Towards Practical Smile Detection

Jacob Whitehill, Gwen Littlewort, Ian Fasel, Marian Bartlett, and Javier Movellan

***Abstract*—Machine learning approaches have produced some of the highest reported performances for facial expression recognition. However, to date, nearly all automatic facial expression recognition research has focused on optimizing performance on a few databases that were collected under controlled lighting conditions on a relatively small number of subjects. This paper explores whether current machine learning methods can be used to develop an expression recognition system that operates reliably in more realistic conditions. We explore the necessary characteristics of the training dataset, image registration, feature representation, and machine learning algorithms. A new database, GENKI, is presented which contains pictures, photographed by the subjects themselves, from thousands of different people in many different real-world imaging conditions. Results suggest that human-level expression recognition accuracy in real-life illumination conditions is achievable with machine learning technology. However, the datasets currently used in the automatic expression recognition literature to evaluate progress may be overly constrained and could potentially lead research into locally optimal algorithmic solutions.**

***Index Terms*—Face and gesture recognition, machine learning, computer vision.**

I. Introduction

Recent years have seen considerable progress in the field of automatic facial expression recognition (see [1], [2], [3], [4] for surveys). Common approaches include static image texture analysis [5], [6], feature point-based expression classifiers [7], [8], [9], [10], 3D face modeling [11], [12], and dynamic analysis of video sequences [13], [14], [15], [16], [17], [18]. Some expression recognition systems tackle the more challenging and realistic problems of recognizing spontaneous expressions [13], [15] – i.e., non-posed facial expressions that occur naturally – as well as expression recognition under varying head pose [19], [11]. However, to date, nearly all automatic expression recognition research has focused on optimizing performance on facial expression databases that were collected under tightly controlled laboratory lighting conditions on a small number of human subjects (e.g., Cohn-Kanade DFAT [20], CMU-PIE [21], MMI [22], UT Dallas [23], and Ekman-Hager [24]). While these databases have played a critically important role in the advancement of automatic expression recognition research, they also share the common limitation of not representing the diverse set of illumination conditions, camera models, and personal differences that are found in the real world. It is conceivable that by evaluating performance on these datasets the field of automatic expression recognition could be driving itself into algorithmic "local maxima."

To illustrate this point, we tested standard linear regression to detect smiles from raw pixel values of face images from one of these databases, DFAT, scaled to a 8×8 pixel size. The system achieved a smile detection accuracy of 97% (cross-validation). However, when evaluated on a large collection of frontal face images collected from the Web, the accuracy plunged to 72%, rendering it useless for real-world applications. This illustrates the danger of evaluating on small, idealized datasets.

This danger became apparent in our own research: For example, in 2006 we reported on an expression recognition system developed at our laboratory [25] based on support vector machines operating on a bank of Gabor filters. To our knowledge, this system achieves the highest accuracy (93% percent-correct on a 7-way alternative forced choice emotion classification problem) reported in the literature on two publicly-available datasets of facial expressions: the Cohn-Kanade [20] and the POFA [26] datasets. On these datasets, the system can also classify images as either smiling or non-smiling with accuracy nearly at 98% (area under the ROC curve). Based on these numbers we expected good performance in real-life applications. However, when we tested this system on a large collection of frontal face images collected from the Web, the accuracy fell to 85%. This gap in performance also matched our general impression of the system: while it performed very well in controlled conditions, including laboratory demonstrations, its performance was disappointing in unconstrained illumination conditions.

Based on this experience we decided to study whether current machine learning methods can be used to develop an expression recognition system that would operate reliably in real-life rendering conditions. We decided to focus on recognizing smiles within approximately 20° of frontal pose faces due to the potential applications in digital cameras (e.g., smile shutter), video games, and social robots. The work we present in this paper became the basis for the first smile detector embedded in a commercial digital camera.

Here we document the process of developing the smile detector and the different parameters that were required to achieve practical levels of performance, including (1) Size and type of datasets, (2) Image registration accuracy (e.g., facial feature detection), (3) Image representations, and (4) Machine learning algorithms. As a test platform we collected our own dataset (*GENKI*), containing over 63,000 images from the Web, which closely resembles our target application: a "smile shutter" for digital cameras to automatically take pictures when people smile. We further study whether an automatic smile detector tested on binary labels can be used to estimate the *intensity* of a smile as perceived by human observers.

