# Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior

Marian Stewart Bartlett[1], Gwen Littlewort[1], Mark Frank[2], Claudia Lainscsek[1], Ian Fasel[1], Javier Movellan[1]
Institute for Neural Computation, University of California, San Diego
[2] Rutgers University, New Brunswick, NJ
mbartlett@ucsd.edu

## Abstract

*We present a systematic comparison of machine learning methods applied to the problem of fully automatic recognition of facial expressions. We report results on a series of experiments comparing recognition engines, including AdaBoost, support vector machines, linear discriminant analysis. We also explored feature selection techniques, including the use of AdaBoost for feature selection prior to classification by SVM or LDA. Best results were obtained by selecting a subset of Gabor filters using AdaBoost followed by classification with Support Vector Machines. The system operates in real-time, and obtained 93% correct generalization to novel subjects for a 7-way forced choice on the Cohn-Kanade expression dataset. The outputs of the classifiers change smoothly as a function of time and thus can be used to measure facial expression dynamics. We applied the system to to fully automated recognition of facial actions (FACS). The present system classifies 17 action units, whether they occur singly or in combination with other actions, with a mean accuracy of 94.8%. We present preliminary results for applying this system to spontaneous facial expressions.*

## 1 Introduction

We present results on a user independent fully automatic system for real time recognition of basic emotional expressions from video. The system automatically detects frontal faces in the video stream and codes each frame with respect to 7 dimensions: Neutral, anger, disgust, fear, joy, sadness, surprise. A second version of the system detects 17 action units of the Facial Action Coding System (FACS). We conducted empirical investigations of machine learning methods applied to this problem, including comparison of recognition engines and 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. The combination of AdaBoost and SVM's enhanced both speed and accuracy of the system. The system presented here is fully automatic and operates in real-time. We present preliminary results for recognizing spontaneous expressions in an interview setting.

## 2 Facial Expression Data

The facial expression system was trained and tested on Cohn and Kanade's DFAT-504 dataset [7]. 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 expressions. 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. For our study, we selected the 313 sequences from the dataset that were labeled as one of the 6 basic emotions. The sequences came from 90 subjects, with 1 to 6 emotions per subject. The first and last frames (neutral and peak) were used as training images and for testing generalization to new subjects, for a total of 626 examples. The trained classifiers were later applied to the entire sequence.

### 2.1 Real-time Face Detection

We developed a real-time face detection system that employs boosting techniques in a generative framework [5] and extends work by [17]. Enhancements to [17] 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. The system presently operates at 24 frames/second on a 3 GHz Pentium IV for 320x240 images.

All faces in the DFAT-504 dataset were successfully detected. The automatically located faces were rescaled to 48x48 pixels. The typical distance between the centers of the eyes was roughly 24 pixels. No further registration was performed. The images were converted into a Gabor magnitude representation, using a bank of Gabor filters at 8 orientations and 9 spatial frequencies (2:32 pixels per cycle at 1/2 octave steps) (See [9] and [10]).

# 3 Classification of Full Expressions

## 3.1 Support Vector Machines

We first examined facial expression classification based on support vector machines (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 $O(10^2)$. The system performed a 7-way forced choice between the following emotion categories: Happiness, sadness, surprise, disgust, fear, anger, neutral. Methods for multiclass decisions with SVM's were investigated in [10].

Here, the seven-way forced choice was performed in two stages. In stage I, support vector machines performed binary decision tasks using one-versus-all partitioning of the data, where each SVM discriminated one emotion from everything else. Stage II converted the representation produced by the first stage into a probability distribution over the seven expression categories. This was achieved by passing the 7 SVM outputs through a softmax competition.

Generalization to novel subjects was tested using leave-one-subject-out cross-validation, in which all images of the test subject were excluded from training. Linear, polynomial, and radial basis function (RBF) kernels with Laplacian, and Gaussian basis functions were explored. Linear and RBF kernels employing a unit-width Gaussian performed best, and are presented here. Results are given in Table 1.

## 3.2 Adaboost

SVM performance was next compared to Adaboost for emotion classification. The features employed for the Adaboost emotion 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 the classifier output distributions for the positive and negative samples were completely separated by a gap proportional to the widths of the two distributions. The union of all features selected for each of the 7 emotion classifiers resulted in a total of 900 features.

Classification results are given in Table 1. The generalization performance with Adaboost was comparable to linear SVM performance. Adaboost had a substantial speed advantage. There was a 180-fold reduction in the number of Gabor filters used. Because the system employed a subset of filter outputs at specific image locations the convolutions were calculated in pixel space rather than Fourier space which reduced the speed advantage, but it nevertheless resulted in a speed benefit of over 3 times faster than the linear SVM.

