Lip AUs Detection by Boost-SVM and Gabor

Xianmei Wang, Yuyu Liang, Xiujie Zhao and Zhiliang Wang
School of Computer and Communication Engineering, University of Science and Technology, Beijing, China
Email: plum-wang@tom.com, ustbly@foxmail.com, snowintheheanven@126.com, wzl@ustb.edu.cn

***Abstract*—Facial expression plays an important role in non-verbal social communication, emotion expression and affective recognition. To make the reorganization of facial expression more effectively, researchers try to recognize facial expression by the recognition of facial action units. In this paper, in order to identify lip AUs, we adopt Gabor wavelet transformation as the feature extraction method and Adaboost-SVM (combination of Adaboost and SVM) as the pattern classifier. Compared with the traditional non-cascaded Adboost algorithms in which the number of weak classifiers actually is fixed beforehand, a process of number optimization is added to ensure the minimum number of the weak classifies combination with the highest recognition rate. To test the effect of the given solution, 150 images from frontal faces for training and 100 images for prediction are collected. Compared with traditional SVM, the proposed system with Adaboost-SVM can not only improve the recognition accuracy from 79.0% to 83.0%, but also speed up AU classification process obviously.**

***Index Terms*—Lip AUs detection, Adaboost, SVM, Gabor filter**

I. INTRODUCTION

Facial expression is one of the basic ways expressing emotion and exchanging information, and plays a crucial role in non-verbal communication. As an effective measure of expression recognition, automatic recognition of action units is significant and challenging. Various application fields include harmonious human-machine interaction system, affective computing, psychology, medical care and cure, distance education, security and others [1].

One of the main streams in facial expression recognition considers muscle actions (action units, AUs) defined in the Facial Action Coding System (FACS) [2]. FACS is currently the most widely used expression coding system in the behavioral sciences. FACS provides an objective and comprehensive language for describing facial expressions and relating them back to what is known about their meaning from the behavioral science literature [3]. In the fields of computer vision and machine learning, in order to use FACS theory for automatic AUs recognition, researchers have done a lot of work. Some are listed below.

Marian Stewart Bartlett et al. [3] presented a systematic comparison of machine learning methods applied to the problem of fully automatic recognition of facial expressions based on AUs, including Adaboost, SVM, and linear discriminate analysis, as well as feature selection methods. In their experiments, best results were obtained by selecting a subset of Gabor filters using Adboost and then training Support Vector Machines on the outputs of the filters selected by Adaboost. The combination of Adaboost and SVM enhanced both speed and accuracy of the system. They also found that PCA was a much less effective representation than Gabor wavelets for facial action recognition.

Khademi et al. [4] presented a novel method which encoded an image sequence as a fourth-order tensor. They used both Gabor representations and the geometric features. Then they applied MBDA (multi-linear biased discriminate analysis) and 2BDA, the dimensionality reduction techniques, to Gabor representations and the geometric features respectively. With SVM as the classifier, an average recognition rate of 89.2% and 96.4% were achieved for upper face and lower face action units respectively.

Littlewort et al. [5] presented the Computer Expression Recognition Toolbox (CERT), a software tool for fully automatic real-time facial expression recognition. CERT applied Gabor filter in the stage of feature extraction and SVM in the stage of action unit recogniton. On a database of posed facial expressions, CERT achieved an average recognition of 90.1% when analyzing facial actions. On a spontaneous facial expression dataset, CERT achieved an accuracy of nearly 80%.

Valstar et al. [6] proved that appearance descriptors offer the opportunity for efficient and robust facial expression recognition. To encode facial expression dynamics, they extended the purely spatial representation LPQ to a dynamic texture descriptor called Local Phase Quantisation from Three Orthogonal Planes (LPQ-TOP), and compared this with the Local Binary Patterns from Three Orthogonal Planes (LBP-TOP). They used SVM classifier to detect the upper face AUs, and results showed that their proposed LPQ-TOP method outperformed all other tested methods.

