Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 66 (2008) 303–317
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Real-time classification of evoked emotions using facial feature tracking and physiological responses

Jeremy N. Bailenson[a,*], Emmanuel D. Pontikakis[b], Iris B. Mauss[c], James J. Gross[d], Maria E. Jabon[e], Cendri A.C. Hutcherson[d], Clifford Nass[a], Oliver John[f]

[a]*Department of Communication, Stanford University, Stanford, CA 94305, USA*
[b]*Department of Computer Science, Stanford University, Stanford, CA 94305, USA*
[c]*Department of Psychology, 2155 South Race Street, University of Denver, Denver, CO 80208, USA*
[d]*Department of Psychology, Stanford University, Stanford, CA 94305, USA*
[e]*Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA*
[f]*Department of Psychology, University of California, Berkeley, CA 94720, USA*

Received 15 February 2007; received in revised form 28 October 2007; accepted 29 October 2007
Communicated by S. Brave
Available online 1 November 2007

Abstract

We present automated, real-time models built with machine learning algorithms which use videotapes of subjects' faces in conjunction with physiological measurements to predict rated emotion (trained coders' second-by-second assessments of sadness or amusement). Input consisted of videotapes of 41 subjects watching emotionally evocative films along with measures of their cardiovascular activity, somatic activity, and electrodermal responding. We built algorithms based on extracted points from the subjects' faces as well as their physiological responses. Strengths of the current approach are (1) we are assessing real behavior of subjects watching emotional videos instead of actors making facial poses, (2) the training data allow us to predict both emotion type (amusement versus sadness) as well as the intensity level of each emotion, (3) we provide a direct comparison between person-specific, gender-specific, and general models. Results demonstrated good fits for the models overall, with better performance for emotion categories than for emotion intensity, for amusement ratings than sadness ratings, for a full model using both physiological measures and facial tracking than for either cue alone, and for person-specific models than for gender-specific or general models.

Keywords: Affective computing; Facial tracking; Emotion; Computer vision

1. Introduction

The number of applications in which a user's face is tracked by a video camera is growing exponentially. Cameras are constantly capturing images of a person's face—on cell phones, webcams, even in automobiles—often with the goal of using that facial information as a clue to understand more about the current state of mind of the user. For example, many car companies (currently in Japan and soon in the US and Europe) are installing cameras in the dashboard with the goal of detecting angry, drowsy, or drunk drivers. Similarly, advertisers on web portals are seeking to use facial information to determine the effect of specific billboards and logos, with the intention of dynamically changing the appearance of a website in response to users' emotions regarding the advertisements. Moreover, video game companies are interested in assessing the player's emotions during game play to help gauge the success of their products.

There are at least two goals in developing real-time algorithms to detect facial emotion using recordings of

*Corresponding author. Tel.: +1 650 7230701; fax: +1 650 7232472.
E-mail addresses: bailenson@stanford.edu (J.N. Bailenson), manos@cs.stanford.edu (E.D. Pontikakis), imauss@psy.du.edu (I.B. Mauss), james@psych.stanford.edu (J.J. Gross), mjabon@stanford.edu (M.E. Jabon), hutcherson@psych.stanford.edu (C.A.C. Hutcherson), nass@stanford.edu (C. Nass), oliver.john@berkeley.edu (O. John).

1071-5819/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2007.10.011

individuals' facial behavior. The first is to assist in the types of human–computer interaction (HCI) applications described above. The second is to advance our theoretical understanding of emotions and facial expression. By using learning algorithms to link rich sets of facial anchor points and physiological responses to emotional responses rated by trained judges, we can develop accurate models of how emotions expressed in response to evocative stimuli are captured via facial expressions and physiological responses. By examining these algorithms, social scientists who study emotion will have a powerful tool to advance their knowledge of human emotion.

2. Related work

2.1. Psychological research on emotion assessment

In the psychological literature, emotion has been defined as an individual's response to goal-relevant stimuli that includes behavioral, physiological, and experiential components (Gross and Thompson, 2007). In the present paper, we focus on the assessment of the first two of these components. There are at least three main ways in which psychologists assess facial expressions of emotions (see Rosenberg and Ekman, 2000, for additional details).

The first approach is to have naïve coders view images or videotapes, and then make holistic judgments concerning the degree to which they see emotions on target faces in those images. While relatively simple and quick to perform, this technique is limited in that the coders may miss subtle facial movements, and in that the coding may be biased by idiosyncratic morphological features of various faces. Furthermore, this technique does not allow for isolating exactly which features in the face are responsible for driving particular emotional expressions.

The second approach is to use componential coding schemes in which trained coders use a highly regulated procedural technique to detect facial actions. For example, the Facial Action Coding System (Ekman and Friesan, 1978) is a comprehensive measurement system that uses frame-by-frame ratings of anatomically based facial features ("action units"). Advantages of this technique include the richness of the dataset as well as the ability to uncover novel facial movements and configurations from data mining the anchor points. The disadvantage of this system is that the frame-by-frame coding of the points is extremely laborious.

The third approach is to obtain more direct measures of muscle movement via facial electromyography (EMG) with electrodes attached on the skin of the face. While this allows for sensitive measurement of features, the placement of the electrodes is difficult and also relatively constraining for subjects who wear them. This approach is also not helpful for coding archival footage.

The use of computer vision algorithms promises to be a solution that maximizes the benefits of the above stated techniques while reducing many of the costs. In the next section, we discuss some of the previous models of detecting facial emotions through computer algorithms.

2.2. Computer vision work

Automatic facial expression recognition and emotion recognition have been researched extensively. One approach has been to evaluate intensity of facial action units (Kimura and Yachida, 1997; Lien et al., 1998; Sayette et al., 2001). Other experiments, such as Essa and Pentland (1997), have represented intensity variation in smiling using optical flow. They measured intensity of face muscles for discriminating between different types of facial actions. Similarly, Ehrlich et al. (2000) emphasized the importance of facial motion instead of the actual face snapshots to recognize emotion in a network environment. While much of the work analyzes the front view of the face, Pantic and Patras (2006) developed a system for automatic recognition of facial action units and analyzed those units using temporal models from profile-view face image sequences.

Many types of algorithms have been employed in this endeavor. For example, Sebe et al. (2002) used video sequences of faces to show that the Cauchy distribution performs better than the Gaussian distribution on recognizing emotions. Similarly, Tian et al. (2000) discriminated intensity variation in eye closure as reliably as did human coders by using Gabor features and an artificial neural network. Zhang et al. (1998) showed that a combination of facial point geometry and texture features, such as Gabor wavelets, led to more accurate estimation of the current facial gesture. Moreover, recent work in Bartlett et al. (2005) has continued to make use of representations based on a combination of feature geometry and texture features. A system developed by Lyons (2004) automatically translated facial gestures to actions using vision techniques. For a more detailed review of the state of the art of current systems, see Li and Jain (2005) or Lyons and Bartneck (2006).

