ORIGINAL PAPER

Posed and spontaneous expression recognition through modeling their spatial patterns

Shangfei Wang · Chongliang Wu · Menghua He · Jun Wang · Qiang Ji

Abstract This paper presents a new method to recognize posed and spontaneous expressions through modeling their spatial patterns. Gender and expression categories are employed as privileged information to further improve the recognition. The proposed approach includes three steps. First, geometric features about facial shape and Action Unit variations are extracted from the differences between apex and onset facial images to capture the spatial facial variation. Second, statistical hypothesis testings are conducted to explore the differences between posed and spontaneous expressions using the defined geometric features from three aspects: all samples, samples given the gender information, and samples given expression categories. Third, several Bayesian networks are built to capture posed and spontaneous spatial facial patterns respectively given gender and expression categories. The statistical analysis results on the USTC-NVIE and SPOS databases both demonstrate the effectiveness of the proposed geometric features. The recognition results on the USTC-NVIE database indicate that the privileged information of gender and expression can help model the spatial patterns caused by posed and spontaneous expressions. The recognition results on both databases outperform those of the state of the art.

Keywords Posed and spontaneous expression recognition · Spatial pattern · Priviledged information · Gender · Expression categories

S. Wang (✉) · C. Wu · M. He · J. Wang
Key Lab of Computing and Communication Software of Anhui
Province, School of Computer Science and Technology,
University of Science and Technology of China, Hefei 230027,
Anhui, People's Republic of China
e-mail: sfwang@ustc.edu.cn

C. Wu
e-mail: clwzkd@mail.ustc.edu.cn

M. He
e-mail: hemh@mail.ustc.edu.cn

J. Wang
e-mail: junwong@mail.ustc.edu.cn

Q. Ji
Department of Electrical, Computer, and Systems Engineering,
Rensselaer Polytechnic Institute, Troy, NY 12180, USA
e-mail: qji@ecse.rpi.edu

1 Introduction

Many areas, including real-life human-robot communications, security, and healthcare, would benefit from a more accurate method to determine when a facial expression is spontaneous versus when it is posed. For example, police can detect deceptive facial expressions for lie detection. Doctors can make a good diagnosis by recognizing patients' genuine pain. Robots can infer users' true emotions by differentiating their posed and spontaneous expressions. Spontaneous expressions reflect one's true emotion, while posed expressions disguise one's true emotion. Till now, only a few works have reported to recognize posed and spontaneous expressions, in spite of its wide application prospects. Most of them only consider one specific expression, such as smile or pain. Only two works [1,2] take all six basic expressions (i.e. happiness, disgust, fear, surprise, sadness, and anger) into account to distinguish posed and spontaneous expressions. The current works recognize posed and spontaneous expressions using different classifiers. Few works explicitly model the spatial and temporal patterns embedded in posed and spontaneous expressions. Furthermore, no reported studies consider gender differences between posed and spontaneous facial expressions, although previous research indicates that

gender differences in facial expression manifestations exist [3,4].

In this paper, we propose a novel method to model spontaneous and posed muscle variations using Bayesian networks (BN). We adopt gender and expression categories as privileged information, which is defined as information that is available only during training but not testing [5,6], to better capture the different spatial facial patterns embedded in posed and spontaneous facial expressions. First, we define 19 geometric features related to AU variations between apex and onset facial images to capture spatial facial variation. Second, we conduct statistical analyses to investigate the effectiveness of these geometric features in differentiating posed and spontaneous expressions from three aspects: all samples without the gender and expression information, samples with the gender information, and samples with expression categories. Third, several BNs are constructed to capture spatial patterns embedded in posed and spontaneous expressions respectively without and with gender and expression categories. The hypothesis testing results on the USTC-NVIE and SPOS databases both demonstrate the effectiveness of the proposed features. The recognition results on both USTC-NVIE and SPOS databases show that the proposed methods outperform the state of the art methods. The experiential results further show that the privileged information of gender and expression can improve modeling the spatial patterns caused by posed and spontaneous expressions.

The outline of this paper is as follows: the related work of posed vs. spontaneous expressions analysis and recognition is described in Sect. 2; our approach of posed and spontaneous expression recognition using gender and expression as privileged information is described in Sect. 3; the experiments and results are revealed in Sect. 4; conclusion and future work are drew in Sect. 5.

2 Related work

Present nonverbal behavior research has shown that spontaneous facial expressions differ from posed ones in both spatial and temporal patterns. Spatial patterns mainly refer to the muscle movements. For example, a spontaneous smile involves contractions of both the zygomatic major and the orbicularis oculi, while a posed smile only involves the zygomatic major, without the orbicularis oculi [7]. The asymmetry of zygomatic major actions (AU 12) occurs more frequently in posed smiles than in spontaneous smiles [8]. Ekman et al. claimed that the absence of muscle movements is a good indicator to whether an expression is posed or spontaneous, since it is difficult to make voluntarily [9,10]. Temporal patterns involve the total duration, trajectory, amplitude, and speed of onset and offset. For example, posed expressions are usually longer than spontaneous expressions [10], the trajectory of

spontaneous expressions appears often smoother than that of posed expressions [10], and the onset of a posed expression is more abrupt than that of a spontaneous expression in most cases [11].

Inspired by the observations from nonverbal behavior research, researchers have begun to pay attention to posed and spontaneous recognition. Just as most nonverbal behavior research investigates only posed versus spontaneous smiles, most computer vision research also focuses on one expression, such as smile or pain. Cohn and Schmidt [12] are the first to distinguish posed and spontaneous expressions using machine learning method. They extracted amplitude, duration, and the ratio of amplitude to duration as the features, and adopted a linear discriminant as the classifier to recognize posed and spontaneous smile. Valstar [13] proposed to distinguish posed and spontaneous smile by fusing head, face, and shoulder modalities. They [14] also studied posed and spontaneous brow actions using velocity, duration, and the order of occurrence. Littlewort et al. [15] proposed to classify real vs. faked pain by detecting 20 facial action units and then putting them into a classifier. Dibeklioglu et al. [16] proposed to distinguish posed and spontaneous smile using the dynamics of eyelid, check, and lip corner movement. Seckington [7] proposed to use dynamic Bayesian network to model the temporal dynamics of expressions, and recognize posed and spontaneous smile.

To the best of our knowledge, only two works classify posed vs. spontaneous expressions for any expressions instead of classifying a specific expression. Pfister et al. [2] proposed a spatiotemporal local texture descriptor (CLBP-TOP) and a generic facial expression recognition framework to differentiate posed from spontaneous expressions from both visible and infrared images using SPOS database. Zhang et al. [1] investigated the performance of a machine vision system for discriminating posed vs. spontaneous versions of six basic expressions using SIFT and FAP features on the NVIE database.

