Multimed Tools Appl (2016) 75:3937–3954
DOI 10.1007/s11042-015-3107-2

# Facial expression recognition through modeling age-related spatial patterns

**Shangfei Wang[1] · Shan Wu[1] · Zhen Gao[1] · Qiang Ji[2]**

Received: 1 April 2015 / Revised: 4 October 2015 / Accepted: 19 November 2015 /
Published online: 28 November 2015


**Abstract** In this paper we tackle the problem of expression recognition by exploiting age-related spatial facial expression patterns, which carry crucial information that have not been thoroughly exploited. First, we conduct two statistic hypothesis tests to investigate age effect on the spatial patterns of expressions and on facial expression recognition respectively. Second, we propose two methods to recognize expressions by modeling age-related spatial facial expression patterns. One is a three-node Bayesian Network to classify expressions with the help of age from person-independent geometric features. The other is to construct multiple Bayesian networks to explicitly capture the spatial facial expression patterns for different ages. For both methods, age information is used as privileged information, which is only available during training, and is exploited during training to construct a better classifier. Statistic analyses on two benchmark databases, i.e. the Lifespan and the FACES, verify the age effect on spatial patterns of expressions and on facial expression recognition. Experimental results of expression recognition demonstrate the effectiveness of the proposed methods in modelling age-related spatial patterns as well as their superior expression recognition performance to existing approaches.

✉ Shangfei Wang
sfwang@ustc.edu.cn

Shan Wu
sa14ws@mail.ustc.edu.cn

Zhen Gao
gzgqllxh@mail.ustc.edu.cn

Qiang Ji
qji@ecse.rpi.edu

[1] Key Lab of Computing and Communication Software of Anhui Province School of Computer Science and Technology, University of Science and Technology of China Hefei, Anhui, 230027, People's Republic of China

[2] Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute Troy, Troy, NY 12180, USA

**Keywords** Expression recognition · Age-related spatial patterns · Privileged information · Bayesian networks

## 1 Introduction

Facial expression recognition has attracted increasing attention in recent years due to its wide application in human-computer interaction [9, 20, 24]. Although much progresses have been achieved in computational facial expression recognition, almost all studies focus on discriminative geometric and appearance features to characterize facial images, and effective classifiers to model the spatial and temporary patterns embedded in facial expressions, ignoring the effects of facial attributes, such as age, on expression recognition even though research indicates that face structures develop with ages and expression manifestation varies with ages. Furthermore, most benchmark facial expression databases, such as the MMI database and CK+ database [18], only consider expressions with a small age ranges. The lack of databases with larger age ranges limits the generality and the performance of current expression recognition studies, and further hinders the development of age-related facial expression recognition.

Recently, researchers in psychology have realized that a large number of faces throughout the adult lifespan carry crucial information for complete understanding of many psychological studies, including perception, attention, memory, social reasoning, emotion, infant and adult development, and neuropsychology [8, 19]. Therefore, two benchmark databases have been constructed: one is Lifespan [19], consisting of 575 faces from ages 18 to 93, and the other is FACES [8], containing 2,052 images from ages 19 to 80. Very recently, Ebner and Johnson's work [7] investigated interference of face-related tasks by irrelevant faces of different ages and with different facial expressions. Their work demonstrates age-group differences in interference from emotional faces of different ages. By reviewing theoretical frameworks and empirical findings of age effects on facial expression decoding, Fölster et al. [10] concluded that the age of the face plays an important role in facial expression decoding. Their review suggests that the expression decoding accuracy for older faces may be reduced by many factors, such as lower expressivity, age-related facial changes, less elaborated emotion schemas, etc. Hess et al. [13] investigated how emotions expressed by the elderly are perceived by others. Their findings suggest that emotions shown on older faces have reduced signal clarity due to wrinkles and folds, and thus may consequently impact on the behavioral inferences that others draw from the emotion expression. Houstis and Kiliaridis [14] quantitatively evaluated the facial expressions of children and adults in order to assess their dependence on age. Their studies on 80 subjects find a trend from childhood to adulthood, showing an increase in the percentage of change in most vertical movements, possibly due to development of the mimic musculature from childhood to adulthood.

To the best of our knowledge, there are only three studies to discover the age effect on facial expression recognition in computer vision. Guo et al. [11] are the first to study the age effect on facial expression recognition computationally. They proposed two methods, i.e. age group constrained facial expression recognition and age-removing facial expression recognition. The former trains a multi-class classifier by considering each expression in each age group as one independent class. The later removes the facial wrinkles and other aging details using an edge-preserving image smoothing technique before expression recognition. Experiments were conducted on the Lifespan and the FACES databases, demonstrating the significant influence of human aging on computational facial expression recognition. Other than focusing on age-invariant expression recognition, Alnajar et al. [1] considered

expression-invariant age estimation. They proposed a graphical model with a latent layer between the age/expression labels and the features to jointly learns the age and the expression. Experimental results on the Lifespan and FACES databases illustrate the improvement in age estimation when the age is jointly learnt with expression in comparison to expression-independent age estimation. In addition, expression recognition performance is improved on the FACES data set, and is comparable on the Lifespan data set by joint-learning. These two studies adopt appearance features, i.e. Gabor features [11] and LBP features [1]. Unlike the two studies, Dibeklioglu et al. [6] analyzed the effect of age on distinguishing posed and spontaneous smile by using age as one feature along with the defined dynamic features. Their experiments on the BBC, MMI, SPOS, and UvA-NEMO databases demonstrate that the performance of posed and spontaneous smile differentiation is improved by using aging information as a feature.

Among the three studies, the first two studies can recognize expression and age jointly, or remove aging details before expression recognition. It means age information is not required during testing. Therefore, age information is used as privileged information, which is only available during training [23], and is exploited during training to construct a better classifier. While in the Dibeklioglu's study, age estimation should be performed before expression recognition during testing. Such sequential approach may propagate the error of age estimation to the subsequent expression recognition. Therefore, we prefer to incorporate age information as privileged information, which is only required during training, in this paper. Furthermore, the first two studies adopted appearance features, which are useful to describe wrinkles, and the third study used dynamic features, which are crucial for posed and spontaneous smile distinction. In this paper, we exploit spatial patterns, which carry crucial information for facial expressions that have not been thoroughly exploited in age-invariant expression recognition. Specifically, we propose two methods. One is a three-node Bayesian Network (BN) [21] to recognize expressions with the help of age from geometric features. During training, we construct a full probabilistic model $P(x, x^\star, y)$ by using the training set $(x_i, x_i^\star, y_i), i = 1, ..., l$, where $x_i$ is geometric features, $x_i^\star$ is age information, and $y_i$ is expression label. During testing, we can obtain $P(y|x)$ by marginalizing over $x^\star$. The other is to construct multiple Bayesian networks to explicitly capture the spatial facial expression patterns for each age group. During testing, only facial geometric features are provided, and the samples are classified into expressions according to the BN with the largest likelihood. Experiments on the Lifespan and FACES databases demonstrate the effectiveness of our proposed approaches.

