Smile Detection by Boosting Pixel Differences

Caifeng Shan, *Member, IEEE*

***Abstract*—Smile detection in face images captured in unconstrained real-world scenarios is an interesting problem with many potential applications. This paper presents an efficient approach to smile detection, in which the intensity differences between pixels in the grayscale face images are used as features. We adopt AdaBoost to choose and combine weak classifiers based on intensity differences to form a strong classifier. Experiments show that our approach has similar accuracy to the state-of-the-art method but is significantly faster. Our approach provides 85% accuracy by examining 20 pairs of pixels and 88% accuracy with 100 pairs of pixels. We match the accuracy of the Gabor-feature-based support vector machine using as few as 350 pairs of pixels.**

***Index Terms*—AdaBoost, facial expression recognition, smile detection.**

I. Introduction

A smile is the most common facial expression that occurs in people's daily life. It often indicates pleasure, happiness, appreciation, or satisfaction. Detecting smiles can be used to estimate a person's mental state. Smile detection has many applications in practice, such as interactive systems (e.g., gaming), product rating, distance learning systems, video conferencing, and patient monitoring. For example, the statistics on the audience smile can be a hint for "how much the audience enjoys" the multimedia content.

The machine analysis of facial expressions in general has been an active research topic in the last two decades [1]–[3]. Most of the existing works have been focused on analyzing a set of prototypic emotional facial expressions, using the data collected by asking subjects to pose deliberately these expressions [4]. However, the exaggerated facial expressions occur rarely in real-life situations. Spontaneous facial expressions induced in natural environments differ from posed expressions, i.e., both in terms of which facial muscles move and how they move dynamically. For example, Cohn and Schmidt [5] observed that posed smiles are of larger amplitude and have a less consistent relationship between amplitude and duration than spontaneous smiles. Recently, research attention has started to shift toward the more realistic problem of analyzing spontaneous facial expressions [5]–[7]. As illustrated in some initial studies [8], [9], this is a challenging problem; it seems very difficult to capture the complex decision boundary among spontaneous expressions [8].

In this paper, we focus on smile detection in face images captured in real-world scenarios. Fig. 1 shows some example images. We present an efficient approach to smile detection, in which the intensity differences between pixels in the grayscale face images are used as simple features. AdaBoost then is adopted to choose and combine weak classifiers based on pixel differences to form a strong classifier for smile detection. Experimental results show that our approach achieves similar accuracy to the state-of-the-art method but is significantly faster.

Fig. 1. Examples of (top two rows) real-life smile faces and (bottom two rows) nonsmile faces, from the GENKI4K database.

Our approach provides 85% accuracy by examining 20 pairs of pixels and 88% accuracy with 100 pairs of pixels. We match the accuracy of the Gabor-feature-based support vector machine (SVM) using as few as 350 pairs of pixels. Varying illumination and pose are the two main difficulties for unconstrained face analysis; we also examine different illumination normalization methods and investigate the impact of pose variation.

The remainder of this paper is organized as follows. First, we review briefly the related work in Section II. Our approach is presented in Section III. Section IV describes in detail our extensive experiments. Finally, Section V concludes the paper.

II. Previous Work

Although there is a large amount of literature on facial expression recognition, few papers have focused specifically on smile detection. Shinohara and Otsu [10] integrated higher order local autocorrelation features with a Fisher weight map for face representation. With the Fisher discriminant analysis, their approach provides the accuracy of 97.9% on 96 face images. Although high accuracy was achieved, very limited data were used in their study. In [11], a six-dimension feature vector is extracted to describe the lip and cheeks, which is used with a perceptron classifier for smile detection. Tested on three video sequences, the method achieves the accuracy of 60%–85%. Kowalik *et al.* [12] developed the BROAFERENCE system, which provides direct feedback of the audience satisfaction level on multimedia content by means of smile detection. A 16-dimension feature vector derived by tracking eight mouth points is used with a neural network classifier to detect smiles. However, performance evaluation are missing. A smile detector based on the Viola–Jones cascade classifier is described in [13]. Trained on 5812 images (i.e., 2436 positive and 3376 negative images), the detector achieves the accuracy of 96.1% on 4928 testing images. However, the used face images are mainly frontal, with limited imaging conditions. In [14], local binary patterns (LBP) and principal component analysis are combined for smile detection.