Yunfeng Zhu et al. [7] proposed novel alternative dynamic cascades with bidirectional bootstrapping (DCBB) to select training samples. Using an iterative approach, DCBB optimally selects positive and negative samples in the training data. They explored the use of the SIFT descriptors to be used as appearance features. With the advantages of feature selection, efficiency and robustness, Cascade Adaboost was used as the classifier.

Manuscript received August 1, 2011; revised January 2, 2012; accepted January 18, 2012.
This project is supported by NSFC with Number 61105120 and No.61170117. Corresponding author is Xianmei Wang.

doi:10.4304/jsw.7.9.1968-1974

They used samples of nonposed behavior recorded during interviews in the RU-FACS database. For all tested action units, DCBB improved AU detection relative to alternative approaches.

Savran, A. et al. [8] presented a technique to automatically recognize lower facial Action Units. Their features are geometry, and consist of a shape model of 83 facial feature points. They compared classification by AdaBoost and SVM. The system was trained on the Cohn and Kanade's DFAT-504 dataset and detected 5 AUs (AU12, AU20, AU23, AU25 and AU27). Performance was measured by the F1 score which can be interpreted as a weighted average of the precision and recall. SVM achieves an average F1 score of 92.77%, while AdaBoost has an average of 90.73%.

From above research, we can obviously draw a conclusion that Gabor filter and SVM classifier (especially the combination of Adaboost and SVM) are used widely in facial AUs recognition.

However, in the traditional non-cascaded Adaboost frame, T actually is designed in advance as a fixed integer. In order to get a high performance, T may be set to a large value. We think this can lead to some problems. One is that the number of fixed selected weak classifier isn't always the best. Sometimes, system performance may have been satisfied before the fixed T classifiers are selected.

In 2008, Manfred Nusseck et al [9] presented the results of four psychophysical experiments in which they systematically manipulated certain facial areas in video sequences of nine conversational expressions to investigate recognition performance and its dependency on the motions of different facial parts. The results showed that different facial parts presented different contribution, with the mouth region being the most critical. So the accurate indentation of lip AUs is very important for the visual perception of natural facial expression. For some basic expression such as happy and surprise, mouth region provide sufficient information for recognition.

However the AUs recognition of mouth regions (called lip AUs) typically causes problems for existing techniques. For example, AU25 and AU26 are similar and easy to be confused [10]. Lip AUs around mouth area does still need further research.

This paper presents an improved solution based on Gabor and Adaboost-SVM for lip AUs recognition. Gabor feature is one kind of features with good characteristics, so it is widely used for pattern presentation and recognition. Actually the dimension of Gabor features is very high and feature transform or selection is followed to decrease the number of Gabor features. Since boosting algorithm is an outstanding method to select good features and enhance the classification accuracy, we try to use boosting algorithm for Gabor feature selection and classification. In the boosting process, SVM acts as the base classifier. Experiments show that, compared by the traditional Gabor and SVM frame, our Gabor, Adaboost and SVM system can achieve higher accuracy rate with faster speed.

The outline of the paper is as follows. Section II provides the structure of our proposed method. Section III shows the utilized approach of the related research. In this part, we first give a short and simple description of our used methods for face detection and facial key point localization. Then the Gabor feature extraction and identification method are presented. In order to obtain the best results within the appointed T weak classifiers, a number optimization step is shown. Finally we extend our two-class identification frame to multi-class inference for lip AUs recognition by pair wise partitioning strategies. Section IV describes the evaluation study and discussed the results, including the minimum number of weak classifiers, the recognition rate of different pairs, the performance of selected weak classifier, and the confused matrix for the recognition of different lip AUs. Section V concludes the paper.