II. Dataset Collection

Crucial to our study was the collection of a database of face images that closely resembled the operating conditions of our target application: a smile detector embedded in digital cameras. The database had to span a wide range of imaging

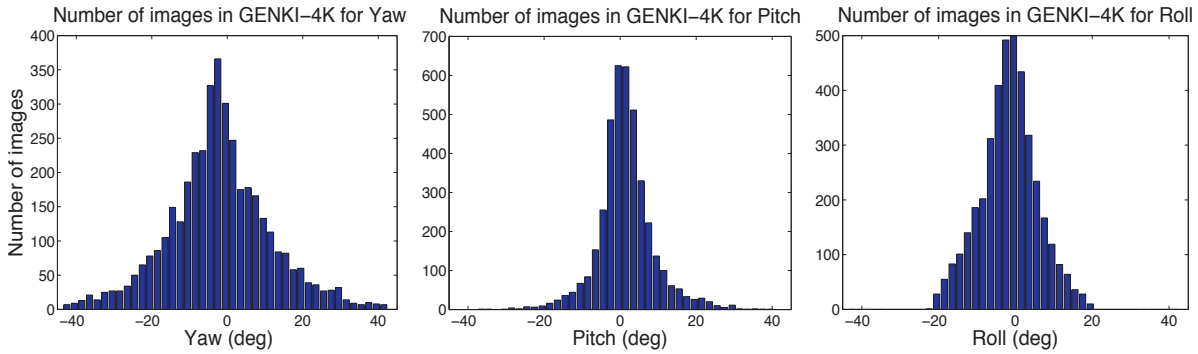


Fig. 1. Histogram of GENKI images as a function of the human-labeled 3-D pose. These histograms were computed for GENKI-4K, which is a representative sample of all GENKI images whose faces could be detected automatically.

conditions, both outdoors and indoors, as well as variability in age, gender, ethnicity, facial hair, and glasses. To this effect we collected a dataset, which we named *GENKI*[1], that consists of 63,000 images, of approximately as many different human subjects, downloaded from publicly available Internet repositories of personal Web pages. The photographs were taken not by laboratory scientists, but by ordinary people all over the world taking photographs of each other for their own purposes – just as in the target smile shutter application. The pose range (yaw, pitch, and roll parameters of the head) of most images was within approximately $\pm 20^\circ$ of frontal (see Figure 1). All faces in the dataset were manually labeled for the presence of prototypical smiles. This was done using three categories, which were named "happy", "not happy", and "unclear". Approximately 45% of GENKI images were labeled as "happy", 29% as "unclear", and 26% as "not happy". For comparison we also employed a widely used dataset of facial expressions, the Cohn-Kanade DFAT dataset.

III. Experiments

Figure 2 is a flowchart of the smile detection architecture under consideration. First the face and eyes are automatically located. The image is rotated, cropped, and scaled to ensure a constant location of the center of the eyes on the image plane. Next, the image is encoded as a vector of real-valued numbers which can be seen as the output of a bank of filters. The outputs of these filters are integrated by the classifier into a single real-valued number which is then thresholded to classify the image as smiling or not-smiling. Performance was measured in terms of area under the ROC curve (A'), a bias-independent measure of sensitivity (unlike the "%-correct" statistic). The A' statistic has an intuitive interpretation as the probability of the system being correct on a 2 Alternative Forced Choice Task (2AFC), i.e., a task in which the system is simultaneously presented with two images, one from each category of interest, and has to predict which image belongs to which category. In all cases, the A' statistic was computed over a set of validation images not used during training. An upper-bound on the uncertainty of the A' statistic was obtained using the formula $s = \sqrt{\frac{A'(1-A')}{\min\{n_p, n_n\}}}$ where n_p, n_n are the number of positive and negative examples [27]. Experiments were conducted to evaluate the effect of the following factors:

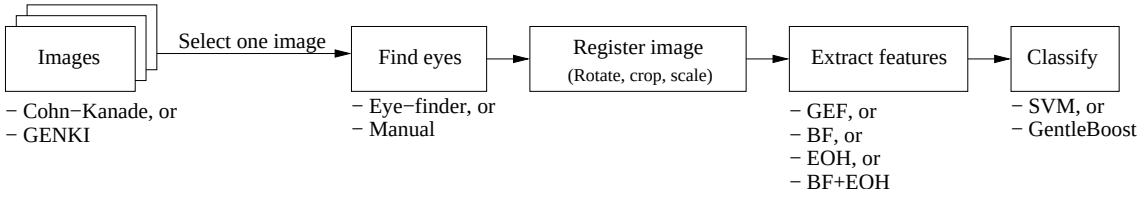
a) **Training Set:** We investigated two datasets of facial expressions: (1) DFAT, representing datasets collected in controlled imaging conditions; and (2) GENKI, representing data collected from the Web. The DFAT dataset contains 475 labeled video sequences of 97 human subjects posing prototypical expressions in laboratory conditions. The first and last frames from each video sequence were selected, which correspond to neutral expression and maximal expression intensity. In all, 949 video frames were selected. (One face could not be found by the automatic detector.) Using the Facial Action codes for each image, the faces were labeled as "smiling," "non-smiling,", or "unclear." Only the first two categories were used for training and testing. From GENKI, only images with expression labels of "happy" and "not happy" were included – 20,000 images labeled as "unclear" were excluded. In addition, since GENKI contains a significant number of faces whose 3D pose is far from frontal, only faces successfully detected by the (approximately) frontal face detector (described below) were included (see Figure 1). Over 25,000 face images of the original GENKI database remained. In summary, DFAT contains 101 smiles and 848 non-smiles, and GENKI contains 17,822 smiles and 7,782 non-smiles.

b) **Training Set Size:** The effect of training set size was evaluated only on the GENKI dataset. First a validation set of 5000 images from GENKI was randomly selected and subsets of different sizes were randomly selected for training from the remaining 20,000 images. The training set sizes were $\{100, 200, 500, 949, 1000, 2000, 5000, 10000, 20000\}$. For DFAT, we either trained on all 949 frames (when validating on GENKI), or on 80% of the DFAT frames (when validating on DFAT). When comparing DFAT to GENKI we kept the training set size constant by randomly selecting 949 images from GENKI.

c) **Image Registration:** All images were first converted to gray-scale and then normalized by rotating, cropping, and scaling the face about the eyes to reach a canonical face width of 24 pixels. We compared the smile detection accuracy obtained when the eyes were automatically detected, using the eye detection system described in [28], to the smile detection accuracy obtained when the eye positions were hand-labeled.

[1]A 4K subset of these images is available at http://mplab.ucsd.edu.

Fig. 2. Flowchart of the smile detection systems under evaluation.



Inaccurate image registration has been identified as one of the most important causes of poor performance in applications such as person identification [29]. In previous work we had reported that precise image registration, beyond the initial face detection, was not useful for expression recognition problems [25]. However, this statement was based on evaluations on the standard datasets with controlled imaging conditions and not on larger, more diverse datasets like GENKI.

d) **Image Representation:** We compared five widely used image representations:

1) *Gabor Energy Filters (GEF)*: These filters [30] model the complex cells of the primate's visual cortex. Each energy filter consists of a real and an imaginary part which are squared and added to obtain an estimate of energy at a particular location and frequency band, thus introducing a non-linear component. We applied a bank of 40 Gabor Energy Filters consisting of 8 orientations (spaced at 22.5° intervals) and 5 spatial frequencies with wavelengths of 1.17, 1.65, 2.33, 3.30, and 4.67 Standard Iris Diameters (SID)[2]. This filter design has shown to be highly discriminative for facial action recognition [24].
2) *Box Filters (BF)*: These are filters with rectangular input responses, which makes them particularly efficient for applications on general purpose digital computers. In the computer vision literature, these filters are commonly referred to as Viola-Jones "integral image filters" or "Haar features." In our work we included 6 types of Box Filters in total, comprising two-, three-, and four-rectangle features similar to those used by Viola and Jones [31], and an additional two-rectangle "center-surround" feature.
3) *Edge Orientation Histograms (EOH)*: These features have recently become popular for a wide variety of tasks, including object recognition (e.g., in SIFT [32]) and face detection [33]. They are reported to be more tolerant to image variation and to provide substantially better generalization performance than Box Filters, especially when the available training datasets are small [33]. We implemented two versions of EOH: "dominant orientation features" and "symmetry" features, both proposed by Levi and Weiss [33].
4) *BF+EOH*: Combining these feature types was shown by Levi and Weiss to be highly effective for face detection; we thus performed a similar experiment for smile detection.

