Facial expression recognition via deep learning

Yadan Lv
School of Computer Science and Technology, Tianjin University
Tianjin 300072, China
Email: yadan_lv@163.com

Zhiyong Feng
School of Computer Science and Technology, Tianjin University
Tianjin 300072, China
Email: zyfeng@tju.edu.cn

Chao Xu*
School of Computer Software, Tianjin University
Tianjin 300072, China
Email: xuchao@tju.edu.cn

***Abstract*—This paper mainly studies facial expression recognition with the components by face parsing (FP). Considering the disadvantage that different parts of face contain different amount of information for facial expression and the weighted function are not the same for different faces, an idea is proposed to recognize facial expression using components which are active in expression disclosure. The face parsing detectors are trained via deep belief network and tuned by logistic regression. The detectors first detect face, and then detect nose, eyes and mouth hierarchically. A deep architecture pretrained with stacked autoencoder is applied to facial expression recognition with the concentrated features of detected components. The parsing components remove the redundant information in expression recognition, and images don't need to be aligned or any other artificial treatment. Experimental results on the Japanese Female Facial Expression database and extended Cohn-Kanade dataset outperform other methods and show the effectiveness and robustness of this algorithm.**

Keywords*—*face parse, Deep Belief Network, Stacked Autoencoder, expression recognition, local feature

I. Introduction

Facial expression plays significant roles in human communication and social interaction, since it is closely related to psychological activities. Facial expression analysis is a challenging problem, and impacts important applications in many areas such as human-computer interaction and data-driven animation. Though much progress has been made, recognizing facial expression with high accuracy remains difficult due to its complexity and variability.

Deriving an effective facial representation from original face image is a vital step for successful facial expression recognition [1][2][3]. The optimal features should minimize within-class variations of expression whilst maximize between-class variations. If features are inadequate, even the best classifier could fail to achieve better performance. Two types of approaches have been proposed to extract facial features for expression recognition: geometric feature-based method and appearance-based method. In geometric feature-based method, shape and location of facial components are considered. Geometric relationships between these components are used to form a feature vector. M.Lyons et al. [4] used geometric positions of 34 fiducial points as facial features to represent facial image. In image sequences, the facial movements can be qualified by measuring geometrical displacement of facial feature points between current frame and initial frame. Valstar et al. [5] presented AU detection with classifying features calculated from tracked fiducial facial points. The appearance-based method extracts features by applying an image filter or filter banks on the whole face or some specific regions of the face such as principal component analysis (PCA), independent component analysis (ICA) [6][7] and Gabor wavelet [8][9]. Among them Gabor-wavelet representations [10] have been widely adopted in face image analysis [11][12].

Different techniques have been proposed for facial expression classification, such as Neural Network [11][13], Support Vector Machine (SVM) [14], Bayesian Network (BN) [15] and rule-based classifiers [16][17]. In Lyons et al.' work [12], the principle components of the feature vectors from training images were analyzed by Linear Discriminant Analysis (LDA) to form discriminant vectors, and facial image classification was performed by projecting the input vector of a testing image along the discriminant vectors. Cohen et al. compared different Bayes classifiers [15], and Gaussian Tree-Augmented-Naïve (TAN) Bayes classifiers performed best. Bartlett et al. [14] performed systematic comparison of different techniques including AdaBoost, SVM and LDA for facial expression recognition. Best results were obtained by selecting a subset of Gabor filters using AdaBoost and then training SVM on the outputs of the selected filters.

Different facial areas contain different amount of facial expression information. Eyes and mouth contain more information than forehead and cheek. Qiao et al. [18] proposed to assign different weights for each part, and the determination of weighted function is the key. However, weighted functions are not the same for different faces, and the shape of the weighted area is based on the facial shape. In our work, the component detectors are generatively trained with deep belief network (DBN), and discriminatively tuned by logistic regression. Then the features of parsed components including eyes, mouth are concentrated for expression recognition. Each feature dimension is treated equally. The main contributions of our work can be summarized as follows:

1) To the best of our knowledge, we are the first to recognize the expression with only facial components;
2) Treat the feature of parsed components equally to avoid adjusting the weighted function and weighted area on various faces;
3) Parse the face via deep belief network, the images can be used directly without preprocessing such as illumination compensation, face alignment.