Almost all the above research regards posed and spontaneous expression recognition as a binary classification problem. Few works explicitly model the spatial and temporal patterns. In this paper, we propose a new method to use BN to model the muscle variation of posed and spontaneous expressions. Furthermore, although there are two works recognizing posed and spontaneous expressions for several expression categories, instead of one specific expression, they have not investigated the effect of gender and expression categories on distinguishing posed and spontaneous expressions. Since previous research indicates that gender differences in facial expression manifestations exist, and different expressions have different spatial patterns [3,4]. Gender and expression categories may provide additional information for posed vs. spontaneous expression analyses and recognition. Thus, in this paper, we regard gender and expression categories as

the privileged information, and build models for posed and spontaneous expressions given gender and expression categories respectively using BN during training. During testing, the sample is assigned a label whose model has the maximum likelihood. Unlike the work in [4], which recognizes expression using gender as a middle-level representation, our paper employs gender and expression categories as privileged information to help construct a better posed vs. spontaneous expression classifier. The work in [4] first recognizes gender, and then recognizes expression. It sequentially combines different tasks. One issue with this sequential approach is that the error with gender recognition will propagate to subsequent expression recognition. While in our work, as privileged information, gender and expression categories are only available during training, we do not need estimate them during testing.

Compared with the related work, our contributions are:

1. We extend current posed and spontaneous expression classification methods from single expression category to multiple expression categories.
2. We incorporate gender and expression category as privileged information into posed and spontaneous expression classification.
3. We are among the first to explicitly model posed and spontaneous expressions by spatial facial patterns.

3 The proposed method

The framework of our proposed approach is shown in Fig. 1, including feature extraction, statistical analyzes for hypothesis testing, and posed or spontaneous expression modeling by BNs which include gender and expression categories as privileged information. The details are described as follows.

3.1 Data preprocessing and feature extraction

First, 29 feature points, as shown in Fig. 2, are automatically detected on the onset and apex expression frames using the algorithm introduced in [17]. The onset frame is the beginning of the onset phase, which is similar to the neutral frame here, and the apex frame is the most exaggerated expression frame during the apex phase. The centers of the eyes are determined as the 28th and 29th points. For geometric normalization, we rotate the face image to make the inter-ocular line horizontal and of fixed length, and change the positions of other facial points accordingly. Then, the facial region is resized to 100×100 using bicubic interpolation [18] and Anti-aliasing filter [19]. Through the face alignment and normalization, the facial features are robust to different subjects and to moderate face pose variation. After that, 19 geometric
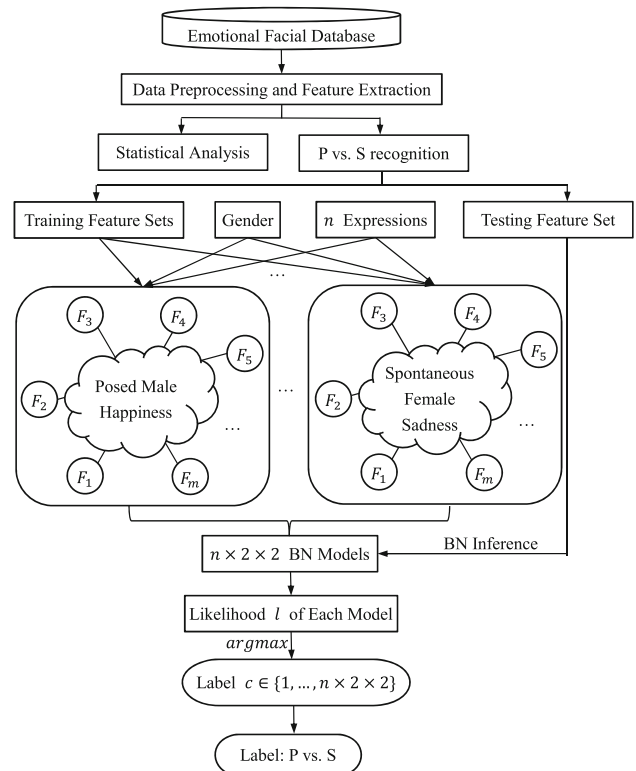

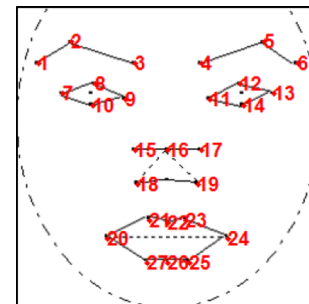**Fig. 1** The framework of our proposed method

Fig. 2 The distribution of feature points

features as discussed below are defined to capture the spatial facial pattern [20].

Since the eye brows and the mouth are salient facial subregions to expression, the displacement ratios of the lip in height, the corners of the mouth in width, and the height of the left and right eyebrows are defined according to Eqs. (1), (2), (3), and (4).

$$\text{HMouth} = \frac{\sum_{i=21}^{23} A_i(y) - \sum_{i=25}^{27} A_i(y)}{\sum_{i=21}^{23} O_i(y) - \sum_{i=25}^{27} O_i(y)} \tag{1}$$

$$\text{WMouth} = \frac{A_{20}(x) - A_{24}(x)}{O_{20}(x) - O_{24}(x)} \tag{2}$$

$$\text{LBrow} = \frac{A_2(y) - A_{28}(y)}{O_2(y) - O_{28}(y)} \tag{3}$$

Table 1 AU-related geometric features

Facial region	Name	Geometric feature	Related AU
Brow	Brow1	$\angle(P9, P7, P3) + \angle(P11, P13, P4)$	AU1
	Brow2	$(P4, P11) + (P3, P9)$	AU1, AU2, AU4
	Brow3	$(P2, P10)y + (P5, P14)y$	AU1, AU2, AU4
	Brow4	$(P3, P4)$	AU4
	Brow5	$((P9, P15) + (P11, P17))/2$	AU4
	Brow6	$\angle(P7, P9, P8) + \angle(P13, P11, P12)$	AU4
Lid	Lid1	$(P8, P10)y + (P12, P14)y$	AU5
	Lid2	$\angle(P7, P10, P9) + \angle(P13, P14, P11)$	AU5
Lip	Lip1	$(P22, P9)y + (P22, P11)y$	AU10
	Lip2	$|(P11, P24) - (P9, P20)|$	AU10
	Lip3	$(P22, (P18 + P19)/2)$	AU10, AU16
LipCorner	LipCorner1	$(P11, P24)$	AU12, AU15, AU20
	LipCorner2	$(P9, P20)$	AU12, AU15, AU20
	LipCorner3	$(P12, P14) + (P8, P10)$	AU12, AU15, AU20
Chin	Chin1	$(P26, (P18 + P19)/2)$	AU17

$$\text{RBrow} = \frac{A_5(y) - A_{29}(y)}{O_5(y) - O_{29}(y)} \tag{4}$$

where $(A_i(x), A_i(y))$ and $(O_i(x), O_i(y))$ are the coordinates of the ith feature point in the apex and onset frame, respectively.