II. System Structure

Fig. 1 gives an overview of our system. The whole system can be divided into two stages: training stage and classification stage. In the training stage, Gabor features are extracted from the images in the database which have been cropped manually and then normalized into fixed size. Then all the features are put into the Adaboost-SVM framework and T best ones are selected from all the features. Because T is appointed beforehand, an analysis is followed to find $\widetilde{T}$, the optimized number of the weak classifiers from T selected classifiers, which will be used to guide the feature extraction and classification in the identification/classification stage.

For the classification stage, in the preprocessing phase, the face is detected by improved discrete Adaboost mechanism using 5 types of Haar features and facial key points are localized by ASM. Then the mouth area is obtained by the key points around the mouth and normalized to the size of 10*20. The extraction of $\widetilde{T}$ -dimension Gabor features is followed. Finally the $\widetilde{T}$ weak classifiers are linearly combined to form a strong classifier and identify the unknown lip AUs which are represented by the real part of Gabor features.

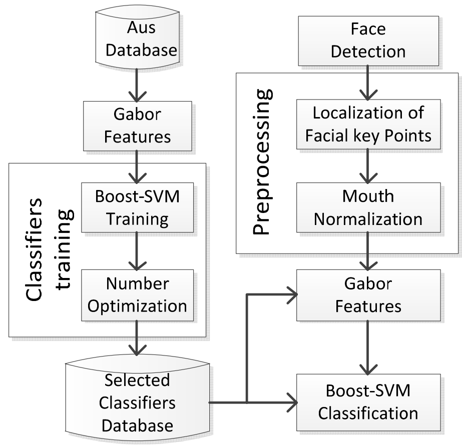


Figure 1. The system overview.

II. Methodology

A. Real-time Facial Detection Technique

We develop a real-time frontal face detection system that employs boosting techniques in a generative framework [11] and extend work by [12]. Enhancements to [11] include employing Binary-Threshold Weak Classifiers (BTWC)", to replace the Unique Threshold Weak Classifiers. With controlled lighting and background, detection accuracy is much higher.

Suppose x_i is the *ith* sample, $f_j(\bullet)$ is the *jth* presentation for x_i, $h_i(\bullet)$ is a weak classifier for the *jth* presentation of all training samples, the format for BTWC in the improved Adaboost [12] can be shown as (1).

$$h_j(x_i \mid \theta_{j1}, \theta_{j2}) = \begin{cases} 1 & when\ \ \theta_{j1} < \theta_{j2}\ ,\ f_j(x_i) \in [\theta_{j1}, \theta_{j2}] \\ & or\ \ when\ \ \theta_{j1} > \theta_{j2}\ ,\ f_j(x_i) \notin [\theta_{j2}, \theta_{j1}] \\ 0 & otherwise \end{cases} \quad (1)$$

θ_{j1} and θ_{j2} are the two thresholds for a weak classifier and they are determined by following condition

$$\theta_{j1}, \theta_{j2} = \arg\min_{\theta_1, \theta_2} \sum_i \left(w[i] * \left| h_j(x_i \mid \theta_1, \theta_2) - y_i \right| \right) \quad (2)$$

Where $w(i) \in [0,1]$ and $y_i = \{+1, -1\}$ are the weight and true label of each training sample respectively.

The features employed for the Adaboost classifier were the 5 types of basic Haar features which are listed in Figure 2. This gives 78460 possible features, shown in Table I.

Figure 2 Five types of Haar features used in our face detection system

TABLE I
THE NUMBER OF 5 BASIC HAAR FEATURES USED IN OUR SYSTEM

Feature style	Feature number
1	21000
2	21000
3	13230
4	13230
5	10000
total	78460

Using 3000 frontal faces with size of 20*20 as positive training set which mainly from FERET and CAS-PEAL [13], we finally get a cascaded classifier with 22 stages. There are totally 336 Haar features to be selected. In the training process, some key parameters are as following:

False positive rate of the cascaded classifier: $F = 1e - 6$.

False positive rate of the *ith* strong classifier: $f_i = 0.5$.