II. Related work

Studies in psychology show that facial features of expressions are located around mouth, nose and eyes, and their locations are essential for categorizing facial expressions. Kotsia

978-1-4799-5711-8/14/$31.00 ©2014 IEEE

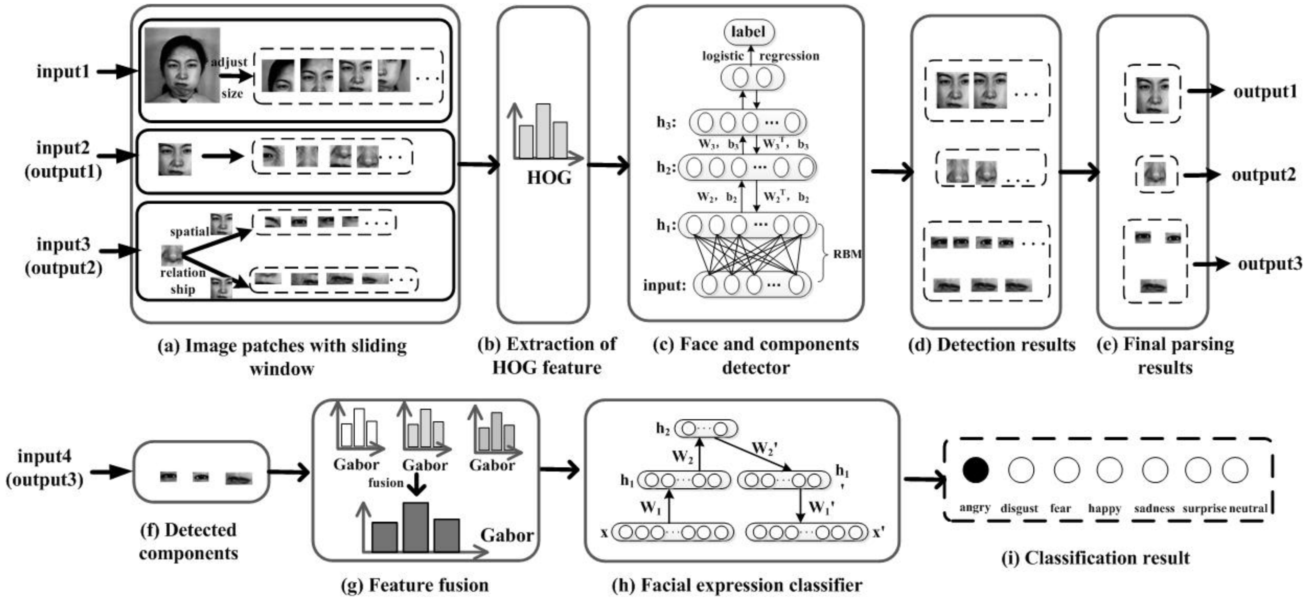

Fig. 1: Facial expression recognition via deep learning

et al. [19] have performed experiments to study how partial occlusion affects the classification of facial expressions. The results demonstrate that left/right facial region occlusion does not affect the recognition accuracy rate, indicating that both facial regions possess similar discriminate information. The results show that mouth and eyes occlusion, in general, causes great decrease in facial expression recognition, especially the mouth.

PCA is a classic feature extraction algorithm, which is widely used in expression recognition. However, with the goal of minimize the sum of each feature reconstruction error in Euclidean distance, the image features are treated indiscriminately. In fact, different areas of face contain different amount of facial expression information, and researchers have proposed improved method to overcome the defects above. Qiao et al. [18] described Weighted Principal component analysis (WPCA), which enhanced the critical features for expression recognition and assigned different weights for each part to strengthen the separability of the image. “Feature function” - which has big intermediate values and small peripheral values, assigns a weight for each feature. To enhance the weight of each part which contains significant recognition information, “multicenter feature function” was put forward. The ordinary feature area of the weighted function was circular. In practical application, the bidirectional scale around the weighted center should be adjustable considering the shape of different parts are not the same. However, these methods assigned weights to different face parts empirically, thus it lacked statistical support for weights settings. Lin et al. [20] learned active facial patches for expression analysis. If too many patches are selected, the performance goes down slightly and fluctuates. It means that only some patches are discriminative for the expressions. When some patches with little importance are selected, they will introduce some noises and influence the discriminative power of the patches. The selected patches are basically around the areas of mouth, eye, and eyebrows, which are consistent with AU-based analysis in FACS [21].

