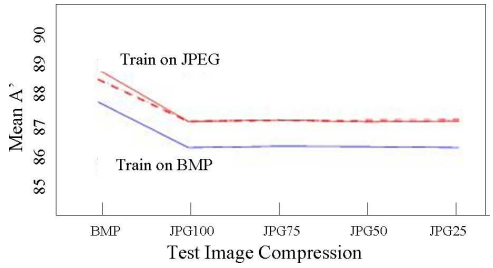

Figure 12. Effects of compression on AU recognition performance.

B. Handling AU combinations

When they occur together, facial actions can take on a very different appearance from when the occur in isolation. A similar effect happens in speech recognition with phonemes, and is called a co-articulation effect. The SVM's in this study were trained to detect an action unit regardless of whether it occurs alone or in combination with other action units. However, as in speech recognition, we may obtain better performance by training dedicated detectors for certain AU combinations.

An AU combination analysis was performed on the three brow action units (1, 2, 4). The analysis was performed on a set of linear SVM's trained on half of the Ekman-Hager database and tested on the other half. AU's 1 and 2 pull the brow up (centrally and laterally, respectively), whereas AU 4 pulls the brows together and down using primarily the corrugator muscle at the bridge of the nose. The appearance of AU 4 changes dramatically depending on whether it occurs alone or in combination with AU 1 and 2. Inspection of Table VII shows that performance may benefit from treating AU4 separately. We also see that it is not necessary or desirable to treat all

© 2006 ACADEMY PUBLISHER

AU combinations separately. For example, performance does not benefit from treating AU 1 and AU 2 separately.

TABLE VII.
COMBINATION ANALYSIS

Performance for a linear SVM trained to detect specific combinations of the brow AU's. Target = target set for training. Non-targets during training were all other AU's in the Ekman-Hager database. There were no spontaneous examples of 2+4.

Target	Posed A'	Spont. A'
AU 1 alone or in comb.	93.5	72.6
Individual AU 1 only	72.7	60.4
AU 2 alone or in comb.	92.8	69.8
Individual AU 2 only	59.2	53.4
AU 4 alone or in comb.	83.2	65.4
Individual AU 4 only	85.0	60.1
1+2	98.2	70.5
1+4	89.4	60.1
2+4	95.8	–
1+2+4	90.8	68.5

IX. CONCLUSIONS

The current state of the art in automatic face and gesture recognition suggests that user independent, fully automatic real time coding of facial expressions in the continuous video stream is an achievable goal with present computer power. The system presented here operates in real time. Face detection runs at 24 frames/second in 320x240 images on a 3 GHz Pentium IV. The AU recognition step operates in less than 10 msec per action unit.

The next step in the development of automatic facial expression recognition systems is to begin to apply them to spontaneous expressions for real applications. Spontaneous facial behavior differs from posed expressions both in which muscles are moved, and in the dynamics of those movements. These differences are described in more detail in the Introduction. The step to spontaneous behavior involves handling variability in head pose, and often the presence of speech, in addition to handling the wide variety of facial muscle constellations that occur in natural behavior.

This paper presented preliminary results for a fully automated facial action detection system on a database of spontaneous facial expressions. These results provide a benchmark for future work on spontaneous expression video. The system was able to extract information about facial actions in this dataset despite substantive differences between the spontaneous expressions and the posed data on which it was trained. While at this time there is insufficient FACS-labeled spontaneous expressions to support training of data-driven systems exclusively on spontaneous expressions, a combined training approach is possible, and an important next step is to explore systems trained on a combined posed and spontaneous dataset. Preliminary results in our lab using such a combined dataset, and testing with cross-validation, show that performance is substantively improved on the spontaneous expression data, while performance declines on the posed data. This finding reinforces the differences between posed and spontaneous expressions.

A significant finding from this paper is that data-driven classifiers such as SVM's learned information about expression intensity. The distance to the separating hyperplane, the margin, was significantly correlated with facial action intensity codes. Current work in our lab is showing a similar relationship with Adaboost, where in the case of AdaBoost, it is the likelihood ratios in the Adaboost discriminant function that correlate with measures of expression intensity. The system therefore is able to provide information about facial expression dynamics in the frame-by-frame intensity information. This information can be exploited for deciding the presence of a facial action and decoding the onset, apex, and offset. It will also enable explorations of the dynamics of facial behavior, as discussed below.

The accuracy of automated facial expression measurement in spontaneous behavior may be considerably improved by 3D alignment of faces. Moreover, information about head movement dynamics is an important component of nonverbal behavior, and is measured in FACS. Members of this group have developed techniques for automatically estimating 3D head pose in a generative model and for aligning face images in 3D [32]. We are also exploring feature selection techniques. Our previous work with expressions of basic emotion showed that feature selection by AdaBoost significantly enhanced both speed and accuracy of SVM's [31]. We are presently exploring whether such advantages found for basic emotion recognition carry over to the task of AU detection in spontaneous expressions.