[2]An SID is defined as 1/7 of the distance between the center of the left and right eyes

TABLE I
CROSS-DATABASE SMILE DETECTION PERFORMANCE (% AREA UNDER ROC ± STDERR) USING AUTOMATIC EYE-FINDER

	Validation	
Training	GENKI	DFAT
GENKI (949 image subset)	**95.1** ± 0.55	**98.4** ± 1.30
DFAT (949 images)	**84.9** ± 0.91	**100** ± 0.00

5) *Local Binary Patterns (LBP)* We also experimented with LBP [34] features for smile detection using LBP either as a preprocessing filter or as features directly.

e) **Learning Algorithm:** We compared two popular learning algorithms: GentleBoost, and Support Vector Machines (SVMs): GentleBoost [35] is a boosting algorithm [36] that minimizes the χ-square error between labels and model predictions [35]. In our GentleBoost implementation, each elementary component consisted of a filter chosen from a large ensemble of available filters, and a non-linear tuning-curve, computed using non-parametric regression [28]. The output of GentleBoost is an estimate of the log probability ratio of the category labels given the observed images. In our experiment, all the GentleBoost classifiers were trained for 500 rounds.

When training with linear SVMs, the entire set of Gabor Energy Filters or Box Filters was used as the feature vector of each image. Bagging was employed to reduce the number of training examples down to a tractable number (between 400 and 4000 examples per bag) [37].

IV. RESULTS

A. Dataset

We compared the generalization performance within and between datasets. The feature type was held constant at BF+EOH. Table I displays the results of the study. Whereas the classifier trained on DFAT achieved only 84.9% accuracy on GENKI, the classifier trained on an equal-sized subset of GENKI achieved 98.4% performance on DFAT. This accuracy was not significantly different from the 100% performance obtained when training and testing on DFAT ($t(115) = 1.28, p = 0.20$), which suggests that for smile detection, a database of images from the Web may be more effective than a dataset like DFAT collected in laboratory conditions.

Figure 3 (left) displays detection accuracy as a function of the size of the training set using the GentleBoost classifier and an automatic eye-finder for registration. With GentleBoost, the performance of the BF, EOH, and BF+EOH feature types mostly flattens out at about 2000 training examples. The

Gabor features, however, show substantial gains throughout all training set sizes. Interestingly, the performance of Gabor features is substantially higher using SVMs than GentleBoost; we address this issue in a later section.

B. Registration

One question of interest is to what extent smile detection performance could be improved by precise image registration based on localization of features like the eyes. We compared accuracy when registration was based on manually versus automatically located eyes. For automatic eye detection, we used an updated version of the eye detector presented in [28]; its average error from human-labeled ground truth was 0.58 SID. In contrast, the average error of human coders with each other was 0.27 SID.

Figure 3 (middle) shows the difference in smile detection accuracy when the image was registered using the human-labeled eye center versus when using the automatically detected eye centers. The performance difference was considerable (over 5%) when the training set was small and diminished down to about 1.7 % as the training size increased. The best performance using hand-labeled eye coordinates was 97.98% compared to 96.37% when using fully automatic registration. Thus, overall it seems that continued improvement in automatic face registration would still benefit the automatic recognition of expression in unconstrained conditions.

C. Representation and Learning Algorithm

Figure 3 compares smile detection performance across the different types of image representations trained using either GentleBoost (left) or a linear SVM (right). We also computed two additional data points: (1) BF features and a SVM with a training set size of 20000, and (2) Linear SVM on EOH features using 5000 training examples.

The combined feature set BF+EOH achieved the best recognition performance over all training set sizes. However, the difference in accuracy compared to the component BF and EOH feature sets was much smaller than the performance gain of 5-10% reported by Levi and Weiss [33]. We also did not find that EOH features were particularly effective with small datasets, as reported by Levi and Weiss [33]. It should be noted, however, that we implemented only two out of the three EOH features used in [33]. In addition, we report performance on smile detection, while Levi and Weiss' results were for face detection.