Detection rate of the *ith* strong classifier: $d_i = 0.995$.

Positive number for the *ith* strong classifier: 3000

Negative number for the *ith* strong classifier: 3000.

In the face detection process, scale is 1.2, and detective window shifting step is 2 pixels. Different from [11], scaling here is achieved by scaling the image, rather than scaling the detector itself.

B. Localization of Facial Key Points

For the facial key point localization, ASM tool [14] is employed in our system. This tool is developed and provided by Wei Yao, which is built on pyramid analysis of image [15]. With this dynamic linkage library tool, we can train and detect facial points easily according to our special requirement. The points cluster with 68 key points is also shown in our previous paper [16]. Based on the facial key points, the rectangle image around mouth area can be obtained and then be normalized into the special size of 20*10.

C. Extraction of Gabor Feature

Gabor filter can capture salient visual properties such as spatial localization, orientation selectivity, and spatial frequency characteristics. Considering these outstanding capacities and its great success in face recognition, we choose Gabor features to represent lip AUs samples. Gabor filter is defined as

$$\Psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{u,v}z} - e^{-\sigma^2/2}] \quad (3)$$

In our system, the parameters in (1) are as following:

$$k_{u,v} = k_v e^{i\phi_u}$$

$$k_v = k_{\max} / f^v$$

$$\phi_u = \frac{u\pi}{8}$$

where

$k_{\max} = \pi/2$, f= $\sqrt{2}$,v=5, u=8, $\sigma = 2\pi$, $\vec{z} = (x, y)$.

The final Gabor features are the convolution of Gabor filters and grey-level information of an image on different values of spatial frequencies v and orientations u. They are calculated by (4)

$$J_{u,v}(z) = \int I(z') \Psi_{u,v}(z - z') d^2 z' \quad (4)$$

Where $I(z')$ is the grey-level information of an image.

As we know, $J_{u,v}(Z)$ is a complex which includes a real part and an imaginary part. In our system, only the real parts of the Gabor features are used as the final features. That is to say, the images are converted into Gabor real part presentations.

D. Feature Selection by Adaboost-SVM Algorithm

The basic idea of boosting algorithm [17] is to produce several weak classifiers which are slightly better than random guessing, then incorporate them into an estimate with high accuracy. Because each dimension of Gabor feature corresponds to a weak classifier, the selection process of classifiers is no other than features selection. Thus we realize the reduction of dimensions. In our system, we use the framework of Real Adaboost [18] which is one of the boosting algorithms. The advantages of Adaboost are as following: (1) Its training process focuses on the data difficult to be classified. (2) Instead of average voting mechanism, weighted voting is adopted during the integration of weak classifiers. (3) Features are selected contingent on the features.

In this paper, we use SVM as the base classifier in the framework of Adaboost. There are some reasons: (1) SVM can deal with problems, such as small size of samples, nonlinearity or high dimensions (2) There exists mature and convenient software package of SVM, such as LIBSVM. On each training round, the Gabor feature with the best classification performance tested by SVMs for the current boosting distribution is chosen. The performance measure is a weighted sum of errors on a binary classification task, where the weighting distribution (boosting) is updated at every step to reflect how well each training vector is classified.

Suppose $X(j)=\{x_i^{(j)}\},(i=1,2,...N, j=1,2,...,n)$ is the features bank for one sample, N is the total dimension of features, $y(j)\in\{1,-1\}(j=1,2,...,n)$ is the label of each samples, T is the appointed feature number which is set beforehand, and n is the total number of training samples. Then the Real Adaboost-SVM process can be described as following:

(1) Initialize weight for each sample:

$$D_i = 1/n \ \ (i=1,2,...,n) \tag{5}$$

(2) For each feature x_i ($i=1,2,...,N$), train a SVM classifier $h(x_i)\in\{-1,+1\}$ which is restricted to using a single feature. Then N SVM models are got.