In terms of using the facial tracking data to predict affective states, the pioneering work of Picard et al. (see Picard, 1997 for an early example, and Picard and Daily, 2005, for a recent review of this work) has demonstrated across a number of types of systems that it is possible to track various aspects of the face, and that by doing so one can gain insight into the mental state of the person whose face is being tracked. More recently, el Kaliouby et al. (el Kaliouby et al., 2003; Michel and el Kaliouby, 2003; el Kaliouby and Robinson, 2005) have developed a general computational model for facial affect inference and have implemented it as a real-time system. This approach used dynamic Bayesian networks for recognizing six classes of complex emotions. Their experimental results demonstrated that it is more efficient to assess a human's emotion by looking at the person's face historically over a two second window instead of just the current frame. Their system was designed to classify discrete emotional classes as opposed to the intensity of each emotion.

More generally, there has been much work in human–computer interaction using learning algorithms to predict human behavior. For example, work by Curhan and Pentland (2007) utilized automatic feature extraction from spoken voice to predict quite reliably the outcome of very complex behavior in terms of performance in negotiations. The models presented in the current paper will aid researchers who seek to use real-time computer vision to predict various types of human behavior by providing accurate, real-time methods for extracting emotional information to use as input for those more elaborate psychological processes.

3. Our approach

There are a number of factors that distinguish the current approach from previous ones. First, the stimuli used as input are videotapes of people who were watching film clips designed to elicit intense emotions. The probability that we accessed actual emotional behavior is higher than in studies that used deliberately posed faces (see Nass and Brave, 2005, for further discussion of this distinction). One example for the importance of the distinction between automatically expressed and deliberately posed emotions is given by Paul Ekman and colleagues. They demonstrated that only "Duchenne Smiles"—automatic smiles involving crinkling of the eye corners—but not deliberately posed smiles correlate with other behavioral and physiological indicators of enjoyment (Ekman et al., 1990). Indeed, there is a large amount of research attempting to detect deception through facial and vocal cues by distinguishing incidental from deliberate behaviors (see Ekman, 2001 for a review). In sum, some emotional facial expressions are deliberate, while others are automatic, and the automatic facial expressions appear to be more informative about underlying mental states than posed ones.

Second, because in our approach the emotions were coded second-by-second by trained coders using a linear scale for two oppositely valenced emotions (amusement and sadness), we are able to train our learning algorithms using not just a binary set of data (e.g., sad versus not-sad), but also a linear set of data spanning a full scale of emotional intensity. Most psychological models of emotion allow for the expression of mixed emotional states (e.g., Bradley, 2000). Our approach allows us to compare approaches that only look at binary values—in our case the two most extreme values on the ends of the linear scale—to approaches that linearly predict the amount of amusement and sadness.

Third, given that we collected large amounts of data from each person (i.e., hundreds of video frames rated individually for amusement and sadness), we are able to create three types of models. The first is a "universal model" which predicts how amused *any* face is by using one set of subjects' faces as training data and another independent set of subjects' faces as testing data. This model would be useful for HCI applications in which lots of people use the same interface, such as bank automated teller machines, traffic light cameras, and public computers with webcams. The second is an "idiosyncratic model" which predicts how amused or sad *a given* face is by using training and testing data from the same subject for each model. This model is useful for HCI applications in which the same person repeatedly uses the same interface—for example, driving in an owned car, using the same computer with a webcam, or any application with a camera in a private home. The third is a *gender-specific* model, trained and tested using only data from subjects of the same gender. This model is useful for HCI applications targeting a specific gender—for example make-up advertisements directed at female consumers, or home repair advertisements targeted at males. It is also theoretically interesting to compare the idiosyncratic, gender-specific, and universal models as such a comparison provides valuable information to social scientists studying how personal differences such as gender effect the expression of emotion. Furthermore, although it has previously been shown that the effectiveness of facial expression recognition systems is usually affected by the subject's skin color, facial and scalp hair, sex, race, and age (Zlochower et al., 1998), the comparison of the various individual model enables us to quantitatively evaluate these differences, and better predict the differences in performance of emotion recognition systems via personal differences.

Fourth, since our data include physiological responses (cardiovascular activity, electrodermal responding, and somatic activity) we are able to quantify the improvement in the fit of our models by the addition of such features. One could easily imagine practical contexts in which physiological data could easily be added, such as in an automobile in which the interface could capture facial features from a camera in the dashboard and measure heart rate from the hands gripping the steering wheel. Comparing fit of the models with and without physiological data offers new information regarding the effectiveness of emotion-detection systems with both facial and physiological inputs. This enables application designers to assess the rewards of building physiological measures into their emotion-detection systems.

Finally, all of the processing (e.g., computer vision algorithms detecting facial features, physiological measures, formulas based on the learning algorithms) used in our study can be utilized in real-time. This is essential for applications that seek to respond to a user's emotion in ways to improve the interaction, for example cars which seek to avoid accidents for drowsy drivers or advertisements which seek to match their content to the mood of a person walking by a billboard.

We targeted amusement and sadness in order to sample positive and negative emotions that recruit behavioral as well as physiological responses. Amusement rather than happiness was chosen, because amusement more clearly allows predictions on which facial behaviors to expect (Bonanno and Keltner, 2004). Sadness was then chosen as

the emotion opposite to amusement on the valence continuum (cf. Watson and Tellegen, 1985). We chose only these two emotions since increasing the number of emotions would come at the cost of sacrificing the reliability of the emotions we induced. Amusement and sadness (in contrast to anger, fear, or surprise) can be ethically and reliably induced using films (Philippot, 1993; Gross and Levenson, 1995), a feature crucial to the present design as films allow for standardization of moment-by-moment emotional context across participants across long enough time periods. The selected films induced dynamic changes in emotional states over the 9-min period, ranging from neutral to more intense emotional states. Because different individuals responded to films with different degrees of intensity we were able to assess varying levels of emotional intensity across participants.

4. Data collection

The training data were taken from a study in which 151 Stanford undergraduates watched movies pretested to elicit amusement and sadness while their faces were videotaped and their physiological responses were assessed. In the laboratory session, participants watched a 9-min film clip that was composed of an amusing, a neutral, a sad, and another neutral segment (each segment was approximately 2 min long). From the larger dataset of 151, we randomly chose 41 to train and test the learning algorithms. We did not use all 151 due to the time involved running the models with such rich datasets. In incremental tests during dataset construction, we determined that the current sample size was large enough such that adding additional subjects did not change the fits of the models.

4.1. Expert ratings of emotions

A total of five trained coders rated facial expressions of amusement and sadness from the video recordings of participants' faces such that each participant's tape was rated by two coders (cf. Mauss et al., 2005). Coders used laboratory software to rate the amount of amusement and sadness displayed in each second of video. The coding system was informed by microanalytic analyses of expressive behavior (Ekman and Friesan, 1978). It was anchored at 0 with neutral (no sign of emotion) and 8 with strong laughter for amusement and strong sadness expression/sobbing for sadness. Coders were unaware of other coders' ratings, of the experimental hypotheses, and of which stimuli participants were watching. Average inter-rater reliabilities were satisfactory, with Cronbach's alphas = 0.89 (S.D. = 0.13) for amusement behavior and 0.79 (S.D. = 0.11) for sadness behavior. We thus averaged the coders' ratings to create one second-by-second amusement and one second-by-second sadness rating for each participant. These average ratings of amusement and sadness were used as criterion in our model.

