Methods of Recognizing True and Fake Smiles by Using AU6 and AU12 in a Holistic Way

Pingping Wu, Wenmin Wang and Hong Liu

Abstract. Smile is one of the simplest forms of expressions that it is easy to recognize for human beings. It will be one of the most natural, straightforward and friendly ways in Human Computer Interaction (HRI) if a computer could catch the subtle expression, understand the inner state of human and meanwhile give its feedback according to the corresponding instance. In this paper, some different methods are proposed, to realize the recognition of true and fake smiles, based on facial action units from the research field of psychology and human behavior. In all of the methods we used, AU6 and AU12 are dealt with together in each example, which is different from AU recognition. Some popular feature extraction and classification methods such as Gabor wavelets, 2DPCA, Adaboost and SVM are used in the holistic way to implement the recognition. Images in our database are all frontal facial images with smiles of different types and levels from subjects of different countries with different colors and ages. Lots of experiments show that the best accuracy of our methods in recognizing true and fake smiles is close to 86%, while people's true-fake-smile recognition ability is much lower.

Keywords: Smile detection • Gabor wavelets • Facial action units • 2DPCA • SVM

1.1 Introduction

Smile is one of the most common facial behaviors in our daily life. It plays 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 such as to be polite, to express his/her inside feelings or even to conceal his/her real feelings, which cause different types of smiles. From researches of Frank et al. [1], however, only one particular type of smile called the enjoyment smile accompanies experienced positive emotions such as happiness, pleasure, or enjoyment. Here, the enjoyment smile is defined as true smile and other types as fake ones.

In the last decades, people have done a lot of research on face detection, face recognition and facial expression analysis. Face detection is a rather important and prerequisite step because we can't get the face recognized or the facial expression analyzed before the face is detected. For recent years, researchers have proposed different kinds of methods and some of them achieve fairly good results. A boosted cascade of Haar features was proposed by Viola et al. [2] to get the face detected and their system was very robust and had the fastest detection speed. After all, face detection approaches are either based on a holistic way or an analytic way [3]. In the holistic way, the face is regarded as a

Pingping Wu • Wenmin Wang(✉) • Hong Liu
Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Peking University Shenzhen Graduate School, Room 105 Building G, Nanshan district, 518055 Shenzhen, China
e-mail: wangwm@pkusz.edu.cn

whole unit. In the analytic way, the face is detected by analyzing some important facial features first (e.g., the eyes and the lips). The overall location of the face is then determined by the location of the features in correspondence with each other.

The research on true and fake smile recognition is closely related to facial expression recognition. When it comes to the facial expression analysis, it is necessary to mention the Facial Action Coding System (FACS). Due to the richness and complexity of facial expression, behavior scientists realize that it is necessary to create an objective coding standard. FACS is the most objective and widely used method for measuring and describing facial behaviors. Most automatic facial expression recognition systems were studied with posed expressions. Nowadays, some studies transfer their attention and focus on spontaneous facial expression which involved muscles and dynamics were different from those posed ones. Lv et al. [4] used head motion and AAM features to realize a spontaneous facial expression recognition. A survey of affect recognition methods for spontaneous expressions has been done by Zeng et al. [5]. A new data set named GENKI was collected by Littlewort et al. [6] which consists of 63,000 images for practical smile detection.

Nakano et al. [7] designed a true smile recognition system using neural networks, in which they didn't give an explicit description about the true smile mentioned in their paper and also didn't give a reason why a smile was a true one or not. If the result of the neural network could be regarded as the classification of true and false smiles, it could also be considered as a cluster of different smile intensities. Zhang et al. did a deceit-detection [8] in facial expressions 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 and got the accuracy of 73.16% in deceit detection in enjoyment. Hoque et al. [9] explored temporal patterns to distinguish delight smiles from frustrated smiles. The best classifier distinguished between the patterns of spontaneous smiles under delighted and frustrated stimuli with 92% accuracy. However, their work is based on video sequences in which the sound information is also added. Differently, we put our hands to static images and have no dynamic information used. The deceit detection in posed smile and spontaneous smile has been done in [10].