Recently, Whitehill *et al.* [15] presented a comprehensive study on practical smile detection. They collected the GENKI database consisting of 63 000 real-life face images from the Web. They investigated

Manuscript received February 04, 2011; revised June 07, 2011; accepted June 30, 2011. Date of publication July 14, 2011; date of current version December 16, 2011. This work was supported by the Visual Context Modeling Project. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mary Comer.

The author is with the Philips Research, 5656AE Eindhoven, The Netherlands. (e-mail: caifeng.shan@philips.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2011.2161587

different parameters, including size and type of data sets, image registration, facial representation, and machine learning algorithms. Their study suggests that high detection accuracy is achievable in real-life situations. They argued that the order of 1000–10 000 images that have a wide range of imaging conditions and personal variables is required for training.

Smile detection has received much interest for commercial applications. For example, in some digital cameras, the "smile shutter" shoots automatically when a smiling face is detected. Omron developed "smile measurement" software [16], which measures the amount of happiness that a person is exhibiting by fitting a 3-D model on the face. The software is claimed to achieve the accuracy of more than 90%.

Fig. 2. Normalized face images: (top two rows) smile and (bottom two rows) nonsmile.

III. Boosting Pixel Differences

A vital step in facial expression analysis is extracting effective features from original face images. Based on feature-point (e.g., mouth corners) detection, geometric features can be exploited [11], [12]. However, accurate and reliable detection and tracking of feature points are difficult to accommodate in real-world unconstrained scenarios. Another kind of features considers the appearance (skin texture) of the face [10], [13]. Appearance features are less sensitive to the errors in feature point detection and can encode changes in skin texture that are important for facial expression modeling. Therefore, appearance features look more promising for unconstrained facial expression analysis, and different features can be exploited. For example, the responses of Gabor filters at multiple spatial scales, orientations, and locations have been used traditionally [17] and have been proven successful for smile detection [15]. However, it is computationally expensive to extract Gabor features. Other features include the Haar wavelet (or boxlet) features [18], LBP [19], and orientation histograms of gradients (edges) [20] (see [15] for comparisons on these features).

In many practical applications, speed or computational efficiency is a key concern. Because of the limited computational resource, it is highly desired that the features used can be computed easily and efficiently. Baluja *et al.* [21], [22] introduced to use the relationship between two pixels' intensities as features. They obtained high accuracy on face orientation discrimination and gender classification by comparing the intensities of a few pixels. More specifically, they used five types of pixel comparison operators (and their inverses) [22]:

1) $\text{pixel}_i > \text{pixel}_j$;
2) pixel_i within 5 units (out of 255) of pixel_j;
3) pixel_i within 10 units (out of 255) of pixel_j;
4) pixel_i within 25 units (out of 255) of pixel_j;
5) pixel_i within 50 units (out of 255) of pixel_j.

The binary result of each comparison, which is represented numerically as 1 or 0, is used as the feature. Thus, for an image of 24 × 24 pixels, there are $2 * 5 * (24 * 24)(24 * 24 - 1)$ or 3 312 000 pixel-comparison features.

In this paper, we also examine the relationship between two pixels' intensities as image features. Instead of utilizing the above comparison operators, we propose to use the intensity difference between two pixels as a simple feature. We think that the intensity difference covers the information provided by all these comparison operators. Furthermore, it contains more information beyond these comparisons. Since there is a feature extracted for each pair of different pixels, for an image of 24 × 24 pixels, there are $(24*24)(24*24-1)$ or 331 200 features extracted. Compared with the above pixel-comparison features, the number of features is reduced by ten times. The feature set can be reduced by a factor of 2 (i.e., 165 600) if restricting to the upper diagonal part of the pairwise feature matrix.