## 3.3 Linear Discriminant Analysis

A previous successful approach to basic emotion recognition used Linear Discriminant Analysis (LDA) to classify Gabor representations of images [11]. While LDA may be optimal when the class distributions are Gaussian, SVM's may be more effective when the class distributions are not Gaussian. Table 1 compares LDA with SVM's and Adaboost. A small ridge term was used in LDA.

The performance results for LDA were dramatically lower than SVMs. Performance with LDA improved by adjusting the decision threshold for each emotion so as to balance the number of false detects and false negatives. This form of threshold adjustment is commonly employed with LDA classifiers, but it uses post-hoc information, whereas the SVM performance was without post-hoc information. Even with the threshold adjustment, the linear SVM performed significantly better than LDA. (See Tables 1 and 2.)

## 3.4 Feature selection using PCA

Many approaches to LDA also employ PCA to perform feature selection prior to classification. For each classifier we searched for the number of PCA components which gave maximum LDA performance, which was typically 40 to 70 components. The PCA step resulted in a substantial improvement. The combination of PCA and threshold adjustment gave performance accuracy of 80.7% for the 7-alternative forced choice, which was comparable to other LDA results in the literature [11]. Nevertheless, the linear SVM outperformed LDA even with the combination of PCA and threshold adjustment. SVM performance on the PCA representation was significantly reduced, indicating an incompatibility between PCA and SVM's for the problem.

### 3.5 Feature selection by Adaboost

Adaboost is not only a fast classifier, it is also a feature selection technique. An advantage of feature selection by Adaboost is that features are selected contingent on the features that have already been selected. In feature selection by Adaboost, each Gabor filter is a treated as a weak classifier. Adaboost picks the best of those classifiers, and then boosts the weights on the examples to weight the errors more. The next filter is selected as the one that gives the best performance on the errors of the previous filter. At each step, the chosen filter can be shown to be uncorrelated with the output of the previous filters [6, 15].

We explored training SVM and LDA classifiers on the features selected by Adaboost. Here, the classifiers were trained on the *continuous* outputs of the selected Gabor features, in contrast to the Adaboost classifier which employed *thresholded* outputs. Adaboost was used to select 900 features from 9x8x48x48=165888 possible Gabor features, which were then classified by the SVM or LDA.

The results are shown in Table 1 and 2. Best performance was obtained with the combination of Adaboost and SVM's. We informally call these combined classifiers AdaSVM. We informally call these combined classifiers AdaSVM. The results are shown in Table 1. AdaSVM's outperformed both Adaboost ($z = 2.1, p = 0.2$) and SVM's ($z = 2.6, p < .01$), where $z$ is the Z-statistic for comparing success rates of Bernoulli random variables, and $p$ is probability that the two performances come from the same distribution. The result of 93.3% accuracy for a user-independent 7-alternative forced choice was encouraging given that previously published results on this database were 81-83% accuracy (e.g. [2]). AdaSVM's also carried a substantial speed advantage over SVM's. The nonlinear AdaSVM was over 400 times faster than the nonlinear SVM.

Regarding LDA, feature selection with Adaboost gave better performance than feature selection by PCA and reduced the difference in performance between LDA and SVM's. Nevertheless, SVM's continued to outperform LDA.

Table 1. Leave-one-out generalization performance of Adaboost,SVM's and AdaSVM's. AdaSVM: Feature selection by AdaBoost followed by classification with SVM's. LDA$_{pca}$: Linear Discriminant analysis with feature selection based on principle component analysis, as commonly implemented in the literature.

| Kernel | Adaboost | SVM | AdaSVM | LDA$_{pca}$ |
|---|---|---|---|---|
| Linear | 90.1 | 88.0 | 93.3 | 80.7 |
| RBF | | 89.1 | 93.3 | |

Table 2. Comparing SVM performance to LDA with different feature selection techniques. The two classifiers are compared with no feature selection, with feature selection by PCA, and feature selection by Adaboost.

| Feature selection | LDA | SVM (linear) |
|---|---|---|
| None | 44.4 | 88.0 |
| PCA | 80.7 | 75.5 |
| Adaboost | 88.2 | 93.3 |

## 4 Application to Spontaneous Behavior

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) [4] is the most objective and comprehensive coding system in the behavioral sciences. A human coder decomposes facial expressions in terms of 46 component movements, which roughly correspond to the 44 facial muscles. Several research groups have recognized the importance of automatically recognizing FACS [3, 16, 14, 8]. Here we apply the system described above to the problem of fully automated facial action coding.