The rest of this paper is organized as follows. Section 2 introduces two benchmark databases and the extracted geometric features. Section 3 analyzes the age effect on spatial pattern of expressions and on expression recognition. Section 4 introduces our two proposed methods. Section 5 presents the results and analyses on the experiments for validating our proposed methods. Section 6 compares our methods with related work. Section 7 summarizes our work.

## 2 Two databases

Currently, only two databases, i.e. the FACES [8] and Lifespan [19] databases, contain a large range of age variations as mentioned in Section 1, therefore, we adopt them in our work.

The FACES database consists of 2052 images, which are divided into two sets. Since the images of the two sets are almost the same, we adopt one set in this work. The FACES

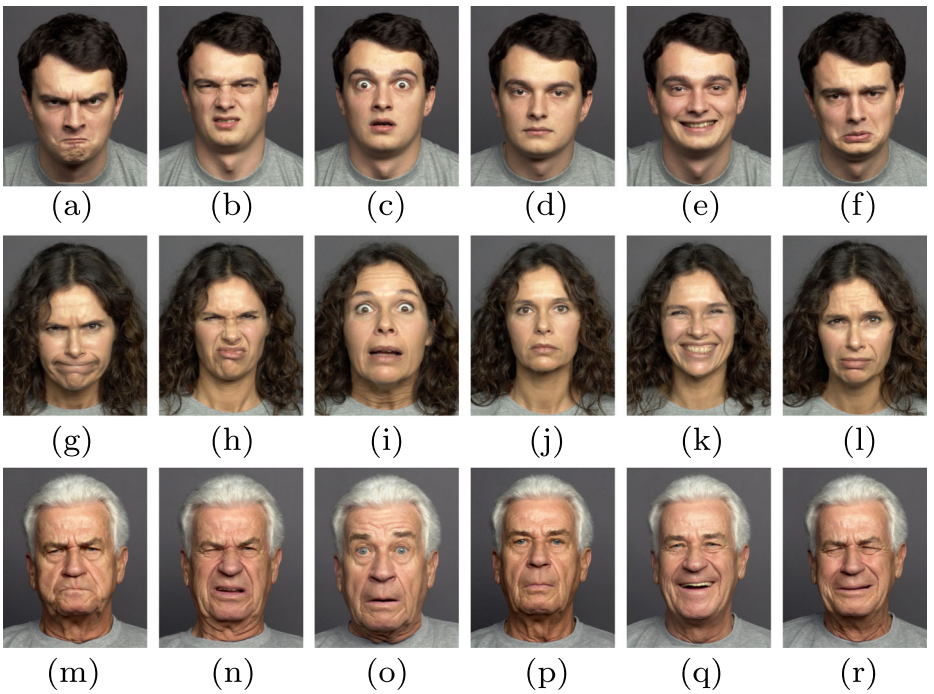


**Fig. 1** Expression samples in the FACES database **a** young-anger; **b** young-disgust; **c** young-fear; **d** young-neutral; **e** young-happy; **f** young-sad; **g** middle age-anger; **h** middle age-disgust; **i** middle age-fear; **j** middle age-neutral; **k** middle age-happy; **l** middle age-sad. **m** old-anger; **n** old-disgust; **o** old-fear; **p** old-neutral; **q** old-happy; **r** old-sad

database includes six expressions, i.e. anger, disgust, fear, happiness, neutral and sadness, as shown in Fig. 1. The Lifespan database consists of images with eight expressions as shown in Table 1. Since the numbers of expression samples with surprise, sadness, anger, annoy, disgust, and grumpy are much fewer than those of neutral and happy facial images, only neutral and happy samples of the Lifespan database are adopted in our work. Therefore, the number of used samples is 835. Figure 2 lists sample faces of happy and neutral expressions for the Lifespan database. Both databases are posed facial expression databases. In our work, we group the samples of both databases into 3 age groups, which are 18–31, 32–59 and 60–93 respectively as shown in Table 1.

In addition, Guo et al. [11] manually labeled the fiducial points for each face image of both databases. (For FACES database, they only labeled 2004 images. So in our experiment, the database we use contains 2004 images). Therefore, we choose 26 fiducial points on the FACES database and 31 fiducial points (include two eye pupils) on the Lifespan database in our work, as shown in Fig. 3 (for the Lifespan database, the 22-th point is only used to extract person-independent features). Since only apex images are provided in the two databases, and neutral faces are not available, we extract person-independent geometric features [2], i.e. ratios of distances, areas and angles, to represent the spatial patterns embedded in expressions, instead of using distances directly or normalizing these distances using neutral faces. The person-independent features are listed in Table 2, where the second column denotes the corresponding formula to calculate the feature $f_j$. Each facial landmark is

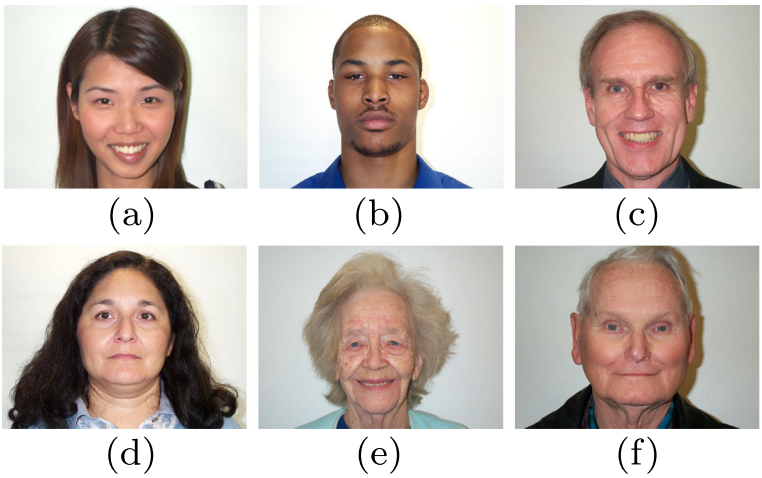


**Fig. 2** Expression samples in the Lifespan database **a** young-happy; **b** young-neutral; **c** middle age-happy; **d** middle age-neutral; **e** old-happy; **f** old-neutral

denoted as $p_i = (x_i, y_i) \in R^2$, and the index $i$ in this table is consistent with the point index in Fig. 3a, b. The features listed in the table represent the spatial relationships of the fiducial points on faces and exhibit discriminative person-independent properties. For example, the first feature $f_1$ is the ratio of the distance $d_1$ (i.e. the distance between the left eye outer corner and the left mouth corner) to the distance $d_2$ (i.e. the distance between the right eye outer corner and the right mouth corner). This ratio almost remains the same for every person for the same expression, thus it is a person-independent feature. Similarly, other features listed