Given the large number of features, the next step is to minimize the number of features that need to be computed for a new face image while still achieving high discrimination accuracy. Similar to [22], we adopt AdaBoost to achieve this. AdaBoost [18] provides a simple yet effective approach for stagewise learning of a nonlinear classification function. It combines the feature selection and classifier training steps in one process. AdaBoost learns a small number of weak classifiers whose performance is just better than random guessing and boosts them iteratively into a strong classifier of higher accuracy. The process of AdaBoost maintains a distribution on the training samples. At each iteration, a weak classifier, which minimizes the weighted error rate, is selected, and the distribution is updated to increase the weights of the misclassified samples and reduce the importance of the others. The weak classifier is designed to select the feature that best separates the positive and negative examples. In this paper, the weak classifier is defined based on the intensity difference of a pair of pixels. More precisely, the weak classifier $h_j(x)$ consists of feature f_j (i.e., the intensity difference), threshold θ_j, and parity p_j indicating the direction of the inequality sign as follows:

$$h_j(x) = \begin{cases} 1, & \text{if } p_j f_j(x) \leq p_j \theta_j \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

IV. Experiments

A. Data

We carried out experiments on the publicly available GENKI4K database,[1] which is a subset of the images used in [15]. The database consists of 4000 images (2162 "smile" and 1828 "nonsmile"); some samples are shown in Fig. 1. As can be seen, the images span a wide range of imaging conditions, i.e., both outdoors and indoors, as well as variability in age, gender, ethnicity, facial hair, and glasses [15]. The ground truth on "smile" and "nonsmile" is provided with the database.

In our experiments, the images were converted to grayscale. Following [15], the faces were normalized to reach a canonical face of 48 × 48 pixels, which was based on the manually labeled eyes positions. Fig. 2 illustrates the normalized faces. We partitioned the image set into four groups of 1000 images, with similar number of "smile" and "nonsmile" samples, and adopted a fourfold cross-validation. That is, one group was used as testing data while the other groups as training data; the process was repeated four times for each group in turn to be used for testing. In the following experiments, we report the average detection rate with the standard deviation.

B. Baseline

As the baseline to compare against, we conducted experiments with several state-of-the-art face representations. In [15], Gabor features achieve superior performance on smile detection. Therefore, Gabor features were considered as the baseline in this paper. We utilized a

[1] http://mplab.ucsd.edu

TABLE I
EXPERIMENTAL RESULTS OF SMILE DETECTION WITH THE BASELINE APPROACHES

Approach		Accuracy (%)
Feature	Dimension	
Gabor	23,040	89.55±0.63
LBP	944	87.10±0.76
Raw Pixel Values	2,304	80.38±1.04

bank of Gabor filters at eight orientations and five spatial frequencies (9:36 pixels per cycle at 1/2 octave steps[2]). The outputs of the 40 Gabor filters were downsampled by a factor of 4 [17]; therefore, the dimensionality of the Gabor feature vector is 23 040. Another representation we considered is LBP. LBP features have shown promising performance for face analysis in recent years. Following [19], we divided face images of 48 × 48 pixels into 16 subregions of 12 × 12 pixels, and the 59-label $LBP(8, 2, u2)$ operator was adopted to extract LBP features. Thus, each face image was described by an LBP histogram of 944 ($16 * 59$) bins. In some literature, the raw pixel values are used directly for face representation. Here, we also examined the raw pixel intensities for smile detection.

An SVM was adopted as the classifier, which has proven effective in the existing studies [15]. For simplicity, a linear kernel was utilized. We used the SVM implementation (SVM-Light) in the public library SPIDER.[3] Each dimension of the feature vector was scaled to be between -1 and 1. We show the experimental results in Table I. As can be seen, Gabor features provide the accuracy of 89.55%, whereas LBP features achieve 87.10%. These rates are lower than the results reported in [15]. This might be because over 25 000 face images were used in [15], with a training set of 20 000 images. Although Gabor features produce better results than LBP, considering that the dimensionality of LBP features $O(10^3)$ is much lower than that of Gabor features $O(10^5)$, LBP is more promising for practical applications. We also observe that, even with the grayscale pixel values, a linear SVM can achieve the performance of 80.38%, although the standard deviation is larger than the Gabor and LBP features.