The most surprising result was a cross-over interaction between the image representation and the classifier. This effect is visible in Figure 3 and is highlighted in Table II for a training set of 20000 GENKI images: When using GentleBoost, Box Filters performed substantially better than Gabor Energy Filters. The difference was particularly pronounced for small training sets. Using linear SVMs, on the other hand, Gabor Energy Filters (and also EOH features) performed significantly better than Box Filters. Overall GentleBoost and linear SVMs performed comparably when using the optimal feature set for each classifier, 97.2% for SVM and 97.9% for GentleBoost (See Table II). Where the two classifiers differ is in their associated optimal features sets.

TABLE II
GENTLEBOOST VS. LINEAR SVMS (% AREA UNDER ROC $\pm$ STDERR) FOR SMILE DETECTION ON GENKI

Human-labeled Eyes		
Features	**SVM**	**GentleBoost**
Gabor Energy Filters	**97.2** $\pm$ 0.23	**95.5** $\pm$ 0.29
Box Filters (BF)	**96.3** $\pm$ 0.27	**97.9** $\pm$ 0.20
Eye-finder-labeled Eyes		
Features	**SVM**	**GentleBoost**
Gabor Energy Filters	**96.3** $\pm$ 0.27	**91.4** $\pm$ 0.40
Box Filters (BF)	**91.6** $\pm$ 0.39	**96.1** $\pm$ 0.27

We suggest two explanations for the observed cross-over interaction between feature set and learning algorithm: (1) Using Box Filters with linear SVMs forces the classification to be linear on the pixel values. In contrast, Gabor Energy Filters (and also EOH features) are non-linear functions of pixel intensities, and GentleBoost also introduces a non-linear tuning curve on top of the linear Box Filters, thus allowing for non-linear solutions. (2) The dimensionality of Gabor filters, 23040, was small when compared to the Box Filter representation, 322945. It is well known that, due to their sequential nature, Boosting algorithms tend to work better with a very large set of highly redundant filters.

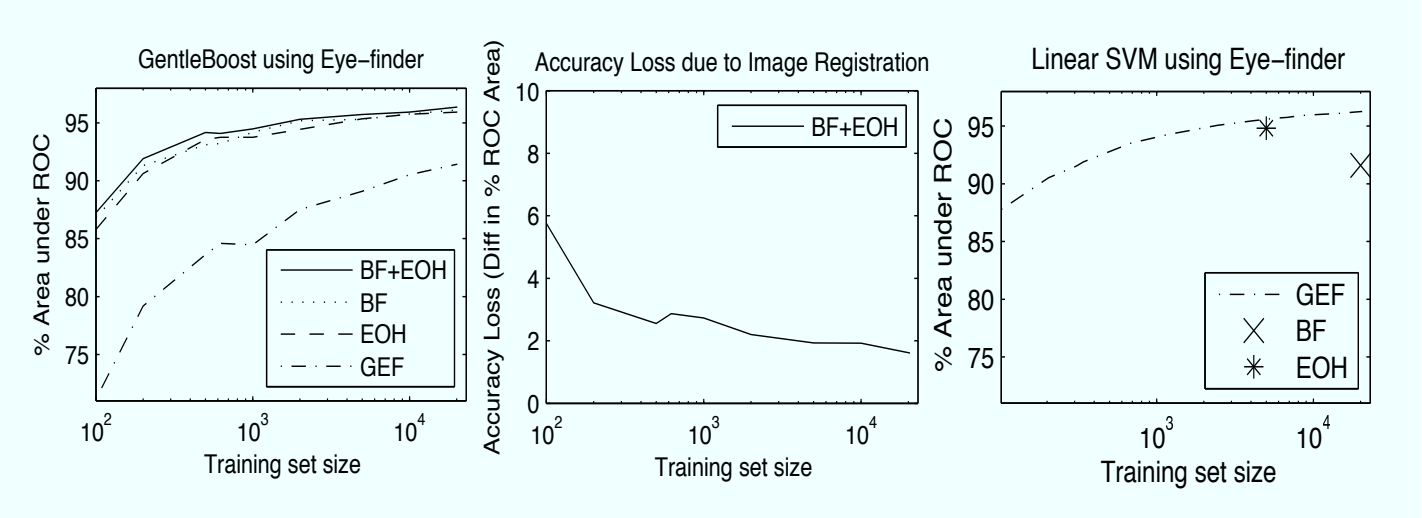
To test hypothesis (1), we trained an additional SVM classifier using a radial basis function (RBF) kernel ($\sigma = 1$) on Box Filters using the eyefinder-labeled eyes. Accuracy increased by 3% to 94.6%, which is a substantial gain. It thus seems likely that a non-linear decision boundary on the face pixel values is necessary to achieve optimal smile detection performance.

