Journal of Computational Science 25 (2018) 289–297

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Macro-to-micro transformation model for micro-expression recognition

Xitong Jia [a], Xianye Ben [a,*], Hui Yuan [a], Kidiyo Kpalma [b], Weixiao Meng [c]

[a] School of Information Science and Engineering, Shandong University, Jinan 250100, PR China
[b] IETR, Institut National des Sciences Appliquées de Rennes, Rennes 35708, France
[c] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, PR China

ARTICLE INFO

Article history:
Received 15 December 2016
Received in revised form 4 March 2017
Accepted 21 March 2017
Available online 24 March 2017

Keywords:
Micro-expression recognition
Macro-to-micro transformation model
Feature selection
Singular value decomposition

ABSTRACT

As one of the most important forms of psychological behaviors, micro-expression can reveal the real emotion. However, the existing labeled training samples are limited to train a high performance model. To overcome this limit, in this paper we propose a macro-to-micro transformation model which enables to transfer macro-expression learning to micro-expression. Doing so improves the efficiency of the micro-expression features. For this purpose, LBP and LBP-TOP are used to extract macro-expression features and micro-expression features, respectively. Furthermore, feature selection is employed to reduce redundant features. Finally, singular value decomposition is employed to achieve macro-to-micro transformation model. The experimental evaluation based on the incorporated database including CK+ and CASME2 demonstrates that the proposed model achieves a competitive performance compared with the existing micro-expression recognition methods.

1. Introductions

Emotion is an expression form of people's psychological activity, which directly reflects people's current psychological and mental states [1]. Under high pressure, people in stimulating source, such as the moment of effective stimulation source of voice, text, images, video and others, undergo rapid psychological changes. This is accompanied by short-term emotional and a subconscious revelation that does not appear under the control of thinking, called "micro-expression" [2]. Emotion is an expression of the inner psychological state, and micro-expression is the physical external manifestations of emotion. Emotion acts like the link that combines closely together the psychological state and the external manifestation. These three elements are interdependent and closely related. This makes it possible for us to judge people's emotions by studying the psychological behavior. So how to interpret the psychological behavior has become an important issue. In traditional sense, the research on micro-expression of psychological behavior depends on the professional personnel trained to judge and interpret, which inevitably adulterates a lot of subjective emotional factors. In this paper, we will try to analyze people's real emotional and psychological states through micro-expression signals relying on computer vision, pattern recognition and machine learning methods, which has a very important influence on theory and application.

As one of the most important forms of psychological behaviors, micro-expression is the revelation of a rapid and transient facial expression [2]. Micro-expression reflects patterns of facial muscle movement during various feelings periods and occurs with characteristics of short continued period, low intensity, and high difficulty to induce [3], which makes it one of the most potential physiological characteristics in the field of spiritual psychology and sentiment analysis.

Although the duration of the micro-expression is short, it can reveal the true feelings of the heart, so as to provide a reliable basis to judge the spiritual state of people. At present, in psychology, micro-expressions are generally classified as six basic types named anger, fear, disgust, sadness, happiness, surprise [4]. Due to the short duration of expression (1/25 s–1/5 s [5]), the weakness of intensity (only reflected on part of the facial motor unit [6]) and other factors, we face difficulties in micro-expression detection and recognition. However, with the rapid development of machine learning algorithms [7–12] and face recognition algorithms [13–15], micro-expression research has been paid more and more attention.

At present, the existing micro-expression databases (CASME I [16], CASME II [17], SMIC [18], Polikovsky [19,20]) of approximately seven kinds of micro-expressions. It is easy to cause the

* Corresponding author at: School of Information Science and Engineering, Shandong University, No. 27, Shanda South Road, Jinan 250100, China.
E-mail address: benxianye@gmail.com (X. Ben).

https://doi.org/10.1016/j.jocs.2017.03.016
1877-7503/

researchers to draw inaccurate conclusion of the psychological state on the subject being observed and not to efficiently find problems or avoid risks. Considering the above reasons, how to solve automatic labeling and recognition problems for the tested micro-expressions becomes a challenging task under a small number of labeled training micro-expression samples condition.

On the contrary, there are large amounts of macro-expression databases, each of which consists of vast labeled training samples compared with micro-expression databases. For instance, the CK+ database, expanded from Cohn-Kanade Dataset, includes 123 subjects, 593 image sequences. The last frame of each image sequence has the label of action units. In addition, among the image sequences, there are 327 sequences which have emotion labels, much more than that in the existing micro-expression databases. Thus, how to take advantage of the macro-expression databases at the area of micro-expression recognition has become an important direction of the research.

Transfer learning [21], a kind of method solves the question that how to take advantage of limited data, it can transfer knowledge learned from the existing data to help the studying in the future. The purpose of transfer learning is to help learning tasks in a new environment by knowledge learned from an available environment, so it won't make the same distribution assumption, as machine learning does. Thus, by using the main idea of transfer learning, it is able to take advantage of the quantitative superiority of macro-expression to recognize the micro-expression.

In this paper, we proposed a macro-to-micro transformation model for micro-expression recognition, which aims at solving the difficulty of micro-expression recognition due to the limitation of data amounts. The proposed model combined LBP features of macro-expression and LBP-TOP features of micro-expression. Furthermore, feature selection was employed in the proposed model for removing redundant information. Finally, singular value decomposition is employed to achieve macro-to-micro transformation model. Extensive experiments on the popular CK+ macro-expression database and CASME2 micro-expression databases have shown that the proposed model outperforms the existing micro-expression recognition methods.

The macro-to-micro-expression model proposed in this paper established a bridge of relationship between macro-expression and micro-expression, which creatively combined the two kinds of features during the training process, and realized the micro-expression recognition by means of a linear transfer learning. Owing to the participation of macro-expression, the training sample set has been expanded and the available information is increased, thus the micro-expression recognition performance significantly improved.

The rest of this paper is organized as follows. Section 2 introduces the related works on recognition of micro-expression. The proposed macro-to-micro transformation model will be described in Section 3. Section 3 also includes the representation and selection of macro-expression and micro-expression features. Experimental results and analysis are given in Section 4. Section 5 concludes with a summary and a further discussion.