4.2. Physiological measures

During the experimental session, 15 physiological measures were monitored at 400 Hz using a 12-channel Grass Model 7 polygraph. Fig. 1 depicts a participant wearing the measurement sensors. The features included: *heart rate* (derived from inter-beat intervals assessed by placing Beckman miniature electrodes in a bipolar configuration on the participant's chest and calculating the interval in ms between successive R-waves), *systolic blood pressure* (obtained from the third finger of the non-dominant hand), *diastolic blood pressure* (obtained from the third finger of the non-dominant hand), *mean arterial blood pressure* (obtained from the third finger of the non-dominant hand), *pre-ejection period* (identified as the time in ms elapsed between the Q point on the ECG wave of the left ventricle contracting and the B inflection on the ZCG wave), *skin conductance level* (derived from a signal using a constant-voltage device to pass 0.5 V between Beckman

Fig. 1. System for recording physiological data.

electrodes attached to the palmar surface of the middle phalanges of the first and second fingers of the non-dominant hand), *finger temperature* (measured with a thermistor attached to the palmar surface of the tip of the fourth finger), *finger pulse amplitude* (assessed using a UFI plethysmograph transducer attached to the tip of the participant's second finger), *finger pulse transit time* (indexed by the time in ms elapsed between the closest previous R-wave and the upstroke of the peripheral pulse at the finger), *ear pulse transit time* (indexed by the time in ms elapsed between the closest previous R-wave and the upstroke of the peripheral pulse at the ear), *ear pulse amplitude* (measured with a UFI plethysmograph transducer attached to the participant's right ear lobe), *composite of peripheral sympathetic activation* (as indexed by a composite of finger pulse transit time, finger pulse amplitude, ear pulse transit time, and finger temperature), *composite cardiac activation* (as indexed by a composite of heart rate, finger pulse transit time reversed, finger pulse amplitude reversed, and ear pulse transit time reversed standardized within individuals and then averaged), and *somatic activity* (assessed through the use of a piezo-electric device attached to the participant's chair, which generates an electrical signal proportional to the participant's overall body movement in any direction). For more detailed descriptions of these measures, see Gross and Levenson (1995), Mauss et al. (2006).

5. System architecture

The videos of the 41 participants were analyzed at a resolution of 20 frames per second. The level of amusement/sadness of every person for every second in the video was measured via the continuous ratings from 0 (less amused/sad) to 8 (more amused/sad). The goal was to predict at every individual second the level of amusement or sadness for every person based on measurements from facial tracking output and physiological responses (Fig. 2).

For measuring the facial expression of the person at every frame, we used the NEVEN Vision Facial Feature Tracker, a real-time face-tracking solution. This software

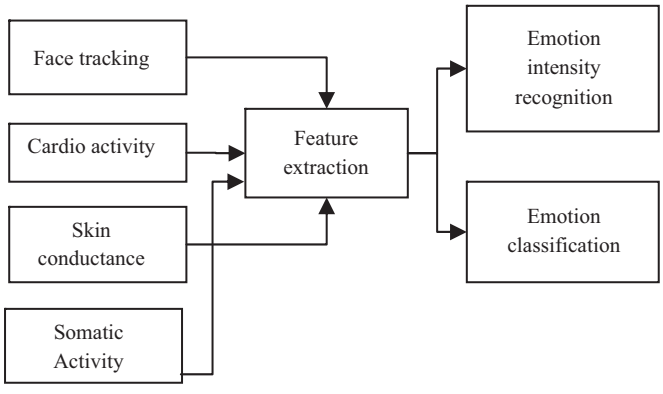

Fig. 2. Emotion recognition system architecture.

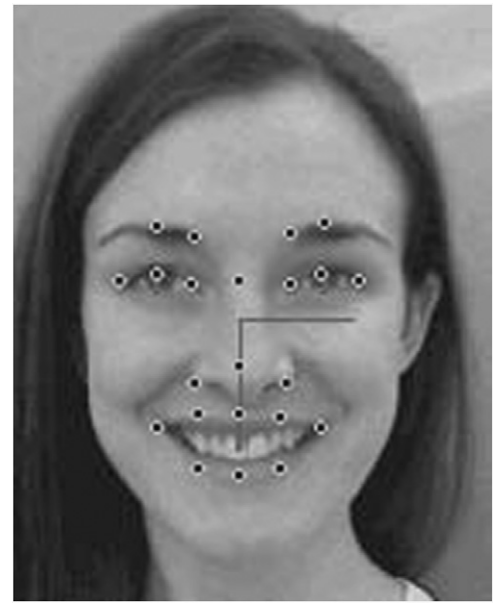

Fig. 3. The points tracked by NEVEN Vision real-time face tracking.

uses patented technology to track 22 points on a face at the fate of 30 frames per second with verification rates of over 95% (Fig. 3).

By plugging our videos into the NEVEN Vision software using Vizard 2.53, a Python-based virtual environment development platform, we extracted 53 measurements of head-centered coordinates of the face at every frame as well as the confidence rating of the face tracking algorithm. All the points were measured in a two-dimensional head-centred coordinate system normalized to the apparent size of the head on the screen; the coordinates were not affected by rigid head movements, and scaled well to different heads. These 53 points included eight points around the contour of the mouth (three on each lip, and one at each corner), three points on each eye (including the pupil), two points on each eyebrow, and four points around the nose. Pitch, yaw and roll of the face, as well the aspect ratio of the mouth and each eye, the coordinates of the face in the image (a loose proxy for posture), and the scale of the face (which is inversely proportional to the distance from the face to the camera, another indication of posture) were also included. Our real-time face-tracking solution required no training, face-markers, or calibration for individual faces, and collected data at 30 Hz. When the confidence rating of the face-tracking algorithm fell below 40%, the data were discarded and the software was told to re-acquire the face from scratch. We used the software on the pre-recorded videos because the experiment in which the subjects had their faces recorded occurred months before the current study. However, given that the NEVEN vision software locates the coordinates at 30 Hz, the models we developed would currently work in real-time (Bailenson et al., 2006).

In our final datasets, we included the 53 NEVEN Vision library facial data points. We excluded the confidence

rating, as it is not a meaningful predictor a priori of emotion. We also included six new features which we created heuristically from linear and non-linear combinations of the NEVEN Vision coordinates: the difference between the right mouth corner and right eye corner, the difference left mouth corner and left eye corner, the mouth height, the mouth width, the upper lip curviness (defined as the difference between the right upper lip Y and the upper lip center Y plus the difference between the left upper lip Y and the upper lip center Y), and the mouth aspect ratio divided by the product of the left and right eye aspect ratios. We analyzed the 2 s history for those 59 features, computing the averages, the velocities, and the variances for each one of them. This totaled to 236 facial features (59 instantaneous, 59 averaged positions, 59 averaged velocities, 59 averaged variances) used as inputs to the models. Finally, we added the 15 physiological and somatic measures utilized by Mauss et al. (2006). So in total, there were 251 features used (236 facial features, 15 physiological features). A complete list of these features can be found in Appendix A.

