# COMPARISON OF METHODS FOR SMILE DECEIT DETECTION BY TRAINING AU6 AND AU12 SIMULTANEOUSLY

*Hong Liu, Pingping Wu*

Key Laboratory of Machine Perception and Intelligence
Shenzhen Graduate School, Peking University, Beijing, China
Email: hongliu@pku.edu.cn, pingpingwu@pku.edu.cn

## ABSTRACT

Smiles play an important role in face to face interaction which is a human-specific direct and naturally preeminent way of communication. People smile out of various reasons. Some smiles are spontaneous while some others are just posed. It is hard for human to catch such subtle differences between the two different smiles. In this paper, different algorithm combinations are tried to realize the deceit detection of smiles by computer, that is to say, to recognize true (spontaneous) and fake (posed) smiles. The detection is based on AUs (facial action units). Moreover, AU6 and AU12 are dealt with together in each example, which is different from AU recognition. Images in our database are all frontal facial images with smiles of different types and levels, which are from subjects of different countries and ages with different colors. Experiments are implemented to find out which algorithm combination is the best one. Results show that the best accuracy of the tried combinations in detecting true and fake smiles is close to 86%, while human true-fake-smile recognition ability is much lower. Our work could be used as a tool for the analysis of smiles in psychological area and other regions.

***Index Terms*****—** Facial action units, Gabor wavelets, 2DPCA, Adaboost, SVM

## 1. INTRODUCTION

Smile is one of the most common facial behaviors in our daily life. People smile in order to be polite, to express his/her inside feelings or to conceal his/her real feelings. However, from researches of Frank et al. [1], only one particular type of smile called the enjoyment or spontaneous smile accompanies experienced positive emotions such as happiness, pleasure, or enjoyment. Here, the enjoyment smiles are defined as true smiles and other types as fake or posed ones.

Although fake smiles often look very similar to true ones, they are actually slightly different because they are brought about by different muscles controlled by different parts of the brain. Fake smiles can be performed at will while true smiles are generated by the unconscious brain, so are automatic. Five makers are proposed by Frank et al. [1] to differentiate the true smile from the fake one, from which the Ducheene's maker and the symmetry maker could be captured in a static image of smiles. According to the Ducheene's maker, one of the differences between true and fake smiles is that when a person really feels happy, zygomatic major contracts together with orbicularis oculi. While, according to the symmetry marker, Zygomatic major action produces symmetrical changes on both side of the face. Associated the two makers with FACS (Facial Action Coding System) [2], which is the most objective and widely used method for measuring and describing facial behaviors, two Action Units, AU6 and AU12 are found to be crucial to distinguish the true and fake smile. AU6 is the cheek raiser and lid compressor while AU12 is the lip corner puller. Corresponding to the two morphology makers above, if a smile is a true one, AU6 and AU12 need to happen together at the same time. Moreover, in feature extraction and classification, AU6 and AU12 are treated together. This makes our work different from AU recognition.

A true smile recognition system is designed by Nakano et al. [3] by using neural networks. However, they didn't give an explicit description about the true smile mentioned in their paper and didn't explain why a smile was a true one or not. If results of their neural network could be regarded as the classification of true and false smiles, it could also be considered as a cluster of different intensities of smiles. A deceit-detection in facial expressions was done by Zhang et al. [4] in which the enjoyment expression was also involved. In order to carry out the detection, they used DBF (distance based features) and TBF (texture based features) corresponding to MCs (Major Components) and got the accuracy of 73.16% in deceit detection in enjoyment.

In this paper, we aim to find a best automatic smile deceit detection method. We define that a smile is a true one only if AU6 and AU12 both happen in a static smile image according to the theory from psychological area. Methods presented in our paper are robust for they could work with different races and ages and could tolerate the face off front in some extent. Moreover, methods in our paper are of data-driven, when there are enough examples, the accuracy could still be improved to some extent.

## 2. ALGORITHMS

In the following part, several related methods are introduced to realize feature representation, extraction and classification of AU6 and AU12, which will be combined and evaluated in our experiments to find which combination is the best one.