2. Related works

During the past years, many researches of micro-expression recognition appeared, most of which discussed the problem of micro-expression detection and recognition respectively. Given an expression video sequence, the purpose of detection is to judge whether the sequence consists of any micro-expressions, while recognition is to estimate the classification of micro-expressions appeared in the sequence. Fig. 1 shows the existing micro-expression recognition methods, which are mainly divided into two categories: feature representation and multi-linear subspace learning methods.

The feature representation methods are used to find efficient and robust micro-expression features and use them in the original micro-expression video sequence. At present, the main micro-expression features include: Gabor features [22,23], 3D HOG features [19], optical flow features [24,25] and LBP or LBP-TOP texture features [26–29]. Gabor features methods are based on Gabor filters for representing face images in different directions and scales of the features; this primitive method can only deal with facial muscles which have great changes of expressions. Thus it is not the true sense of the detection and identification [22,23]. 3D HOG method [19] captures micro-expression video sequences with a high-speed video camera and extracts 3D gradient histograms of ROI according to Facial Action Coding System (FACS). Optical flow method [24,25] extracts muscle movement information according to optical flow changes of facial action unit caused by muscle movement when micro-expression occurs. The classification and identification of the micro-expressions can be processed by this information. LBP-TOP [26–29] is a kind of texture features to describe the micro-expression for the extraction of micro-expression from video stream texture features to be a sequence of three directions, using micro-expression feature fusion in three directions in the classification process. These methods greatly enhance the effect of micro-expression, but due to micro-expressions' short duration and their low intensity, these methods are not robust in the micro-expression identification and classification.

Subspace learning method is based on the needs of micro-expressions in video sequences interpolation into a certain number of frames, and then directly handle the stream video sequence of micro-expressions, which can mainly be divided into two categories: tensor based subspace learning method [30–32] and sparse representation based on the tensor subspace learning theory [33]. Tensor-based subspace learning method extracts directly on the entire video sequence feature and retains the features of micro-expressions in video sequences of different directions, and to maintain the structure of micro-expression sequence, this method is vulnerable to face structure, pixel characteristics influence and may cause the wrong classification of micro-expression. In order to make feature extraction robust, the sparse technology needs to be used. Through the sparsity constraint, the micro-expressions which have been mapped have better classification ability, but this binding force is limited, still largely influenced by features of human faces, light, noise effects and it cannot really extract the micro-expression's feature robustly.

To sum up, micro-expression's duration is short and its intensity is weak. Hence the micro-expression features are vulnerable to the appearance features of faces, illuminations, noises and other factors. Thus, it results in difficulties on the feature representation and extraction of the micro-expression. At present, there are few samples in current micro-expression databases, and the number of existing methods of automatic micro-expression recognition is largely limited by the lack of labeled training samples and the difficulty of sample labeling. In this paper, a macro-to-micro transformation model was proposed to solve the problems above. The proposed model creatively used the macro-expression database for micro-expression recognition by means of transfer learning, which took advantage of the existing macro-expression database well. The model combined macro-expression features with micro-expression features in the process of training. Micro-expression samples were classified in the process of testing. Thus the relation between macro-expression and micro-expression was established to improve the micro-expression recognition accuracy.

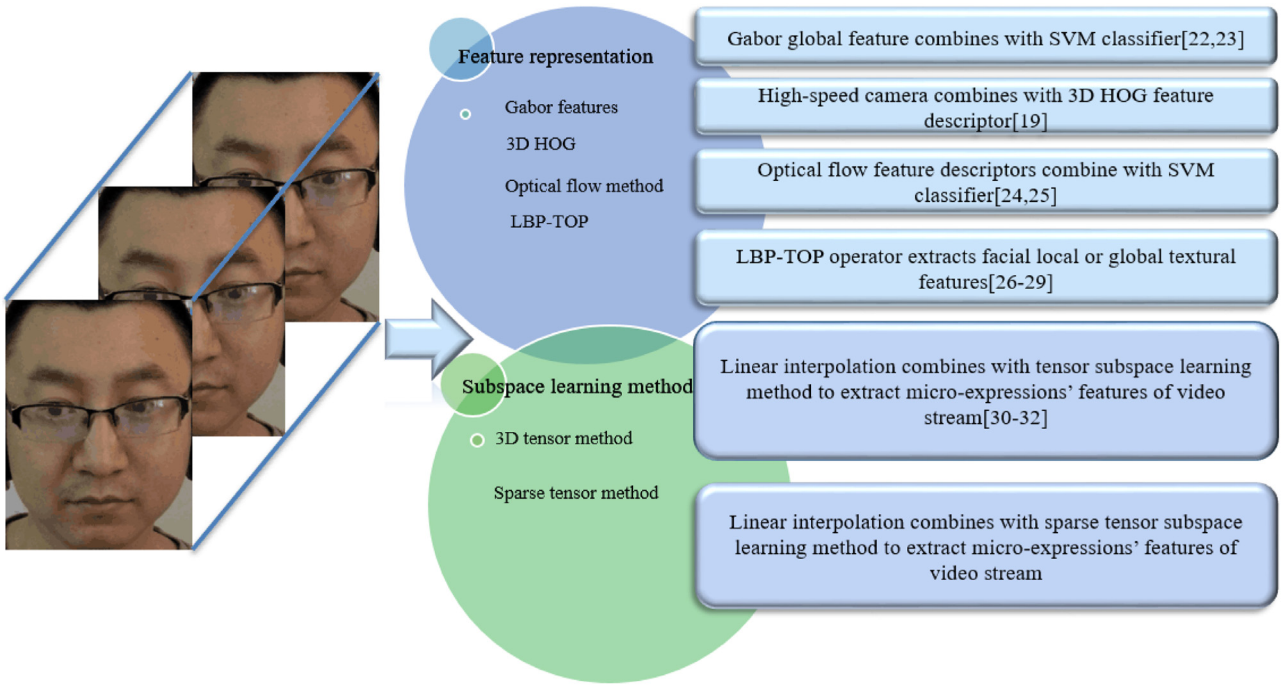


Fig. 1. Micro-expression recognition methods.