In addition, we define 15 AU-related geometric features [21] and calculate their differences between the apex and onset frames to represent the variations of facial AUs. The relations between these features and AUs are listed in Table 1, where $Pi (i = 1, 2, \ldots, 27)$ represents the ith feature point. Supposing $P9$ is A, $P7$ is B, and $P3$ is C, then $\angle(P9, P7, P3)$ stands for the angle degree of $\angle ABC$; $(P4, P11)$ stands for the Euclidean distance between the 4th and 11th points; $(P2, P10)y$ stands for the distance between the 2nd and 10th points on the y axis.

3.2 Feature selection through statistical analysis

Given the 19 geometric features defined above, we conducted statistical tests to investigate the differences between posed and spontaneous expressions using the extracted features under three conditions: all samples without gender and expression information, samples with the gender information, and samples with expression categories.

In this section, the null hypothesis (H0) is that the median difference between posed and spontaneous facial expressions for each feature is zero. The alternative hypothesis (H1) is that the median difference between posed and spontaneous facial expressions for each feature is not zero [22]. In statistical significance testing, the p value is the probability that the null hypothesis is true [23]. We may reject H0 when the p value is less than the significant level. Such a result indicates that the observed result would be highly unlikely under H0.

If the posed and spontaneous samples are in pairs, which means that the samples are in the same expression category of the same subject, a Wilcoxon signed-rank test is used to analyze the differences between posed and spontaneous facial expressions for each feature. Wilcoxon signed-rank test is a nonparametric method and can be used to assess whether the population means of the paired samples' rank differ. In this case, the subject effect can be reduced. Otherwise a Kolmogorov–Smirnov test is adopted, which is one of the most useful and general nonparametric methods for comparing the distributions of two samples. Both tests belong to nonparametric tests, which are suitable for continuous variables and do not require the normality assumption, and the difference between the two tests is whether the samples are in pairs [24].

3.3 Posed and spontaneous expression modeling using Bayesian network

A Bayesian Network is adopted to capture the spatial patterns embedded in posed or spontaneous expression. A BN is a directed acyclic graph (DAG) that represents a joint probability distribution among a set of variables. In our work, each node of the BN is a geometric feature, and the links and their conditional probabilities capture the probabilistic dependencies among the selected features. Figure 3 shows the BN models using 15 selected significant features in Table 4 with gender as prior knowledge.

The BN learning consists of structure learning and parameter learning, respectively. The structure consists of the directed links among the nodes, while the parameters are the conditional probabilities of each node given its parents. The structure learning is to find a structure G that maximizes a

Fig. 3 Learned BN, representing probabilistic dependencies among features

(a) Posed Male

(b) Spontaneous Male

(c) Posed Female

(d) spontaneous Female
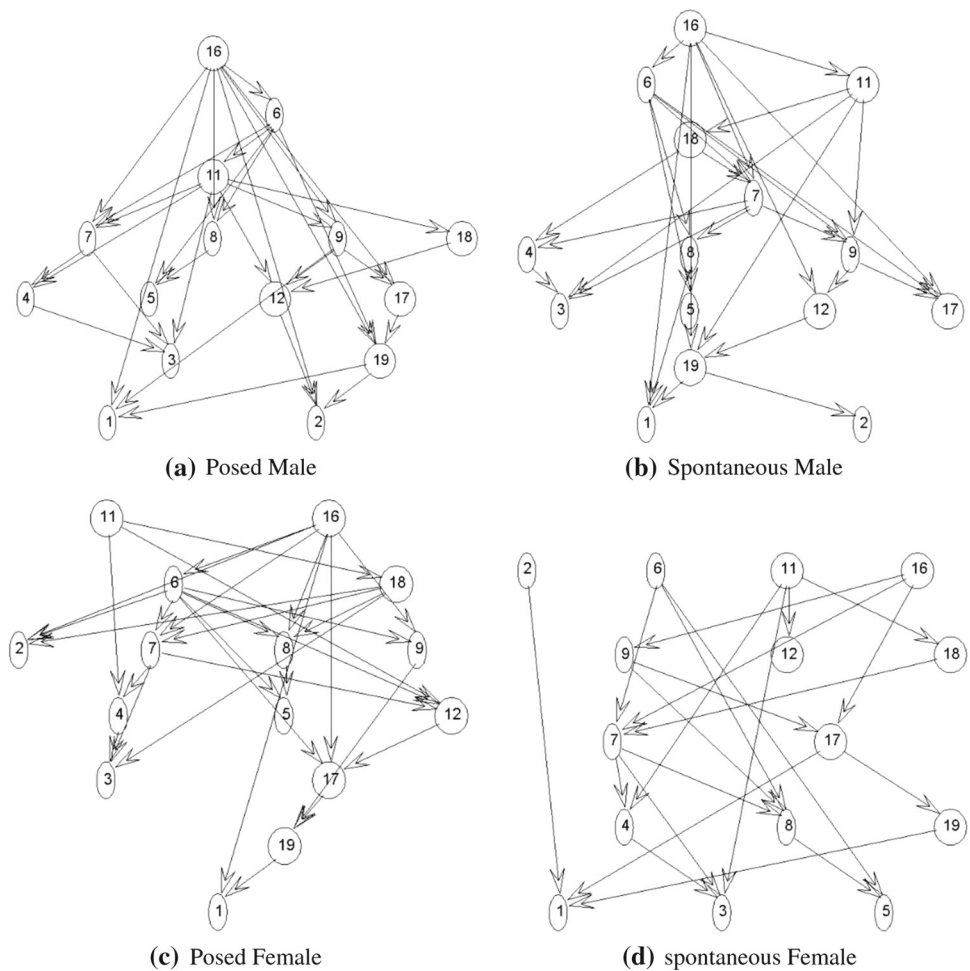
score function. In this work, we employ the Bayesian Information Criterion (BIC) score function which is defined as follows:

$$\text{Score}(G) = \max_{\theta} \log(p(DL|G, \theta)) - \frac{\text{Dim}_G}{2} \log m \quad (5)$$

where the first term is the log-likelihood function of parameters θ with respect to data DL and structure G, representing the fitness of the network to the data. And p represents the joint probability of data under BN model. The second term is a penalty relating to the complexity of the network, Dim_G is the number of independent parameters, and m indicates the number of samples in data. The K2 algorithm [25] is adopted to learn the BN structure. After the BN structure is constructed, parameters can be learned from the training data. Because complete training data are provided in this work, Maximum Likelihood Estimation (MLE) method is used to estimate the parameters. The algorithms of BN structure and parameter learning are well implemented in Bayes Net Toolbox (BNT) [26]. In our experiment, we employed the BNT directly. Since we do not have any prior knowledge about the order of the extracted features, the order of the BN nodes is determined randomly. The upper bound on the number of parents is 2, and we assume that the variables in BN follow Gaussian distribution. We randomly determined the order a few times, and the BN structures which produce the best recognition performance are shown in Fig. 3.