In this paper, we aim to find an automatic true/fake smile recognition method. As the dynamic information couldn't be derived from a static image, there is not enough information to differentiate the true and fake smile, which increases the difficulty of recognition. In this paper, we treat AU6 and AU12 together in each example to realize the recognition of true and fake smiles. Here, we define that a smile is a true one only if AU6 and AU12 both happen in a static smile image. The methods for recognition 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. In addition, sufficient theoretical foundations are given from the viewpoint of psychology to explain why the true smile is different from the fake one and how to distinguish them.

The rest of this paper is organized as follows: Section 1.2 describes how to extract the features from the true and fake smiles images. In Section 1.3, the experimental results are analyzed, and the conclusions are drawn in Section 1.4.

1.2 Feature Extraction and Classification

1.2.1 Feature Representation

Gabor wavelets were widely used in image processing, pattern recognition and other fields due to their biological relevance and computational properties. Gabor filters are robust since it has the ability to hold with the rotation and deformation of images in some degree. For this advantage, the accuracy of our experiments is ensured because some smile images in our database are slightly yawed, pitched or rolled. Here, five different scales $v \in \{0, ..., 4\}$ and eight different orientations $u \in \{1, ..., 8\}$ are chosen to realize the Gabor filter. The image $I(x, y)$ is convolved with the 40 Gabor kernels $g_{\mu,v}(z)$ separately (5 scales $\times$ 8 orientations),

$$W_{u,v}(x, y) = I(x, y) * g_{u,v}(x, y) \tag{1.1}$$

The magnitude response $\|W_{u,v}(x, y)\|$ is used to represent the feature.

After Gabor filtering, the dimension is increased by 40 times. Using them as the feature directly will lead to high computational complexity and memory requirements. Furthermore, it is difficult to analyze such high-dimensional data accurately. In the following part, the dimension reduction is done by 2DPCA and Adaboost.

1.2.2 Feature Extraction Using 2DPCA

Principal component analysis (PCA) is probably one of the most popular techniques used for dimension-reducing. Yang et al. proposed the 2D-PCA approach whose basic idea is directly using 2D matrices to construct the corresponding covariance matrix instead of a 1D vector set, which improves the computational efficiency. The projection of a sample on each principal orthogonal vector is a vector and the problem of over-compression is alleviated in the 2D-PCA case. As outputs of Gabor filters are 2D matrices, it is more suitable to apply 2DPCA directly on them than PCA.

The convolution output of each Gabor filter contains different local, scale and orientation features. Instead of transforming each 2D convolution output into a vector, it makes sense to operate on them directly. Different from PCA, the covariance matrix C is defined as follows [11]:

$$C = \frac{1}{N}\sum_{k=1}^{N}(x_k - \bar{x})(x_k - \bar{x})^T \tag{1.2}$$

Where $\bar{x} = 1/N\sum_{k=1}^{N} x_k$ is the mean of the total training samples.

According to the generalized total scatter criterion:

$$J(v) = v^T C\, v \tag{1.3}$$

where v is a unitary column vector. The unitary vector v that maximizes the criterion is called the optimal projection axis. Intuitively, this means that the total scatter of the projected samples is maximized after the projection of an image matrix onto v.

$$\begin{cases} \{v_1, ..., v_d\} = \arg\max J(v) \\ v_i^T v_j = 0, i \neq j, \; j = 1, ..., d. \end{cases} \tag{1.4}$$

In fact, the optimal projection, $v_1, ..., v_d$ are orthonormal eigenvectors of *C* corresponding to the first *d* largest eigenvalues. For any output of Gabor filters *W*, its projection to this group of optimal projection vectors is：

$$y_i = W\ v_i, \quad i = 1, ..., d \tag{1.5}$$

Therefore, $y_1, ..., y_d$ are called the principal component (vector) of the sample image.