3. Macro-to-micro transformation model

In this section, the proposed macro-to-micro transformation model is introduced in detail. In the data preprocessing, Local Binary Pattern (LBP) is employed to extract the macro-expression features and Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) is employed to extract the micro-expression features from the video sequences. Then Group LASSO is used to respectively select useful features from LBP features and LBP-TOP features. Then, the two kinds of features were combined by the proposed model in the training process, thus the transformation between macro-expression and micro-expression was accomplished by SVD [34].

3.1. Macro-expression feature representation

LBP is a kind of operator to describe the local textural features of an image, which has rotation and grayscale invariance. LBP operator is defined as follows,

$$LBP_{P,R} = \sum_{k=0}^{P} 2^k s(i_k - i_c)\ ,\ where\ \ s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases} \tag{1}$$

where P represents the number of adjacent pixels, R represents the radius of the local region, i_k represents the grey value of the adjacent pixel and i_c represents the grey value of center pixel. ic

In a 3*3 region, assume that the center pixel is regarded as threshold, then the grey value of adjacent 8 pixels was compared with it respectively. If the grey value of the surrounding pixel is larger than that of the center pixel, mark the position to 1, otherwise, to 0. After the comparison, 8 pixels in the 3*3 neighborhood can produce an eight-digit binary number, and it usually converted to a decimal number called LBP code, a total of 256 kinds. Then the LBP code of the center pixel is obtained, which is used to reflect the textural information of that region.

In order to adapt to the different scales of textural feature, the region of 3*3 is expanded to an arbitrary neighborhood, and the square neighborhood was replaced by a circular neighborhood.

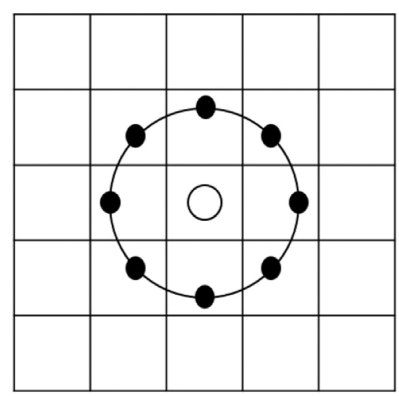


Fig. 2. the circular (8, 1) neighborhood, there are 8 adjacent pixels in a 3*3 region as the black points, the center pixel as the white point, and the radius of the local neighborhood is 1.

Thus the improved LBP operator allowed arbitrary number of pixels in the circular neighborhood within radius R. Then the LBP operator with P sampling points can be employed. Fig. 2 shows $LBP_{8,1}$.

In an actual image, the vast majority of LBP patterns consist of only twice jumps as from 0 to 1 or from 1 to 0. Thus the Uniform Pattern is defined as, when the specific circulation binary number of a LBP operator only consists of only twice jumps, the binary of the LBP is defined as a Uniform Pattern LBP.

The Uniform Pattern LBP is defined as follows,

$$LBP_{(P,R)_{u2}} = \begin{cases} \sum_{k=0}^{P-1} s(i_k - i_c),\ U \leq 2 \\ P(P-1) + 2, \text{otherwise} \end{cases} \tag{2}$$

where $U = |s(i_{P-1} - i_c) - s(i_0 - i_c)| + \sum_{k=1}^{P-1} |s(i_k - i_c) - s(i_{k-1} - i_c)|$.

In this paper, the uniform $LBP_{8,1}$ is employed to extract the macro-expression feature. The dimension of each expression sample can be reduced to 59.

3.2. Micro-expression feature representation

The macro-expression features extracted by LBP were from each frame of the sequences, which ignores the temporal information of the expression. To solve this problem, LBP-TOP can extract the concatenation of LBP histograms on 3 orthogonal planes named *XY*, *XT* and *YT*. Generally, an image sequence is considered as a stack of *XY* plane centered on time axis *T*, however, the image sequence can also be regarded as a stack of *XT* plane centered on axis *Y* or a stack of *YT* plane centered on axis *X*. Plane *XT* and *YT* provide plenty of temporal transition information, i.e., the facial movement displacement, where the position changes of eyes, nose, lips, cheek can be obtained. And plane *XY* includes the spatial information, i.e., expression, identity, emotion.

The micro-expression sequences include a video sequence which indicates several kinds of changes on one's face. The video sequence is short, and the micro-expression remains in a short duration and does not change significantly between each frame. The features extracted from the micro-expression sequences are supposed to be time-related, from which we can analyze the feature relationship between two adjacent frames. The Uniform Pattern of LBP-TOP histogram is cascaded according to the order of *XY*, *XT*, *YT*, thus the feature dimension will be 3×59 for each sample, while the dimension of macro-expression feature is 1×59 mentioned above. In order to keep both dimensions of the macro-expression and micro-expression consistent, the features of each plane are added together and the average of them is viewed as the final feature of micro-expression. Thus, the dimension of micro-expression feature is also 1×59. Fig. 3 shows the description of micro-expression feature extracted by LBP-TOP.

3.3. Feature selection using group LASSO

For macro-expression and micro-expression sequences, the final purpose is to test the performance of features transferred from macro-expression to micro-expression. Obviously, some redundant features exist in the feature vectors. For instance, the typical features used to distinguish different kinds of emotions are mostly the movement changes of eyes, lips and cheek muscles, not all changes of a whole face. Thus, some areas of the face may be redundant for the recognition.

LASSO can get the globally optimal solutions of important factors for accurate estimation, which is defined as:

$$\hat{\beta}^{LASSO}(\lambda) = \underset{\beta}{argmin}(\|Y - \beta X\|^2 + \lambda\|\beta\|_{\ell_1}), \tag{3}$$

where $\|\|_{\ell_1}$ represents the ℓ1-norm, and λ represents a parameter for adjusting.