**Table 1** Facial Expressions with Age Group Divisions on two databases

| DB | Expression | Age group | | | Total |
|---|---|---|---|---|---|
| | | 18–31 | 32–59 | 60–93 | |
| FACES | Anger | 58 | 55 | 54 | 1002 |
| | Disgust | 58 | 55 | 54 | |
| | Fear | 58 | 55 | 54 | |
| | Happy | 58 | 55 | 54 | |
| | Neutral | 58 | 55 | 54 | |
| | Sad | 58 | 55 | 54 | |
| Lifespan | Neutral | 225 | 99 | 253 | 1043 |
| | Happy | 145 | 29 | 84 | |
| | Anger | 4 | 4 | 2 | |
| | Annoyed | 22 | 11 | 7 | |
| | Disgust | 2 | 5 | 0 | |
| | Grumpy | 1 | 5 | 3 | |
| | Sad | 34 | 21 | 9 | |
| | Surprise | 43 | 25 | 10 | |

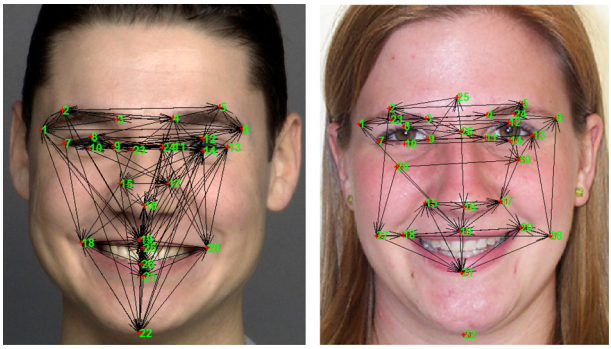


**Fig. 3** The face fiducial points on two databases. **a** FACES; **b** Lifespan

in Table 2 also exhibit discriminative person-independent properties. More details can be found in [2].

Before feature extraction, we normalize the images according to the coordinates of two pupils.

## 3 Statistical analyses of age effect on expressions

Two kinds of statistical analyses are conducted to investigate the age effect on expressions. The first one is to discover whether there is any aging difference on spatial patterns embedded in expressions. The second one is to analyze age effect on expression recognition. Both analyses use person-independent features.

For the first study, a one-way ANOVA [3, 15] with age as an independent variable and the geometry features as dependent variables is adopted. The null hypothesis (H0) is that the mean value of geometric features among three age groups for each expression are equal. The alternative hypothesis (H1) is that the mean value of geometrical features among age groups for each expression are not exactly the same. The significance level is set at 0.05.

Statistical analysis results are listed in Table 3. From Table 3, we can find that for most expressions, more than half features are age-related, since their p-values are less than 0.05. It proves the age effect on spatial patterns embedded in expressions. For both databases, happy and neutral expressions have the largest number of features with significant difference. The age effect on the neutral expression may indicate that face structures develop with ages, since neutral expression mainly represents face structures but rarely expression. Compared with other non-neutral expressions, the happy expression shows more variations across age groups. It may demonstrate the changes of happy expression manifestation are much more significant than those of other expressions with ages. The reason may be that happy expression, a kind of smile expressions, is the most frequent displayed expressions during our daily life. This kind of frequently display may enhance the change of expression manifestation with ages. In addition, the features with significant difference among age groups for the same expression are not exact the same on the two databases. For example, for happy and neutral expressions, the p-values of $f_3$ on the FACES database are lower than 0.05, but larger than 0.05 on the Lifespan database. It may caused by the database bias.

**Table 2** Person Independent Geometric Features [2]

| Feature | Equation |
|---|---|
| $f_1$ | $d_1 = \|p_{18} - p_{20}\|, d_2 = \|p_{19} - p_{21}\|, f_1 = d_1/d_2$ |
| $f_2$ | $A_1 = area\{\triangle(p_{16}p_{18}p_{20})\}, A_2 = area\{\triangle(p_{18}p_{20}p_{22})\}, f_2 = A_1/A_2$ |
| $f_3$ | $line_t = \overline{p_{18}p_{20}}; d_1 = distance(p_{19}, line_t), d_2 = distance(p_{21}, line_t), f_3 = d_1/d_2$ |
| $f_4$ | $f_4 = \angle(\overrightarrow{p_{21}p_{18}}, \overrightarrow{p_{22}p_{18}})$ |
| $f_5$ | $p_t = (p_{18} + p_{20}) * 0.5; d_1 = \|p_{19} - p_t\|; d_2 = \|p_{21} - p_t\|; f_5 = d_1/d_2$ |
| $f_6$ | $p_t = (p_{19} + p_{21}) * 0.5; f_6 = \angle(\overrightarrow{p_{18}p_t}, \overrightarrow{p_{20}p_t})$ |
| $f_7$ | $p_t = (p_{19} + p_{21}) * 0.5; line_t = \overline{p_{18}p_{20}}, f_7 = distance(p_t, line_t)$ |
| $f_8$ | $A_1 = area\{\triangle(p_{19}p_{18}p_{20})\}, A_2 = area\{\triangle(p_{18}p_{20}p_{22})\}, f_8 = A_1/A_2$ |
| $f_9$ | $d_1 = \|p_{17} - p_{21}\|, d_2 = \|p_{17} - p_{22}\|, f_9 = d_1/d_2$ |
| $f_{10}$ | $d_1 = \|p_7 - p_9\|, d_2 = \|p_{18} - p_{20}\|, f_{10} = d_1/d_2$ |
| $f_{11}$ | $A_1 = area\{\triangle(p_8p_7p_9)\}, A_2 = area\{\triangle(p_2p_7p_9)\}, f_{11} = A_1/A_2$ |
| $f_{12}$ | $f_{12} = \angle(\overrightarrow{p_1p_2}, \overrightarrow{p_1p_9})$ |
| $f_{13}$ | $d_1 = \|p_3 - p_9\|, d_2 = \|p_2 - p_9\|, f_{13} = d_1/d_2$ |
| $f_{14}$ | $A_1 = area\{\triangle(p_1p_3p_9)\}, A_2 = area\{\triangle(p_1p_7p_9)\}, f_{14} = A_1/A_2$ |
| $f_{15}$ | $f_{15} = \angle(\overrightarrow{p_8p_9}, \overrightarrow{p_{10}p_9})$ |
| $f_{16}$ | $d_1 = \|p_8 - p_{10}\|, d_2 = \|p_7 - p_9\|, f_{16} = d_1/d_2$ |
| $f_{17}$ | $f_{17} = \angle(\overrightarrow{p_2p_3}, \overrightarrow{p_4p_5})$ |
| $f_{18}$ | $f_{18} = \angle(\overrightarrow{p_{15}p_{18}}, \overrightarrow{p_{15}p_7})$ |