(3) For t=1,...,T:

For each $h(x_i)$, calculate the weighted error

$$\varepsilon_t^{(i)} = \sum_{j=1}^{n} D_i \left| h(x_j) - y(j) \right| \tag{6}$$

Choose a feature with the lowest weighted error rate ε_j and save its corresponding SVM model.

Calculate the selected weak classifier's weight

$$a_t = \frac{1}{2}\ln(1-\varepsilon_t/\varepsilon_t) \tag{7}$$

Update sample weights according to a_t:

$$D_{t+1}(i) = D_t(i)\exp(-a_t y_i h_t(x_i))/Z_t \tag{8}$$

(Z_t is a normalization factor).

E. Number Optimization

Since the parameter T is determined in advance, it may not the best for the final performance and the best value $\widetilde{T}$ for the final system may no more than T. So parameter optimization for the number of feature selection is added in our system. Given T selected weak classifiers $h(x_j)$, this process can be described as below.

(1) Calculate the recognition accuracy with different weak classifiers' number from 1 to T.

For $t=1,....T$

For $i=1,....n$

Use strong classifier H integrated by t SVM weak classifiers to test training set.

$$H(t) = sign\left(\sum_{j=1}^{t} a_j h(x_j)\right) \tag{9}$$

$$h(x_j) = \begin{cases} -1 & svm\_h(x_j) < 0 \\ +1 & svm\_h(x_j) > 0 \end{cases}$$

Here $svm\_h(x_j)$ presents the SVM classifier whose output is a real data between [-1, 1].

Calculate the error rate (not the weighted error rate) of AU recognition by strong classifier H.

$$\varepsilon_t = \varepsilon_{t-1} + \sum_{j=1}^{n}\left|h(x_j) - y(j)\right|/n \tag{10}$$

(2) Choose the best $\widetilde{t}$ with the highest accuracy rates.

$$\widetilde{t} = \arg(\min_{t=1,2,...T} \varepsilon_t) \tag{11}$$

Finally, save the spatial frequencies and orientations of the $\widetilde{t}$ selected Gabor features, as well as their corresponding SVM models and weights.

F. Multi-class AU Recognition

Above is the algorithm for the recognition of only two classes. In our system, 5 types of AUs need to be detected in total. So we must extend the algorithm from binary classes to multi-classes.

There are a number of strategies for partitioning the classification task into binary decisions. The simplest strategy is to train 1 versus all ($\omega_i/\overline{\omega}_i$). Pairwise partitioning strategies (ω_i/ω_j) also have been used widely. The first strategy is simpler but may require larger storage space. For ω_i/ω_j partitioning, the number of training samples for each SVM may be relatively small. Considering the storage limitation, ω_i/ω_j partitioning strategies are used in this paper which means that SVM are trained to discriminate all pairs of lip AUs.

If the type of lip AUs for prediction is m, then there will be $m-1$ pairs for one class ω_i and C_m^2 2-class pairs in all for all classes. In training stage, the parameters for each pair ω_i/ω_j are developed by the algorithms shown in D and E. In the recognition stage, we use voting mechanism to get the final decision.

Suppose $H_{ij} = \{+1, -1\}$ is the strong classifier established for a pair ω_i / ω_j, X_i is the sample set for ω_i and acts as the positive set, X_j is the training set for w_j and acts as the negative set, P_{ij} is the label output by H_{ij}, x^k is the feature from the k^{th} sample, $l(i)$ is the voting time for class ω_i, P is the final label which is achieved by voting mechanism, $c(i) = \{1,2,3...m\}$ is the label for each class. The decision can be described as

$$H_{ij}(x^k) = sign(\sum_{l=1}^{\tilde{T}} a_{lj} h_{ij}(x_k))$$

$$P_{ij} = \begin{cases} i & H_{ij}(x^k) = 1 \\ j & H_{ij}(x^k) = -1 \end{cases} \tag{12}$$

$$l(i) = \sum_{j=1}^{C_m^2} (P_{ij} == i)$$

$$P = c(\arg \max_{i=1,2,\dots,C_m^2} (l(i)))$$

Finally, the label for x^k is the one with the largest voting times.