C. Illumination Normalization

As can be observed in Figs. 1 and 2, varying illumination is one of the difficulties for smile detection in real-life faces. In the literature, many methods have been proposed for illumination-invariant face recognition. Here, we apply some illumination normalization methods to see the impact of illumination variation, and how much the accuracy can be improved by removing them with the state-of-the-art methods.

We examine the following illumination normalization methods:

1) Histogram equalization (HE): HE is a simple and widely used technique for normalizing illumination effects.
2) Single-scale retinex (SSR): SSR [23] is a basic photometric normalization technique, which smooths the image with a Gaussian filter to estimate the luminance function.
3) Discrete cosine transform (DCT): the DCT-based normalization [24] sets a number of DCT coefficients corresponding to low frequencies as zero to achieve illumination invariance.
4) LBP: one important property of LBP operators is their tolerance against monotonic illumination changes; therefore, LBP can be used as a preprocessing filter to remove illumination effects.
5) Tan–Triggs: Tan and Triggs [25] proposed a series of steps to counter the effects of illumination variation, local shadowing, and highlights while still preserving the essential elements of visual appearance.

[2]i.e., 9, $9\sqrt{2}$, 18, $18\sqrt{2}$, and 36 pixels per cycle.

[3]www.kyb.tuebingen.mpg.de/bs/people/spider/index.html

TABLE II
EXPERIMENTAL RESULTS OF DIFFERENT ILLUMINATION NORMALIZATION METHODS

Illumination Normalization	Gabor (%)	LBP (%)	Raw Pixel Values (%)
None	89.55±0.63	**87.10**±0.76	80.38±1.04
HE	**89.68**±0.62	86.33±1.41	**81.45**±0.32
SSR	86.50±1.35	85.98±0.54	78.85±1.67
DCT	87.05±0.96	85.50±1.21	78.45±0.90
LBP	88.70±0.57	82.15±1.13	78.85±1.26
Tan-Triggs	85.75±0.47	80.65±0.83	78.45±1.37

TABLE III
COMPARISON OF DIFFERENT APPROACHES FOR SMILE DETECTION

Approach			Accuracy (%)
Feature	Dimension	Classifier	
Gabor	23,040	SVM	89.68±0.62
LBP	944	SVM	87.10±0.76
Raw Pixel Values	2,304	SVM	81.45±0.32
Pixel Comparisons	500	Adaboost	89.70±0.45

We applied each of the above methods to the grayscale face image and then performed smile detection with the baseline approaches. The results obtained are shown in Table II. Surprisingly, for Gabor features and raw pixel values, only the HE technique achieves improved accuracy, whereas all other methods fail to improve. Furthermore, for LBP features, none of these illumination normalization methods appear to work. One possible reason could be is that illumination variation in real-life images is very complex. Although these normalization methods show superior performance in the existing studies on some data sets, it seems difficult for them to handle the complex illumination variation in this data set. Since it performs best in the experiments, the HE technique is used in the rest of our experiments.

D. Boosting Pixel Intensity Differences

After extracting intensity difference features from face images preprocessed by HE, we run AdaBoost to choose the discriminative features and combine the selected weak classifiers as a strong classifier. With the selected top 500 features, the trained AdaBoost achieves the accuracy of 89.7%. As compared in Table III, the proposed approach obtains the similar performance to the Gabor-feature-based SVM but uses much less features. We plot in Fig. 3 the accuracy of our approach as the function of the number of weak classifiers. It can be seen that our approach achieves 85% accuracy by using 20 pairs of pixels and 88% accuracy with 100 pairs of pixels. We match the accuracy of the Gabor-feature-based SVM using as few as 350 pairs of pixels.