While LASSO has great computational advantages of feature selection and wonderful performance, it aims at selecting several single variables which represent the pixels of each image, not the general factor selection. Group LASSO [35] can solve the problem of feature selection with grouped variables based on Eq. (3). Given positive definite matrices $K_1, \ldots, K_J$, the group LASSO estimation is defined as:

$$\hat{\beta}^{Group-LASSO}(\lambda) = \frac{1}{2}||Y - \sum_{j=1}^{J} X_j\beta_j||^2 + \lambda\sum_{j=1}^{J}\|\beta_j\|_{K_j} \tag{4}$$

where λ is a parameter for adjusting and $\|\beta_j\|_{K_j} = \left(\beta_j{}' K_j \beta_j\right)^{\frac{1}{2}}$,

Group LASSO is stable with kinds of features and usually reaches reasonable convergence tolerance within iterations. Thus, it is used to select LBP features of macro-expression and LBP-TOP features of micro-expression in this paper.

3.4. Macro-to-micro transformation model

Now there are two kinds of features, i.e. X_m represents a macro-expression feature, and Y_m represents a micro-expression feature, both of which are *m*-dimensional features.

We know that there are some differences between macro-expression and micro-expression features, thus a connection between them should be explored. In this way, these features are associated with the training objects and target objects for recognition. Assuming that there are *N* pairs of samples, the proposed model is constructed by arraying the training and target objects' features. Suppose the training matrix *M* is given by

$$M = \begin{bmatrix} X_m^1 \ldots X_m^N \\ Y_m^1 \ldots Y_m^N \end{bmatrix}$$

where the rows indicate the features and the columns indicate the samples of each kind of features.

For a 2×2 matrix *D*, two mutually orthogonal unit vector v_1 and v_2 are chosen to make vectors Dv_1 and Dv_2 orthogonal, and u_1 and u_2 represent the unit vector of Dv_1 and $Dv2$, respectively. $\sigma 1$and $\sigma 2$ respectively represent the mould of the two vectors with different direction, which are also called as the singular value of matrix *D*. Thus

$$\begin{aligned} Dv_1 &= \sigma_1 u_1 \\ Dv_2 &= \sigma_2 u_2 \end{aligned} \tag{5}$$

Now the representation of vector *x* after a linear transformation *D* can be described as

$$x = (v_1 \cdot x)v_1 + (v_2 \cdot x)v_2 \tag{6}$$

where vectors v_1 and v_2 are orthogonal unit vectors, which means

$$\begin{aligned} Dx &= (v_1 \cdot x)Dv_1 + (v_2 \cdot x)Dv_2 \\ Dx &= (v_1 \cdot x)\sigma_1 u_1 + (v_2 \cdot x)\sigma_2 u_2 \end{aligned} \tag{7}$$

The vector inner product can be represented with vector transposition

$$v \cdot x = v^T x$$

Thus,

$$\begin{aligned} Dx &= u_1\sigma_1 v_1{}^T x + u_2\sigma_2 v_2{}^T x \\ D &= u_1\sigma_1 v_1{}^T + u_2\sigma_2 v_2{}^T \\ &= USV^T \end{aligned} \tag{8}$$

This suggests that an arbitrary matrix *D* can be transformed into three matrix as follows,

$$\begin{aligned} D &= USV^T \\ &= [u^1 \ldots u^N] \begin{bmatrix} s^1 & 0 & \ldots & 0 \\ 0 & s^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \ldots & 0 & s^N \end{bmatrix} [v^1 \ldots v^N] \end{aligned} \tag{9}$$

where $U \in \Re^{2d \times N}$ is constructed by the concatenated features in a joint subspace, and *d* represents the dimension of each feature. $S \in \Re^{N \times N}$ is a diagonal matrix representing singular values. $V \in \Re^{N \times N}$

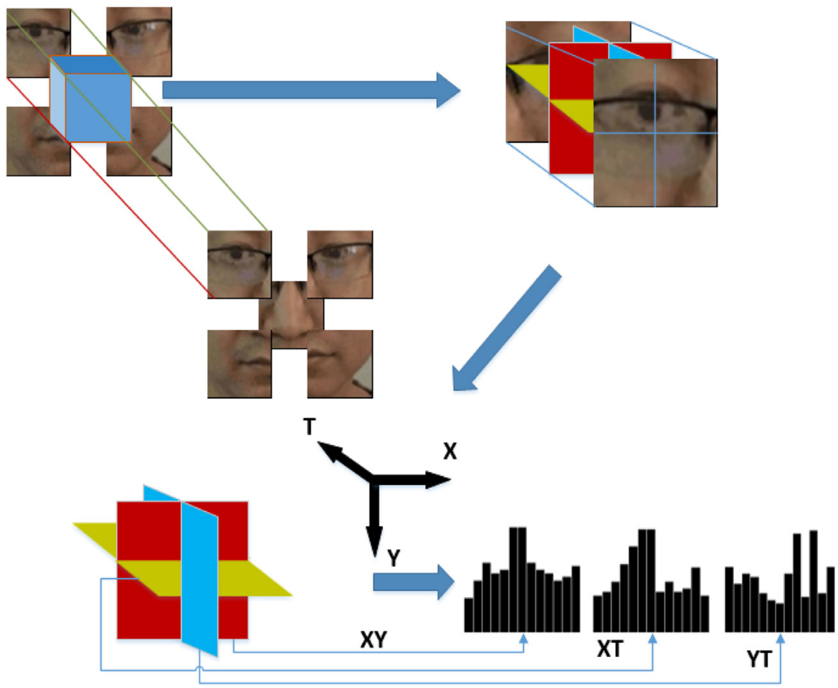


Fig. 3. Description of facial micro expression feature extracted by LBP-TOP.