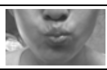
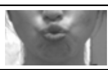
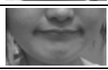
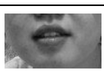
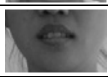
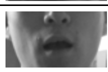
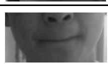
IV. EXPERIMENTS

A. Database

Our image database contains 5 types of mouth/lip AUs, including AU18, AU20, AU25, AU26 and AU28. The training set consists of 30 samples of each single AU, which is cropped by nearly frontal face. In the testing set the number of each pattern is equal that is 20. In all, the total number of testing samples in testing set is 250.

Table II gives some samples used in our training process.

TABLE II
SOME SAMPLES IN OUR TRAINING SET

Patterns	Image samples			
AU18				
AU20				
AU25				
AU26				
AU28				

B. Results

In this paper, we recognize 5 single AUs, so according to the above description we can get 10 models in the training process. Since each mouth is normalized to 10*20, the total amount of Gabor feature set for one sample is 21860.T is set to 100 for all ω_i / ω_j pairs. Table III presents the optimization number of selected features for each ω_i / ω_j classifier.

TABLE III
OPTIMIZATION NUMBER OF SELECTED WEAK CLASSIFIERS PER AU PAIR

	AU18	AU20	AU25	AU26	AU28
AU18	-	5	7	11	6
AU20	5	-	7	3	9
AU25	7	7	-	5	9
AU26	11	3	5	-	9
AU28	6	9	9	9	-

From Table III, we can clearly observe that the final number of the selected weak classifiers is much smaller than the original dimension 100.

Table IV shows the comparative result of each pair.

TABLE IV
THE RECOGNITION RESULT FOR BINARY CLASSIFIERS

AU Pairs	Accuracy Rate(%)
AU20/AU18	87.5
AU20/AU28	75.0
AU20/AU25	90.0
AU20/AU26	95.0
AU18/AU26	85.0
AU18/AU28	87.5
AU18/AU25	85.0
AU26/AU28	90.0
AU26/AU25	97.5
AU28/AU25	85.0
Mean	87.0

From table IV, it is obviously observed that AU26/AU25 can be separated easily. For AU20/AU28 the effect is bad and the recognition accuracy is only 75%. There may be several reasons for this result. Firstly they are very similar. Secondly the solution is not good at separating the two type lip AUs. Thirdly, the samples are not good or typical, because it is difficult for the people who are not an expert in emotion behavior to present AU20/AU28.

It is said that over 7000 action unit combinations have been reported. For some AUs pairs, since it is difficult to classify, it seems a good idea to incorporate them into one AU if their meanings are similar. Or we can also try different classification methods for different pairs.

Table V gives the average identification rate for each class by Adaboost-SVM.

TABLE V
THE RECOGNITION RESULT FOR EACH CLASS

AU Pairs	Accuracy Rate(%)
AU18	85
AU20	95
AU25	80
AU26	80
AU28	75

The mean recognition results on testing set are shown in Table VI.

TABLE VI
RECOGNITION RESULTS FOR TESTING DATABASE

	Ad boost-SVM	SVM
Accuracy Rate (%)	83.0	79.0
Speed (s)	0.04	0.45

From Table VI, we can apparently see that not only the recognition rate based on Adaboost-SVM for testing set is much higher compared to SVM solution only, but also the speed for former is much faster than the latter. The reasons may be as following:

(1) During training process, Adaboost-SVM focuses on incorrectly classified samples which can make learning machines more adapted to incorrectly classified samples. Thus the recognition rate is improved.