The system presented here is fully automated, and performance rates for posed expressions compare favorably with other systems tested on the Cohn-Kanade dataset that employed varying levels of manual registration. A standard database of FACS coded spontaneous expressions would be of great benefit to the field and we are preparing to make the RU-FACS spontaneous expression database available to the research community.

A. Applications

Man-machine interaction for education: A major thrust of research in human-computer and human-robot interaction is the development of tools for education. Expression measurement tools enable automated tutoring systems that recognize the emotional and cognitive state of the pupil and respond accordingly. Such systems would also assist robots and animated agents to establish social resonance. Research has shown that behaviors including

© 2006 ACADEMY PUBLISHER

mirroring facial expressions and head movements assists in generating social rapport and can lead to increased information transfer between humans (e.g. [5], [7]. Such behaviors may increase the effectiveness of animated agents and robots designed for education environments. In addition, there is evidence suggesting that automatic tutors can become more effective if they use information about the eye movements of the students [1], [43].

Behavioral Science and Psychiatry: Tools for automatic expression measurement would enable tremendous new research activity not only in emotion, but also social psychology, development, cognitive neuroscience, psychiatry, education, human-machine communication, and human social dynamics. These tools will bring about paradigmatic shifts in these fields by making facial expression more accessible as a behavioral measure. New research activities enabled by this technology include studying the cognitive neuroscience of emotion, mood regulation, and social interaction; measuring the efficacy of psychiatric treatment including new medications, and studying facial behavior in psychiatric and developmental disorders.

Dynamics of facial behavior: Automated expression measurement tools developed in projects such as the one presented here will enable investigations into the dynamics of human facial expression that were previously infeasible with manual coding. This would allow researchers to directly address a number of questions key to understanding the nature of the human emotional and expressive systems, and their roles interpersonal interaction, psychopathology, and development. Previous research with manual coding has shown differences in the dynamics of spontaneous expressions compared to posed, as well as differences in the facial dynamics of patients with neuropathology (e.g. schizophrenia). Research has also shown that subtle movement differences between felt and unfelt expressions can be a critical indicator of social functioning, of the progress and remission of depression, and of suicide potential, as well as provide signs of deception [21], [26]. There are very few experiments of this nature because of the time burden of manual coding of dynamics, and the coding that has been done measures only coarse information about dynamics.

Security: Automatic facial action measurement has profound consequences on law enforcement and counter terrorism. Careful laboratory studies show that many of the clues to concealed emotion and deceit currently used in law enforcement training programs may be quite unreliable. Moreover, research based on facial action coding showed that more reliable cues exist in facial behavior [26]. Extracting this information requires detailed analysis of facial expression. Real-time automated coding can non-obtrusively supplement the other information available to interviewers, screeners, and law enforcement agents by identifying subtle or conflicted expressions that may betray someone's true emotional state. Automatic Expression coding will also enable more thorough investigation of the role of facial expression in deception.

Acknowledgments

Support for this work was provided by NSF IIS-0220141, NSF CNS-0454233, and NRL/HSARPA Advanced Information Technology 55-05-03 to Bartlett, Movellan, and Littlewort. Support for M.G. Frank was provided by NSF 0220230 and NSF 0454183. This material is based upon work supported by the National Science Foundation under a Cooperative Agreement/Grant. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] J. R. Anderson. Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science*, 26:85–112, 2002.

[2] M.S. Bartlett, G.L. Donato, J.R. Movellan, J.C. Hager, P. Ekman, and T.J. Sejnowski. Image representations for facial expression coding. In S.A. Solla, T.K. Leen, and K.-R. Muller, editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 2000.

[3] M.S. Bartlett, G. Littlewort, I. Fasel, and J.R. Movellan. Real time face detection and expression recognition: Development and application to human-computer interaction. In *CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction*, 2003.

[4] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Recognizing facial expression: Machine learning and application to spontaneous behavior. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 568–573, 2005.

[5] F. Bernieri, J. Davis, R. Rosenthal, and C. Knee. Interactional synchrony and rapport: Measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin*, 20:303–311, 1994.

[6] A. Brodal. *Neurological anatomy: In relation to clinical medicine*. Oxford University Press, New York, 1981.

[7] T.L. Chartrand and J. A. Bargh. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76:893–911, 1999.

[8] K.D. Craig, S.A. Hyde, and C.J. Patrick. Genuine, suppressed, and faked facial behavior during exacerbation of chronic low back pain. *Pain*, 46:161–172, 1991.