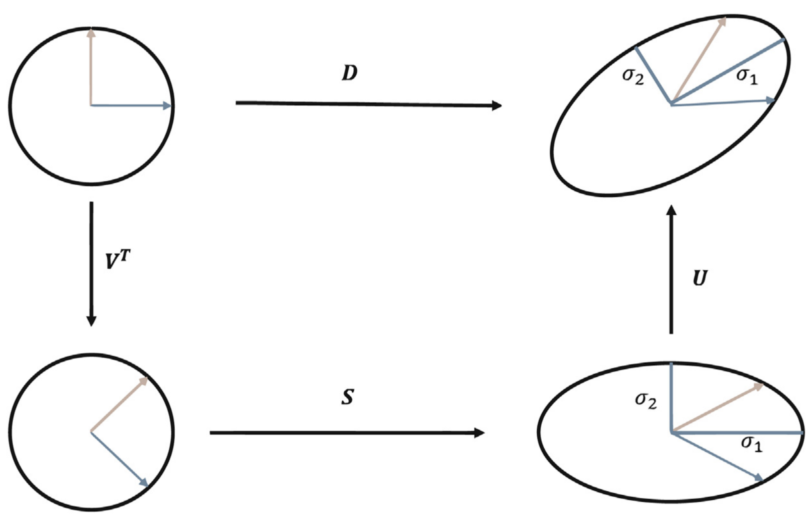


Fig. 4. Visualization of SVD.

represents the standard orthogonal basis of the subspace. Fig. 4 shows the visualization of SVD.

For the training matrix D, the two different features can be projected into a joint subspace, where they are associated with each other.

As is shown above, U consists of the two different features and it can be decomposed as follows

$$U = \begin{bmatrix} R_x \\ R_y \end{bmatrix}, \tag{10}$$

where R_x and R_y represent the sub-matrices, and they are feature-independent matrices.

Assuming that x represents a gallery macro-expression feature and y represents a probe micro-expression feature, the gallery macro-expression feature can be transformed to the probe micro-expression feature.

Let $K = SV^T$ and k^m be the m-th row vector of K, thus a point within a joint subspace λ can be calculated by

$$\begin{aligned} \lambda &= arg\ \min_k \|x - R_x k\|_2^2 \\ &= \left(R_x{}^T R_X\right)^{-1} R_x{}^T x \end{aligned} \tag{11}$$

Then the gallery micro-expression feature can be estimated by projecting R_y into the joint subspace λ

$$\hat{y} = R_y \lambda = R_y (R_x{}^T R_x)^{-1} R_x{}^T x \tag{12}$$

Finally, the nearest neighbor (NN) classifier is used to classify the probe micro-expression samples. Given a gallery macro-expression feature set $\{x_i, i = 1, \cdots, N\}$, any probe micro-expression feature sample y can be assigned to class

$$\pi_i \text{where} i = \underset{i}{argmin} dis\left(y, \hat{y}_i\right) = \underset{i}{argmin} dis\left(y, R_y(R_x{}^T R_x)^{-1} R_x{}^T x_i\right) \quad (13)$$

where π_i is the label of the gallery macro-expression sample x_i.

The proposed macro-to-micro transformation model is summarized in Algorithm 1.

Algorithm 1. Macro-to-micro transformation model

Input: LBP features X_m^i, $i = 1,2,\cdots N$ of macro-expression samples and LBP-TOP features Y_m^i, $i = 1,2,\cdots N$of micro-expression samples.
Output: The label of a probe micro-expression feature sample y
Procedure:
1: Construct the training matrix$M = \begin{bmatrix} X_m^1 \cdots X_m^N \\ Y_m^1 \cdots Y_m^N \end{bmatrix}$;
2: Decompose M to be USV^T ;
3: Extract R_x and R_y from matrix U according to Eq. (10);
4: Calculate a point within a joint subspace λ according to Eq. (11);
5: Estimate the gallery micro-expression feature according to Eq. (12);
6: Recognize the probe micro-expression feature sample y according to Eq. (13).

4. Experimental results and analysis

4.1. Database

In this paper, there are two popular databases, i.e, Extended Cohn-Kanade (CK+) [36] expression database and CASME2 micro-expression database [17] are used in the experiments.

The CK+ database contains 327 image sequences, and there are 7 kinds of expressions, i.e., anger, contempt, disgust, fear, happy, sadness and surprise in this database. There is only one frame in each sequence that is labeled as a specific expression, which is called the peak frame. For each expression sample, the last 5 frames are selected to represent the specific expression, thus we can collect more experimental expression samples.

The CASME2 database contains expression- labeled 157 image sequences in 6 kinds of micro-expressions, such as disgust, fear, happiness, repression, sadness and surprise.

In our experiments, some repetitive samples for the CASME2 micro-expression database were added to train the macro-to-micro transformation model in order to guarantee the same data amount as the selected samples from the CK+ database, because the samples in the CASME2 database are fewer than that in the CK+ database. Therefore, we can test the performance of micro-expression recognition by transfer learning from the macro-expression sample, at the same time, the number of training samples is expanded in another way. Three kinds of expressions including disgust, happiness, and surprise were selected from the CK+ expression database and micro-expression database, respectively. In addition, the facial regions across these two databases were aligned to remove the positional and scale variance based on eye positions, and further cropped to the resolution of 231*231 pixels. Each cropped image was then divided into different patches, the numbers of which were 5*5, 6*6, 7*7, 8*8, 9*9, respectively. In the next step, a uniform $LBP_{8,1}$ was used to extract the features of each patch. Therefore, a histogram with 59 bins was calculated for each patch. Thus, one expression sample in the CK+ database was transformed to a feature vector containing n*n*59-dimentional features, where n*n denotes the number of patches. Similarly, each image from sequences of the CASME2 database were cropped to the same resolution as CK+ database and also divided into the same number of patches as CK+ database. LBP-TOP was adopted to extract the micro-expression features. Although each sequence was transformed into a 3*59-dimentional vector, it was further transformed to the size of 59 dimensions by average processing of 3 planes. Thus, micro-expression sequences in the CASME2 database were transform to a feature of n*n*59 dimensions. In all, 885 samples from the CK+ database and 57 stochastic samples from the CASME2 database were selected for training. Moreover, the number of samples in the CASME2 database was fictitiously expanded to 885 by means of duplication in order to ensure the same amount between macro-expression and micro-expression samples. We used the remaining 55 samples for testing from the CASME2 database, and repeat the test randomly for 20 times. We reported the average value of all test results.