Deep belief network proposed by Hinton [22] provides effective method for approximate reasoning, and efficient greedy algorithms for learning and approximate inference have allowed these models to be applied successfully in many application domains such as recognition [23] and denoising [24]. At present, facial parsing [25] based on deep belief network has successfully applied in face alignment, facial keypoints extraction and face sketching.

In our work, expression recognition based on hierarchical face parsing is proposed. Deep belief networks are trained as detectors. They first detect the face, and then hierarchically detect the eyes, nose and mouth in this area which are used for expression recognition. Stacked autoencoder is employed as classifier to recognize the expression. Experimental results strongly support our algorithm and show the validity of it.

III. Facial expression recognition via deep learning

The crux of facial expression recognition using components is to extract each component accurately. As the labeled data is limited in fact, learning the model is essential. Deep belief networks are probabilistic generative models that contain many layers of hidden variables, in which each layer captures strong high-order correlations between the activities of hidden features in the layer below. In this paper, DBN is employed to establish strong correlations between images and shapes of each component by estimating the label maps directly from the detected image patches. Sec.A gives an overview of the algorithm. After that, we describe the learning of detectors in sec.B and the learning of classifier in sec.C.

A. General framework

Fig.1 describes basic framework of facial expression recognition via deep learning. In an image, adjust the image size first, and divide the image into several patches with sliding

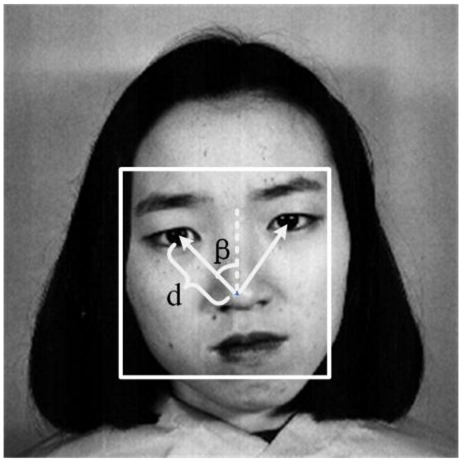

Fig. 2: Spatial relationship of face (d represents the distance from eye to nose, β represents the angle between the vertical dashed line and the solid line from nose to one eye).

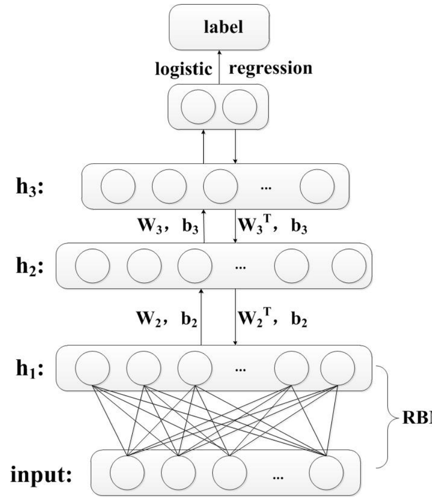

Fig. 3: The learning process of detectors

window (Fig.1(a)). Extract HOG features [26]in each window as the input of deep belief network for recognition (Fig.1(c)). After the training of face detector, face patches (Fig.1(d)) will be detected and the best patches are selected(Fig.1(e)) for future process. Then the detectors parse eyes, nose and mouth in face region. All the detectors are trained by deep belief network and tuned by logistic regression. While detecting eyes and mouth, their spatial relationship with nose is used to narrow the detection range (Fig.2). Finally, concentrate the Gabor feature of eyes and mouth (Fig.1(g)), and treat each dimension indiscriminately. A deep architecture pretrained with stacked autoencoder (Fig.1(h)) is employed as the classifier for expression recognition (Fig.1(i)).