$p_t$ represents the mid point of a line $\overline{p_i p_j}$

$\triangle(p_i p_j p_k)$ indicates a triangle formed by three points $p_i$, $p_j$ and $p_k$

Vector $\overrightarrow{p_i p_j}$ represents a vector pointing from $p_i$ to $p_j$. Angular features (i.e. $\angle$) are calculated by employing vector inner products

For the second study, we compare the performance of expression recognition within age group with that of cross age group by using person-independent features and SVM. Ten-fold cross validation is adopted. Experimental results on the FACES database and the Lifespan database are listed in Table 4. From this table, we can obtain the following observations: first, the recognition accuracies of within age group are much higher than those of a cross age group in most cases, which clearly demonstrates the age effect on expression recognition. Second, in most cases, the accuracies of cross age group decrease with the increase of age difference between age groups. For example, on the Lifespan database, when training on age group 18–31, the expression recognition accuracy of within age group is 93.3 %, while the accuracy drops to 84.57 and 83 % respectively when testing on age group of 32–59 and 60–93. Third, the accuracy of within age group decreases with aging for both databases, suggesting the challenge of expression recognition for the old. This may be caused by the wrinkles and the facial muscle elasticity reduction developed with aging. Another possible reason is different expression manifestations for different age groups. For example, old people tend to express their expressions in a subtle way, while the young are inclined to show expressions exaggeratedly. This difficulty of expression recognition for the old may lead to the lower accuracy of within age group 60–93, compared with those of a cross age groups for both databases.

Last, comparing the performance on two databases, the accuracies on the Lifespan database are higher than those on the FACES database for both within age group or cross

**Table 3** Results of statistic hypothesis test on two databases

| | FACES | | | | | | Lifespan | |
|---|---|---|---|---|---|---|---|---|
| | Anger | Disgust | Fear | Happy | Neutral | Sad | Happy | Neutral |
| $f_1$ | 0.009* | 0.000* | 0.000* | 0.000* | 0.000* | 0.000* | 0.000* | 0.000* |
| $f_2$ | 0.001* | 0.015* | 0.092 | 0.000* | 0.000* | 0.006* | N/A | N/A |
| $f_3$ | 0.112 | 0.025* | 0.016* | 0.000* | 0.016* | 0.443 | 0.355 | 0.134 |
| $f_4$ | 0.185 | 0.055 | 0.000* | 0.000* | 0.000* | 0.009* | N/A | N/A |
| $f_5$ | 0.444 | 0.052 | 0.020* | 0.000* | 0.000* | 0.146 | 0.357 | 0.369 |
| $f_6$ | 0.528 | 0.048* | 0.733 | 0.000* | 0.067 | 0.097 | 0.001* | 0.797 |
| $f_7$ | 0.144 | 0.064 | 0.585 | 0.000* | 0.004* | 0.416 | 0.000* | 0.888 |
| $f_8$ | 0.019* | 0.693 | 0.561 | 0.021* | 0.105 | 0.222 | N/A | N/A |
| $f_9$ | 0.511 | 0.122 | 0.001* | 0.001* | 0.001* | 0.025* | N/A | N/A |
| $f_{10}$ | 0.000* | 0.000* | 0.000* | 0.003* | 0.000* | 0.000* | 0.007* | 0.000* |
| $f_{11}$ | 0.010* | 0.001* | 0.002* | 0.062 | 0.047* | 0.000* | 0.517 | 0.010* |
| $f_{12}$ | 0.018* | 0.002* | 0.008* | 0.098 | 0.226 | 0.012* | 0.015* | 0.000* |
| $f_{13}$ | 0.000* | 0.000* | 0.859 | 0.768 | 0.002* | 0.962 | 0.032* | 0.000* |
| $f_{14}$ | 0.089 | 0.129 | 0.023* | 0.048* | 0.189 | 0.404 | 0.451 | 0.586 |
| $f_{15}$ | 0.677 | 0.008* | 0.029* | 0.624 | 0.603 | 0.733 | 0.012* | 0.000* |
| $f_{16}$ | 0.058 | 0.360 | 0.014* | 0.266 | 0.631 | 0.536 | 0.053 | 0.004* |
| $f_{17}$ | 0.154 | 0.866 | 0.091 | 0.003* | 0.316 | 0.547 | 0.000* | 0.000* |
| $f_{18}$ | 0.057 | 0.281 | 0.066 | 0.423 | 0.025* | 0.047* | 0.157 | 0.000* |
| Percentage (%) | 38.89 | 50.00 | 61.11 | 66.67 | 61.11 | 44.44 | 57.14 | 64.29 |

*N/A* means the features that are not available

*indicates $P \leq 0.05$

age group. Since the number of expression categories of the FACES is six, while that of the Lifespan database is two, obviously it is easier to classify two expressions than to classify six expressions.

To further analyze age effect on expression recognition, we conduct the above within age group and cross age group facial expression recognition experiment for twenty times, and employ Wilcoxon test [22] to investigate whether there are significant differences between the performance of within age group and cross age group. Wilcoxon signed-rank test is a

**Table 4** Facial expression recognition of within age group and cross age group on two databases

| DB | Train Group | Test Group | | |
|---|---|---|---|---|
| | | 18–31 (%) | 32–59 (%) | 60–93 (%) |
| FACES | 18–31 | 76.56 | 73.44 | 58.94 |
| | 32–59 | 72.39 | 74.44 | 62.50 |
| | 60–94 | 71.28 | 75.39 | 63.00 |
| Lifespan | 18–31 | 93.30 | 84.57 | 83.00 |
| | 32–59 | 90.32 | 91.89 | 84.18 |
| | 60–93 | 91.77 | 90.91 | 87.55 |

nonparametric method and can be used to assess whether the population means of the paired samples' rank differ. The null hypothesis (H0) is that the difference between two age groups comes from a distribution whose median is zero. The alternative hypothesis (H1) is that the difference between two age groups comes from a distribution whose median is not zero. The significance level is set at 0.05 in our work. The results is that all the p-value are 1.9E-6, which is much lower than the significant level 0.05. It means the age influence on facial expression recognition is statistically significant.