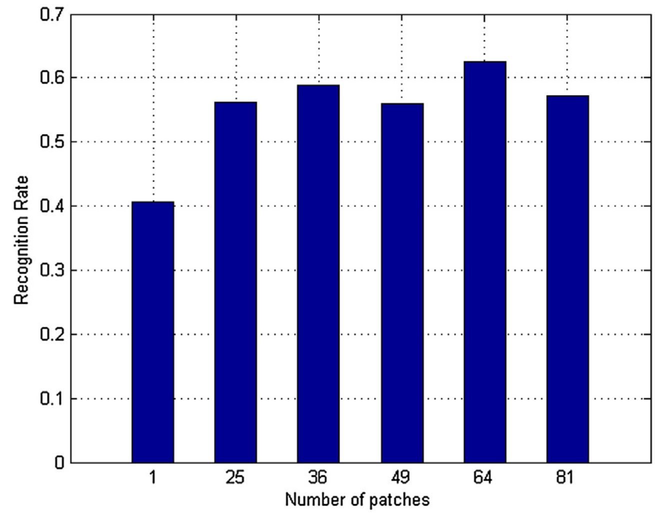


Fig. 5. Recognition performance of the proposed model with different number of patches.

4.2. Analysis of patch size results on the cross-database

The face images in the two databases were divided into 5*5, 6*6, 7*7, 8*8, 9*9 patches respectively, and the proposed method also implement feature selection with increasing amount of 40% to 90% of the original features. The results of different numbers of patches are shown in Fig. 5. We can see that the recognition accuracy is not satisfactory if there is only one patch. This is because that the holistic feature is not conducive to recognize expressions. Both of the dimensions of macro-expression and micro-expression features extracted by LBP and LBP-TOP were 59, and the features will be disturbed by the holistic redundant information. To solve the problem, each image of the sequences was divided into more patches. When the number of patches reached 8*8, the recognition accuracy reached 0.624, improved about 22% compared with the situation of one patch.

4.3. Analysis of feature dimension results on the cross-database

The influence on the recognition accuracy of the patch size is shown in Section 4.1. The division of the images cannot wipe off the redundant information. Fig. 6 shows the recognition performance of the proposed model with different feature dimensions. To reduce the redundant information, SVD was employed for dimension reduction. From Fig. 6 we can see that compared to the recognition accuracy of without dimensionality reduction, the dimensionality reduction has a particular effectiveness for micro-expression recognition in most cases.

4.4. Analysis of patch selection results on the cross-database

Table 1 shows the recognition performance of the proposed model with varying patch selection degree and different numbers

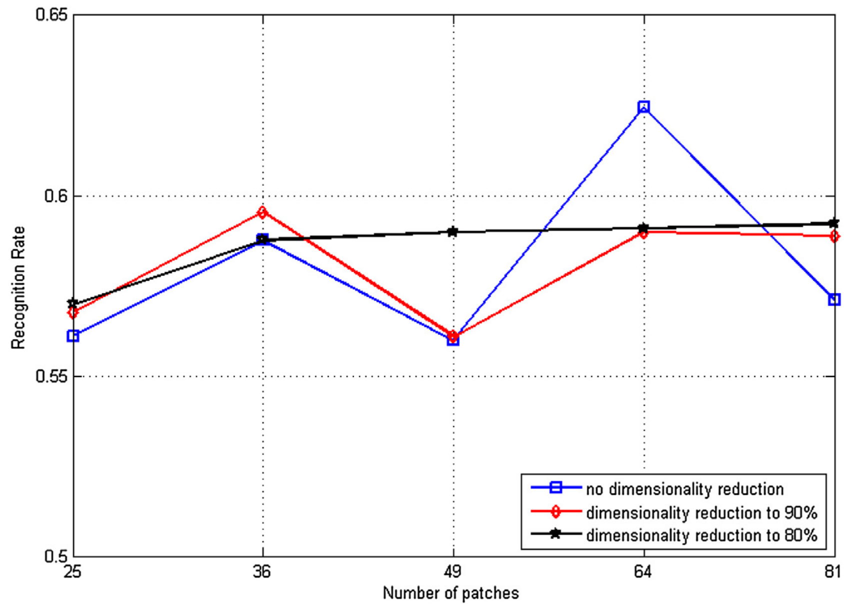


Fig. 6. Recognition performance of the proposed model with different reduced dimensions.

Table 1
Recognition rates of the proposed model with feature selection.

	90%	80%	70%	60%	50%	40%
5×5	0.586	0.589	0.576	0.563	0.565	0.591
6×6	0.600	0.608	0.611	0.611	0.613	0.589
7×7	0.588	0.562	0.599	0.612	0.588	**0.655**
8×8	0.625	0.598	0.596	0.633	0.598	**0.637**
9×9	0.573	0.561	0.628	**0.637**	0.603	0.578

The bold values aims to highlight the best recognition rates.

Table 2
Recognition rates of four methods.

The proposed model	FDM	LBP-TOP	DTSA
0.655	0.426	0.418	0.366

of patches. Group LASSO were used to select different numbers of patch features according to the degree of patches' importance. Table 1 gives the recognition accuracy results with different degrees of feature selection via Group LASSO, i.e., from the top 90% patches to 40% patches of all. From Table 1 we can see that the recognition accuracy rose in proportion after feature selection via Group LASSO. The highest recognition rate appeared in varying degrees of feature selection because of the quantitative difference among the patches. Mostly, when the degree of feature selection rose to 40%, the recognition rate reached a maximum. When the number of patches became 7×7, the recognition rate reached 0.655 accordingly.

4.5. Comparison with other methods

We evaluated the performance of the proposed model (setting 7×7 patches and 40% of feature selection) by comparing with three state-of-the-art methods, such as Discriminant Tensor Subspace Analysis (DTSA) [32], Facial Dynamics Map (FDM) [37], and LBP-TOP [26] on the CASME2 micro-expression database. The nearest neighbor classifier was used for classification. The recognition rates of four methods are listed in Table 2.