6. Relevant feature extraction

We applied Chi-square feature selection (which evaluates the contribution of each feature by computing the value of the Chi-squared statistic with respect to the emotion ratings) using the freely distributed machine learning software package Waikato Environment for Knowledge Analysis (WEKA; Witten and Fank, 2005) to find the most relevant features for the amusement dataset. For this experiment, we processed 19,625 instances and we discretized the expert's ratings into two classes (amused and neutral) where each rating above 3 is considered to be amused and each rating below 0.5 is considered as neutral. We repeated the same methodology for finding the most relevant features for the sadness dataset. The top 20 results are shown in Tables 1 and 2. For amusement, the facial characteristics were the most informative (compared to the physiological measures) according to Chi-square, with only two of the physiological features appearing in the top 20. In contrast, for predicting sadness the physiological features seemed to play much more of a role, with 6 out of the top 20 features being physiological. This would indicate that the facial features by themselves are not as strong an indicator of sadness as the physiological characteristics. It is important to note that while the Chi-square analysis is important to understand the features which contribute most to the model fit, we used all facial and physiological features when building our models.

Table 1
Chi-square values for top 20 features in amusement analysis

Chi-square value	Features from amusement analysis
5833.16	Average difference right mouth corner eye corner
5402.68	Average difference left mouth corner eye corner
5392.56	Difference right mouth corner eye corner
5285.98	Average face left mouth corner Y
5043.99	Difference left mouth corner eye corner
4971.89	Face left mouth corner Y
4631.21	Somatic activity
4449.13	Average mouth aspect ratio, divided with eyes aspect ratio
4407.18	Average face right mouth corner Y
4233.76	Face right mouth corner Y
4198.46	Average upper lip curviness
3944.40	Finger temperature
3868.17	Mouth aspect ratio, divided with eyes aspect ratio
3710.65	Upper lip curviness
3635.23	Average face left upper lip Y
3600.06	Average mouth aspect ratio
3581.85	Average face right upper lip Y
3348.80	Left upper lip Y
3314.28	Average face left nostril Y
3241.55	Mouth aspect ratio

Table 2
Chi-square values for top 20 features in sadness analysis

Chi-square value	Features from sadness analysis
5882.53	Finger temperature
4903.57	Skin conductance level
3391.84	Average face Y
3356.33	Average face X
3314.70	Face X
3282.54	Face Y
2601.23	Average Face Scale
2321.38	Average Face Euler Y
2152.49	Average upper lip curviness
2066.28	Face Scale
2031.30	Face Euler Y
1995.66	Heart rate
1975.78	Average face left nostril Y
1930.09	Ear pulse transit time
1802.30	Average face left mouth corner Y
1743.33	Average difference left mouth corner eye corner
1657.78	Face left nostril Y
1656.34	Average face nose tip Y
1622.59	Finger pulse transit time
1615.12	Ear pulse amplitude

7. Predicting emotion intensity

We began with the more challenging task of assessing emotion intensity before turning to the more commonly reported task of classifying emotion. Research by Schiano et al. (2004) has demonstrated that people perceive emotions of others in a continuous fashion, and that merely representing emotions in a binary (on/off) manner is problematic. Consequently, we used linear regression and neural networks for predicting experts' ratings in a continuous manner for every second in the face video.

We used the WEKA software package linear regression function using the Akaike criterion for model selection and used no attribute selection. The linear neural nets were Multilayer Perceptrons configured to have two hidden layers. Two-fold cross-validation was performed on each

Table 3
Linear classification results for all-subject datasets

Category	Regression type	Emotion	Correlation coefficient	Mean absolute error	Mean squared error
Face	Linear regression	Amusement	0.53	1.02	1.44
Face	Linear regression	Sadness	0.23	0.62	0.79
Face	Neural network	Amusement	0.58	0.99	1.51
Face	Neural network	Sadness	0.20	0.57	0.80
Face and physio	Linear regression	Amusement	0.58	1.05	1.45
Face and physio	Linear regression	Sadness	0.30	0.61	0.79
Face and physio	Neural network	Amusement	0.58	0.90	1.38
Face and physio	Neural network	Sadness	0.21	0.61	0.85
Physio	Linear regression	Amusement	0.48	1.03	1.34
Physio	Linear regression	Sadness	0.08	0.68	0.93
Physio	Neural network	Amusement	0.37	1.91	2.27
Physio	Neural network	Sadness	0.08	0.68	0.93

dataset using two non-overlapping sets of subjects. We performed separate tests for both sadness and amusement, using face video alone, physiological features alone, as well as face video in conjunction with the physiological measures to predict the expert ratings. All classifiers were trained and tested on the entire nine minutes of face video data. Our intention in doing so was to demonstrate how effective a system predicting emotion intensity just from a camera could be and to allow application designers to assess the rewards of building in physiological measures. The results are shown in Table 3.

As can be seen, the classifiers using only the facial features performed substantially better than the classifiers using only the physiological features, having correlation coefficients on average nearly 20% higher. Yet combining the two sets of data yielded the best results; with both facial and physiological data included the correlation coefficients of the linear regressions increased by 5% over the next best model in the amusement dataset and by 7% in the sadness dataset, and the neural networks performed slightly better as well.

Table 3 also demonstrates that predicting the intensity of sadness is not as easy as predicting the intensity of amusement. The correlation coefficients of the sadness neural nets were consistently 20–40% lower than those for the amusement classifiers. One possible explanation for the discrepancies between the models' performance on amusement in sadness, however, is that amusement dataset had a mean rating of 0.876 (S.D. = 1.50) while the sadness dataset had mean rating of 0.555 (S.D. = 0.73). This difference was significant in a paired t-test, $t(41) = 1.23$, $p < 0.05$, and could partly account for the lower performance of the sadness classifiers; given the lower frequency and intensity of the rated sadness in our subject pool, the models may have had more difficulty in detecting sadness.

8. Emotion classification

The previous section presented models predicting linear amounts of amusement and sadness. This is unique because most work predicting facial expressions of emotion has not utilized a training set rich enough to allow such a fine-grained analysis. However, in order to compare the current work to previous models, which often presented much higher statistical fits than those we presented above with the linear intensity levels of emotion, we processed our dataset to discretize the expert ratings for amusement and sadness. In the amusement datasets all the expert ratings less than or equal to 0.5 were set to neutral, and ratings of 3 or higher were discretized to become amused. In the sadness datasets all the expert ratings less than or equal to 0.5 were discretized to become neutral, and ratings of 1.5 or higher were discretized to become sad. All the other instances (intermediate ratings) were discarded in these new datasets. Other threshold values (e.g., everything below 1.0 being neutral, etc.) were experimented with, but the thresholds of 0.5 and 3 for amusement and 0.5 and 1.5 for sadness yielded the best fits in our models. The percentage of amused instances in the final amused dataset was 15.2% and the percentage of sad instances in the final sad dataset was 23.9%. We applied a Support Vector Machine classifier with a linear kernel and a Logitboost with a decision stump weak classifier using 40 iterations (Freund and Schapire, 1996; Friedman et al., 2000) to each dataset using the WEKA machine learning software package (Witten and Fank, 2005). As in the linear analyses, we split the data into two non-overlapping datasets and performed a two-fold cross-validation on all our classifiers.

$$\mathrm{Precision} = \frac{C_i}{C_i + C'_i}, \ \mathrm{Recall} = \frac{C_i}{N_i}$$

$$F1 = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Fig. 4. The formulas for precision, recall, and F1.