### 2.1. Feature Extraction

Gabor filters are used to extract features as they have excellent performance in texture representation and discrimination. Image $I(x, y)$ is convolved with 40 Gabor kernels $g_{\mu,\nu}(z)$ separately,

$$W_{\mu,\nu}(x,y) = I(x,y) * g_{\mu,\nu}(x,y) \quad (1)$$

The magnitude response $\|W_{\mu,\nu}(x,y)\|$ is used to represent the feature. After Gabor filtering, the dimension is increased. Using them as the feature directly will lead to high computational complexity and memory requirements.

### 2.2. Dimension Reduction Using 2DPCA

Yang et al. [5] proposed the Two Dimensional Principal Component Analysis approach (2DPCA). Its basic idea is to directly use 2D matrices to construct the corresponding covariance matrix instead of a 1D vector set as in PCA, which improves the computational efficiency. The details of applying 2DPCA are shown in algorithm 1.

**Algorithm** 1: **2DPCA**

**Data:** $W^i_{\mu,\nu}(x,y)$ is the 2D Gabor convolution output, $i \in \{1,\dots,N\}$, $\mu \in \{0,\dots,O\}, \nu \in \{0,\dots,S\}$, where *N*, *O* and *S* are corresponding to the number of examples, orientations and scales, Compute the average matrix $\overline{W} = 1/NOS \sum_{i,\mu,\nu} W^i_{\mu,\nu}(x,y)$ .

**Do the row dimension reduction**

1. *Compute the covariance matrix* $C_r$ *:*
   $C_r = \frac{1}{NOS}\sum_{i,\mu,\nu}(W^i_{\mu,\nu} - \overline{W})(W^i_{\mu,\nu} - \overline{W})^T$
2. *Compute* $d_1$ *largest eigenvalues of C:* $\lambda_1 > \lambda_2 > \cdots > \lambda_{d_1}$
3. *Compute* $d_1$ *corresponding eigenvectors:* $u_1, u_2, \ldots, u_{d_1}$ , $U_{d_1} = [u_1, u_2, \ldots, u_{d_1}]$
4. *Compute the feature matrix* $Y^i_{u,v}$ *of the ith image at orientation* $\mu$ *and scale* $\nu$ *:* $Y^i_{u,v} = U^T_{d_1}(W^i_{u,v} - \overline{W})$

**Do the column dimension reduction**

5. *Put* $(Y^i_{u,v})^T$ *back to step 1 to step 4 as* $W^i_{u,v}$ *to get* $d_2$ *eigenvectors:* $v_1, v_2, \ldots, v_{d_2}$ , $V_{d_2} = [v_1, v_2, \ldots, v_{d_2}]$
6. *Compute the feature matrix* $Z^i_{u,v}$ *of the ith image at orientation* $\mu$ *and scale* $\nu$ *:* $Z^i_{u,v} = (Y^i_{u,v} - \overline{Y})V^T_{d_2}$

Automatic face detection is used to get face regions which are resized with size $M \times M$ pixels. After applying Gabor filters to the face region, a feature with size $40 \times M \times M$ is derived. Then 2DPCA is applied to reduce the dimension. Finally, a feature with size $40 \times d_1 \times d_2$ is obtained.

### 2.3. Dimension Reduction Using Adaboost

The basic idea of Adaboost algorithm is that a strong classifier could be expressed as the linear combination of a series of weak classifiers with different weights on the training set. Here Adaboost is used to extract features and treat each Gabor filter as a weak classifier. Adaboost picks up the best one of these classifiers and boosts the weights on the error examples. The next filter is selected which gives the best performance on the errors of the previous one. After *T* rounds of iteration, *T* features are selected out. The weak classifier could be expressed as follows [6]:

$$h_j(x) = \begin{cases} +1, & p_j\phi_j(x) < p_j\theta_j \\ -1, & \text{otherwise} \end{cases} \quad (2)$$

where $x$ is an example, $\phi_j(x)$ represents extracting a feature from $x$ and $p_j$ is the sign which maintains the direction of the inequality. The detailed steps of attribute sorting with Adaboost algorithm could be seen in [7].