### 4.1 Spontaneous Expression Database

Our collaborators at Rutgers University have collected a dataset of spontaneous facial behavior consisting 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. This paradigm has been shown to elicit a wide range of emotional expressions as well as speech-related facial expressions. 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.

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). 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.

### 4.2 FACS Training

The system was trained on FACS-coded images from 2 datasets. The first dataset was the Cohn Kanade dataset, which contains FACS scores by two certified FACS coders in addition to the basic emotion labels. The second dataset consisted of directed facial actions collected by Hager and Ekman. (See [3].) The combined dataset contained 2568 training examples from 119 subjects. As above, the system was fully automated. Automatic eye detection [5] was employed to align the eyes in each image. Images were scaled to 192x192, passed through a bank of Gabor filters at 8 orientations and 7 spatial frequencies (4:32 pixels per cyc). Output magnitudes were then passed to nonlinear support vector machines using RBF kernels. No feature selection was performed, although we plan to evaluate feature selection by AdaBoost in the near future.

Separate support vector machines, one for each AU, were trained to perform context-independent recognition. In context-independent recognition, the system detects the presence of a given AU regardless of the co-occurring AU's. 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.

## 4.3 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. The results are shown in Table 3. All system outputs above threshold were treated as detections. Performance was evaluated for thresholds of 0 in the SVM, and then evaluated again for the optimal threshold that maximized percent correct.

The system obtained a mean of 94.8% agreement with human FACS labels. System outputs for full image sequences of test subjects are shown in Figure 1. 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, enabling applications for measuring the magnitude and dynamics of facial expressions.

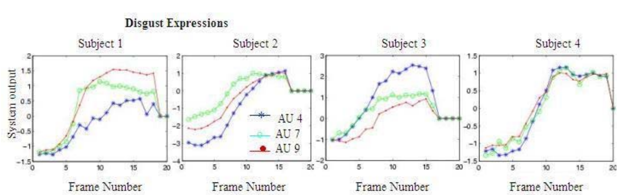


Figure 1. Automated FACS measurements for full image sequences. Shown are 4 subjects from the Cohn-Kanade dataset posing disgust containing AU's 4,7 and 9. These are test sequences not used for training.

Over 7000 action unit combinations have been reported in the psychology literature, and the problem of how to handle recognition of action unit combinations has received considerable discussion (e.g. [16, 13]). Here we address recognition of combinations by training a data-driven system to detect a given action regardless of whether it appears singly or in combination with other actions (context independent recognition). A strength of data-driven systems is that they learn the variations due to combinations, and they also learn the most likely contexts of an action. Nonlinear support vector machines have the added advantage of being able to handle multimodal data distributions which can arise with action combinations.[1] It is an open question whether building classifiers for specific combinations improves recognition performance, and that is a topic of future work.

Table 3. Performance for fully automatic recognition of 17 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). $P_{opt}$: Same with optimal threshold. FA, Hit: Hit and false alarm rates with optimal threshold.

| AU | Name | N | P | $P_{opt}$ | FA | Hit |
|---|---|---|---|---|---|---|
| 1 | Inn. brow raise | 409 | 90.3 | 92.9 | 0.4 | 71.3 |
| 2 | Out. brow raise | 315 | 91.8 | 92.8 | 1.6 | 62.6 |
| 4 | Brow lower | 412 | 82.7 | 86.8 | 6.9 | 41.0 |
| 5 | Upper lid raise | 286 | 91.2 | 92.9 | 2.1 | 61.9 |
| 6 | Cheek raise | 278 | 92.8 | 93.5 | 1.4 | 70.1 |
| 7 | Lower lid tight | 403 | 85.7 | 88.5 | 4.6 | 52.1 |
| 9 | Nose wrinkle | 68 | 98.7 | 98.8 | 0.04 | 85.3 |
| 10 | Lip Raise | 50 | 97.7 | 98.1 | 13.9 | 26.0 |
| 12 | Lip crnr. pull | 196 | 97.8 | 98.0 | 0.04 | 93.4 |
| 15 | Lip crnr. depr. | 100 | 97.0 | 97.2 | 1.0 | 72.0 |
| 17 | Chin raise | 203 | 87.0 | 92.8 | 7.0 | 40.4 |
| 20 | Lip stretch | 99 | 94.4 | 96.2 | 6.6 | 41.4 |
| 23 | Lip tighten | 57 | 97.0 | 97.9 | 11.0 | 36.8 |
| 24 | Lip press | 49 | 98.4 | 98.5 | 1.7 | 61.2 |
| 25 | Lips part | 376 | 89.7 | 91.2 | 2.2 | 64.9 |
| 26 | Jaw drop | 86 | 96.7 | 97.1 | 5.9 | 45.3 |
| 27 | Mouth stretch | 81 | 99.2 | 99.2 | 0.04 | 97.5 |
| | Mean | | 93.4 | 94.8 | 3.9 | 60.2 |