In all the experiments we conducted, we calculated the precision, the recall and the F_1 measure, which is defined as the harmonic mean between the precision and the recall. For a multi-class classification problem with classes A_i,

Table 4
Discrete classification results for all-subject datasets

Category	Classifier	Emotion	Neutral precision	Emotion precision	Neutral F_1 measure	Emotion F_1 measure
Face	SVMs	Amusement	0.93	0.64	0.93	0.62
Face	SVMs	Sadness	0.78	0.35	0.82	0.20
Face	LogitBoost	Amusement	0.93	0.66	0.93	0.63
Face	LogitBoost	Sadness	0.78	0.31	0.79	0.26
Face and physio	SVMs	Amusement	0.95	0.63	0.93	0.66
Face and physio	SVMs	Sadness	0.81	0.51	0.84	0.37
Face and physio	LogitBoost	Amusement	0.94	0.75	0.95	0.69
Face and physio	LogitBoost	Sadness	0.79	0.36	0.79	0.28
Physio	SVMs	Amusement	0.88	0.77	0.93	0.41
Physio	SVMs	Sadness	0.78	0.32	0.79	0.25
Physio	LogitBoost	Amusement	0.90	0.49	0.91	0.49
Physio	LogitBoost	Sadness	0.78	0.43	0.76	0.24

$i = 1, \ldots, M$ and each class A_i having a total of N_i instances in the dataset, respectively, if the classifier predicts correctly C_i instances for A_i and predicts C_i' instances to be in A_i where in fact those belong to other classes (misclassifies them), then the former measures are defined as following (Fig. 4):

$$\text{Precision} = \frac{C_i}{C_i + C_i'}, \qquad \text{Recall} = \frac{C_i}{N_i}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Maximizing precision or recall individually does not result in a perfect classifier; F_1 gives the optimal accuracy score by relating precision to recall. The results of our analyses are shown in Table 4.

In these analyses both classifiers performed equally as well, with precisions nearing 70% for amusement, 50% for sadness, and 94% for neutral in the face and physiological datasets, a substantial improvement over the precisions of the linear classifiers. We noted too, that the addition of the physiological features offered much greater improvement in the discrete classifiers than in the linear classifiers. The addition of physiological features increased the SVM sadness precision by over 15% and the LogitBoost amusement precision by 9%. Also, just as in the linear analyses, the precisions of the sadness classifiers were consistently over 15% worse than the precisions of the amusement classifiers.

9. Experimental results within subjects

In addition to creating general models applicable to any subject, we ran experiments in which we trained and tested individual models specifically for each one of the 41 subjects. We expected the linear prediction and the classification accuracy to be better within the same subject, since the models are optimized for the facial characteristics of each specific subject as well as his or her levels of expressivity.

Table 5
Linear classification results for individual subjects (average results)

Categories	Emotion	Correlation coefficient (average)	Mean absolute error (average)	Mean squared error (average)
Face only	Amusement	0.85	0.38	0.64
Face only	Sadness	0.73	0.20	0.33
Face and physio	Amusement	0.90	0.29	0.52
Face and physio	Sadness	0.83	0.15	0.26
Physio only	Amusement	0.84	0.48	0.69
Physio only	Sadness	0.83	0.16	0.24

9.1. Predicting continuous ratings within subjects

We built 41 different Multilayer Perceptron neural nets with two hidden layers and individualized them by training and testing them only within the same subject. We chose Multilayer Perceptron neural nets over regression formulas as the previous analyses indicated better fits with the neural nets. For each subject, we used two-fold cross-validation for training and testing. In Table 5, we present the average of the results of the 41 neural nets.

Using these idiosyncratic methods of building specialized models for particular subjects, we noted a number of important trends.

First, building specialized models for each subject significantly increased the prediction accuracy. With sadness in particular, we saw an improvement in the correlation coefficient of more than 50%. This is especially remarkable given that the input set was reduced 20-fold; the all-subject training sets had on average 12,184 instances

Table 6
Discrete classification results for individual subjects (average results)

Categories	Emotion	Neutral accuracy (average)	Emotion accuracy (average)	Neutral F_1 measure (average)	Emotion F_1 measure (average)
Face only	Amusement	0.99	0.93	0.99	0.90
Face only	Sadness	0.97	0.89	0.97	0.86
Face and physio	Amusement	0.99	0.94	0.99	0.92
Face and physio	Sadness	0.99	0.98	0.99	0.95
Physio only	Amusement	0.96	0.82	0.97	0.73
Physio only	Sadness	0.95	0.82	0.95	0.77

while the individual training sets had on average only 595 instances. So even though the within-subject models only had about 300 trials to train on, the fits remained quite high.

Second, like the universal models, using physiological measures improved the fit for all models. Interestingly, the classifier for sadness using only physiological data slightly out-performed the classifier using only facial features. This supports our earlier findings that physiological features seem to be more important in the detection of sadness compared to amusement.

The mean absolute error and the mean squared error were not comparable between the amusement and sadness cases, however since mean ratings of the two datasets were unequal; the majority of expert ratings on sadness did not go beyond the scale of 3.5, while the amusement ratings fluctuated in scale from 0 to 7.

9.2. Classification results within subjects

We performed a similar analysis by building an individual Support Vector Machine classifier with a linear kernel for each one of the 41 subjects. In Table 6 we present those results.

As can be seen by comparing the prediction success in Table 6 to all other tables in the paper, the discrete classifiers designed to predict emotion within subjects performed by far the best, with average accuracies nearing 95%.

10. Experimental results by gender

Given that previous research has identified systematic gender differences in facial expressions of emotion, with women appearing somewhat more accurate in expressing some emotions than men (see Hall, 1984, for a review; Timmers et al., 1998; Kring, 2000), we separated our dataset into two parts, with one part containing only male subjects ($n = 17$) and the other part only female subjects ($n = 24$). We created individual classifiers for each of the datasets in order to compare their performance. We expected the linear prediction and the classification accuracy to be better for the female classifiers given the greater facial expressiveness of women. We also performed relevant feature extraction on each of the datasets to examine any differences in the most informative features between the two genders.

10.1. Relevant feature extraction within gender

We applied Chi-square feature selection using the WEKA machine learning software package (Witten and Fank, 2005) to find the most relevant features for both the male and female amusement datasets. For this experiment, we discretized the expert's ratings into two classes (amused and neutral) where each rating above 3 was considered to be amused and each rating below 0.5 was considered as neutral. We repeated the same methodology for finding the most relevant features for the male and female sadness datasets. The top 20 results for each gender are shown in Tables 7 and 8. We observed that in the male dataset the physiological measures were more informative according to Chi-square than for females. This is especially noticeable in the sadness analysis where 8 of the top 20 male features are physiological whereas only 3 of the top 20 female characteristics are physiological.

10.2. Predicting continuous ratings within gender

We created separate Multilayer Perceptron neural network models with two hidden layers for each gender and measured the correlation coefficient, mean absolute error, and root mean squared error. As in previous analyses the subjects were split into two non-overlapping datasets in order to perform two-fold cross-validation on all classifiers. The results are shown in Table 9.