To understand the features being used in the classifier, we visualize the features that have been learned. We plot in Fig. 4 the spatial distribution of the pixels involved in the selected features in the fourfold cross-validation experiments. That is, the weight for each pixel is accumulated, and the grayscale intensity in pictures in Fig. 4 is proportional to the times of that pixel being used. It is evident that the involved pixels are distributed mainly in the regions around mouth, with a few from the eye areas. This is reasonable, considering that the major difference between smile and nonsmile faces is the mouth or the lips. To validate this further, we derive the "mean faces" of smile and nonsmile by averaging all smile faces and nonsmile faces in the data set, which are shown in Fig. 5. We can see in the mean faces that, visually, the main difference lies in the mouth region and the eyes, where smile faces have open mouth, whereas nonsmile faces have mouth closed. Comparing the left and right pictures in Fig. 4, we can also see the top 251–500 features lies in the similar locations of the top 250 features. Although faces are (on average) symmetric, the selected features are not symmetric because of the pose and illumination variations in the

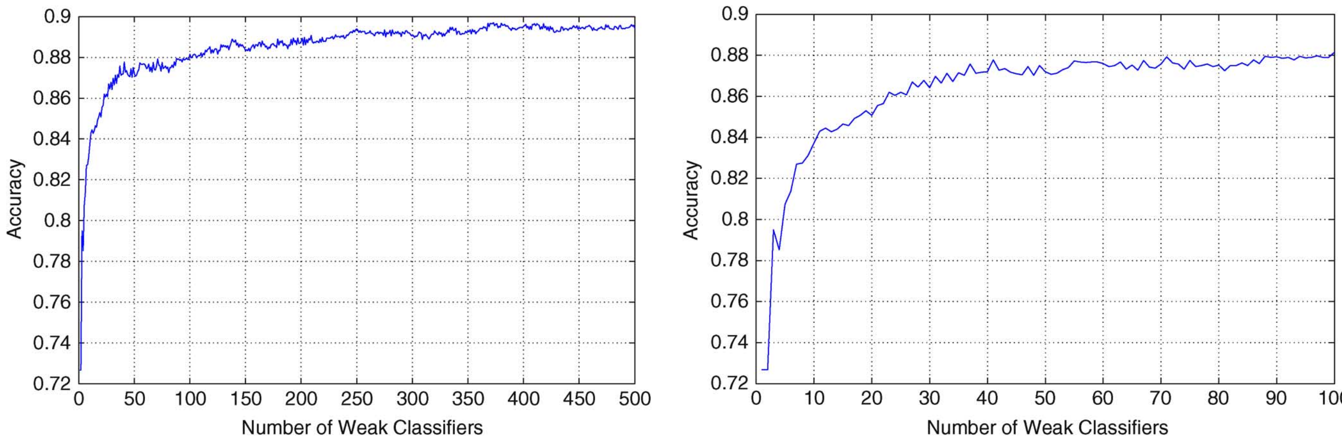

Fig. 3. Average accuracy as a function of the number of weak classifiers used. Left: full graph with 500 weak classifiers. Right: enlarged graph of the first 100 weak classifiers.

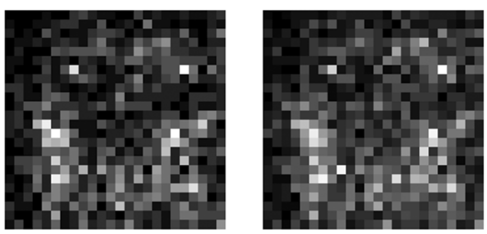

Fig. 4. Distribution of the pixels involved in the selected intensity difference features. Left: top 250 features. Right: top 500 features.

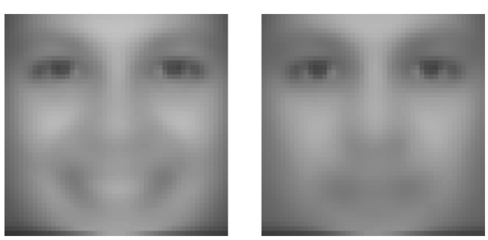

Fig. 5. Average of (left) all smile faces and (right) all nonsmile faces.