## 4 Expression recognition enhanced by ages

We propose two methods to recognize expressions by modeling age-related spatial expression patterns. One is a three-node Bayesian Network to classify expressions with the help of age from person-independent geometric features. During training, we construct a full probabilistic model of features, age groups, and expression labels. During testing, we can infer the posterior probability of expression labels given geometric features by marginalizing over ages. For such a method, the age-related spatial patterns are represented in geometric features. The other is to construct multiple Bayesian networks to explicitly capture the spatial patterns embedded in expressions from feature points for different ages. During training, the age-related spatial patterns are modeled through structure and parameter learning of multiple Bayesian networks. During testing, only feature points are provided, and the samples are classified into expressions according to the BN with the largest likelihood. For such method, the spatial expression patterns are represented in the structure and parameters of learned BNs. The framework of our proposed method are shown in Fig. 4.

### 4.1 3-node Bayesian network for age-augmented expression recognition

The proposed 3-node Bayesian network for expression recognition enhanced by age is shown in Fig. 5b.

During training, we construct a full probabilistic model $P(x, x^\star, y)$ by using the training set $(x_i, x_i^\star, y_i), i = 1, ..., l$, where $x_i$ is geometric features, $x_i^\star$ is age information, and $y_i$ is expression label. The label prior probability $P(y = k)(k = 1, 2, \cdots, m)$ and the Conditional Probability Distribution(CPD) $P(x|y = k)$ and $P(x|y = k, x_i^\star)$ are estimated through the Maximum Likelihood Estimation (MLE) [16] method from the training data $(x_i, x_i^\star, y_i), i = 1, ..., l$, where $m$ is the number of expressions, and $l$ is the number of training samples. During testing, the posterior probability $P(y = k|x)$ is computed for each class $y$, and the class is recognized as the one with the highest posterior probability, according to (1):

$$\begin{aligned} y^\star &= \underset{k}{argmax}\, P(y = k|x) \\ &= \underset{k}{argmax}\, \frac{\sum_{x^\star} P(y = k, x, x^\star)}{P(x)} \\ &= \underset{k}{argmax}\, \frac{P(y = k) \sum_{x^\star} P(x^\star|y = k) P(x|x^\star, y = k)}{P(x)} \end{aligned} \tag{1}$$

where $P(x^\star|y = k)$ is a tabular probability and the CPD $P(x|x^\star, y = k)$ can be represented as Gaussian distribution: $P(x|x^\star, y = k) \sim \mathcal{N}\left(x|\mu_i^{(k)}, \Sigma_i^{(k)}\right) (i = 1, 2, \cdots, n)$ for each

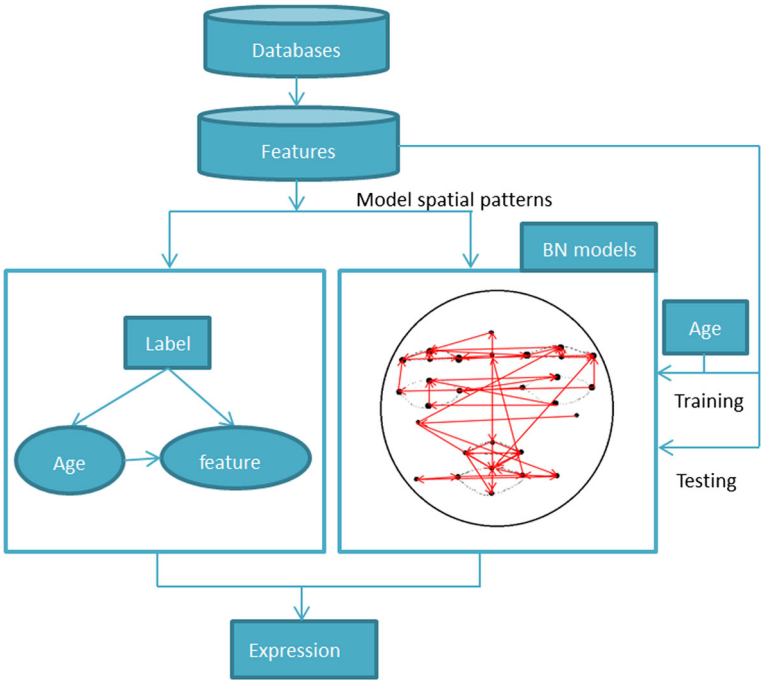


**Fig. 4** The flowdiagram of our methods

given value of $x^\star$, suppose $x^\star$ has $n$ states. In our work, $n$ represents the number of age groups.

It is clear from (1) that $x^\star$ is encoded into $p(y|x)$. Furthermore, according to the definition of mixture Gaussian, we find that $P(x|y=k) = \sum_{x^\star} P(x^\star|y=k)P(x|x^\star, y=k)$ follows a mixture of Gaussian distribution, while $P(x|y=k)$ of the native two-node BN

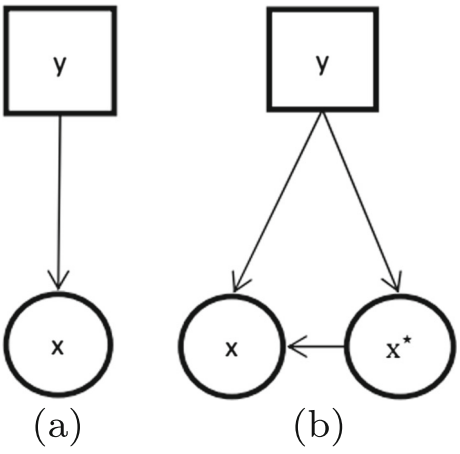


**Fig. 5** Two kinds of Bayesian Network. **a** two-node BN; **b** three-node BN

structure (shown in Fig. 5a) often obeys a single Gaussian distribution. Since mixture Gaussian distribution can fit the data better than a single Gaussian distribution, this three-node BN structure with discrete $x^\star$ can better the class distribution of $P(x|y)$.