1.2.3 Feature Extraction Using Adaboost

Adaboost is not only a fast classifier but also an effective feature selection method. 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 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:

$$h_j(x) = \begin{cases} +1, & p_j\phi_j(x) < p_j\theta_j \\ -1, & \text{other}wise \end{cases} \tag{1.6}$$

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 [12].

1.2.4 Feature Classification

Adaboost, SVM, LDA and BP Network are all familiar classifiers. Both SVM and Adaboost could deal with high dimensional space and are simple to train and perform in real time. Their generalization ability is well. A deep inside analysis has been done in [13] to narrate the similarities and differences between them. Special examples (support vectors) are selected by SVM while particular features are selected by Adaboost. It is very important to know that Adaboost is not only a fast classifier but also an effective feature selection method. SVM has been shown to perform better when the feature space is dense which means the features are highly relevant to each other [14]. Experiments have been done in [13] to explore training SVM with the features selected by Adaboost. And the results show that training SVM on the continuous outputs of the selected filters of Adaboost outperforms Adaboost and SVM individually. Linear Discriminant Analysis (LDA) has been shown to be also widely used in facial expression recognition in many works. 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 [15]. SVM shows better generalization ability than BP network. Here, SVM, LDA and BP Neural Network are chosen to complete the task of classification. Experiments are done in 1.3 to find out which classifier is best for the feature selected above in this specific mission.

The overall procedure of our true/fake smile recognition system is shown in Fig. 1.1, which will be stated in detail in the following section.

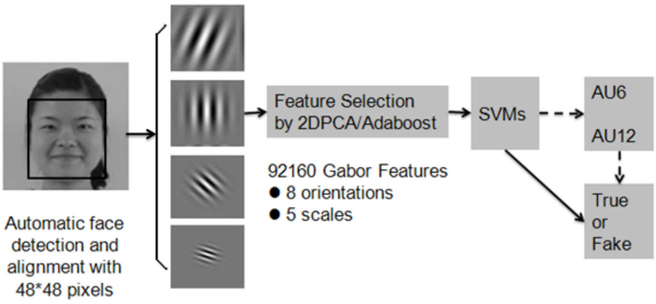


Fig. 1.1 Automated true and fake smile recognition system

1.3 Experiments and Analysis

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

1.3.1 Database

Images in database are all of smiling-face images from front view. 220 images of different kinds of smiles are gathered with half true and half fake. All the images are colored and saved with resolution of 256×256. 100 of them are captured from public database (BBC: Human Body & Mind) from 20 objects, 7 females and 13 males with 5 pictures per person of different ages and races. The rest 120 are created by ourselves for 12 subjects whom are all Chinese aged from 20 to 25 with 10 pictures per person. In the 12 subjects, 5 of them are female and 7 male. The 100 images collected from the public database have been analyzed and labeled by the author already, which is shown in Fig. 1.2. 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. 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.