From Table 2 we can see that the proposed model outperformed the other three state-of-the-art methods. Just as mentioned at the beginning, the limited micro-expression samples greatly reduced the micro-expression recognition and classification accuracy. However, the proposed model takes advantage of the macro-expression samples in the training stage, and exploits the ideas of transfer learning, which can solve the problem of the limited micro-expression samples.

5. Conclusion

In this paper, we proposed a macro-to-micro transformation model for micro-expression recognition to overcome the difficulty in micro-expression recognition due to the limitation of labeled trained data. LBP and LBP-TOP are employed for representing macro-expression features and micro-expression features, respectively. In the proposed model, each image of expression image sequences is divided into different sizes of patches to dilute the features in order to reduce the influence of redundant information. Afterwards, Group LASSO is used to select the most efficient patches from LBP and LBP-TOP features. This way, the most discriminative features for the micro-expression recognition can be selected and used in the subsequent processing. Finally, singular value decomposition is employed to achieve macro-to-micro transformation model. According to our experiments, we determined the prime parameter settings and defined the rules of patch design and feature selection. As demonstrated by the experimental results, the proposed model outperforms the compared state-of-the-art methods. More importantly, the proposed model can take advantages from the trained macro-expression databases to the micro-expression recognition: this aspect will be investigated more deeply in future work to check its efficiency. Besides this study, the feasibility of nonlinear transfer learning and complicated classifiers [38–43] will be explored.

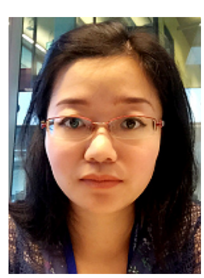
Acknowledgments

We sincerely thank the Institute of Psychology, Chinese Academy of Sciences for granting us permission to use the CASME database. This project is supported by the Natural Science Foundation of China (Grant Nos. 61571275, 61571274, 61379095).

References

[1] E.M. Bankó, V. Gál, Z. Vidnyánszky, Flawless visual short-term memory for facial emotional expressions, J. Vis. 9 (1) (2009) 1–13.
[2] X. Ben, M. Yang, P. Zhang, et al., Survey on automatic micro expression recognition methods, Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/J. Compu.-Aided Des. Comput. Graphics 26 (9) (2014) 1385–1395.
[3] X.B. Shen, Q. Wu, X.L. Fu, Effects of the duration of expressions on the recognition of microexpressions, J. Zhejiang Univ.-Sci. B 13 (3) (2012) 221–230.
[4] P. Ekman, W.V. Friesen, The repertoire of nonverbal behavior: categories, origin, usage and coding, Semiotica 1 (1) (1969) 49–98.
[5] P. Ekman, Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage, 2nd edition, Norton, New York, 2001.
[6] P. Ekman, Micro Expression Training Tools (METT), 2017, From http://www.paulekman.com, 2002. Retrieved April 15, 2009.
[7] B. Du, L. Zhang, A discriminative metric learning based anomaly detection method, IEEE Trans. Geosci. Remote Sens. 52 (11) (2014) 6844–6857.
[8] B. Du, L. Zhang, Target detection based on a dynamic subspace, Pattern Recogn. 47 (1) (2014) 344–358.
[9] C. Gong, D. Tao, K. Fu, et al., Fick's law assisted propagation for semisupervised learning, IEEE Trans. Neural Netw. Learn. Syst. 26 (9) (2014) 2148–2162.
[10] C. Gong, T. Liu, D. Tao, et al., Deformed graph laplacian for semisupervised learning, IEEE Trans. Neural Netw. Learn. Syst. 26 (10) (2015) 2261.
[11] C. Gong, D. Tao, S.J. Maybank, et al., Multi-modal curriculum learning for semi-supervised image classification, IEEE Trans. Image Process. A Publ. IEEE Signal Process. Soc. 25 (7) (2016) 3249–3260.
[12] W. Yang, Z. Wang, C. Sun, A collaborative representation based projections method for feature extraction, Pattern Recogn. 48 (1) (2015) 20–27.
[13] C. Ding, J. Choi, D. Tao, et al., Multi-directional multi-level dual-cross patterns for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2014) 518–531.
[14] C.X. Ren, D.Q. Dai, Incremental learning of bidirectional principal components for face recognition, Pattern Recogn. 43 (1) (2010) 318–330.
[15] W. Yang, Z. Wang, B. Zhang, Face recognition using adaptive local ternary patterns method, Neurocomputing 213 (2016) 183–190.
[16] J. Yan W, Q. Wu, J. Liu Y, et al., CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces, IEEE Int. Conf. Workshops on Automatic Face Gesture Recognit. (2013) 1–7.
[17] W.J. Yan, X. Li, S.J. Wang, et al., CASME II: an improved spontaneous micro-expression database and the baseline evaluation, PLoS One 9 (1) (2014) e86041.
[18] X. Li, T. Pfister, X. Huang, et al., A Spontaneous Micro-expression Database: Inducement, Collection and Baseline IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, Shanghai, 2013, pp. 1–6, http://dx.doi.org/10.1109/FG.2013.6553717.
[19] S. Polikovsky, Y. Kameda, Y. Ohta, Facial Micro-expressions Recognition Using High Speed Camera and 3D-gradient Descriptor IET International Conference on Crime Detection and Prevention (2009), 16–16.
[20] S. Polikovsky, Y. Kameda, Y. Ohta, Facial micro-expression detection in hi-Speed video based on facial action coding system (FACS), IEICE Trans. Inf. Syst. E96 (D(1)) (2013) 81–92.
[21] B. Tan, Y. Song, E. Zhong, et al., Transitive transfer learning, in: The, ACM SIGKDD International Conference, ACM, 2015, pp. 1155–1164.
[22] Q. Wu, X. Shen, X. Fu, The Machine Knows What You Are Hiding: An Automatic Micro-expression Recognition System. Affective Computing and Intelligent Interaction, Springer, Berlin, Heidelberg, 2011, pp. 152–162.
[23] P. Zhang, X. Ben, R. Yan, et al., Micro-expression recognition system, Optik 127 (2016) 1395–1400.
[24] S.T. Liong, C.W. Phan, J. See, et al., Optical Strain Based Recognition of Subtle Emotions IEEE International Symposium on Intelligent Signal Processing and Communication Systems, Kuching, 2014, pp. 180–184.
[25] Y.J. Liu, J.K. Zhang, W.J. Yan, et al., A main directional mean optical flow feature for spontaneous micro-expression recognition, IEEE Trans. Affective Comput. (2015), http://dx.doi.org/10.1109/TAFFC.2015.2485205.
[26] T. Pfister, X. Li, G. Zhao, et al., Recognising spontaneous facial micro-expressions, in: International Conference on Computer Vision, IEEE Computer Society, 2011, pp. 1449–1456.
[27] S.J. Wang, W.J. Yan, G. Zhao, et al., Micro-Expression Recognition Using Robust Principal Component Analysis and Local Spatiotemporal Directional Features ECCV 2014 Workshops, Springer International Publishing, 2014, pp. 325–338.
[28] X. Huang, S.J. Wang, G. Zhao, et al., Facial Micro-Expression Recognition Using Spatiotemporal Local Binary Pattern with Integral Projection The Workshop on Computer Vision for Affective Computing at ICCV, Santiago, 2015, pp. 1–9, http://dx.doi.org/10.1109/ICCVW.2015.10.
[29] Y. Wang, J. See, R.C. Phan, et al., Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition, PLoS One 10 (5) (2014) e0124674.
[30] S.J. Wang, W.J. Yan, X. Li, et al., Micro-expression recognition using color spaces, IEEE Trans. Image Process. 24 (12) (2015) 6034–6047.
[31] X. Ben, P. Zhang, R. Yan, et al., Gait recognition and micro-expression recognition based on maximum margin projection with tensor representation, Neural Comput. Appl. 27 (8) (2016) 2629–2646.
[32] S.J. Wang, H.L. Chen, W.J. Yan, et al., Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine, Neural Process. Lett. 39 (1) (2014) 25–43.
[33] S.J. Wang, M.F. Sun, Y.H. Chen, et al., STPCA: Sparse tensor Principal Component Analysis for feature extraction, in: 21st International Conference on Pattern Recognition (ICPR 2012), Tsukuba, Japan, 2012, pp. 2278–2281.
[34] D. Muramatsu, Y. Makihara, Y. Yagi, View transformation model incorporating quality measures for cross-view gait recognition, IEEE Trans. Cybern. 46 (7) (2015) 1602–1615.
[35] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 68 (1) (2006) 49–67.
[36] P. Lucey, J.F. Cohn, T. Kanade, et al., The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, in: Computer Vision and Pattern Recognition Workshops, IEEE, 2010, pp. 94–101.
[37] F. Xu, J. Zhang, J. Wang, Microexpression identification and categorization using a facial dynamics map, IEEE Trans. Affective Comput. 2016 (2016), http://dx.doi.org/10.1109/TAFFC.2016.2518162, 1–1.
[38] B. Gu, V.S. Sheng, A. Robust, Regularization path algorithm for ν-support vector classification, IEEE Trans. Neural Netw. Learn. Syst. 1 (2016) 1–8.
[39] B. Gu, X. Sun, V.S. Sheng, Structural minimax probability machine, IEEE Trans. Neural Netw. Learn. Syst. 1 (2016) 1–11.
[40] B. Gu, V.S. Sheng, K.Y. Tay, et al., Incremental support vector learning for ordinal regression, EEE Trans. Neural Netw. Learn. Syst. 26 (7) (2015) 1403–1416.
[41] B. Gu, V.S. Sheng, Z. Wang, et al., Incremental learning for ν-support vector regression, Neural Netw. 67 (C) (2015) 140–150.
[42] B. Gu, V.S. Sheng, S. Li, Bi-parameter space partition for cost-sensitive SVM, International Joint Conferences on Artificial Intelligence (2015) 3532–3539.
[43] X. Wen, L. Shao, Y. Xue, et al., A rapid learning algorithm for vehicle classification, Inf. Sci. 295 (2015) 395–406.