### 4.2 Expression recognition by modeling age-related spatial patterns using multiple BNs

The proposed multiple BNs for expression recognition by modeling age-related spatial patterns are shown in Fig. 3. As a directed acyclic graph, a BN represents a joint probability distribution among a set of variables. In this figure, each node of a BN represents the coordinates of a feature point, and the links between nodes and their conditional probabilities capture the probabilistic dependencies among the feature points. The BN hence captures the spatial relationships among facial landmark points. We further assume the spatial relationships vary with facial expression and age. Different BNs are constructed to capture the spatial facial patterns under different age and expression.

In our work, the age group information is regarded as privileged information, thus $m \times n$ BN models $G_c, c = 1, \ldots, m \times n$ are established during training, where $m$ is the number of expressions, and $n$ is the number of age groups. For every BN model $G_c$, the learning procedure includes structure learning and parameter learning from the training data set $x_c = (x_{ci})_{i=1}^{l_c}$ where $x_{ci} = \left(f_{ci}^1, f_{ci}^2, \ldots, f_{ci}^p\right)$, and $p$ is the dimension of features. The structure learning is to find the network with the highest score, so that the learned network can represent the training data $x_c$ best. In our work, the Bayesian Dirichlet equivalence(BDe) criterion score function is adopted [5], as defined in (2). Supposing that the prior probability of $G_c, c = 1, \ldots, m \times n$ are uniform, we get $P(G_c|x) \propto P(x|G_c)$ when testing on the test set $x$. For the continuous nodes, the local probability distribution are linear Gaussian of the continuous parents. The parameters for each node are defined as $f_j \sim N\left(b_j + W_j^T Pa(f_j), \delta_j^2\right) (j = 1, \ldots, p)$, where $Pa(f_j)$ is the states of node $f_j$'s parents, $W_j$ is the regression coefficients, $b_j$ is the regression intercept, and $\delta_j^2$ is the variance. We use $\theta_c$ to represent the parameters given $G_c$.

$$\begin{aligned} Score(G_c) &= \log P(x|G_c) \\ &= \max_{\theta_c} \log P(x|G_c, \theta_c) \end{aligned} \tag{2}$$

Given the score function, the search strategy *greedy search with random restarts* [12] was employed to learn $G_c$. After the BN structure is constructed, the parameters can be learned from the training data. The parameter learning is to determine the conditional probability of each node given the structure of Bayesian Network. And we use Maximum Likelihood Estimation (MLE) method to estimate the parameters:

$$\theta_c^* = arg\max_{\theta_c} \log P(x|\theta_c), \tag{3}$$

where $\theta_c$ denotes the parameter set for $c_{th}$ BN model. The algorithms of BN structure and parameter learning for continuous variables are already implemented in DEAL package [4]. In our experiment, we employed the DEAL directly. After training, the learned BNs capture the spatial patterns embedded in expressions respectively given age groups.

During testing, the posterior probability of every testing sample represents the fitness on each BN model. And the sample is given the label of the BN that best fits the sample. Thus

we use the following equation to classify the testing set into expression with the maximum log-likelihood:

$$\begin{aligned} c^{\star} &= arg \max_{c\in[1,m\times n]} \frac{P(E_T|G_c)}{Complexity(G_c)} \\ &= arg \max_{c\in[1,m\times n]} \frac{\prod_{j=1}^{p} P_c(F_j|pa(F_j))}{Complexity(G_c)} \\ &\propto arg \max_{c\in[1,m\times n]} \sum_{j=1}^{p} \log(P_c(F_j|pa(F_j))) \\ &\quad - \log(Complexity(G_c)), \end{aligned} \tag{4}$$

where $E_T$ represents the features of a sample, $G_c$ stands for the $c_{th}$ model where $c$ ranges from 1 to $m \times n$, $P(E_T|G_c)$ denotes the likelihood of the sample given the $c_{th}$ model, $F_j$ is the $j_{th}$ node in the BN, and $pa(F_j)$ denotes the parent nodes of $F_j$, and $Complexity(G_c)$ represents the complexity of $G_c$. Because of the diversity among different spatial structures, the model likelihood $P(E_T|G_c)$ will be divided by the model complexity for balance. In our work, the total number of the links in BN is used as the model complexity.

## 5 Experiments and analyses

To validate our proposed methods, expression recognition experiments are conducted on the FACES and Lifespan databases, and ten-fold subject-independent cross validation is adopted. For both methods, two experiments are conducted, one is to recognize expressions without considering age information, denoted as Exp model, and the other is to recognize expression using age information as privileged information, denoted as Exp_age model. For the first method, two-node BN is used as Exp model, and our proposed three-node BN is adopted as Exp_age model. For the second method, Exp model is performed by constructing $m$ BNs using samples for each expression category, while Exp_age model is conducted by constructing $m \times 3$ BN models to recognize expressions using samples for each age group respectively. Thus, we can obtain 6 Exp models and 18 Exp_age models on the FACES database and 2 Exp models and 6 Exp_age models on the Lifespan database. Figure 3a, b show a example of BN model on the FACES and Lifespan database respectively.

Experimental results on the FACES and Lifespan databases are shown in Tables 5 and 6 respectively. From Tables 5 and 6, we can find follows:

First, for both methods, experimental results demonstrate clear performance improvement with the help of age information, since the accuracy and F1-score of Exp_age model are higher than those of Exp model in most cases. Specifically, for the first method, the average accuracy increases 0.3 percent on the FACES database as well as 1.0 percent on the Lifespan database, and the F1-score increases 1.0 percent on both databases by using age information as privileged information. For the second method, the average accuracy is improved by 0.7 percent on the FACES database and 0.5 percent on the Lifespan database, and the F1-score increases 2.5 percent and 0.6 percent on the FACES and Lifespan database respectively. It indicates that by modeling the age-related spatial patterns embedded in expressions, our proposed methods not only improve the recognition accuracy, but also make the recognition results more balanced.