## 4.4 Generalization to Spontaneous Expressions

The system described in Section 4.2 was then tested on the spontaneous expression database. Preliminary results

[1] when the class of kernel is well matched to the problem. The distribution of facial expression data is not well known, and this question requires empirical study. Several labs in addition to ours have found a range of RBF kernels to be effective for face classification tasks.

are presented for 12 subjects. This data contained a total of 1689 labeled events, consisting of 33 distinct action units, 16 of which were AU's for which we had trained classifiers. The face detector operates for frontal faces of $\pm 10\,\mathrm{deg}$, whereas unconstrained head movements during discourse can rotate outside that range. 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. All detected faces were passed to the AU recognition system.

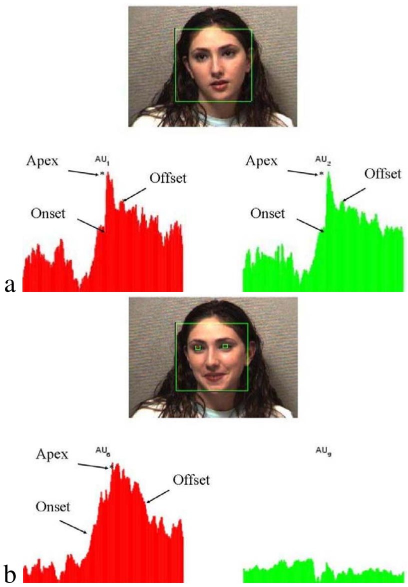


Figure 2. Sample system outputs for a 10-second segment containing a brow-raise (FACS code 1+2). System output is shown for AU 1 (left) and AU 2 (right). Human codes are overlayed for comparison (onset, apex, offset).

Here we present benchmark performance of the basic frame-by-frame system on the video data. Figure 2 shows sample system outputs for one subject, and performance is shown in Table 4. 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 considered correct. This gave an overall accuracy of 90.5% correct across AU's.[2]

Next an interval analysis was performed which measured percent correct 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. An SVM threshold of

[2]Overall percent correct can give high numbers since the AU's are present for a small percentage of frames.

Table 4. Recognition of spontaneous facial actions. AU: Action unit number. N: Total number of testing examples. Dur.: Mean duration of the AU in frames. P: percent correct over all frames; $\mathrm{Hit}_{apex}$: Hit rate for AU apex frame. $\mathrm{P}_\Delta$: Percent correct for interval analysis (see text). FA, Hit: Hit and false alarm rates for interval analysis.

| AU | N | Dur. | P | $\mathrm{P}_\Delta$ | FA | Hit |
|---|---|---|---|---|---|---|
| 1 | 166 | 30 | 84 | 81 | 17 | 48 |
| 2 | 138 | 23 | 88 | 79 | 20 | 55 |
| 4 | 33 | 23 | 93 | 78 | 22 | 55 |
| 5 | 34 | 26 | 98 | 80 | 20 | 33 |
| 6 | 56 | 112 | 91 | 86 | 13 | 79 |
| 7 | 48 | 78 | 83 | 76 | 22 | 33 |
| 9 | 2 | 12 | 100 | 79 | 21 | 100 |
| 10 | 53 | 69 | 95 | 76 | 23 | 29 |
| 12 | 112 | 102 | 86 | 84 | 11 | 58 |
| 15 | 73 | 18 | 98 | 80 | 19 | 40 |
| 17 | 88 | 39 | 93 | 78 | 20 | 48 |
| 20 | 8 | 8 | 99 | 80 | 20 | 18 |
| 23 | 29 | 46 | 94 | 79 | 21 | 36 |
| 24 | 66 | 27 | 92 | 77 | 22 | 17 |
| 25 | 131 | 65 | 65 | 74 | 21 | 34 |
| 26 | 105 | 55 | 92 | 73 | 23 | 27 |
| Mean | | | 90.5 | 78.8 | 19.7 | 44.4 |

1 standard deviation above the mean was employed. 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. Mean percent correct for the interval analysis was 79%, with hit and false alarm rates of 44% and 20% respectively.