[9] C. Darwin. *The expression of the emotions in man and animals*. Oxford University Press, New York, 1872/1998. 3rd Edition, w/ commentaries by Paul Ekman.

[10] P. Doenges, F. Lavagetto, J. Ostermann, I.S. Pandzic, and E. Petajan. Mpeg-4: Audio/video and synthetic graphics/audio for real-time, interactive media delivery. *Image Communications Journal*, 5(4), 1997.

[11] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

[12] Torres E. and Andersen R. Space-time separation during obstacle-avoidance learning in monkeys. submitted.

[13] P. Ekman. The argument and evidence about universals in facial expressions of emotion. In D.C. Raskin, editor, *Psychological methods in criminal investigation and evidence*, pages 297–332. Springer Publishing Co, Inc., New York, 1989.

[14] P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. W.W. Norton, New York, 2nd edition, 1991.

© 2006 ACADEMY PUBLISHER

[15] P. Ekman. Facial expression of emotion. *American Psychologist*, 48:384–392, 1993.
[16] P. Ekman, R. J. Davidson, and W. V. Friesen. The duchenne smile: Emotional expression and brain physiology ii. *Journal of Personality and Social Psychology*, 58:342–353, 1990.
[17] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.
[18] P. Ekman and W. V. Friesen. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6:238–252, 1982.
[19] P. Ekman, R.W. Levenson, and W.V. Friesen. Autonomic nervous system activity distinguishes between emotions. *Science*, 221:1208–1210, 1983.
[20] P. Ekman and E.L. Rosenberg. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)*. Oxford University Press, New York, 1997.
[21] P. Ekman and E.L. Rosenberg. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS). 2nd Edition*. Oxford University Press, New York, 2005.
[22] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36:259–275, 2003.
[23] Ian R Fasel, Bret Fortenberry, and Javier R Movellan. A generative framework for real-time object detection and classification. *Computer Vision and Image Understanding*, 98, 2005.
[24] M. G. Frank and P. Ekman. Not all smiles are created equal: The differences between enjoyment and other smiles. *Humor: The International Journal for Research in Humor*, 6:9–26, 1993.
[25] M.G. Frank. *International Encyclopedia of the Social and Behavioral Sciences*, chapter Facial expressions. Elsevier, Oxford, 2002.
[26] M.G. Frank and P. Ekman. Appearing truthful generalizes across different deception situations. *Journal of personality and social psychology*, 86:486–495, 2004.
[27] M. Heller and V. Haynal. The faces of suicidal depression (translation). les visages de la depression de suicide. *Kahiers Psychiatriques Genevois (Medecine et Hygiene Editors)*, 16:107–117, 1994.
[28] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the fourth IEEE International conference on automatic face and gesture recognition (FG'00)*, pages 46–53, Grenoble, France, 2000.
[29] Qi Y. Kapoor, A. and R.W. Picard. Fully automatic upper facial action recognition. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
[30] Y. Lin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3d facial expression database for facial behavior research. In *Proceedings, International Conference Automatic on Face and Gesture Recognition*, 2006.
[31] G. Littlewort, M.S. Bartlett, I. Fasel, J. Susskind, and J.R. Movellan. An automatic system for measuring facial expression in video. *Image and Vision Computing*, in press.
[32] T. K. Marks, J. Hershey, J. Cooper Roddey, and J. R. Movellan. 3d tracking of morphable objects using conditionally gaussian nonlinear filters. In *CVPR Workshop on Generative-Model Based Vision*, 2004.
[33] A. Meihlke. *Surgery of the facial nerve*. Saunders, Philadelphia, 1973.
[34] AJ O'Toole, J Harms, SL Snow, DR Hurst, MR Pappas JH Ayyad, and H Abdi. A video database of moving faces and people. *IEEE Transactions on Pattern Analysis and Machine*, 27(5):812–816, 2005.
[35] M. Pantic and I. Patras. Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man and Cybernetics-Part B*, 36(2):433–449, 2006.
[36] M. Pantic and J.M. Rothkrantz. Automatic analysis of facial expressions: State of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.
[37] M. Pantic and J.M. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man and Cybernetics*, 34(3):1449–1461, 2004.
[38] M. Pantic, M.F. Valstar, R. Rademaker, and L. Maat. 'web-based database for facial expression analysis. In *Proc. IEEE Int'l Conf. on Multimedia and Expo (ICME '05)*, pages 317–321, 2005.
[39] R.W. Picard. *Affective Computing*. MIT Press, 1997.
[40] W. E. Rinn. The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95(1):52–77, 1984.
[41] Morecraft RJ, Louie JL, Herrick JL, and Stilwell-Morecraft KS. Cortical innervation of the facial nucleus in the non-human primate. a new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. *Brain*, 124:176–208, 2001.
[42] E.L. Rosenberg, P. Ekman, W. Jiang, M. Babyak, et al. Linkages between facial expressions of anger and transient myocardial ischemia in men with coronary artery disease. *American Psychological Assn, US. Emotion*, 1(2):107–115, 2001.
[43] D. D. Salvucci and J. R. Anderson. Intelligent gaze-added interfaces. in human factors in computing systems. In *CHI Conference Proceedings*, pages 273–280, New York, 2000. ACM Press.
[44] M. A. Sayette, D. W. Smith, M.J. Breiner, and G. T Wilson. The effect of alcohol on emotional response to a social stressor. *Journal of Studies on Alcohol*, 53:541–545, 1992.
[45] Y.L. Tian, T. Kanade, and J.F. Cohn. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:97–116, 2001.
[46] Y. Tong, W. Liao, and Q. Ji. Inferring facial action units with causal relations. In *Proceedings, Computer Vision and Pattern Recognition*, 2006.
[47] P. Viola and M. Jones. Robust real-time object detection. In *ICCV Second International Workshop no Statistical and Conceptual Theories of Vision.*, 2001.