### 2.4. Feature Classification

Some popular classifiers in pattern recognition like SVM, LDA and BP Neural Network are chosen to complete the classification task. Experiments are done in Section 3 to find out which classifier is the best for features selected above in this specific mission. The overall procedure of our smile deceit detection is shown in Fig. 1.

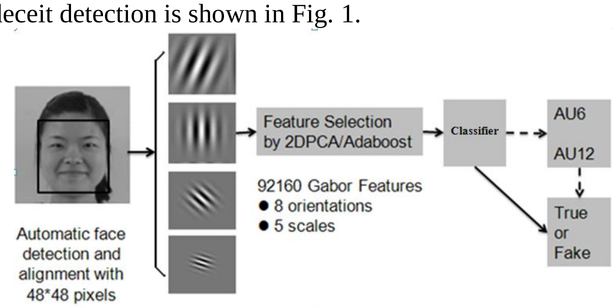


Fig. 1 the overall framework of the smile deceit detection

## 3. EXPERIMENTS AND ANALYSIS

Our experiments are implemented in Matlab and C++. LIBSVM [8] from C. Lin is used and the linear kernel is chosen. The free source code of face detector available at mplab.ucsd.edu is used, which is an improved version and has been shown to perform rather well.

### 3.1. Database

Images in our database are all of smiling-face images from front view. 220 images of different kinds of smiles are gathered with half true smiles and half fake. All the images are colored and saved with resolution of 256×256. 100 of them are collected from Internet [9] from 20 objects, 7 females and 13 males with 5 pictures per person of different ages and races. The rest 120 images are created by ourselves from 12 subjects which are all Chinese aged from 20 to 25 with 10 pictures per person. In 12 subjects, 5 of them are female and 7 male. The 100 images collected from Internet have been analyzed and labeled by the author already. For the rest 120 images, some were taken when the subjects were watching some funny films and some of them were taken when the subject just posed a smile. Examples are shown in Fig. 2. We analyze and label them according to FACS. All the smile images are captured from the video sequences at the smile apexes manually. All the used subjects are healthy people without any disease of the facial muscle.

### 3.2. 2DPCA-SVM

After obtaining the face region and resizing it to 48×48 pixels, Gabor filters are applied, which keep unchanged in other methods' implementation. Different numbers of principal components are tried as shown in Fig.3. When $d_1 = d_2 =$ 8, the hit rate is tending to reach the apex. Therefore, $d_1 = d_2 =$ 8 are used, which makes the dimensionality reduced by 97.2%. The total dimensionality of the extracted feature is 40×8×8.

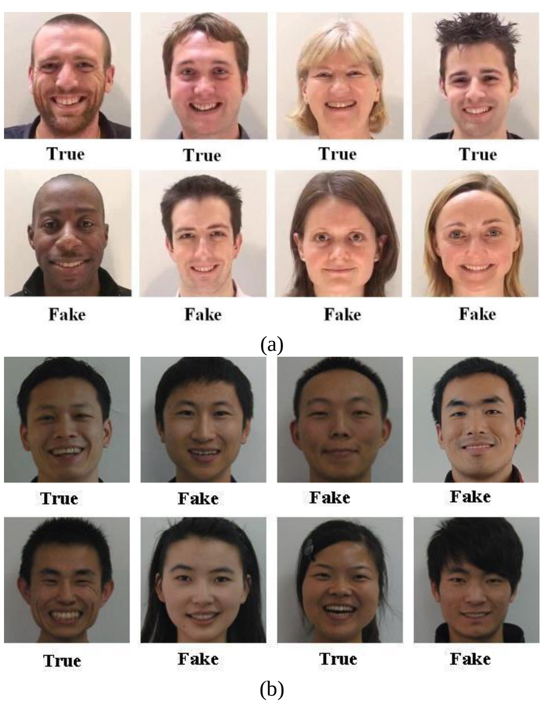


Fig.2 (a) Part of the smile faces collected from BBC and (b) part of the smile faces captured by ourselves