## 5 Conclusions

We presented a systematic comparison of machine learning methods applied to the problem of fully automatic recognition of facial expressions, including AdaBoost, support vector machines, and linear discriminant analysis, as well as feature selection methods. 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. The combination of Adaboost and SVM's enhanced both speed and accuracy of the system. The full system operates in real time. Face detection runs at 24 frames/second in 320x240 images on a 3 GHz Pentium IV. The expression recognition step operates in less than 10 msec.

The generalization performance to new subjects for recognition of full facial expressions of emotion in a 7-way forced choice was 93.3%, which is the best performance reported so far on this publicly available dataset. Our results suggest that user independent, fully automatic real time coding of facial expressions in the continuous video stream is an achievable goal with present computer power, at least for applications in which frontal views can be assumed.

The machine-learning based system presented here can be applied to recognition of any facial expression dimension given a training dataset. Here we applied the system to fully automated facial action coding, and obtained a mean agreement rate of 94.8% for 17 AU's from the Facial Action Coding System. The outputs of the expression classifiers change smoothly as a function of time, providing information about expression dynamics that was previously intractable by hand coding. The system is fully automated, and performance rates are similar to or better than other systems tested on this dataset that employed varying levels of manual registration. The approach to automatic FACS coding presented here, in addition to being fully automated, also differs from approaches such as [13] and [16] in that instead of designing special purpose image features for each facial action, we explore general purpose learning mechanisms for data-driven facial expression classification. The approach detects not only changes in position of feature points, but also changes in image texture such as those created by wrinkles, bulges, and changes in feature shapes.

Here we presented preliminary results for the performance of the system on spontaneous expressions. The system was able to detect facial actions in this database despite the presence of speech, out-of-plane head movements that occur during discourse, and the fact that many of the action units occurred in combination. These results provide a benchmark for frame-by-frame analysis by a system trained for frontal views. The output sequence contains information about dynamics that can be exploited for deciding the presence of a facial action [1]. Future work will explore these dynamics, and compare improvement to the benchmark provided here. The accuracy of automated facial expression measurement may also 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 [12] and for aligning face images in 3D. These techniques will be integrated into future versions of our system.

**Acknowledgments.** Support for this work was provided by NSF-ITR IIS-0220141, California Digital Media Innovation Program DiMI 01-10130, and NRL Advanced Information Technology 55-05-03.

## References

[1] M.S. Bartlett, G. Littlewort, B. Braathen, T.J. Sejnowski, and J.R. Movellan. A prototype for automatic recognition of spontaneous facial actions. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 1271–1278, Cambridge, MA, 2003. MIT Press.

[2] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, and T. Huang. Learning baysian network classifiers for facial expression recognition using both labeled and unlabeled data. *Computer Vision and Pattern Recognition.*, 2003.

[3] 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.

[4] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[5] I. R. Fasel, B. Fortenberry, and J. R. Movellan. GBoost: A generative framework for boosting with applications to real-time eye coding. *Computer Vision and Image Understanding*, in press.

[6] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting, 1998.

[7] 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.

[8] A. Kapoor, Y. Qi, and R.W.Picard. Fully automatic upper facial action recognition. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures.*, 2003.

[9] M. Lades, J. Vorbrüggen, J. Buhmann, J. Lange, W. Konen, C. von der Malsburg, and R. Würtz. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[10] G. Littlewort, M.S. Bartlett, I. Fasel, J. Susskind, and J.R. Movellan. Dynamics of facial expression extracted automatically from video. In *IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video*, 2004.

[11] M. Lyons, J. Budynek, A. Plante, and S. Akamatsu. Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis. In *Proceedings of the 4th international conference on automatic face and gesture recognition*, pages 202–207, 2000.

[12] T. K. Marks, J. Hershey, J. Cooper Roddey, and J. R. Movellan. 3d tracking of morphable objects using conditionally gaussian nonlinear filters. *Computer Vision and Image Understanding*, under review. See also CVPR04 workshop: Generative-Model Based Vision.

[13] 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.

[14] 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.

[15] R. E. Schapire. A brief introduction to boosting. In *IJCAI*, pages 1401–1406, 1999.

[16] 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.

[17] Paul Viola and Michael Jones. Robust real-time object detection. Technical Report CRL 20001/01, Cambridge Research-Laboratory, 2001.