In this work, gender and expression categories are regarded as privileged information, thus $n \times 2 \times 2$ models $\Theta_c, c = 1, \ldots, n \times 2 \times 2$ are established during training, where n is the number of expression categories, the first 2 means gender categories, and the last 2 indicates posed and spontaneous models. After training, the learned BNs capture the muscle spatial movement pattern for posed and spontaneous expressions respectively given gender and expression categories.

During testing, the samples are classified into posed or spontaneous expressions according to

$$c^{\star} = \arg\max_{c \in [1,n]} \frac{P(E_T|\Theta_c)}{\text{Complexity}(M_c)}$$

$$= \arg\max_{c \in [1,n]} \frac{\prod_{i=1}^{19} P_c(F_i|pa(F_i))}{\text{Complexity}(M_c)}$$

Fig. 4 Image samples from USTC-NVIE and SPOS databases. In both **a** and **b**, images of the *first row* are the posed apex samples, and those of the *second row* are the spontaneous apex samples. Each row contains images from six different emotion categories, i.e. happiness, disgust, fear, surprise, anger, and sadness, respectively. Each column contains facial images of the same subject
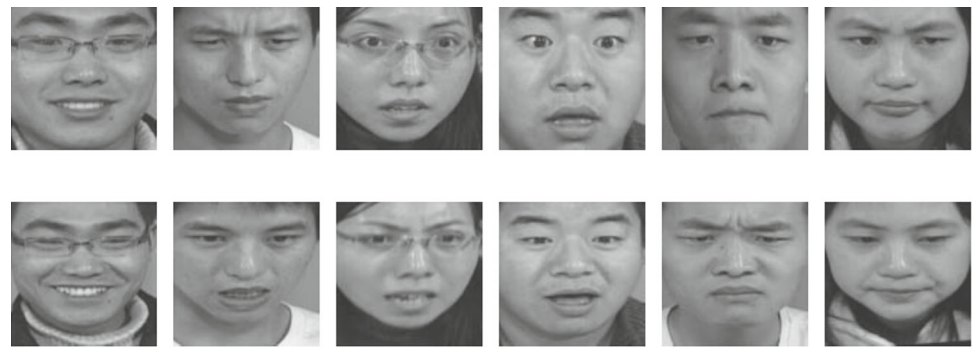
(a) samples from USTC-NVIE database
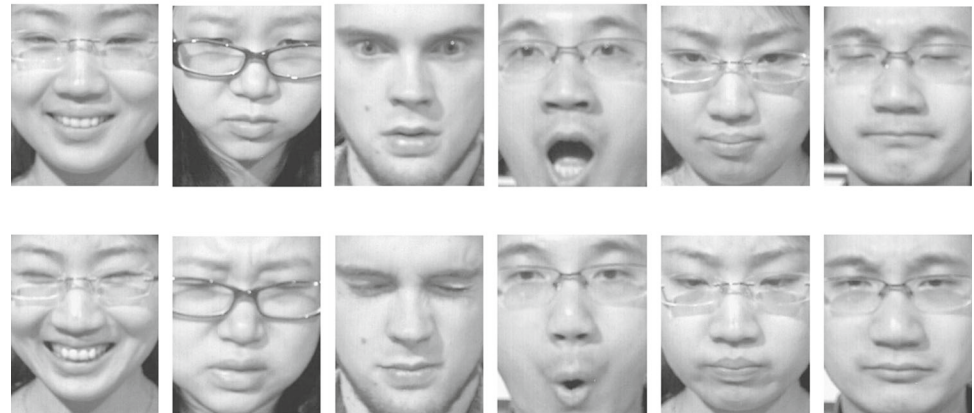
(b) samples from SPOS database

$$\propto \arg\max_{c \in [1,n]} \sum_{i=1}^{19} \log(P_c(F_i | pa(F_i)))$$
$$-\log(\text{Complexity}(M_c)) \qquad (6)$$

where E_T represents the features of a sample, $P(E_T|\Theta_c)$ denotes the likelihood of the sample given the cth model, F_i is the ith node in the BN, and $pa(F_i)$ denotes the parent nodes of F_i, and M_c stands for cth model and Complexity(M_c) represents the complexity of M_c. Since different models may have different spatial structures, the model likelihood $P(E_T|\Theta_c)$ will be divided by the model complexity for balance. We use the total number of the links as the model complexity.

4 Experiments and analyses

4.1 Experimental conditions

Till now, there are several databases available for posed and spontaneous expressions recognition, i.e., BBC Smile Dataset [27], MAHNOB-Laughter database [28], UvA-NEMO smile database [16], SPOS database [2], and USTC-NVIE database [29]. Among them, the BBC, MAHNOB-Laughter and UvA-NEMO databases only contain posed and spontaneous expressions for smiles, while the USTC-NVIE

and SPOS databases consist posed and spontaneous expression for six basic expression categories. Thus, these two databases are adopted in our experiments.

The USTC-NVIE database [29] is a natural visible and thermal infrared facial expression database, which contains both spontaneous and posed expressions with six basic categories (i.e. happiness, disgust, fear, surprise, anger, and sadness) of more than 100 subjects. The onset and apex frames are provided for both posed and spontaneous subsets. The SPOS database [2] is a visible and near-infrared expression database, including both posed and spontaneous expressions with six basic categories from seven subjects (four males and three females). The image sequences in this database start from onset frame and end with apex frame. Figure 4 provides image samples from these two databases.

For USTC-NVIE database, both the apex and onset frames of all posed and spontaneous expression samples, which come in pairs from the same subject, are selected. In this procedure, we discarded spontaneous samples whose maximum evaluation value on the six expression categories is zero, since there is no expression in these samples.