As can be seen, the female classifiers generally yielded a greater correlation coefficient, suggesting that our models more accurately predict emotions in women than in men. Also, adding the physiological features increased the correlation coefficient in males by almost 20%, whereas it only increased the correlation coefficient in females by 10%. This indicates that physiological features may be

Table 7
Chi-square values for top 20 features in male and female amusement analyses

Chi-square	Features from male amusement analysis	Chi-square	Features from female amusement analysis
2390.29	Average difference left mouth corner eye corner	3363.06	Average left mouth corner Y
2192.56	Difference left mouth corner eye corner	3236.11	Average mouth aspect ratio divided with eyes aspect ratio
2173.04	Average left mouth corner Y	3209.90	Finger temperature
2149.07	Average difference right mouth corner eye corner	3196.37	Average difference left mouth corner eye corner
2107.02	Somatic activity	3137.08	Left mouth corner Y
1996.80	Left mouth corner Y	3115.47	Somatic activity
1925.40	Skin conductance level	3099.96	Average right mouth corner Y
1787.31	Average upper lip curviness	2974.77	Right mouth corner Y
1785.56	Difference right mouth corner eye corner	2966.45	Difference left mouth corner eye corner
1533.58	Finger temperature	2890.10	Average left upper lip Y
1454.53	Upper lip curviness	2823.22	Average mouth aspect ratio
1369.33	Average right mouth corner Y	2789.79	Average right upper lip Y
1347.73	Average left nostril Y	2659.73	Average mouth height
1346.67	Average left upper lip Y	2610.82	Mouth aspect ratio divided with eyes aspect ratio
1240.03	Composite of cardiac activation	2562.26	Left upper lip Y
1217.84	Right mouth corner Y	2526.48	Mouth aspect ratio
1203.93	Left upper lip Y	2470.99	Average upper lip curviness
1174.46	Mouth aspect ratio divided with eyes aspect ratio	2442.35	Mouth height
1096.93	Average right upper lip Y	2387.87	Right upper lip Y
1095.57	Left nostril Y	2333.37	Skin conductance level

Table 8
Chi-square values for top 20 features in male and female sadness analyses

Chi-square	Features from male sadness analysis	Chi-square	Features from female sadness analysis
3907.62	Skin conductance level	3363.06	Average left mouth corner Y
3412.75	Finger temperature	3236.11	Average mouth aspect ratio divided with eyes aspect ratio
2372.60	Average X position	3209.90	Finger temperature
2325.77	X position	3196.37	Average difference left mouth corner eye corner
2169.87	Average Euler Y	3137.08	Left mouth corner Y
2148.95	Y position	3115.47	Somatic activity
2114.66	Average scale	3099.96	Average right mouth corner Y
2065.24	Average Y position	2974.77	Right mouth corner Y
1984.77	Euler Y	2966.45	Difference left mouth corner eye corner
1939.58	Scale	2890.10	Average left upper lip Y
1779.59	Heart rate	2823.22	Average mouth aspect ratio
1691.42	Pre-ejection period	2789.77	Average right upper lip Y
1682.98	Average left pupil X	2659.73	Average mouth height
1646.60	Ear pulse transit time	2610.82	Mouth aspect ratio divided with eyes aspect ratio
1523.33	Ear pulse amplitude	2562.26	Left upper lip Y
1467.22	Diastolic blood pressure	2526.48	Mouth aspect ratio
1448.72	Average upper lip curviness	2470.99	Average upper lip curviness
1380.15	Average left nostril Y	2442.35	Mouth height
1347.30	Finger pulse transit time	2387.87	Right upper lip Y
1339.80	Average left eye aspect ratio	2333.37	Skin conductance level

more important in detecting male emotional responses than female responses.

10.3. Classification results by gender

We performed a similar analysis by building an individual Support Vector Machine classifier with a linear kernel for both males and females. As in other classifications, two-fold cross-validation was used. We present those results in Table 10.

Again we see a significantly higher accuracy in our female models over our male models. Interestingly, the only classifier that performed better in the male dataset than the female dataset was the sadness classifier using only physiological data. Also, when adding the physiological data to the facial data we only saw improvements in

Table 9
Linear classification results for gender-specific datasets

Categories	Emotion	Gender	Correlation coefficient (average)	Mean absolute error (average)	Mean squared error (average)
Face only	Amusement	Male	0.25	1.40	1.94
Face only	Amusement	Female	0.63	0.94	1.43
Face only	Sadness	Male	0.23	0.61	0.82
Face only	Sadness	Female	0.33	0.56	0.79
Face and physio	Amusement	Male	0.45	0.92	1.48
Face and physio	Amusement	Female	0.73	0.77	1.17
Face and physio	Sadness	Male	0.20	0.57	0.78
Face and physio	Sadness	Female	0.04	0.65	0.82
Physio only	Amusement	Male	0.24	2.26	2.71
Physio only	Amusement	Female	0.20	2.06	2.40
Physio only	Sadness	Male	0.02	0.71	0.97
Physio only	Sadness	Female	0.03	0.78	1.01

Table 10
Discrete classification results for gender-specific datasets

Categories	Emotion	Gender	Neutral accuracy (average)	Emotion accuracy (average)	Neutral F_1 measure (average)	Emotion F_1 measure (average)
Face only	Amusement	Male	0.93	0.53	0.93	0.62
Face only	Amusement	Female	0.94	0.69	0.93	0.68
Face only	Sadness	Male	0.77	0.15	0.84	0.46
Face only	Sadness	Female	0.80	0.39	0.80	0.39
Face and physio	Amusement	Male	0.93	0.28	0.85	0.34
Face and physio	Amusement	Female	0.95	0.82	0.96	0.79
Face and physio	Sadness	Male	0.85	0.38	0.85	0.38
Face and physio	Sadness	Female	0.84	0.42	0.84	0.47
Physio only	Amusement	Male	0.90	0.43	0.90	0.25
Physio only	Amusement	Female	0.89	0.62	0.88	0.41
Physio only	Sadness	Male	0.87	0.29	0.87	0.34
Physio only	Sadness	Female	0.71	0.19	0.56	0.22

performance in the male classifiers. These results support the findings of the Chi-square analysis, suggesting physiological data are more important for males than females.

11. Conclusion and future work

We have presented a real-time system for emotion recognition and showed that this system is accurate and easy to implement. The present study is unique for a number of reasons, perhaps most notably because of the unusually rich data set. A relatively large number of subjects watched videos designed to make them feel amused or sad while having their facial and physiological responses recorded, and we then produced second-by-second ratings of the intensity with which they expressed amusement and sadness using trained coders. By having this level of detail in both input and output, we were able to make a number of important advances in our learning algorithms.

11.1. Summary of findings

First, we demonstrated the ability to find good statistical fits on algorithms to predict the emotion from the natural facial expressions of everyday people, rather than from discrete and deliberately created facial expressions of trained actors, as in many previous studies. This is important, because people in their day-to-day lives may not produce extreme facial configurations such as those displayed by actors used in typical experimental stimuli. Consequently, previous work may be overestimating the utility of emotion prediction based on the novelty of the stimulus set.