B. Learn detectors

This paper models the detectors with deep belief network, which is probabilistic model contains some hidden variables. It consists of several Restricted Boltzmann Machine (RBM) [27] layers, the connection between the network layers, but there is no connection between the units inside the layer. Restricted Boltzmann Machines are undirected generative models that use a layer of hidden variables to model a distribution over visible variables. The bipartite structure contains two binary random nodes: the visual layer nodes $v \in \{0,1\}^{N_v}$, hidden layer nodes $h \in \{0,1\}^{N_h}$. For all nodes, the energy function is defined as:

$$E_{RBM}(v,h;\theta) = -\sum_{i}^{N_v} a_i v_i - \sum_{j}^{N_h} b_j h_j - \sum_{i,j}^{N_v,N_h} W_{ij} v_i h_j \quad (1)$$

where $\theta = \{W, a, b\}$ is model parameter. The joint distribution of visual layer and hidden layer is:

$$p(v,h;\theta) = \frac{p^*(v,h)}{Z(\theta)} = \frac{exp^{-E(v,h)}}{Z(\theta)} \quad (2)$$

where $p^*(\cdot)$ is non-standard probability distribution, $Z(\theta) = \sum_{v,h} exp^{-E(v,h)}$ is the normalized parameter.

To train the detectors, we get the weights by layer-wise pre-train first. Given an image patch I_b, the extracted feature f_b is the input of DBN, and the visual layer generate a vector v, that is $v = f_b$, then pass the value to the hidden layer,

$$p(h_j = 1 \mid v) = sigm(b_j + \sum_i v_i w_{ij}) \quad (3)$$

In turn, select the visual layer input at random to reconstruct the original input,

$$p(v_i = 1 \mid h) = sigm(a_i + \sum_j h_j w_{ij}) \quad (4)$$

At last, the new visual neuron activation unit transfers to reconstruct hidden layer activation unit. The correlation difference between hidden layer activation unit and visual input layer is the main basis of weight updating.

We construct 3-layers deep architectures with 2 outputs. The weights in the top two layers of DBN connect together, and the output of lower levels will provide a clue or reference to the top floor, which link it to the memory, as shown in Fig.3. After the pre-train, parameters are tuned with logistic regression discriminatively.

C. Learn expression classifier

In this paper, we model the classifier with a deep architecture of stacked autoencoder (SAE). Stacked autoencoder introduced by Bengio et al. [28] is a deep network consisting of an autoencoding neural network in each layer. In the single-layer case (Fig.4(a)), in response to an input pattern $x \in R^n$, the hidden layer h_1 is computed as:

$$h_1 = \tanh(W_1 x + b_1) \quad (5)$$

where $h_1 \in R^m$ is the vector of neuron activations, $W_1 \in R^{m\times n}$ is a weight matrix, $b_1 \in R^m$ is a bias vector, and $\tanh$ is the hyperbolic tangent applied componentwise. The network output is then computed as:

$$x' = \tanh(W_1' h_1 + b_2) \quad (6)$$

where $x' \in R^n$ is a vector of output values, $W_1' \in R^{n\times m}$ is a weight matrix, and $b_2 \in R^n$ is a bias vector. Given a set of component features as the input patterns $x^{(i)}, i = 1, \cdots, p$, the