**Table 5** Experimental results on FACES

| Experiment | Parameter | Method | Anger | Disgust | Fear | Happy | Neutral | Sadness | Average |
|---|---|---|---|---|---|---|---|---|---|
| First | Accuracy(%) | Exp | 89.82 | 86.53 | 94.01 | 96.00 | 88.52 | 87.03 | 90.32 |
| | | Exp_age | 90.02 | 88.22 | 93.71 | 96.11 | 88.22 | 87.43 | 90.62 |
| | | Guo's+SVM | 91.42 | 90.62 | 94.31 | 97.00 | 89.12 | 88.22 | 91.78 |
| | | Guo's+BN | 88.22 | 87.53 | 94.41 | 96.11 | 85.53 | 85.63 | 89.62 |
| | F1-score(%) | Exp | 62.22 | 66.50 | 83.70 | 88.83 | 63.72 | 54.86 | 69.97 |
| | | Exp_age | 65.28 | 68.28 | 82.64 | 89.08 | 65.09 | 55.94 | 71.05 |
| | | Guo's+SVM | 73.78 | 72.02 | 83.19 | 91.18 | 67.07 | 64.24 | 75.25 |
| | | Guo's+BN | 56.93 | 63.77 | 83.13 | 89.01 | 62.63 | 54.72 | 68.37 |
| Second | Accuracy(%) | Exp | 94.61 | 93.11 | 98.40 | 99.00 | 94.21 | 90.72 | 95.00 |
| | | Exp_age | 94.81 | 94.31 | 98.7 | 99.20 | 94.61 | 92.81 | 95.74 |
| | | Guo's+SVM | 92.81 | 93.81 | 96.51 | 99.20 | 90.52 | 90.42 | 93.88 |
| | | Guo's+BN | 89.72 | 89.82 | 96.31 | 98.60 | 86.13 | 83.93 | 90.75 |
| | F1-score(%) | Exp | 84.66 | 76.77 | 95.32 | 97.06 | 83.71 | 70.66 | 84.70 |
| | | Exp_age | 84.80 | 82.67 | 96.12 | 97.59 | 83.83 | 78.31 | 87.22 |
| | | Guo's+SVM | 78.82 | 81.10 | 89.68 | 97.60 | 72.46 | 69.81 | 81.58 |
| | | Guo's+BN | 69.25 | 70.18 | 89.21 | 95.76 | 53.51 | 54.65 | 70.93 |

Second, the method of multiple BNs outperforms three-node BN method on both databases with higher accuracy and F1-score. It may indicate that the age-related spatial patterns represented by links and parameter of BNs may be more effective in capture expression spatial pattern than those represented in geometric features.

Third, when comparing the results on two databases, we find that the performance of the Lifespan database is better than that of the FACES database. This further proves that multi-class recognition is more challenging than binary classification.

Finally, we find that for both methods, the improvement margin of disgust and sad expression is the biggest. We think this is because that the baseline performance of these two expressions are lower than other expressions, so it is easier to achieve an improvement.

## 6 Comparison with related work

We compare our methods with the most related work, Guo et al's work [11]. Since Guo et al. use Gabor features, not geometric features, we can not compare our experimental results with theirs directly. So we perform a comparison experiment by using their recognition method [11] and our features. Guo et al's proposed to perform age group classification and facial expression recognition jointly. Specifically, each expression in each age group is considered as one independent class. Thus, the number of classifiers is equal to the product of the number of expressions and the number of age groups, and a multi-class classification is performed. In our work, we use two classifier to conduct experiment, one is SVM (Support Vector Machine) [17], the other is two-node Bayesian network. The experimental results are shown in Tables 5 and 6, denoted as Guo's.

**Table 6** Experimental results on Lifespan

| Experiment | Parameter | Method | Happy | Neutral | Average |
|---|---|---|---|---|---|
| First | Accuracy(%) | Exp | 91.38 | 91.38 | 91.38 |
| | | Exp_age | 92.46 | 92.46 | 92.46 |
| | | Guo's+SVM | 91.02 | 91.02 | 91.02 |
| | | Guo's+BN | 91.02 | 91.02 | 91.02 |
| | F1-score(%) | Exp | 85.94 | 93.78 | 89.86 |
| | | Exp_age | 87.43 | 94.61 | 91.02 |
| | | Guo's+SVM | 85.60 | 93.47 | 89.54 |
| | | Guo's+BN | 84.60 | 94.67 | 89.13 |
| Second | Accuracy(%) | Exp | 96.05 | 96.05 | 96.05 |
| | | Exp_age | 96.53 | 96.53 | 96.53 |
| | | Guo's+SVM | 94.85 | 94.85 | 94.85 |
| | | Guo's+BN | 92.46 | 92.46 | 92.46 |
| | F1-score(%) | Exp | 93.54 | 97.15 | 95.35 |
| | | Exp_age | 94.50 | 97.46 | 95.98 |
| | | Guo's+SVM | 91.62 | 96.28 | 93.95 |
| | | Guo's+BN | 86.45 | 94.77 | 90.61 |

From the tables, we can find for both databases, the proposed multiple BN method outperforms Guo's in terms of both accuracy and F1-score despite using SVM or BN. Specifically, the average accuracy and F1-score of our method is 4 percent and 5 percent higher than Guo's by using BN on Lifespan database. And for the FACES database, ours is 5 percent and 17 percent higher on the accuracy and F1-score separately. Likewise, Table 6 shows that modeling spatial pattern for each expression in each age group generally improves both the accuracy and F1-score by 2 percent when applying SVM in Guo's method. What's more, compared to Guo's by using SVM, the average F1-score of our method is 2.0 percent and 6.0 percent higher on the FACES database. This further demonstrates that our proposed multiple BN models systematically captures the age-related spatial patterns embedded in expressions. This also empirically shows that spatial expression pattern is more discriminative than appearance pattern.

The performance of the proposed three-node BN method is better than that of Guo's not only on the FACES database but also on the Lifespan database when using Bayesian Network. This indicates that our method is really better than Guo's when applying the same classifier. However, when using SVM, our method is comparable to Guo's, since it is superior to Guo's on the Lifespan database, but not on the FACES database. This is because as a discriminative classifier SVM is stronger than the generative Bayesian Network classifier.

The above comparison demonstrates the advantages of our approaches compared with state of the art. Our approaches can successfully capture the age-related spatial patterns embedded in expressions through the parameters and structure of Bayesian networks. The age information, which is available during training, further enhance expression classifiers.

As discussed in Section 1, both Guo et al. [11] and Alnajar et al.[1] conducted experiments on the FACES database and the Lifespan database. Although the former focused on

**Table 7** The accuracy of facial expression recognition in [11] and [1]

| Database | [11] | [1] | ours_3-node BN | ours_multi-BNs |
|---|---|---|---|---|
| Lifespan | 96.79 % | 93.68 % | 92.46 % | 96.53 % |
| FACES | 97.89 % | 92.19 % | 90.62 % | 95.74 % |

age-invariant expression recognition, and the latter considered expression-invariant age estimation, they both adopted appearance features, i.e. Gabor features [11] and LBP features [1] respectively, and provided expression recognition results as shown in Table 7. From this table, we can find that the expression recognition performance of our method using geometric features are comparable with those using texture features. It further demonstrates the importance of spatial patterns for expression recognition.