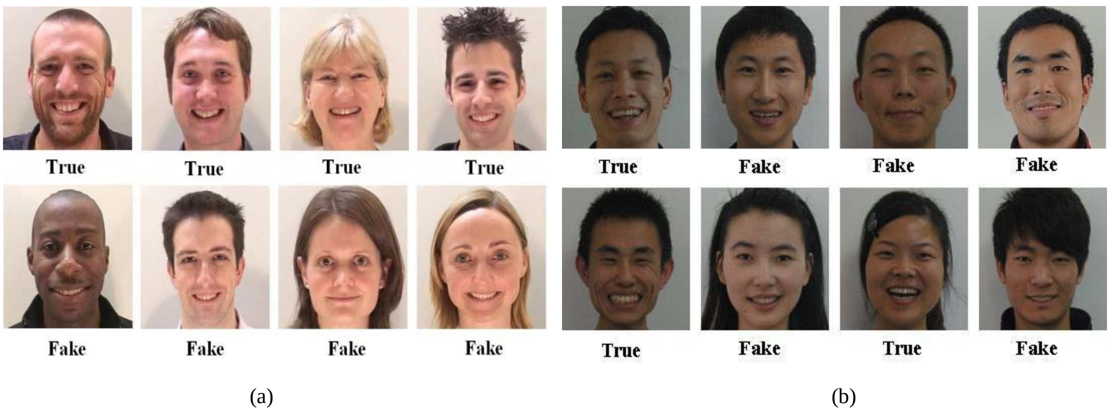


Fig. 1.2 (a) Part of the smile faces collected from BBC and (b) part of the smile faces captured by ourselves

The Alignment is realized in the automatic way, in which automatic eye detection is used to find the centers of the two eyes and then the image is rotated to make the eyes horizontal. An updated version of the eye detector presented in [16] is used. All the faces and eyes in our database have been successfully detected.

1.3.2 2DPCA-SVM

After obtaining the face region and resizing it to 48×48 pixels, the Gabor filter is applied. This will keep unchanged in other methods' implementation. Different number of principal components $d_1 = d_2 = 1$, 2, 4, 6, 8, 10, 12, 16, 20 and 24 are tried as shown in Fig.1.3. When $d_1 = d_2 =$ 8, the hit rate is tending to reach the apex. Therefore, $d_1 = d_2 = 8$ are used, which make the dimensionality reduced by 97.2%. The total dimensionality of the extracted feature is 40×8×8.

The 220 images are divided into 5 subsets of equal size randomly. Sequentially, one subset is tested using the classifier trained on the remaining four subsets. Therefore, each instance of the whole set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. This procedure can prevent the over-fitting problem. Results of 2DPCA-SVM are shown in Table 1.1, from which it could be found that the method has better performance when recognizing fake smiles.

Table 1.1 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)

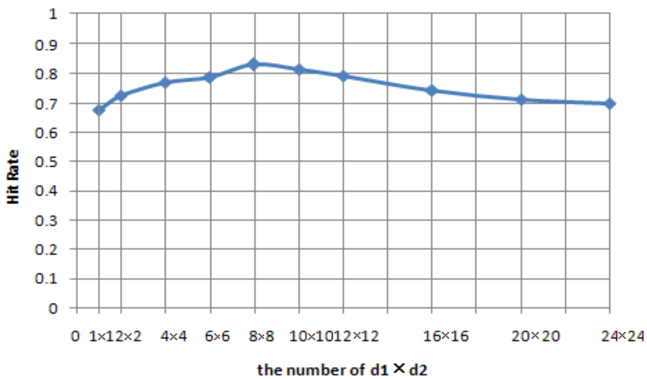


Fig. 1.3 Recognition accuracy with different number of d_1 and d_2

1.3.3 Adaboost-SVM

Instead of 2DPCA, the Adaboost algorithm introduced in 1.2.3 is tried to get better results. Different feature numbers are selected to find out the most proper amount of features. As shown in Fig.1.4, the hit rate comes to its peak around 2,500. Therefore, 2,500 features are chosen as the input features of the classifiers. Results with cross validation are also shown in Table 1.1. It could be found that all of the rates have been improved through the method. In the meantime, the difference in TPR and TNR is smaller than 2DPCA-SVM. With Adaboost-SVM method, 189 smile pictures are correctly recognized while with 2DPCA-SVM method, 180 pictures are rightly classified. In the two correctly recognized sets, 168 pictures are in common. 81 true smile pictures classified by 2DPCA-SVM are also recognized as true with Adaboost-SVM while 87 fake smile pictures which are correctly recognized are the same in both methods. Therefore, if covering the true recognized pictures in both methods, the hit rate will be improved to 91.4%.