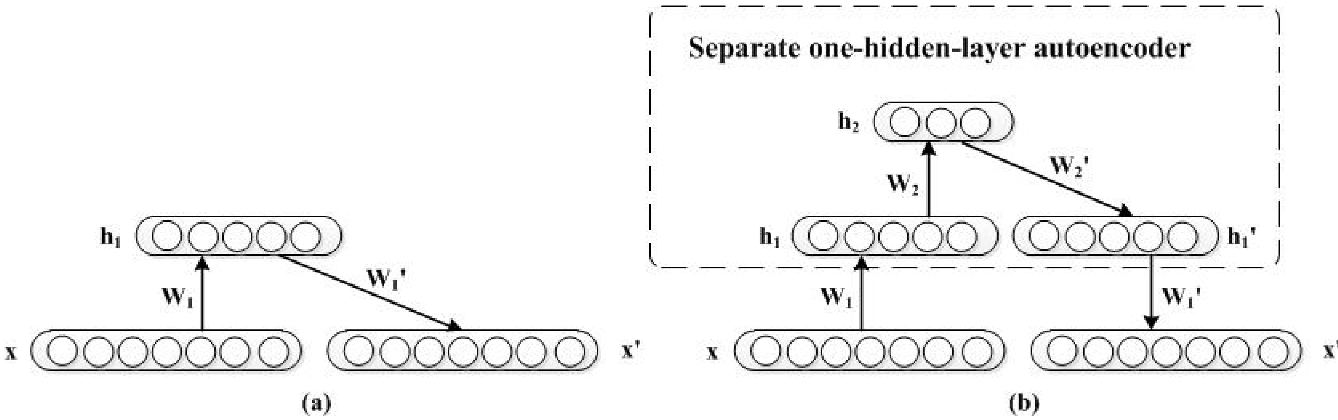

Fig. 4: Stacked Autoencoder

weight matrices W_1 and W_1' are adapted using backpropagation to minimize the reconstruction error $\sum_{i=1}^{p} \|x^{(i)} - x'^{(i)}\|^2$.

Stacked autoencoder models are trained layerwise by learning an autoencoder that reconstructs the previous layer (Fig.4(b)). A separate one-hidden-layer network consisting of input layer h_1, output layer h_1' and hidden layer h_2 is stacked onto the existing autoencoder. Since the size and the value of input and output layers of an autoencoder are the same, output layer h_1' is the same as input layer h_1. Hidden layer h_2 is a new hidden layer added onto the autoencoder. The weights of W_2 and W_2' can be trained by using backpropagation.

After training a stack of encoders as explained in Fig.4, an output layer is added on the top of the stack. The classifier for expression recognition is a 3 hidden layer neural network pretrained with stacked autoencoders. The parameters of all layers are simultaneously fine-tuned using a gradient-based procedure to converge to a global minimum.

IV. Experiment

We employ the facial components for facial expression recognition, and study its performance on Japanese Female Facial Expression (JAFFE) database [29] and extended Cohn-Kanade (CK+) database [30], which are widely used for facial expression recognition algorithms. The experiment carries out a 7-fold cross-validation scheme where each dataset is randomly partitioned into seven groups separately. Six groups are used as a training dataset to train the classifiers, while the remaining group is used as testing dataset. The above process repeats seven times, and the average recognition rate is calculated.

A. Data

The JAFFE database contains 213 images of female facial expression expressed by 10 subjects. Each image has a resolution of 256×256 pixels with almost the same number of image for each categories of expression. Each person has seven types of facial expressions: angry, disgust, fear, happy, neutral, sadness, and surprise.

The extended Cohn-Kanade dataset contains facial expression from 210 adults in which 69% are female, 81% are Euro-American, 13% are Afro-American and 6% are from other groups. Participants are 18 to 50 years of age. A total of 7 expressions are labeled in the dataset, including anger, contempt, disgust, fear, happy, sadness and surprise.

B. Experimental results and analysis

A characteristic of deep network is that the number of neurons in training set and testing set is fixed. The detected conponent regions will be the input of next network, so these regions should be scaled to the same size. In order to minimize human intervention, this paper proposes the face scaling algorithm with horizontal and vertical gradient to extract face region. In JAFFE, the face region, nose, eyes and mouth are normalized to 100×100, 40×40, 30×20, 55×25 respectively, and the standardized sizes in CK+ are 200×200, 70×70, 80×40, 100×90. We construct four 3-layers deep architectures to model face, nose, eyes and mouth detectors in each database, and then tune four DBNs with 2 outputs at the top layer respectively. We extract the HOG feature as the input of DBNs. As the high dimension of HOG feature, the standardized sizes in CK+ are reduced to half of the original. The number of nodes of each hidden layer is 100, 1000, 10000, from the highest layer to the lowest one; and the visual input is 72900, which is the dimension of HOG feature in both two databases.

During the training of the expression classifier, this paper extracts Gabor feature of the parsed components including eyes and mouth, and then concentrate the features as the input layer. We train the classifier for expression recognition by using a 3 hidden layer neural network pretrained with stacking autoencoders. The numbers of nodes are 100, 500, 2000, and 70, 400, 1200, from the highest hidden layer to the lowest one in JAFFE and CK+ respectively; and the numbers of nodes in the visual layers are 5150 and 3850. The average recognition rate with FP in JAFFE is 90.47% using 7-fold cross-validation and 91.11% in CK+ database.