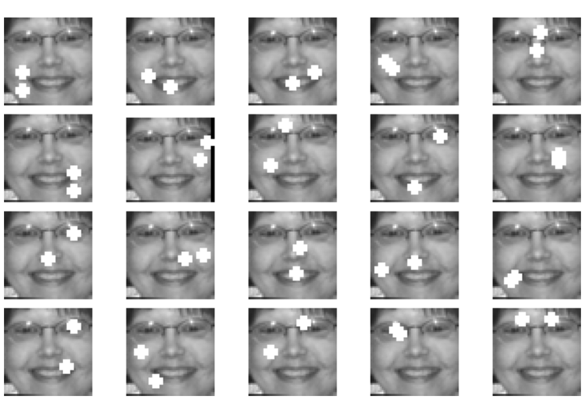

Fig. 6. Top 20 pairs of pixel being selected in one cross-validation experiment.

data set. In Fig. 6, we plot the top 20 pairs of pixels being selected in one cross-validation experiment.

We examine further different illumination normalization methods for the AdaBoost approach. Table IV shows the results obtained. Similar

TABLE IV
EXPERIMENTAL RESULTS OF THE ADABOOST APPROACH WITH DIFFERENT ILLUMINATION NORMALIZATION METHODS

Illumination Normalization	Accuracy (%)
None	89.45±0.25
HE	**89.70**±0.45
SSR	87.50±0.25
DCT	86.88±0.93
LBP	85.73±0.38
Tan-Triggs	81.40±0.45

TABLE V
TIME COST FOR SMILE DETECTION USING THE SVM AND ADABOOST (WITH 20, 100, 300, AND 500 INTENSITY DIFFERENCE FEATURES)

Approach		Time (*ms*)
SVM		1768.7
Adaboost	20 features	0.6
	100 features	2.4
	300 features	7.2
	500 features	10.9

to our early observation in Table II, among all normalization methods, only HE achieves slightly improved performance. This reinforces our early argument that illumination variation in real-life images is very complex, and most of the existing illumination normalization methods cannot handle it effectively.

E. Computational Efficiency

We compare the computational cost of two smile detectors: the SVM (Gabor) and AdaBoost (intensity difference). Given a new test image, the first method extracts Gabor features and then performs classification with the trained linear SVM classifier. In our experiments, which are trained on 3000 face images, the SVM classifier has around 750 support vectors. For the second method, intensity differences of a limited number of pixel pairs from the given image are computed; classifications based on these weak classifiers are combined to get the final decision. Table V shows the average time cost for classifying a new image, using the nonoptimized MATLAB code on a standard personal computer. We can see that the proposed detector is significantly faster than the Gabor-feature-based SVM. Even with 500 features, our approach has a 162-times improvement in speed. SVM classification requires time that is linear in both the number of features as well as the number of support vectors; thus it can be hundreds of times more expensive.

Fig. 7. Examples of failure on smile detection.

Fig. 8. Examples of correct detection.

TABLE VI
DISTRIBUTION OF THE 4000 FACES BASED ON THE POSE RANGE (YAW, PITCH, AND ROLL)

Pose (absolute)	$(20^\circ, 50^\circ)$	$(15^\circ, 20^\circ]$	$(10^\circ, 15^\circ]$	$(5^\circ, 10^\circ]$	$[0^\circ, 5^\circ]$
# Face	689	625	943	1,216	527

F. Impact of Pose Variation

We show in Figs. 7 and 8 some examples of failure and correct detection, respectively. We can see that illumination variation exists in both cases. However, it seems that most of correctly classified faces are frontal, whereas for the faces that are wrongly classified, some of them have large pose variation. Therefore, we investigate further the impact of pose variation on smile detection. Table VI shows the distribution of the 4000 faces based on the pose range (yaw, pitch, and roll). Although the pose range of most images is within $\pm 20^\circ$ of frontal, there are 689 faces with a larger head pose. We run experiments on the faces with the pose range of within $\pm 15^\circ$ and $\pm 5^\circ$ of frontal, respectively. Fig. 9 shows the results. We observe only slight improvements by reducing the pose range. The small difference indicates that the pose variation is not the only main difficulty.