According to the rule, we finally select 1,028 samples, including 514 posed and 514 spontaneous expression samples from 55 male and 25 female subjects. The distribution of posed and spontaneous expression samples is shown in Table 2.

Table 2 The distribution of samples on USTC-NVIE database

	Happiness	Disgust	Fear	Surprise	Anger	Sadness
Posed	104	93	68	78	91	80
Spontaneous	104	93	68	78	91	80

Table 3 The distribution of samples on SPOS database

	Happiness	Disgust	Fear	Surprise	Anger	Sadness
Posed	14	14	14	14	14	14
Spontaneous	66	23	32	11	13	5

Given the databases, we first conduct statistical analyses to explore the differences of the proposed 19 geometric features between posed and spontaneous expressions from three aspects: all samples without gender and expression information, samples with the gender information, and samples with expression categories.

We then construct $6 \times 2 \times 2$ BN models to recognize posed and spontaneous expressions using posed and spontaneous samples for each expression and each gender respectively, denoted as "PS_gender_exp model". Our experimental results are obtained by applying a ten-fold cross validation method on all samples according to the subjects. To demonstrate the effectiveness of gender and expression categories as privileged information, we conducted three other experiments. The first one builds 2 BN models, which is denoted as "PS model", using posed and spontaneous samples without the gender and expression labels, respectively. The second one builds 4 BN models, which is denoted as "PS_gender model", using male posed, male spontaneous, female posed, and female spontaneous samples, respectively. The third one builds 12 BN models, which is denoted as "PS_exp model", using posed and spontaneous samples for each expression, respectively.

For SPOS database, the first and the last frames of all the posed and spontaneous samples are selected, including 84 posed expression samples and 150 spontaneous expression samples, as shown in Table 3. Since SPOS database consists of images from only seven subjects, and it does not include all six expression images for certain subjects, we did not select samples in pairs from SPOS database as we did on USTC-NVIE database. Furthermore, the subject number and the sample number of each expression in SPOS database are not large enough to do hypothesis test given gender and expression categories, so we only conduct statistical analysis to explore the difference between posed and spontaneous expressions for all samples. Due to the same reason, only two BN models are built for recognizing posed and spontaneous expressions. To compare with [2], leave-one-subject-out cross-validation is used.

4.2 Statistic feature analysis

From Table 2, we find that the numbers of the six expressions are different, and so do the genders. To avoid the influence of imbalance sample distribution under gender and six expressions on the statistical analysis results, we select samples randomly from a larger number of samples to ensure data balance and conduct the statistical analysis experiments. Thus, $68*6*2 = 816$ and $25*2*6*2 = 600$ samples are obtained from the NVIE database for the statistic tests with expression categories and gender information, respectively.

The statistical tests results are summarized in Table 4. We set the significant value to 0.05. If the result is less than the significant value, there is a significant difference. The "Percentage" row stands for the proportion of the number of features with significant differences to that of all the features. From Table 4, we can find as follows:

First, most geometric features have significant differences between posed and spontaneous expressions on both USTC-NVIE and SPOS databases. The ratios of the significant features to all the features are 73.68 and 68.42%, respectively. This demonstrates the effectiveness and the generalization ability of our defined features.

Second, for the NVIE database, most features in brow regions are significant given male and female, which confirms the effectiveness of these features for both male and female. The lip features are effective to recognize posed vs. spontaneous expressions for male, while the lip corner features are effective for female. Therefore, the significant features for male and female are different, which confirm the gender difference in expressions.

Third, for the NVIE database, the distributions of features with significant difference for six expressions are different. For example, both WMouth and HMouth are significant given happiness and surprise, but only WMouth is significant for the remaining four expressions. The ratios of features with significant difference for six expressions range from 42.11 to 73.68%. It indicates the effect of expression categories on posed vs. spontaneous expression manifestation.

Fourth, the change of mouth width (i.e. WMouth) and the movements of the lip corner (i.e. LipCorner1, LipCorner2, LipCorner3) show significant differences for distinguishing posed and spontaneous expressions for both NIVE and SPOS databases. This is consistent with [10,12,30]. Just as in [14], we conclude that the movements of brows-related AUs are helpful for differentiating between posed and spontaneous expressions, since the features Brow4 and Brow5 are with significant difference for both databases.

Last, the features with significant differences for the NIVE database are not the exactly same as those for the SPOS database. The reason may be the database bias, since the setup conditions of the two databases are not exactly same.

Table 4 Statistical analysis results of geometric features on USTC-NVIE and SPOS database

| Feature | USTC-NVIE | | | | | | | | | SPOS |
| | P–S | P–S-gender | | P–S-expression | | | | | | P–S |
		Male	Female	Happiness	Disgust	Fear	Surprise	Anger	Sadness	
HMouth	2e−04*	0.002*	0.148	3e−04*	0.255	0.603	6e−04*	0.992	0.075	0.020*
WMouth	3e−54*	1e−15*	1e−14*	7e−12*	3e−07*	2e−07*	6e−08*	1e−10*	1e−04*	6e−05*
LBrow	4e−14*	1e−05*	0.006*	0.214	3e−06*	9e−04*	0.116	0.003*	0.142	0.475
RBrow	3e−15*	2e−12*	3e−04*	0.049*	4e−09*	1e−04*	0.069	1e−06*	0.004*	0.274
Brow1	8e−22*	8e−08*	0.002*	0.399	3e−10*	0.009*	0.612	1e−08*	4e−05*	0.555
Brow2	1e−53*	4e−17*	5e−13*	9e−07*	1e−12*	1e−05*	5e−04*	8e−12*	1e−09*	0.055
Brow3	2e−20*	4e−09*	3e−04*	0.036*	2e−10*	0.016*	0.316	6e−05*	0.016*	0.054
Brow4	1e−14*	9e−05*	2e−05*	1e−07*	0.537	0.001*	1e−06*	0.310	9e−04*	0.049*
Brow5	2e−30*	3e−10*	1e−11*	4e−08*	4e−04*	6e−04*	0.012*	3e−08*	0.005*	0.002*
Brow6	0.602	0.395	0.506	0.638	5e−04*	0.052	0.013*	0.835	0.010*	9e−04*
Lid1	4e−08*	0.008*	0.027*	0.026*	0.173	1e−05*	1e−06*	0.026*	0.607	9e−04*
Lid2	0.170	0.038*	0.868	0.032*	0.353	0.005*	0.005*	0.992	4e−04*	0.310
Lip1	0.867	0.070	0.075	0.047*	0.902	0.372	0.395	0.188	0.120	0.032*
Lip2	0.533	0.440	0.225	0.263	0.174	0.907	0.985	0.013*	0.616	0.050*
Lip3	0.660	0.228	0.145	0.217	0.625	0.353	0.402	0.147	0.603	0.017*
LipCorner1	8e−06*	0.400	3e−04*	1e−05*	0.058	0.010*	0.835	0.582	0.545	0.009*
LipCorner2	0.014*	0.648	0.002*	0.009*	0.474	0.084	0.893	0.956	0.660	3e−04*
LipCorner3	1e−08*	0.003*	0.036*	0.021*	0.145	8e−06*	1e−06*	0.032*	0.616	0.001*
Chin1	0.004*	0.004*	0.976	0.018*	0.096	0.830	0.002*	0.156	0.199	3e−04*
Percentage	73.86	68.42	63.16	73.68	42.11	63.16	52.63	52.63	47.37	68.42