For further evaluation of the results, we compare them with other methods. TABLE I shows the comparison in JAFFE database. [31] and [32] all divided facial image into several sub-regions. Each region is assigned a weight, the larger a weight is, the more important is the region. Taskeed et al. [32] assigned for the weighted χ^2 measure and described a new local facial descriptor based on Local Directional

TABLE I: Comparison in JAFFE

Method	Accuracy(%)
Boosted-LBP + SVM [31]	81
LDP + template matching [32]	82.6
LDP + SVM [32]	85.4
Gabor + LVQ [9]	87.51
Gabor + FLD [33]	86.1
Log-Gabor + FLD [33]	85.72
FP + SAE	**90.47**

Pattern (LDP). The recognition accuracy of LDP+SVM was 85.4%. Shishir et al. [9] conducted the experiment with Gabor and Learning Vector Quantization (LVQ). A graphical user interface application has been developed to graphically select the 34 fiducial points in the image where Gabor filter responses were sampled, and the recognition rate was 87.51%. Nectarios et al. [33] proposed feature vectors composed of facial points convolved with Gabor and Log-Gabor filters, while the recognition rate were 86.1% and 85.72% respectively.

TABLE II: Comparison in CK+

Method	Accuracy(%)
LBP + template matching [31]	79.1
LBP + SVM (RBF) [31]	88.9
Gabor + SVM (RBF) [31]	86.8
LDP + template matching [32]	86.9
CPL + SVM [20]	88.42
CSPL + SVM [20]	89.89
ITBN [34]	86.3
BDBN [35]	96.7
FP + SAE	**91.11**

TABLE II compares our result with other methods in CK+ database. Among all the works we find, [20] is the most similar to FP, in which CPL stands for their methods that only use common patches that are active ones for all expressions, CSPL is their method that use common and specific patches. Our result outperforms theirs by 1.22% at least. Wang et al. [34] modeled the facial expression as a complex activity that consists of temporally overlapping or sequential primitive facial event, and proposed the Interval Temporal Bayesian Network to capture these relations. They got the recognition rate of 86.3%. Ping Liu et al. [35] presented a novel Boosted Deep Belief Network for performing feature learning, feature selection and classifier construction iteratively in a unified loopy framework. The average classification rate was 96.7%, however, the experiment was conducted in 6 expressions and ours was in 7 expressions. Also, it needed to construct 80 DBNs and ours trained 4 DBNs and 1 stacked autoencoder. The computational complexity of our work is relatively lower. Overall we can see that expression recognition with parsed components has better performance and lower computional complexity. The effectiveness of the method is verified by the experimental results.

V. Conclusion

In this paper, expression recognition with parsed components is proposed. Deep belief networks parse the components precisely, and the parsing components remove the redundant information in expression recognition. Different from previous work, we treat the feature of parsed components equally to avoid adjusting the weighted function on various faces. The effectiveness of these components is evaluated by deep architectures pretrained with stacked autoencoders. Experiments on the JAFFE and CK+ database demonstrate the power of the proposed method in expression recognition and show that eyes and mouth can discriminate the expressions generally.

VI. Acknowledgement

This work is partially supported by National Natural Science Foundation of China (No. 61304262).

References

[1] C.-D. Caleanu, *Face expression recognition: A brief overview of the last decade*, in 2013 IEEE 8th International Symposium on Applied Computational Intelligence and Informatics (SACI), 2013, pp. 157-161.

[2] M. A. Turk and A. P. Pentland, *Face recognition using eigenfaces*, in Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.

[3] A. Geetha, V. Ramalingam, S. Palanivel, and B. Palaniappan, *Facial expression recognition - A real time approach*, Expert Syst. Appl., vol. 36, no. 1, pp. 303-308, Jan. 2009.

[4] M. Lyons, M. Schuster, and S. Akamatsu, *Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron*, in Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 454-459.

[5] M. F. Valstar, I. Patras, and M. Pantic, *Facial Action Unit Detection using Probabilistic Actively Learned Support Vector Machines on Tracked Facial Point Data*, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, 2005, vol. 3, pp. 76-76.

[6] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, *Face recognition by independent component analysis*, IEEE Trans. Neural Netw., vol. 13, no. 6, pp. 1450-64, Jan. 2002.

[7] C.-F. Chuang and F. Y. Shih, *Recognizing facial action units using independent component analysis and support vector machine*, Pattern Recognit., vol. 39, no. 9, pp. 1795-1798, Sep. 2006.