© 2006 ACADEMY PUBLISHER

Marian Stewart Bartlett is Associate Research Professor at the Institute for Neural Computation, UCSD, where she co-directs the Machine Perception Lab. She studies learning in vision, with application to face recognition and expression analysis. She has authored over 30 articles in scientific journals and refereed conference proceedings, as well as a book, Face Image Analysis by Unsupervised Learning, published by Kluwer in 2001. Dr. Bartlett obtained her Bachelor's degree in Mathematics in 1988 from Middlebury College, and her Ph.D. in Cognitive Science and Psychology from University of California, San Diego, in 1998. Her thesis work was conducted with Terry Sejnowski at the Salk Institute. She has also published papers in visual psychophysics with Jeremy Wolfe, neuropsychology with Jordan Grafman, perceptual plasticity with V.S. Ramachandran, machine learning with Javier Movellan, automatic recognition of facial expression with Paul Ekman, cognitive models of face perception with Jim Tanaka, and the visuo-spatial properties of faces and American Sign Language with Karen Dobkins.

Gwen C. Littlewort holds an N.S.F. Advance fellowship at the Machine Perception Laboratory, U.C.S.D., where she investigates automatic facial expression recognition using machine learning techniques. She has a BSc in Electrical Engineering from U. Capetown and a PhD in Physics from Oxford.

Mark Frank is Associate Professor, Communication Department, University at Buffalo State University of New York. His research interests include nonverbal behavior in communication, with an emphasis on facial expression. He received a B.A. in Psychology from SUNY Buffalo in 1983, and his Ph.D. in Social Psychology from Cornell University in 1989. He completed a postdoc with Paul Ekman at the University of California, San Francisco Department of Psychiatry in 1992. Dr. Frank has studied nonverbal behavior in deception for many years, and has worked extensively with law enforcement groups, including helping them develop their interviewing and training programs in the context of counter-terrorism. He has also given workshops to courts nationally and internationally.

Claudia Lainscsek received the M.S. and PhD degrees in Physics and Technical Sciences from the University of Technology in Graz, Austria in 1992 and 1999, respectively. She has been at the Institute for Neural Computation since 2004, working on Nonlinear Dynamical Systems theory and Facial Action recognition.

Ian Fasel is a postdoctoral researcher at the Institute for Neural Computation, University of California, San Diego. His primary research interests are in learning and vision, and in particular what its role is in human social interaction and development. He received his B.S. in Electrical Engineering at University of Texas, Austin, in 1999, and his Ph.D. in Cognitive Science at the University of California, San Diego in 1996.

Javier R. Movellan is Full Project Scientist at the Institute for Neural Computation and Director of the Machine Perception Laboratory at the University of California San Diego. Previous to this he was a Fulbright fellow at UC Berkeley (1984-1989), a Carnegie-Mellon University research associate (1989-1993) and an assistant professor at UCSD (1993-2001). Javier has been studying learning and perception by human and machine for more than 20 years and has published over 50 articles in scientific journals and peer-reviewed conference proceedings. His articles span studies in probability theory, machine learning, machine perception, experimental psychology, and developmental psychology. Javier founded UCSD's Machine Perception Laboratory in 1997 with the goal of developing machine perception systems that combine multiple sources of information (e.g. audio and video) and interact naturally with people, reading their lips. recognizing their facial expressions, and making inferences about cognitive and affective states. Javier started the Kolmogorov project that provides a collection of open source tutorials on topics related to Machine Learning, Machine Perception and Statistics (http://markov.ucsd.edu/ kolmogorov/history.html). In 2004 he chaired the 3rd International Conference on Development and Learning in San Diego.

© 2006 ACADEMY PUBLISHER