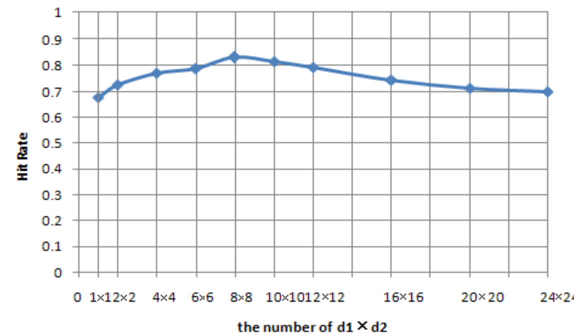


Fig.3 Recognition accuracy with different number of $d_1$ and $d_2$

Table I Results of 2DPCA-SVM and Adaboost-SVM with cross-validation

| | True Positive Rate | True Negative Rate | Hit Rate |
|---|---|---|---|
| 2DPCA-SVM | 80 % (88/110) | 83.6% (92/110) | 81.8% (180/220) |
| Adaboost-SVM | 86.4% (95/110) | 85.4% (94/110) | 85.9% (189/220) |

The 220 images are divided into 5 subsets of equal size randomly in order to implement cross validation. Results of 2DPCA-SVM are shown in Table I, from which it can be found that the method has better performance when recognizing fake smiles.

### 3.3. Adaboost-SVM

In the last part, 2DPCA is used to reduce the dimension and results have been shown in Table I. Instead of 2DPCA, the Adaboost algorithm introduced in Section 2 of Part 3 is tried to get better results. Different feature numbers are selected to find out the most proper amount of features. As shown in Fig.4, the hit rate comes to its peak around 2,500. Therefore, the 2,500 features are chosen as the input features. Results with cross validation are also shown in Table I. It could be found that all of the rates have been improved. In the meantime, the difference in TPR and TNR is less than 2DPCA-SVM.

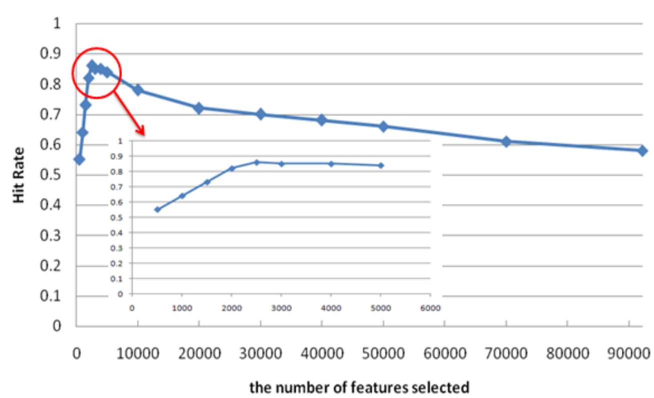


Fig. 4 Recognition accuracy for different number of features selected by Adaboost

For different feature selection methods and classifiers, we try to dig out which feature selection method is better and which classifier is more suitable. The best combination needs to be found out. Corresponding to 2DPCA, 2,560 PCs are selected by PCA. The BP network is of three layers and the input layer has 64 nodes. The output layer is of 2 nodes to represent a true or fake smile. Twelve nodes are used as the hidden layers. As shown in Table II, results from the SVM classifier are usually better than the other two classifiers. One of the reasons is that the SVM classifier is more suitable for small sample data analysis. When using PCA to select the feature, LDA outperforms the other two. As 2DPCA is applied, SVM achieves the best result. When it comes to Adaboost, the three classifiers' performances are all improved and still SVM overcomes the other two. From the experiments, Adaboost is the best feature selection method and SVM the best classifier for this task, which makes them the best combination.