Finally, we tested Local Binary Pattern features for smile detection using two alternative methods: (1) Each face image was pre-filtered using an LBP operator, similar in nature to [38], and then classified using BF features and GentleBoost; or (2) LBP features were classified directly by a linear SVM. Results for (1): Using 20000 training examples and eyefinder-based registration, smile detection accuracy was 96.3%. Using manual face registration, accuracy was 97.2%. Both of these numbers are slightly lower than using GentleBoost with BF features alone (without the LBP preprocessing operator). Results for (2): Using 20000 training examples and eyefinder-based registration, accuracy was 93.7%. This is substantially higher than the 91.6% accuracy for BF+SVM. As discussed above for GEF features, both the relatively low dimensionality of the LBP features ($24*24 = 576$) and the non-linearity of the LBP operator may have been responsible for relatively high performance when using linear SVMs.

V. ESTIMATING SMILE INTENSITY

We investigated whether the real-valued output of the detector, which is an estimate of the log-likelihood ratio of the smile versus non-smile categories, agreed with human perceptions of intensity of a smile. Earlier research on detecting facial actions using SVMs [39] has shown empirically that the distance to the margin of the SVM output is correlated with

Fig. 3. **Left**: A' statistics (area under the ROC curve) versus training set size using GentleBoost for classification. Face registration was performed using an automatic eye-finder. Different feature sets were used for smile detection: Gabor Energy Filter (GEF) features, Box Filter (BF) features, Edge Orientation Histograms (EOH), and BF+EOH. **Middle**: The loss in smile detection accuracy, compared to using human-labeled eye positions, incurred due to face registration using the automatic eye-finder. **Right**: Smile detection accuracy (A') using a linear SVM for classification.



the expression intensity as perceived by humans; here, we study whether a similar result holds for the log-likelihood ratio output by GentleBoost. We used the smile detector trained with GentleBoost using BF+EOH features.

Flashcards: Our first study examined the correlation between human and computer labels of smile intensity on a set of 48 "flashcards" containing GENKI faces of varying smile intensity (as estimated by our automatic smile detector). Five human coders sorted piles of 8 flash-cards each in order of increasing smile intensity. These human labels were then correlated with the output of the trained smile detector. The average caomputer-human correlation of smile intensity was 0.894, which is quite close to the average inter-human correlation of 0.923 and the average human self-correlation of 0.969.

Video: We also measured correlations over five short video sequences (11 to 57 seconds) collected at our laboratory of a subject watching comedy video clips. Four human coders dynamically coded the intensity of the smile frame-by-frame using continuous audience response methods [40]. The smile detector was then used to label the smile intensity of each video frame independently. The final estimates of smile intensity were obtained by low-pass filtering and time shifting the output of GentleBoost. The parameter values (5.7 sec width of low-pass filter; 1.8 sec lag) of the filters were chosen to optimize the inter-human correlation.

On video sequences, the average human-machine correlation was again quite high, 0.836, but smaller than the human-human correlation, 0.939. While this difference was statistically significant ($t(152) = 4.53, p < 0.05$), in practice it was very difficult to differentiate human and machine codes. Figure 4 displays the human and machine codes of a particular video sequence. As shown in the figure, the smile detector's output is well within the range of human variability for most frames. Sample images from every 100th frame are shown below the graph.