"*" indicates a significant difference

Given gender or expression, the features with significant difference are different. It means the spatial patterns of posed and spontaneous expression are different while using gender and emotion category as prior knowledge. Therefore, in the following sections, given gender or expressions, we used all 19 features or the selected significant features separately to learn BN models. Specifically, "PS model" uses the significant features in column 2 in Table 4; "PS_gender model" uses the union of the significant features in columns 3 and 4 in Table 4; "PS_exp model" uses the union of significant features in columns 5–10; "PS_gender_exp model" adopts the union of the significant features from column 3 to 10 in Table 4.

4.3 Experimental results of posed vs. spontaneous expression recognition

After statistic feature analyses and selection, we construct 24 BNs for each expression and each gender. Four of the learned BN structures which can achieve best recognition performance are shown in Fig. 3.

The structure of a Bayesian network captures the dependences among variables. One way of ascertaining the importance of different variables is to use node complexity. Node

Table 5 Average node complexity for both posed and spontaneous expressions from Fig. 3

| Feature | Average | |
	Posed	Spontaneous
1. HMouth	2.5	3
2. WMouth	3	1
3. LBrow	3	3
4. RBrow	3	3
5. Brow1	2	2
6. Brow2	7	4.5
7. Brow3	5.5	6.5
8. Brow4	4	4
9. Brow5	4	4
11. Lid1	5.5	4.5
12. Lid2	4	2
16. LipCorner1	7.5	5
17. LipCorner2	3.5	3.5
18. LipCorner3	3.5	2.5
19. Chin1	4	3.5

complexity is measured by the number of links connected to a node. It is expected that the more complex a node, the more important it is for the model. Based on this measure, we

Table 6 P vs. S recognition results on USTC-NVIE database without feature selection

Model	PS		PS_gender		PS_exp		PS_gender_exp	
	P	S	P	S	P	S	P	S
P	448	66	448	66	458	56	460	54
S	20	494	14	500	17	497	28	496
Accuracy (%)	91.63		92.22		92.90		93.00	
F1-score	0.9124		0.9180		0.9262		0.9173	

P represents posed expression
S represents spontaneous expression

Table 7 P vs. S recognition results on USTC-NVIE database with feature selection

Model	PS		PS_gender		PS_exp		PS_gender_exp	
	P	S	P	S	P	S	P	S
P	448	66	456	58	459	55	464	50
S	16	498	13	501	14	500	15	499
Accuracy (%)	92.02		93.09		93.29		93.68	
F1-score	0.9162		0.9278		0.9301		0.9345	

P represents posed expression
S represents spontaneous expression

list the average link number for both posed and spontaneous models in Table 5.

From the table, in terms of average node complexity for both posed and spontaneous expressions, we can divide features into three categories: high, low, and medium. The features that rank consistently higher than others include brown2, brown3, Lid1, Lipcorner1, and Chin1. These features together encode most of the relationships in the BN models and hence are most important to distinguish the posed and spontaneous expressions. On the other hand, certain features like brown1 and Wmouth rank consistently low. Their contributions to expression recognition are hence marginal because of limited feature relationships they capture. The remaining relationships in the models are contributed by the remaining 8 features and hence have moderate contributions to distinguish posed and spontaneous expressions.

Given the BNs, we can then perform expression recognition. Since we randomly determined the order of BN nodes a few times, the averaged recognition results are adopted in this section. The recognition results on USTC-NVIE database without feature selection and with feature selection are shown in Tables 6 and 7. Comparing Tables 6 and 7, we can find that for all the BN models, not only accuracy but also F1-score with feature selection is higher than those without feature selection. It further confirms that the spatial pattern embedded in posed and spontaneous expressions varies with gender and expressions. Compared the results of PS model with those of the remaining models, we can find that the gender and expression categories existing in training set can help model the muscle variation in posed and spontaneous expression, since both accuracy and F1 score of PS_gender, PS_exp, PS_gender_exp models are higher that those of PS model. The improvement by considering expression cate-

Table 8 P vs. S recognition results on SPOS database with and without feature selection

Model	Without feature selection		With feature selection	
	P	S	P	S
P	55	29	61	23
S	34	116	36	114
Accuracy (%)	73.08		74.79	
F1-score	0.6358		0.6740	

P represents posed expression
S represents spontaneous expression

gories is larger than that by considering gender information. This further confirms that the expression difference is larger than gender difference, which is consistent with the statistical results in Sect. 4.2. Our proposed model PS_gender_ exp achieves the best performance, demonstrating that our proposed multiple BN models can capture the spatial patterns effectively.

Experimental results on the SPOS database are shown in Table 8. From Table 8, we can find that both the accuracy and F1-score increase after feature selection, demonstrating the effectiveness of selected significant features. The results are acceptable, but not as good as those on the USTC-NIVE database. Since the number of samples from USTC-NVIE database vastly exceeds that from SPOS database, it is reasonable that the accurate rate and F1-score obtained on SPOS Database are a little lower than that on the NVIE database.

4.4 Comparison with related work

To further demonstrate the effectiveness of our proposed method further, we compare our work with the only two

Table 9 Comparison with [1] on USTC-NVIE database

Accuracy (%)	Happiness	Disgust	Fear	Surprise	Anger	Sadness	Average
PS	93.27	93.01	91.32	92.31	89.56	90.63	91.68
PS_gender	94.23	91.94	91.18	94.87	91.76	91.87	92.65
PS_exp	94.23	94.84	91.76	95.00	90.11	91.25	92.87
PS_gender_exp	95.37	93.76	92.21	95.51	92.68	91.88	93.60
Zhang et al. [1]	80.5	76.1	79.7	83.4	77.2	79.7	79.43

works, which classify posed and spontaneous expressions for six expressions, i.e Zhang's [1] and Pfister's [2] works. In addition, we conducted experiments using linear support vector machine (SVM) as a baseline.