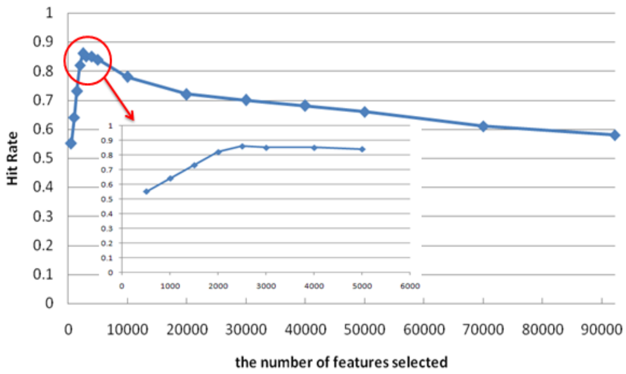
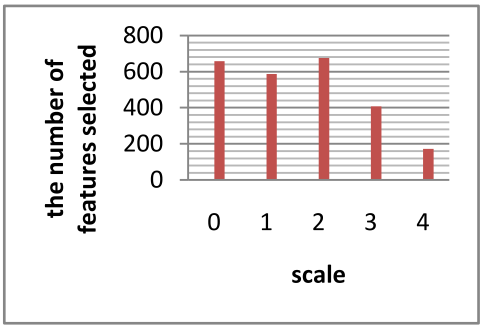


Fig. 1.4 Recognition accuracy for different number of features selected by Adaboost

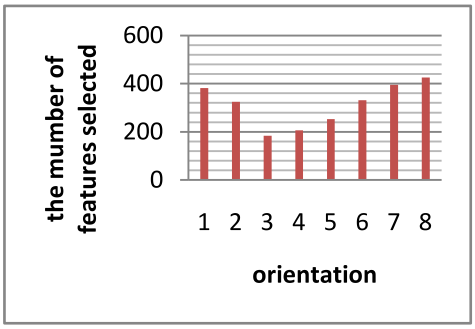
For different feature selection methods and classifiers, we try to dig out which feature selection method is better and which classifier is more suitable. The best combination needs to be found out. Corresponding to 2DPCA, 2,560 PCs are selected by PCA. The BP network is of three layers and the input 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 1.3, results from the SVM classifier are 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. In the meantime, from the data in Table 1.2, it is found that they are the best combination for the task.

Table 1.2 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%



(a)



(b)

Fig. 1.5 Distributions of Gabor features (a) in different scales and (b) in different orientations

In the 2,500 features selected by Adaboost from the Gabor features, we try to learn the distribution of them in each channel which means different scales and orientations. Apparently, different channels of Gabor filters do different levels of contribution for the task. Fig.1.5 shows the distributions of the extracted features in 5 scales and 8 orientations. We could see that the scales v= 0, 1 and 2 are more contributive than the other two. And orientations u=1, 7 and 8 are more dedicative than the rest.

1.3.4 Analysis

Experiments also have been done to check out the methods' ability in recognizing the true/fake smile of different races and genders. In the comparison of different races, the database is divided into two categories. The first category is 120 pictures with subjects all Asians and the second one is 100 pictures with subjects non Asians. When comparing two different genders, the database is divided into two parts with one part all female and the other all male. Table 1.3 shows the results (Hit Rate) from two methods of cross-validation with Gabor filters of 5 scales and 8 orientations. It is found that both methods prefer Asians. Both methods' Hit Rate dropped when dealing with non Asians. One of the reasons leading to the phenomenon is that the coverage of the 120 pictures is not wide enough and the subjects' age concentrates in 20s. On the other hand, the other 100 pictures are from subjects with different ages, colors and races. The hit rate of female pictures and the hit rate of male pictures are quite similar with both two methods. The data shows that the recognition of true and fake smile is irrespective of gender to some extent.