VI. Summary and Conclusions

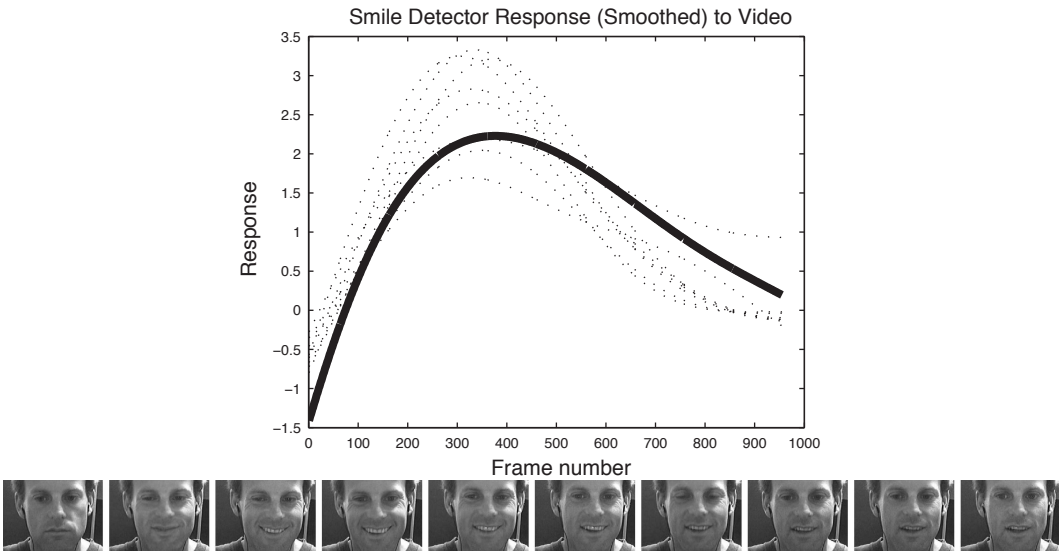
Datasets: The current datasets used in the expression recognition literature are too small and lack variability in imaging conditions. Current machine learning methods may require on the order of 1000 to 10,000 images per target facial expression. These images should have a wide range of imaging conditions and personal variables including ethnicity, age, gender, facial hair, and presence of glasses.

Incidentally, an important shortcoming of contemporary image databases is the lack of ethnic diversity. It is an open secret that the performance of current face detection and expression recognition systems tends to be much lower when applied to individuals with dark skin. In a pilot study on 141 GENKI faces (79 white, 52 black), our face detector achieved 81% hit rate on white faces, but only 71% on black faces (with 1 false alarm). The OpenCV face detector, which has become the basis for many research applications, was even more biased, with 87% hit rate on white faces, and 71% on black faces (with 13 false alarms). Moreover, the smile detection accuracy on white faces was 97.5% whereas for black faces was only 90%.

Image Registration: We found that, when operating on datasets with diverse imaging conditions, such as GENKI, precise registration of the eyes is useful. We have developed one of the most accurate eye-finders for standard cameras to-date, yet it is still about half as accurate as human labelers. This loss in alignment accuracy resulted in a smile detection performance penalty from 1.7 to 5 percentage points. Image registration is particularly important when the training datasets are small.

Image Representation and Classifier: The image representations that have been widely used in the literature, Gabor Energy Filters and Box Filters, work well when applied

Fig. 4. Humans' (dotted) and smile detector's (solid bold) ratings of smile intensity for a video sequence.



to realistic imaging conditions. However there were some surprises: (1) A very popular Gabor filter bank representation did not work well when trained with GentleBoost, even though it performed well with SVMs. Moreover, Box Filters worked well with GentleBoost but performed poorly when trained with SVMs. We explored two explanations for this cross-over interaction, but more research is needed to understand this interaction fully. We also found that Edge Orientation Histograms, which have become very popular in the object detection literature, did not offer any particular advantage for the smile detection problem.

Expression Intensity: We found that the real-valued output of GentleBoost classifiers trained on binary tasks is highly correlated with human estimates of smile intensity, both in still images and video. This offers opportunities for applications that take advantage of expression dynamics.

Future Challenges: In this paper we focused on detecting smiles in poses within approximately $\pm 20^{\circ}$ from frontal. Developing expression recognition systems that are robust to pose variations will be an important challenge for the near future. Another important future challenge will be to develop comprehensive expression recognition systems capable of decoding the entire gamut of human facial expressions, not just smiles. One promising approach that we and others have been pursuing [5], [7], [13] is automating the Facial Action Coding System. This framework allows coding all possible facial expressions as combinations of 53 elementary expressions (Action Units) [26]. Our experience developing a smile detector suggests that robust automation of the Facial Action Coding system may require on the order of 1,000 to 10,000 examples images per target Action Unit. Datasets of this size are likely to be needed to capture the variability in illumination and personal characteristics likely to be encountered in practical applications.