## 7 Conclusion and future Work

Current studies of facial expression recognition pay little attention to the age effect on the performance of expression recognition. In this paper, we propose to enhance expression recognition by modeling age-related spatial expression patterns. First, we conduct two statistical analyses to investigate the age effect on spatial patterns of expressions and on facial expression recognition respectively. Analysis results demonstrate that the spatial expression patterns are significantly different among age groups, and age information has a significant effect on the facial expression recognition. Second, we propose two methods to recognize expression with the help of age. One is a three-node Bayesian Network to classify expressions from person-independent geometric features. The age-related spatial patterns are represented in geometric features. The other is to construct multiple Bayesian networks to explicitly capture the spatial patterns embedded in expressions from feature points for different ages. The spatial expression patterns are represented in the structure and parameters of learned BNs. For both methods, age information is used as privileged information, and is exploited during training to construct a better classifier. Experimental results on two databases demonstrated the power of the proposed model in capturing age-related spatial patterns embedded in expressions as well as its advantage over existing approaches for expression recognition.

In addition to age-related spatial patterns, age-related temporal patterns is crucial for expression recognition. This work only exploits age-related spatial patterns embedded in expressions. Therefore, we will further investigate age-related temporal patterns for expression recognition in the future. Furthermore, we will also consider combing spatial and appearance expression pattern for expression recognition. Although age recognition and facial expression recognition are typically done separately and independently, they may help each other. Specifically, as demonstrated in our paper, age information could help expression recognition, expression information may also help age recognition. Therefore, another possible future work is to use expression as privilege information to improve age recognition. Currently, only two benchmark facial expression databases, i.e. the Lifespan and the FACES, contain a large range of age variations. A large scale facial expression database with multi-ethnic, multi-age, multi-personality, and multi-occupation subjects should be constructed, since the size and the diversity of a database are crucial for the research of expression recognition.

## References


1. Alnajar F, Lou Z, Alvarez J, Gevers T Expression-invariant age estimation
2. Bayramoglu N, Zhao G, Pietikainen M (2013) Cs-3dlbp and geometry based person independent 3d facial action unit detection. In: 2013 International Conference on Biometrics (ICB). IEEE, pp 1–6
3. Bennett J, Fisher RA (1995) Statistical methods, experimental design, and scientific inference. Oxford University Press
4. Bøttcher SG, Dethlefsen C Deal: a package for learning bayesian networks
5. Cooper GF, Herskovits E (1992) A bayesian method for the induction of probabilistic networks from data. Mach Learn 9(4):309–347
6. Dibeklioglu H, Salah AA, Gevers T (2015) Recognition of genuine smiles. IEEE Trans Multimedia 17(3):279–294
7. Ebner NC, Johnson MK (2010) Age-group differences in interference from young and older emotional faces. Cognition and Emotion 24(7):1095–1116
8. Ebner NC, Riediger M, Lindenberger U (2010) Faces-a database of facial expressions in young, middle-aged, and older women and men: development and validation. Behav Res Methods 42(1):351–362
9. Ekman P (1993) Facial expression and emotion. American psychologist 48(4):384
10. Fölster M, Hess U, Werheid K (2014) Facial age affects emotional expression decoding. Frontiers in psychology, vol 5
11. Guo G, Guo R, Li X (2013) Facial expression recognition influenced by human aging
12. Heckerman D (2008) A tutorial on learning with bayesian networks. In: Innovations in Bayesian Networks. Springer, pp 33–82
13. Hess U, Adams RB, Simard A, Stevenson MT, Kleck RE (2012) Smiling and sad wrinkles: Age-related changes in the face and the perception of emotions and intentions. J Exp Soc Psychol 48(6):1377–1380
14. Houstis O, Kiliaridis S (2009) Gender and age differences in facial expressions. Eur J Orthod 31(5):459–466
15. John PWM, John PW (1971) Statistical design and analysis of experiments. SIAM
16. Koller D, Friedman N (2009) Probabilistic graphical models: principles and techniques. MIT press
17. Lee Y, Lin Y, Wahba G (2001) Multicategory support vector machines. In: Proceedings of the 33rd Symposium on the Interface. Citeseer
18. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) 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 (CVPRW). IEEE, pp 94–101
19. Minear M, Park DC (2004) A lifespan database of adult facial stimuli. Behav Res Methods Instrum Comput 36(4):630–633
20. Pantic M, Rothkrantz LJM (2000) Automatic analysis of facial expressions: the state of the art. IEEE Trans Pattern Anal Mach Intell 22(12):1424–1445
21. Shan W, Shangfei W, Jun W (2015) Enhanced facial expression recognition by age. FG2015 accepted
22. Siegel S (1956) Nonparametric statistics for the behavioral sciences
23. Vapnik V, Vashist A (2009) A new learning paradigm: learning using privileged information. Neural Netw 22(5):544–557
24. Zeng Z, Pantic M, Roisman GI, Huang TS (2009) A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Trans Pattern Anal Mach Intell 31(1):39–58

**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 in 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 senior member and ACM member. Her research interests cover affective computing and probabilistic graphical models. She has authored or co-authored over 70 publications.

**Shan Wu** received her BS in computer science from Anhui University of Technology in 2014, and she is currently pursuing her MS in Computer Science in the University of Science and Technology of China, Hefei, China. Her research interesting is affective computing.

**Zhen Gao** received his BS in computer science from Nanjing University of Science and Technology in 2013, and he is currently pursuing his MS in Computer Science in the University of Science and Technology of China, Hefei, China. His research interesting is affective computing.

**Qiang Ji** received his Ph.D degree 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). From 2009 to 2010, he served as a program director at the National Science Foundation (NSF), Arlington, VA, USA, 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, Urbana, IL, USA; the Robotics Institute at Carnegie Mellon University, Pittsburgh, PA, USA; the Dept. of Computer Science at University of Nevada, Reno, Nevada, USA; and the Air Force Research Laboratory, Rome, NY, USA. Prof. 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, machine learning, and their applications in various fields. He has published over 200 papers in peer-reviewed journals and conferences, and has received multiple awards for his work. Prof. 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 for numerous international conferences/workshops. Prof. Ji is a fellow of the IEEE and the IAPR.