Zhang et al. used both geometric and appearance features to recognize posed and spontaneous expression for six basic expressions on USTC-NIVE database. They selected 3,572 posed and 1,472 spontaneous images, after removing the images whose face or facial point cannot be detected correctly. Since they did not explicitly state which images were selected, it is hard for us to select the same images as theirs. Therefore, we can only compare the experimental results as a reference.

To compare with their work [1], we calculate the posed and spontaneous expression recognition rate for each basic expression using "PS", "PS_gender", "PS_exp", and "PS_gender_exp" models. The results of our work and the best results of [1] are shown in Table 9.

From this table, we can find that our proposed four models outperform Zhang et al.'s although we only use geometric features and the number of samples is smaller than Zhang et al.'s. This further demonstrates that our proposed BN models systematically capture the spatial patterns embodied in posed and spontaneous expressions. Even though the average performance is improved with the use of the gender and expression information, the gender and expression information cannot consistently improve the performance for all expressions. For some expressions, they cannot contribute any additional information to distinguish posed and spontaneous expressions. For example, the result of disgust expression from "PS_gender_ exp" is worse than "PS", "PS_gender", and "PS_exp". This demonstrates that gender and expression category's impact on posed–spontaneous classification varies with expression.

Pfister et al.'s paper [2] proposed the spatiotemporal local texture descriptor, CLBP-TOP, to distinguish posed and spontaneous expression from both visible and near-infrared image sequences on their constructed database, SPOS. Here, we only compare our work with their work on visible images, as shown in Table 10. From this table, we can find that the accuracy of our method is 2 % higher than Pfister et al.'s. Their proposed CLBP-TOP texture features were extracted from facial expression sequence, while our geometric fea-

Table 10 Comparison with [2] on SPOS database

Method	Ours	Pfister et al. [2]
Accuracy (%)	74.79	72.0

Table 11 P vs. S recognition results on both databases using linear SVM as classifier

Model	USTC-NVIE database		SPOS database	
	P	S	P	S
P	438	76	46	38
S	73	441	21	129
Accuracy (%)	85.51		74.79	
F1-score	0.8546		0.6093	

P represents posed expression
S represents spontaneous expression

tures are extracted from the apex frames and onset frames. It indicates that we achieve better results using less information, demonstrating once again that our proposed BN model captured the pattern embodied in posed and spontaneous expression successfully.

Since the data and features used in our paper are not the exactly same as those in Zhang's [1] and Pfister's [2] works, it is difficult to make the comparison completely fair. We try to demonstrate that under unfavorable conditions, our method still performs better. For example, compared with [1] and [2], our method still outperforms them even they used more powerful features. It also suggests that the improvement comes from the method instead of from the features. Also, compared with [1], our experiments use less data yet has better performance. It further suggests that the superiority of our proposed method results from the technique instead of from data.

For posed and spontaneous expression recognition using linear SVM, the same cross-validation as that for our proposed method is used, i.e. tenfold-subject cross-validation on the NVIE database, and leave-one-subject-out cross-validation on the SPOS database. Model selection is adopted to select the learning parameters of SVM. Table 11 lists the recognition results using linear SVM on both databases.

From Table 11, we can see that the accuracy on the USTC-NVIE database is 85.51 %, lower than that of our method,

93.00 %. The accuracy on the SPOS database is the same as that of our method. However, the F1-score of our method is 0.6740, which is higher than that of linear SVM. It further demonstrates that our proposed BN model captured the pattern embodied in posed and spontaneous expression successfully.

The above comparison demonstrates that the performance of our approach is better than the state of the art. The most important reason that contributes the superior performance of our method to the existing methods is the use of the Bayesian network to systematically capture the relationships among the spatial movements of different facial landmark points. It is our belief that the relationships among the spatial movements of different parts of the face are more important than any appearance-based image features in characterizing the differences between posed and spontaneous expressions. Furthermore, with the use of additional information from gender and expression category, the performance of our method is further improved. This explains why our method outperforms [1,2] in spite of their use of more powerful features and classifiers.

5 Conclusion and future work

In this paper, we propose a new method to recognize posed and spontaneous expressions by explicitly capturing the different facial spatial patterns between posed and spontaneous expressions using BN. Meanwhile, we employ gender and expression categories as privileged information, to better capture the facial spatial variation with respect to different gender and expression and further improve the recognition performance. First, we propose 19 geometric features related to AU variations between apex and onset facial images to represent spatial patterns. Second, we conduct statistical analyses to investigate their differences between posed and spontaneous expressions using the defined features from three aspects: all samples without expression and gender labels, samples with the gender information, and samples with expression categories. Third, we build several BNs to explore spatial patterns embedded in posed and spontaneous expressions from four aspects: all samples, samples given the gender information, samples given expression categories, and samples given both gender and expression categories.

The statistical analysis results on the USTC-NVIE and SPOS databases both demonstrate the effectiveness of the proposed features. The recognition results demonstrate that the proposed method outperforms the state of the art on both databases. Furthermore, results on the USTC-NVIE database indicate that the privileged information of gender and expression can improve the recognition performance of posed vs. spontaneous expression.

Although significant performances are achieved by explicitly modeling the geometric patterns between posed and spontaneous expressions, it is still unknown the influence of temporal patterns embedded in posed and spontaneous expressions using similar modeling approach. Therefore, in future, we will model the temporal patterns embedded in posed and spontaneous expressions to recognize posed and spontaneous expressions.