(2) Adaboost-SVM algorithm selects a small set of features with low error rates. These selected features are used for identification, and then through integration we get the final label of an unknown sample. Therefore, our system not only avoids dimension disaster, but also enhances the robustness and the recognition rate. Moreover, only selected features are used for pattern prediction, so the recognition speed can be largely improved. It should be noted that the speed doesn't include the time for face detection and facial key point localization.

Table VII presents the confused matrix for Ababoost-SVM framework with the selected weak classifiers shown in Table III.

Table VII
CONFUSED MATRIX OF AU RECOGNITION BASED ON ADABOOST-SVM

	AU18	AU20	AU25	AU26	AU28
AU18	17	0	1	2	0
AU20	0	19	0	1	0
AU25	2	2	16	0	0
AU26	3	0	1	16	0
AU28	1	2	0	2	15

Table VII shows that after voting, the average recognition rate for the whole is improved from 79% to 83%. And the recognition performance for AU20 and AU28 is also enhanced by voting mechanism.

V. CONCLUSION

In this paper, we present an Adaboost-SVM+Gabor framework to classify lip AUs. Because the dimension of Gabor features is very high, this classification algorithm gets a strong classifier through integration of SVM weak classifiers by boost process. Since the important parameter T is set in advance, an extra number optimization step is added after traditional boost. The experimental results demonstrate the effect of our proposed system. The combination of Adaboost and SVMs enhanced both the speed and accuracy of the system. In future, more images will be added for training and testing and more lip AUs will be recognized.

ACKNOWLEDGMENT

This research is partially sponsored by the Natural Science Foundation of China (NSFC) under contract No. 61105120 and No.61170117.

REFERENCES

[1] Chao-Fa Chuang, Frank Y. Shih, "Recognizing facial action units using independent component analysis and support vector machine", *Pattern Recognition*, vol. 39, no. 9, pp. 1795-1798, 2006.

[2] P. Ekman W. Friesen, J. Hager, "The Facial Action Coding System: A Technique for the Measurement of Facial Movement", A Human Face, Salt Lake City, UT, 2002.

[3] Marian Stewart Bartlett, Gwen C. Littlewort, Mark G. Frank, et al, "Automatic Recognition of Facial Actions in Spontaneous Expressions", *Journal of Multimedia*, Vol. 1, No. 6, pp.22-35, 2006.

[4] Mahmoud Khademi, Mehran Safayani, Mohammad T. Manzuri-Shalmani, "Mutilinear Biased Discriminate Analysis: A novel method for facial action unit representation", In: *Proceedings of 16th Korea-Japan Joint Workshop on Frontiers of Computer Vision.* Hiroshima, Japan, pp. 221-225, 2010.

[5] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, et al, "The computer expression recognition toolbox (CERT)", In: *Proceedings of IEEE International Conference on Automatic & Gesture Recognition and Workshops*, pp. 298-305, 2011.

[6] Bihan Jiang, Valstar M.F., Pantic M., "Action unit detection using sparse appearance descriptors in space-time video volumes", In: *proceedings of IEEE International Conference on Automatic & Gesture Recognition and Workshops*, pp. 314-321, 2011.

[7] Yunfeng Zhu, Fernando De la Torre, Jeffrey F. Cohn, et al, "Dynamic cascades with bidirectional bootstrapping for spontaneous facial action unit detection", In: *Proceedings of International Conference on Affective Computing and Intelligent Interaction*, IEEE Computer Society, Los Alamitos, pp. 1-8, Sept 2009.

[8] Savran, A., Sankur, B., Bilge, M.T., "Facial action unit detection: 3D versus 2D modality", In: *Proceeding of 2010*

IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp.71-78, June 2010.

[9] Nusseck, M., Cunningham, D. W., Wallraven, C., Bülthoff, H. H., "The contribution of different facial regions to the recognition of conversational expressions", *Journal of Vision*, vol. 8, no. 8, pp. 1–23, 2008.