Second, in the current study, we demonstrated that amusement is more easily detected than sadness, perhaps due to the difficulty in eliciting true sadness. In our dataset, facial expressions of people watching sad movies and receiving high sadness ratings tended to not have the stereotypical "long face", but were predominantly characterized by a lack of movement or any

expressivity at all. Consequently, we are demonstrating the importance of examining people experiencing emotions in a naturalistic element. Previous work has also demonstrated that sadness is a difficult emotion to capture using analysis of facial feature points (Deng et al., 2006).

Third, we provided evidence that applications for which a single user occupies an interface over time, models tailored to that user, show significant advances over more general models. While in many ways this is intuitive, quantifying the exact level of improvement is an important first step in designing these systems. Specifically with categorizing emotions, the tailored individual models performed extremely well compared to the other models.

Fourth, we have shown that both amusement and sadness are more easily detected in female subjects than in male subjects. This finding is consistent with the research done by Hall (1984) suggesting that women are more facially expressive than men, and provides new quantitative data for social scientists studying the differences in emotional response among individuals of opposite gender.

Fifth, we have demonstrated that by incorporating measures of physiological responding into our model, we get more accurate predictions than when just using the face. Indeed, when we analyze physiological features as the sole inputs to the model, the fit is often extremely high at predicting the coded emotion ratings of the face. Such measurements can be used in real systems with relatively easy installments of sensors (e.g., on a person's chair or on the steering wheel of a car). In fact, the Chi-square analysis indicates that some of the physiological levels in the detection of sadness outperformed facial tracking, especially for males. Given that real-time computer vision algorithms are not yet as reliable as physiological measurement techniques in terms of consistent performance, augmenting facial tracking with physiological data may be crucial.

11.2. Limitations and future work

Of course there are a number of limitations to the current work. First, our models' accuracy is closely related to the quality of the vision library that we are using as well as the accuracy of our physiological measures. As these tools improve, our system will become much more useful. Moreover, while the psychologists trained to code amusement and sadness demonstrated high inter-coder reliability, it could be the case that their ratings were not actually picking up the "true emotion" but were picking up on other types of behavioral artifacts. Our model is only as good as the input and output used to train, and while we are confident that this dataset is more robust than most that have been used previously, there are many ways to improve our measurements.

Second, we only examined two emotions, while most models posit there are many more than two emotions (Ekman and Friesan, 1978). In pilot testing, we examined the videos of subjects and determined that there were very few instances of all seven emotions, such as fear, disgust, and surprise. Consequently, we decided to focus on creating robust models which were able to capture the two oppositely valenced emotions which occurred most frequently in our dataset. We also decided to begin with the most conservative models, which were binary comparisons between amused and neutral and sad and neutral rather than a general comparison of all emotions. We acknowledge, however, that these decisions limit the scope of our experiment. In future work, we can expand the models to include other emotions and to compare emotions directly.

Third, our study was based upon coders' labels of subjects' emotions. Thus, although we are confident in the validity of our coders' ratings based upon their high inter-coder reliability, we cannot claim to be detecting the actual expression of the emotions sadness and amusement; rather, we can only claim to be detecting the expressions of sadness and amusement as evaluated by coders. A possibility for future work would be to repeat the study using reports of emotion from the subjects themselves rather than coders' ratings.

Fourth, our some of the algorithms in our study depend upon physiological features collected through use of electrodes and transducers which may be too intrusive for some applications. In future work alternate ways of obtaining physiological data could be explored.

Finally, all of our results are tied to the specific learning algorithms we utilized as well as to the ways in which we divided the data. The fact that we discarded any data points rated between 0.5 and 3 in our discrete amusement datasets and between 0.5 and 1.5 in our discrete sad datasets makes our models more applicable to subjects with greater facial motion since subjects whose expressions tend to fall in the intermediate range tend to be less represented in the data. It may be the case that different techniques of modeling would produce different patterns of results.

In the future, we plan to use our emotion-recognizer model for analyzing data from other studies; for example, assessing how emotion is related to driving safety and how emotions can affect social interaction during a negotiation setting.

Acknowledgments

We thank Rosalind Picard, Jonathan Gratch, and Alex Pentland for helpful suggestions in early stages of this work. Moreover, we thank Dan Merget for software development, Amanda Luther and Ben Trombley-Shapiro for assistance in data analysis, and Keith Avila, Bryan Kelly, Alexia Nielsen, and Alice Kim for their help in processing video. This work was in part sponsored by NSF Grant 0527377, as well as a grant from OMRON Corporation.

Appendix A. Facial and physiological features

Full list of the 53 facial features from NEVEN Vision, 6 heuristically created facial metafeatures, and 15 physiological features used as inputs for the learning algorithms.

Facial features				Physiological features	
1	X position	28	Left mouth corner X	1	Composite of cardiac activation
2	Y position	29	Left mouth corner Y	2	Skin conductance level
3	Scale	30	Left outer eye corner X	3	Finger temperature
4	Euler X	31	Left outer eye corner Y	4	Systolic blood pressure
5	Euler Y	32	Left inner eye corner X	5	Composite of peripheral sympathetic activation
6	Euler Z	33	Left inner eye corner Y	6	Finger pulse transit time
7	Left eye aspect ratio	34	Right inner eye corner X	7	Ear pulse transit time
8	Right eye aspect ratio	35	Right inner eye corner Y	8	Pre-ejection period
9	Mouth aspect ratio	36	Right outer eye corner X	9	Finger pulse amplitude
10	Right pupil X	37	Right outer eye corner Y	10	Mean arterial blood pressure
11	Right pupil Y	38	Right upper lip X	11	Heart rate
12	Left pupil X	39	Right upper lip Y	12	Ear pulse transit time
13	Left pupil Y	40	Left upper lip X	13	Ear pulse amplitude
14	Right inner eye brow X	41	Left upper lip Y	14	Diastolic blood pressure
15	Right inner eye brow Y	42	Right lower lip X	15	Somatic activity
16	Left inner eye brow X	43	Right lower lip Y		
17	Left inner eye brow Y	44	Left lower lip X		
18	Nose root X	45	Left lower lip Y		
19	Nose root Y	46	Right eye brow center X		
20	Nose tip X	47	Right eye brow center Y		
21	Nose tip Y	48	Left eye brow center X		
22	Upper lip center X	49	Left eye brow center Y		
23	Upper lip center Y	50	Right nostril X		
24	Lower lip center X	51	Right nostril Y		
25	Lower lip center Y	52	Left nostril X		
26	Right mouth corner X	53	Left nostril Y		
27	Right mouth corner Y				

Metafeatures

1	Difference right mouth corner eye corner (right outer eye corner Y-right mouth corner Y)
2	Mouth height (lower lip center Y-upper lip center Y)
3	Difference left mouth corner eye corner (left outer eye corner Y-left mouth corner Y)
4	Mouth width (left mouth corner X—right mouth corner X)
5	Upper lip curviness ((upper lip center Y-right mouth corner Y) + (upper lip center Y-left mouth corner Y))
6	Mouth aspect ratio, divided with eyes aspect ratio (mouth aspect ratio/(right eye aspect ratio × left eye aspect ratio))

Appendix B. Software packages

Software package	Description	Developer	Latest release	OS	License	Website
NEVEN Vision	Real-time face-tracking solution	Nevenvision, Inc.	NEVEN Vision 1.0	Cross-platform	Google, Inc.	www.nevenvision.com
Vizard 2.53 VR Toolkit	Platform for developing 3-D virtual reality worlds	WorldViz, Inc.	Vizard 3.0	Cross-platform	WorldViz, Inc.	http://www.worldviz.com/products/vizard/index.html

Waikato Environment for Knowledge Analysis (WEKA)	Open source collection of visualization tools and algorithms for data analysis and predictive modeling	University of Waikato	WEKA 3.5.6 (developer) WEKA 3.4.11 (book)	Cross-platform	GPL http://www.cs.waikato.ac.nz/~ml/weka/

References

Bailenson, J., Yee, N., Merget, D., Schroeder, R., 2006. The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Non-/verbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction, Presence, vol. 15, No. 4, August 2006, pp. 359–372.

Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J., 2005. Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior, IEEE CVPR 2005.

Bonanno, A.G., Keltner, D., 2004. The coherence of emotion systems: comparing "on-line" measures of appraisal and facial expressions, and self-report. Cognition and Emotion 18, 431–444.

Bradley, M., 2000. Emotion and motivation. In: Cacioppo, J., Tassinary, L., Brenston, G. (Eds.), Handbook of Psychophysiology. Cambridge University Press, New York.

Curhan, J.R., Pentland, A., 2007. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 mins. Journal of Applied Psychology 92, 802–811.

Deng, Z., Bailenson, J.N., Lewis, J.P., Neumann, U., 2006. Perceiving visual emotions with speech. In: Proceedings of the 6th International Conference on Intelligent Virtual Agents.

Ehrlich, S.M., Schiano, D.J., Sheridan, K., 2000. Communicating facial affect: it's not the realism, it's the motion. CHI '00. ACM Press, New York, NY, pp. 251–252.

Ekman, P., 2001. Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. Norton, New York.

Ekman, P., Friesan, W., 1978. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press.

Ekman, P., Davidson, R.J., Friesen, W.V., 1990. The Duchenne smile: emotional expression and brain physiology II. Journal of Personality and Social Psychology 58, 342–353.

el Kaliouby, R., Robinson, P., 2005. Real-time inference of complex mental states from facial expressions and head gestures. In: Real-time Vision for HCI. Springer, pp. 181–200. ISBN 0-387-27697-1.

el Kaliouby, R., Robinson, P., Keates, S., 2003. Temporal context and the recognition of emotion from facial expression. In: Proceedings of HCI International, vol. 2, Crete, June 2003, pp. 631–635. ISBN 0-8058-4931-9.

Essa, I., Pentland, A., 1997. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7), 757–763.

Freund, Y., Schapire, R., 1996. Experiments with a New boosting algorithm. In: Machine Learning—Proceedings of the Thirteenth International Conference. Morgan Kaufmann, San Francisco, pp. 148–156.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. Annals of Statistics 28, 337–407.

Gross, J.J., Levenson, R.W., 1995. Emotion elicitation using films. Cognition and Emotion 9, 87–108.

Gross, J.J., Thompson, R., 2007. Emotion regulation: conceptual foundations. In: Gross, J.J. (Ed.), Handbook of Emotion Regulation. Guilford Press, New York, pp. 3–24.

Hall, J.A., 1984. Nonverbal Sex Differences: Communication Accuracy and Expressive Style. The Johns Hopkins University Press, Baltimore, MD.

Kimura, S., Yachida, M., 1997. Facial expression recognition and its degree estimation. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 295–300.

Kring, A.M., 2000. Gender and anger. In: Fischer, A.H. (Ed.), Gender and Emotion: Social Psychological Perspectives. Cambridge University Press, New York, pp. 211–231.

Li, S.Z., Jain, A.K., 2005. Handbook of Face Recognition. Springer, New York. ISBN# 0-387-40595 (Chapter 11).

Lien, J.-J., Kanade, T., Cohn, J., Li, C., 1998. Subtly different facial expression recognition and expression intensity estimation. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 853–859.

Lyons, M.J., 2004. Facial gesture interfaces for expression and communication. In: IEEE International Conference on Systems, Man, and Cybernetics, vol. 1, pp. 598–603.

Lyons, M.J., Bartneck, C., 2006. HCI and the face. In: CHI '06 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 1671–1674.

Mauss, I.B., Levenson, R.W., McCarter, L., Wilhelm, F.H., Gross, J.J., 2005. The tie that binds? Coherence among Emotion Experience, Behavior, and Physiology 5 (2), 175–190.

Mauss, I.B., Evers, C., Wilhelm, F.H., Gross, J.J., 2006. How to bite your tongue without blowing your top: Implicit evaluation of emotion regulation predicts affective responding to anger provocation. Personality and Social Psychology Bulletin 32, 389–602.

Michel, P., el Kaliouby, R., 2003. Real time facial expression recognition in video using support vector machines. In: Proceedings of the 5th International Conference on Multimodal Interfaces, pp. 258–264. ISBN: 1-58113-621-8

Nass, C., Brave, S., 2005. Wired for Speech: How Voice Activates and Advances the Human–Computer Relationship. The MIT Press. ISBN: 0262140926.

Pantic, M., Patras, I., 2006. 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, vol. 36, no. 2, pp. 433–449, April 2006.

Picard, R.W., 1997. Affective Computing. MIT Press, 1997, Cambridge, MA.

Picard, R.W., Daily, S.B., 2005. Evaluating affective interactions: alternatives to asking what users feel. In: CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches, April 2005, Portland, OR.

Philippot, P., 1993. Inducing and assessing differentiated emotion-feeling states in the laboratory. Cognition and Emotion 7 (2), 171–193.

Rosenberg, E.L., Ekman, P., 2000. Emotion: methods of study. In: Kazdin, A.E. (Ed.), Encyclopedia of Psychology, vol. 3. American Psychological Association, pp. 171–175.

Sayette, M., Cohn, J., Wertz, J., Perrott, M., Parrott, D., 2001. A psychometric evaluation of the facial action coding system for assessing spontaneous expression. Journal of Nonverbal Behavior 25, 167–186.

Schiano, D.J., Ehrlich, S.M., Sheridan, K., 2004. Categorical Imperative NOT: Facial Affect is Perceived Continuously. CHI, pp. 49–56.

Sebe, N., Lew, M.S., Cohen, I., Garg, A., Huang, T.S., 2002. Emotion recognition using a cauchy naive bayes classifier. In: Proceedings of ICPR, vol. 1, pp. 17–20.

Tian, Y.-L., Kanade, T., Cohn, J., 2000. Eye-state action unit detection by Gabor wavelets. In: Proceedings of International Conference on Multimodal Interfaces (ICMI), pp. 143–150.

Timmers, M., Fischer, A.H., Manstead, A.S.R., 1998. Gender differences in motives for regulating emotions. Personality and Social Psychology Bulletin 24, 974–985.

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

Witten, H.I., Fank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, second ed. Morgan Kaufmann, San Francisco.

Zhang, Z., Lyons, M., Schuster, M., Akamatsu, S., 1998. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In: 3rd International Conference on Face and Gesture Recognition, p. 454.

Zlochower, A., Cohn, J., Lien, J., Kanade, T., 1998. A computer vision based method of facial expression analysis in parent–infant interaction. In: International Conference on Infant Studies.