References

1. Zhang, L., Tjondronegoro, D., Chandran, V.: Geometry vs. appearance for discriminating between posed and spontaneous emotions. In: Neural Information Processing, pp. 431–440. Springer, Berlin (2011)
2. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 868–875. IEEE (2011)
3. Lithari, C., Frantzidis, C.A., Papadelis, C., Vivas, A.B., Klados, M.A., Kourtidou-Papadeli, C., Pappas, C., Ioannides, A.A., Bamidis, P.D.: Are females more responsive to emotional stimuli? a neurophysiological study across arousal and valence dimensions. Brain Topogr. **23**(1), 27–40 (2010)
4. Yunus, S., Christopher, T.: Cascaded classification of gender and facial expression using active appearance models. In: IEEE international conference on automatic face and gesture recognition, pp. 393–400 (2006)
5. Vapnik, V., Vashist, A.: A new learning paradigm: learning using privileged information. Neural Netw. Off. J. Int. Neural Netw. Soc. **22**(5–6), 544 (2009)
6. Chen, J., Liu, X., Lyu, S.: Boosting with side information. In: Computer vision-ACCV 2012, pp. 563–577. Springer, Berlin (2013)
7. Seckington, M.: Using dynamic bayesian networks for posed versus spontaneous facial expression recognition. Mater Thesis, Department of Computer Science, Delft University of Technology (2011)
8. Ekman, P., Hager, J.C., Friesen, W.V.: The symmetry of emotional and deliberate facial actions. Psychophysiology **18**(2), 101–106 (1981)
9. Ekman, P., Rosenberg, E.L.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
10. Ekman, P., Friesen, W.V.: Felt, false, and miserable smiles. J. Nonverbal Behav. **6**(4), 238–252 (1982)
11. Ekman, P.: Darwin, deception, and facial expression. Ann. N. Y. Acad. Sci. **1000**(1), 205–221 (2003)
12. Cohn, J.F., Schmidt, K.L.: The timing of facial motion in posed and spontaneous smiles. Int. J. Wavelets Multiresolut. Inform. Process. **2**(02), 121–132 (2004)
13. Valstar, M.F., Gunes, H., Pantic, M.: How to distinguish posed from spontaneous smiles using geometric features. In: Proceedings of the 9th international conference on multimodal interfaces, pp. 38–45. ACM, Nagoya (2007)
14. Valstar, M.F., Pantic, M., Ambadar, Z., Cohn, J.F.: Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In: Pro-

ceedings of the 8th international conference on multimodal interfaces, pp. 162–170. ACM, New York (2006)

15. Littlewort, G.C., Bartlett, M.S., Lee, K.: Automatic coding of facial expressions displayed during posed and genuine pain. Image Vis. Comput. **27**(12), 1797–1803 (2009)

16. Dibeklioğlu, H., Salah, A.A., Gevers, T.: Are you really smiling at me? spontaneous versus posed enjoyment smiles. In: Computer vision-ECCV 2012, pp. 525–538. Springer, Berlin (2012)

17. Tong, Y., Wang, Y., Zhu, Z., Ji, Q.: Robust facial feature tracking under varying face pose and facial expression. Pattern Recognit. **40**(11), 3195–3208 (2007)

18. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Signal Process. **29**(6), 1153–1160 (1981)

19. Mitchell, D.P., Netravali, A.N.: Reconstruction filters in computer-graphics. In: ACM Siggraph Computer Graphics, vol. 22, pp. 221–228. ACM, Newyork(1988)

20. He, M., Wang, S., Liu, Z., Chen, X.: Analyses of the differences between posed and spontaneous facial expressions. In: 2013 Humaine association conference on affective computing and intelligent interaction (ACII), pp. 79–84, Sept 2013

21. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 699–714 (2005)

22. Wilcoxon, F.: Individual comparisons by ranking methods. Biom. Bull. **1**(6), 80–83 (1945)

23. Goodman, S.N.: Toward evidence-based medical statistics. 1: The p value fallacy. Ann. Intern. Med. **130**(12), 995–1004 (1999)

24. Lowry, R.: Concepts and Applications of Inferential Statistics. R. Lowry, Poughkeepsie (1998)

25. Cooper, G.F., Herskovits, E.: A bayesian method for the induction of probabilistic networks from data. Mach. Learn. **9**, 309–347 (1992)

26. Murphy, K.: How to use bayes net toolbox. Avaliable: https://code.google.com/p/bnt/ (2004)

27. Paul, E.: Bbc-dataset. http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/

28. Petridis, S., Martinez, B., Pantic, M.: The mahnob laughter database. Image Vis. Comput. **31**(2), 186–202 (2013)

29. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., Wang, X.: A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Trans. Multimed. **12**(7), 682–691 (2010)

30. Schmidt, K.L., Ambadar, Z., Cohn, J.F., Reed, L.I.: Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. J. Nonverbal Behav. **30**(1), 37–52 (2006)

Shangfei Wang received her BS in Electronic Engineering from Anhui University, Hefei, Anhui, China, in 1996. She received her MS in Circuits and Systems, and the PhD in Signal and Information Processing from University of Science and Technology of China (USTC), Hefei, Anhui, China, in 1999 and 2002. From 2004 to 2005, she was a postdoctoral research fellow at Kyushu University, Japan. Between 2011 and 2012, Dr. Wang was a visiting scholar at Rensselaer Polytechnic Institute in Troy, NY, USA. She is currently an Associate Professor of School of Computer Science and Technology, USTC. Dr. Wang is an IEEE and ACM member. Her research interests cover computation intelligence, affective computing, and probabilistic graphical models. She has authored or co-authored over 70 publications.
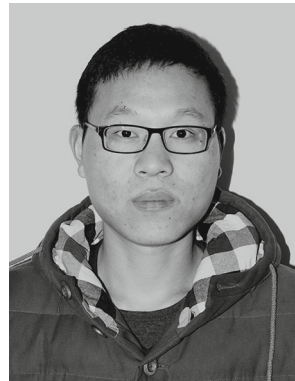
Chongliang Wu received his BS in Computer Science from Wuhan Textile University, Wuhan, China, in 2013. He is currently pursuing his MS in Computer Science at the University of Science and Technology of China, Hefei, China. His research interest is in affective computing.

Menghua He received her BS in Information and Computation Science from Anhui University, Hefei, China, in 2011. She received her MS in Computer Science at the University of Science and Technology of China (USTC), China, in 2014. She is currently researching Iflytek.

Jun Wang received her BS in Computer Science from China University of Geosciences, Wuhan, China, in 2012. She is currently pursuing the MS degree in Computer Science at the University of Science and Technology of China, Hefei, China. Her research interest is in affective computing.

Qiang Ji received his PhD in Electrical Engineering from the University of Washington. He is currently a Professor with the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He recently served as a program director at the National Science Foundation (NSF), where he managed NSF's computer vision and machine learning programs. He also held teaching and research positions with the Beckman Institute at University of Illinois at Urbana-Champaign, the Robotics Institute at Carnegie Mellon University, the Department of Computer Science at University of Nevada at Reno, and the US Air Force Research Laboratory. Professor Ji currently serves as the Director of the Intelligent Systems Laboratory (ISL) at RPI. Prof. Ji's research interests are in Computer Vision, Probabilistic Graphical Models, Information Fusion, and their applications in various fields. He has published over 160 papers in peer-reviewed journals and conferences. His research has been supported by major governmental agencies including NSF, NIH, DARPA, ONR, ARO, and AFOSR as well as by major companies including Honda and Boeing. Professor Ji is an editor on several related IEEE and international journals and he has served as a general chair, program chair, technical area chair, and program committee member in numerous international conferences/workshops. Professor Ji is a fellow of IAPR, and IEEE.