[10] Sander Koelstra, Maja Pantic, Ioannis Patras, "A dynamic texture based approach to recognition of facial actions and their temporal models", *IEEE Transaction on Pattern Analysis and Machine Intelligent*, vol. 32, no. 11, pp. 1940 - 1954 , 2010.

[11] P.Viola and M. Jones, "Rapid object detection uses a boosted cascade of simple features", In: *Proceedings of the IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*, vol.1, pp. 511-518, 2001.

[12] XiaoFeng Zhu, "*Improvement to face detection algorithm and its application to video"*, Master dissert, Department of Computer Science and Technology, FuDan University, Shang Hai, China, 2009.

[13] Wen Gao, Bo Cao, Shiguang Shan, et al, "The CAS-PEAL large-scale Chinese face database and baseline evaluations", *IEEE Transavtions on Systems, Man and Cybernetics- Part A: System and Humans*, Vol. 38, No. 1, pp. 149-161, 2008.

[14] T.F. Cootes, A. Hill, C. J. Taylor and J. Haslam, "The use of active shape models for locating structures in medical images", *Image and Vision Computing*, vol. 12, no. 6, pp. 355-366, 1994.

[15] YAO Wei, "*Research on facial expression recognition and synthesis"*, Master Thesis, Department of Computer Science and Technology, Nanjing University, Jiang Su, China, 2009.

[16] Wang Xianmei, Deng Ti, Liang Lingyan, Wang Zhiliang, "A novel smart household control system by computer vision", *Advanced Materials Research*, vol. 179-180, pp. 264-269, 2011.

[17] Valiant L.G., "A theory of the learnable". *Communications of the ACM*, vol. 27, no. 11, pp.1134-1142, 1984.

[18] Freund Y., Schapire R.E.A., "Decision-Theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, vol. 55, no.1, pp. 119-139, 1997.

Xianmei Wang received the M.S degree and Phd degree in control theory and application from University of Science and Technology Beijing, Beijing, China in 1999 and 2007 respectively. Her research covers machine learning, image processing and computer vision.

She is currently an Associate Professor with School of Computer and Communication Engineering at the University of Science and Technology Beijing, China. Her current research is focusing on facial image analysis-related research topics, including age estimation and mental states inference**.**

Yuyu Liang received the B.S. degree in electronic and information engineering from University of Science and Technology Beijing, Beijing, China in 2011. She is currently working toward the M.S degree in electronic science and technology, at School of Computer and Communication Engineering, University of Science and Technology Beijing.

Between 2009 and 2010, she joined in Robot Team in University of Science and Technology, Beijing. Her research interests lie in the areas of intelligent information processing, human-computer interaction. She is currently with the laboratory for artificial psychology and digital human. Her current research is facial AUs recognition.

Xiujie Zhao received the B.S. degree in electronic and information engineering from University of Science and Technology Beijing, Beijing, China in 2010. She is currently working toward the M.S degree at School of Computer and Telecommunication Engineering, University of Science and Technology Beijing.

In the spring of 2011, she went to Chaoyang University of Technology Taiwan as an exchange student for 4 months. Her research interests lie in the areas of human-computer interaction. She is currently with the laboratory for artificial psychology and digital human. Her current research is facial features localization.

Zhiliang Wang received the B. S degree in industrial automation from Yanshan University, the M. S degree in electronic engineering and Electrical engineering and from Haerbin Institute of Technology in 1985 and 1988 respectively. In 1999, he was a postdoctoral research fellow with the Joint Development Laboratory in Zhejiang University. His research interest includes digital family, human and computer interface technology.

He is currently a professor with School of Computer and Communication Engineering at the University of Science and Technology Beijing. He has published several books and over 100 technical articles in journals and proceedings in the areas of HCI, digital family and artificial psychology.

Dr. Wang is the director of electrical and information engineering department and Artificial Affection Professional Committee of Chinese Association for Artificial Intelligence. He has received several awards, including the National second-class teaching award and China's Metallurgical Science and Technology Award.