ACKNOWLEDGEMENT

This work was supported by the NSF IIS INT2-Large 0808767 grant and by the UC Discovery Grant 10202.

REFERENCES

[1] Y.-L. Tian, T. Kanade, and J. Cohn, "Facial expression analysis," in *Handbook of face recognition*, S. L. . A. Jain, Ed. Springer, October 2003.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: Survey," *Pattern Recognition*, vol. 36, pp. 259–275, 2003.

[3] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1424–1445, 2000.

[4] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, 2009.

[5] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Fully automatic facial action recognition in spontaneous behavior," in *Proceedings of Automatic Facial and Gesture Recognition*, 2006.

[6] Y. Wang, H. Ai, B. Wu, and C. Huang, "Real time facial expression recognition with adaboost," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, 2004.

[7] M. Pantic and J. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, no. 3, 2004.

[8] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, 2001.

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

[10] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, 2007.

[11] Z. Wen and T. Huang, "Capturing subtle facial motions in 3D face tracking," in *ICCV*, 2003.

[12] N. Sebe, Y. Sun, E. Bakker, M. Lew, I. Cohen, and T. Huang, "Towards authentic emotion recognition," in *Proc. IEEE Int'l Conf. Multmedia and Expo (ICME'05)*, The Hague, Netherlands, 2004.

[13] J. Cohn and K. Schmidt, "The timing of facial motion in posed and spontaneous smiles," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, pp. 1–12, 2004.

[14] I. Cohen, N. Sebe, L. Chen, A. Garg, and T. Huang, "Facial expression recognition from video sequences: Temporal and static modelling," *CVIU Special Issue on Face Recognition*, vol. 91, 2003.

[15] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 699–714, 2005.

[16] P. Yang, Q. Liu, and D. Metaxas, "Boosting coded dynamic features for facial action units and facial expression recognition," in *Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[17] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, 2007.

[18] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007.

[19] Z. Zhu and Q. Ji, "Robust real-time face pose and facial expression recovery," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006.

[20] T. Kanade, J. Cohn, and Y.-L. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, March 2000, pp. 46 – 53.

[21] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.

[22] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proc. IEEE Int'l Conf. Multmedia and Expo (ICME'05)*, Amsterdam, The Netherlands, 2005.

[23] A. OToole, J. Harms, S. Snow, D. Hurst, M. Pappas, and H. Abdi, "A video database of moving faces and people," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 812–816, 2005.

[24] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

[25] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," *Image and Vision Computing*, vol. 24, no. 6, pp. 615–625, 2006.

[26] P. Ekman and W. Friesen, "Pictures of facial affect," Photographs, 1976, available from Human Interaction Laboratory, UCSF, HIL-0984, San Francisco, CA 94143.

[27] C. Cortes and M. Mohri, "Confidence intervals for the area under the roc curve," in *Advances in Neural Information Processing Systems*, 2004.

[28] I. Fasel, B. Fortenberry, and J. Movellan, "A generative framework for real time object detection and classification," *Computer Vision and Image Understanding*, vol. 98, 2005.

[29] A. Pnevmatikakis, A. Stergiou, E. Rentzeperis, and L. Polymenakos, "Impact of face registration errors on recognition," in *3rd IFIP Conference on Artificial Intelligence Applications & Innovations (AIAI)*, 2006.

[30] J. Movellan, "Tutorial on gabor filters," MPLab Tutorials, UCSD MPLab, Tech. Rep., 2005.

[31] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, 2004.

[32] D. Lowe, "Object recognition from local scale-invariant features," in *Intl. Conference on Computer Vision*, 1999.

[33] K. Levi and Y. Weiss, "Learning object detection from a small number of examples: The importance of good features," in *Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[34] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[35] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, no. 2, 2000.

[36] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *European Conference on Computational Learning Theory*, 1995, pp. 23–37.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Verlag, 2001.

[38] G. Heusch, Y. Rodriguez, and S. Marcel, "Local binary patterns as an image preprocessing for face authentication," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[39] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *Journal of Multimedia*, 2006.

[40] I. Fenwick and M. Rice, "Reliability of continuous measurement copy-testing methods," *J. of Advertising Research*, 1991.