Table II Results of different feature selection methods with different classifiers with cross validation

| | LDA | SVM (linear) | BP network |
|---|---|---|---|
| PCA | 80.0% | 79.5% | 70.9% |
| 2DPCA | 79.1% | 81.8% | 74.1% |
| Adaboost | 84.6% | 85.9% | 77.8% |

### 3.4. Analysis

Why the combination of Adaboost and SVM obtains the best result? We try to explore the difference between features selected by the above two algorithms. Statistical results show that there are more features concentrated in the eye area for features selected by Adaboost while features selected by 2DPCA are less. It is crucial to distinguish the true and fake smile whether AU6 happens while features of AU6 are around the eye area. The analysis above is consistent with the result that the Adaboost the better feature selection method which derives more detailed features of AU6. Meanwhile, SVM has been shown to perform better when the feature space is dense which means the features are highly relevant to each other [10]. LDA is more suitable to classify the examples of Gaussian distribution while SVM not. BP network is based on empirical risk minimization and easy to trap in local optimum while SVM is based on structure risk minimization and considers the sample error and the model complexity. Local optimum is global optimum in SVM. SVM shows better generalization ability than BP network.

### 3.5. Detect AU6 and AU12 Separately

In the previous discussion, AU6 and AU12 are treated as a whole unit to extract features, train and test. The results are acquired straightforwardly from the classifiers as shown in Fig.2 with the solid line with arrows. AU6 and AU12 detectors are also derived by using our database to train AU6 and AU12 separately. In an image, if both AU6 and AU12 detectors get the positive results which means AU6 and AU12 happen together in the image, the smile is determined to be true. This strategy is shown in Fig.2 with dotted line with arrows. However, the hit rate with detecting AU6 and AU12 separately is only 71.8%, which is much lower than the previous best result we got. One of reasons may be that the AU6 and AU12 detectors we train are not accurate and robust enough while the examples are small.

## 4. CONCLUSIONS

In this paper, we first realize the deceit detection of smiles by training AU6 and AU12 simultaneously, that is to say, AU6 and AU12 are treated as a whole unit. Different combinations of methods are employed to find the best one for this task. When dealing with AU6 and AU12 separately, it is verified that the accuracy of the detection decreases. In order to reduce the dimension to make the analysis and computation easier, 2DPCA and Adaboost are applied respectively. Finally, different classifiers are tried to recognize true and fake smiles. Comparison experiments show that the best combination is Adaboost and SVM. Furthermore, analysis has been done to explain why Adaboost and SVM the best combination. Ultimately, our work of deceit detection in smiles could be used as a tool for human robot interaction, behavioral science and psychology research, which is worth studying.

## 6. REFERENCES


[1] M. Frank and P. Ekman. "Not all smiles are created equal: the difference between enjoyment and nonenjoyment smiles", *International Journal of Humor Research*. Vol. 6, No. 1, pp. 9-26, 1993

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

[3] M. Nakano, Y. Mituskura, M. Fukumi and N. Akamatsu. "True Smile Recognition System Using Neural Networks", *Proceddings of the 9th International Conference on Neural Informaion Processing*. Vol. 2, pp. 650 – 654, 2002

[4] Zhi Zhang, Vartika Singh, Thomas E. Slowe, Sergey Tulyakov and Venuopal Govindaraju. "Real-time Automatic Deceit Detection from Involuntary Facial Expressions", *IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 26, pp. 131–137, 2007

[5] J. Yang, D. Zhang, A. Frangi and J. Yang. "Two dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition". *IEEE Transaction on Pattern Analysis and Machine Intelligence*. Vol. 26, No. 1, pp. 131-137, 2004

[6] P. Viola and M. Jones. "Robust Real-Time Face Detection", *International Journal of Computer Vision*. Vol. 57, No. 2, pp. 137-154, 2004

[7] Yuwen Wu, Hong Liu and Hongbin Zha. "Modeling Facial Expression Space for Recognition", *International Conference on Intelligent Robots and System*. Pp. 1814-1820, 2005

[8] C. Chang and C. Lin. LIBSVM: a Library for Support Vector Machines. Available: http://www.csie.ntu.edu.tw/~cjlin/

[9] BBC-Dataset. Available: http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/

[10] D. Roth, M-H. Yang, and N. Ahuja. "Learning to recognize three dimensional objects", *Neural Computation*. Vol. 14, 2002