Table 1.3 Comparison of performance of different races and genders through different methods with 5-scale and 8-orientation Gabor filters

	Adaboost-SVM	2DPCA-SVM
Asians (120 pictures)	86.7 % (104/120)	84.2 % (101/120)
Non Asians (100 pictures)	85.0 % (85/100)	79.0 % (79/100)
Mixed (220 pictures)	85.9 % (189/220)	81.8% (180/220)
12 Females(85 pictures)	85.9 % (73/85)	82.3% (70/85)
20 Males(135 pictures)	85.9 % (116/135)	81.5 % (110/135)

1.4 Conclusions

In this paper, we combine with the facial action units of FACS to analyze true and fake smiles. The recognition is realized in the holistic way by dealing with AU6 and AU12 together, which is different from AU recognition. Gabor filters are used to present features as they have excellent performance in texture representation and discrimination. In order to reduce the dimension to make the analysis and computation easier, 2DPCA and Adaboost are applied respectively. Finally, different classifiers such as SVM, LDA and BP neural network are used to recognize true and fake smiles. Comparison experimental results show that the best combination of methods is Adaboost+SVM.

From the experiments, it is found that the recognition of true and fake smiles is independent to the gender. The hit rate could still be improved by analyzing more detailed features. Anyway, the true-fake-smile recognition system would improve the human robot interaction and make the interaction more friendly and deeply. Furthermore, it could also be used as a tool for behavioral science and psychology research, which is worth studying.

1.5 Acknowledgment

This work is supported by National Natural Science Foundation of China (NSFC, No.60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Scientific and Technical Innovation Commission of Shenzhen Municipality (No.JC201005280682A, CXC201104210010A).

1.6 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. Viola and M. Jones. “Robust Real-Time Face Detection”, *International Journal of Computer Vision*. Vol. 57, No. 2, pp. 137-154, 2004

[3] M. Pantic and L. Rothkrantz. “Automatic Analysis of Facial Expressions: the State of the Art”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1424-1444, 2000

[4] Y. Lv and S. Wang. “A Spontaneous Facial Expression Recognition Method Using Head Motion and AAM Features”, *2010 Second World Congress on Nature and Biologically Inspired Computing.* Pp. 334 – 339, Dec. 15-17, 2010

[5] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. “A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions”, *IEEE Trans. Pattern Analysis and Machine Intelligence*. Vol. 31, No. 1, pp. 39-58, Jan.2009

[6] J. Whitehill, G. Littlewort, I. Fasel and M. Bartlett. “Toward Practical Smile Detection”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*. Vol. 31, No. 11, pp. 2106-2111, 2009

[7] 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

[8] 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

[9] M. E. Hoque, D. J. McDuff, R. W Picard, “Exploring Temporal Patterns towards Classifying Frustrated and Delighted Smiles”, *IEEE Transactions on Affective Computing*, Vol. 3, pp. 323–334, 2012

[10] H. Liu and Pingping Wu. “Comparison of Methods for Smile Deceit Detection by Training AU6 and AU12 Simultaneously”, *Int. Conf. on Image Processing*, Pp. 1805–1808, 2012

[11] 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

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

[13] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind and J. Movellan. “Dynamics of Facial Expression Extracted Automatically from Video”, *Computer Vision and Pattern Recognition Workshop*. Vol. 24, pp. 615-625, 2006

[14] D. Roth, M-H. Yang, and N. Ahuja. “Learning to recognize three dimensional objects”, *Neural Computation*. Vol. 14, pp. 1071-1103, 2002

[15] D. Niu, Y. Wang and Z. Gu. “Support Vector Machines Based on Lyapunov Exponents in Power Load Forecasting Model”, *IEEE Asia Pacific Conf. on Circuits and Systems*, pp. 368 - 371, 2008

[16] I. Fasel, B. Fortenberry and J. Movellan. “A Generative Framework for Real Time Object Detection and Classification”, *Computer Vision and Image Understanding.* Vol. 98, pp. 182-210, 2005