[8] Weifeng Liu and Zengfu Wang, *Facial Expression Recognition Based on Fusion of Multiple Gabor Features*, in 18th International Conference on Pattern Recognition (ICPR06), 2006, vol. 3, pp. 536-539.

[9] S. Bashyal and G. K. Venayagamoorthy, *Recognition of facial expressions using Gabor wavelets and learning vector quantization*, Eng. Appl. Artif. Intell., vol. 21, no. 7, pp. 1056-1064, Oct. 2008.

[10] S. M. Lajevardi and M. Lech, *Averaged Gabor Filter Features for Facial Expression Recognition*, in 2008 Digital Image Computing: Techniques and Applications, 2008, pp. 71-76.

[11] Ying-Li Tian, *Evaluation of Face Resolution for Expression Analysis*, p. 82, 2004.

[12] M. J. Lyons, J. Budynek, and S. Akamatsu, *Automatic classification of single facial images*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 12, pp. 1357-1362, 1999.

[13] C. Padgett and G.W. Cottrell, *Representing face images for emotion classification*, Advances in Neural Information Processing Systems,vol. 9. Cambridge, MA: MIT Press, 1997, pp. 894

[14] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, *Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior*, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, vol. 2, pp. 568-573.

[15] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, *Facial expression recognition from video sequences: temporal and static modeling*, Comput. Vis. Image Underst., vol. 91, no. 1-2, pp. 160-187, Jul. 2003.

[16] M. Pantic and L. J. Rothkrantz, *Expert system for automatic analysis of facial expressions*, Image Vis. Comput., vol. 18, no. 11, pp. 881-905, Aug. 2000.

[17] M. Pantic and I. Patras, *Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences*, IEEE Trans. Syst. Man Cybern. Part B, vol. 36, no. 2, pp. 433-449, Apr. 2006.

[18] Qiqo yu, Huang xi yue, Chai yi, Deng jin cheng, and Chen hong yu, *Facial recognition based on Weighted Principle Component Analysis*, Chinese Journal of Chongqing University: Science Edition, vol. 27, no. 3, pp. 28-31, 2004.

[19] I. Kotsia, I. Buciu, and I. Pitas, *An analysis of facial expression recognition under partial facial image occlusion*, Image Vis. Comput., vol. 26, no. 7, pp. 1052-1067, Jul. 2008.

[20] D. N. Metaxas, *Learning active facial patches for expression analysis*, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2562-2569.

[21] P. Ekman and W. V. Friesen. *Facial action coding system*, Consulting Psychologists Press, 1, 1978. 1,5

[22] G. E. Hinton, S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets*, Neural Comput., vol. 18, no. 7, pp. 1527-54, Jul. 2006.

[23] M. Ranzato, J. Susskind, V. Mnih, and G. Hinton, *On deep generative models with applications to recognition*, in CVPR 2011, 2011, pp. 2857-2864.

[24] R. Salakhutdinov and G. Hinton, *Robust Boltzmann Machines for recognition and denoising*, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2264-2271.

[25] Ping Luo, Xiaogang Wang, and Xiaoou Tang, *Hierarchical face parsing via deep learning*, in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2480-2487.

[26] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05), 2005, vol. 1, pp. 886C893.

[27] H. Larochelle and Y. Bengio, *Classification using discriminative restricted Boltzmann machines*, in Proceedings of the 25th international conference on Machine learning - ICML '08, 2008, pp. 536-543.

[28] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. *Greedy layer-wise training of deep networks*, in NIPS, 2007.

[29] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, *Coding facial expressions with Gabor wavelets*, in Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 200-205.

[30] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*, in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 94-101.

[31] 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.

[32] T Jabid, M Kabir, O Chae, *Robust facial expression recognition based on local directional Pattern*, ETRI J. 2010, 32, 784-794.

[33] N. Rose, *Facial Expression Classification using Gabor and Log-Gabor Filters*, in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006, pp. 346-350.

[34] Z. Wang, S. Wang, and Q. Ji, *Capturing Complex Spatio-temporal Relations among Facial Muscles for Facial Expression Recognition*, in 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3422-3429.

[35] P. Liu, S. Han, Z. Meng, and Y. Tong, *Facial Expression Recognition via a Boosted Deep Belief Network*, in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1805C1812.