V. CONCLUSION

Smile detection in face images captured in real-world scenarios is an interesting problem with many applications. In practice, because of the limited computational resource, it is desired that the features used can be computed easily and efficiently. In this paper, the intensity differences between pixels in the grayscale face images are used as simple features. We adopt AdaBoost to choose and combine intensity differences to form a strong classifier. Experiments illustrate that our approach matches the accuracy of the state-of-the-art method (i.e., the

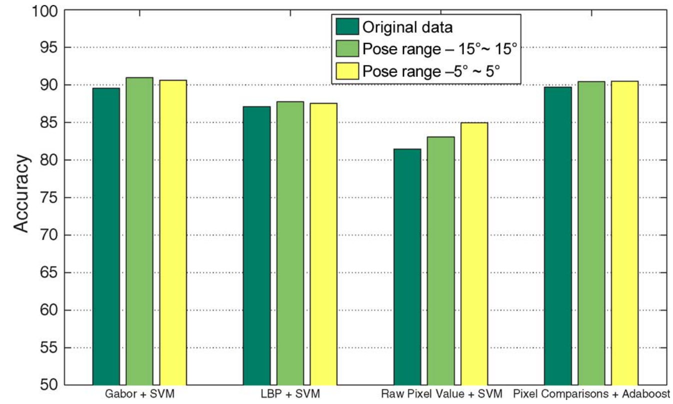

Fig. 9. Smile detection on faces with less pose variation.

Gabor feature-based SVM) using as few as 350 pairs of pixels, allowing significantly faster detection. We also examine different illumination normalization methods and investigate the impact of pose variation.

REFERENCES

[1] M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1424–1445, Dec. 2000.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognit.*, vol. 36, no. 1, pp. 259–275, Jan. 2003.

[3] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[4] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2000, pp. 46–53.

[5] J. F. Cohn and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 2, pp. 1–12, 2004.

[6] Z. Zeng, Y. Fu, G. I. Roisman, Z. Wen, Y. Hu, and T. S. Huang, "Spontaneous emotional facial expression detection," *J. Multimedia*, vol. 1, no. 5, pp. 1–8, Aug. 2006.

[7] M. Bartlett, G. Littlewort, M. Frank, C. Lainscseck, I. Fasel, and J. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *J. Multimedia*, vol. 1, no. 6, pp. 22–35, Sep. 2006.

[8] N. Sebe, M. S. Lew, I. Cohen, Y. Sun, T. Gevers, and T. S. Huang, "Authentic facial expression analysis," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 517–522.

[9] J. F. Cohn, L. I. Reed, Z. Ambadar, J. Xiao, and T. Moriyama, "Automatic analysis and recognition of brow actions in spontaneous facial behavior," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2004, pp. 610–616.

[10] Y. Shinohara and N. Otsu, "Facial expression recognition using fisher weight maps," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 499–504.

[11] A. Ito, X. Wang, M. Suzuki, and S. Makino, "Smile and laughter recognition using speech processing and face recognition from conversation video," in *Proc. Int. Conf. Cyberworlds*, 2005, pp. 437–444.

[12] U. Kowalik, T. Aoki, and H. Yasuda, "Broaference—a next generation multimedia terminal providing direct feedback on audience's satisfaction level," in *Proc. IFIP TC13 Int. Conf. Human-Comput. Interact. (INTERACT)*, 2005, pp. 974–977.

[13] O. Deniz, M. Castrillon, J. Lorenzo, L. Anton, and G. Bueno, "Smile detection for user interfaces," in *Proc. Int. Symp. Adv. Vis. Comput.*, 2008, pp. 602–611.

[14] D. Freire, M. Castrillon, and O. Deniz, "Novel approach for smile detection combining LBP and PCA," in *Proc. Int. Conf. Comput. Aided Syst. Theory (EUROCAST)*, 2009.

[15] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Towards practical smile detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2106–2111, Nov. 2009.

[16] Omron Okao Vision System [Online]. Available: www.omron.com/r_d/coretech/vision/okao.html

[17] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 974–989, Oct. 1999.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 511–518.

[19] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, May 2009.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[21] S. Baluja, M. Sahami, and H. Rowley, "Efficient face orientation discrimination," in *Proc. IEEE Int. Conf. Image Process.*, 2004, pp. 589–592.

[22] S. Baluja and H. A. Rowley, "Boosting set identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, 2007.

[23] D. Jobson, Z. Rahman, and G. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[24] W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 2, pp. 458–466, Apr. 2006.

[25] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, 2007, pp. 168–182.