Xitong Jia was born in Shandong, China, in 1993. He received the B.S. degree in communication engineering from the School of Information Science and Engineering, Shandong University, Jinan, China, in 2015. He is currently a master student in the School of Information Science and Engineering, Shandong University. His current research interests include micro-expression recognition and gait recognition.



Xianye Ben was born in Harbin, China, in 1983. She received the B.S. degree in electrical engineering and automation from the College of Automation, Harbin Engineering University, Harbin, China, in 2006, and the Ph.D. degree in pattern recognition and intelligent system form the College of Automation, Harbin Engineering University, Harbin, in 2010. She is currently working as an Associate Professor in the School of Information Science and Engineering, Shandong University, Jinan, China. She has published more than 70 papers in major journals and conferences. Her current research interests include pattern recognition, digital image processing and analysis, machine learning.

Hui Yuan was born in Shandong, China, in 1984. He received the B.E. and Ph.D. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2006 and 2011, respectively. From 2011.04, he joined Shandong University (SDU), Ji'nan, China, as a Lecturer and Postdoctor. From 2013.01 to 2014.12, he went to City University of Hong Kong (CityU) for Postdoctor research by the grant of the 2012th "Hong Kong Scholar" Project. He was promoted to be Associate Professor in 2015. 01, Shandong University. His current research interests include video coding, multimedia communication, computer vision, image/video processing, etc.

Kidiyo Kpalma was born in Togo in 1962. He received the M.Sc. degree in signal processing from the University of Rennes 1, France, in 1988. He joined the Institut National des Sciences Appliquées de Rennes (INSA), Rennes, for his Ph.D. studies in image processing, and received the Ph.D. degree in 1992. Since 1994, he has been an Associate Professor (Maître de conferences) with INSA, where he teaches signal and systems, signal processing, and DSP. He received the H.D.R. (Habilitation à diriger des recherches) degree in signal processing and telecommunications from the University of Rennes 1, in 2009. He has been a Professor with INSA since 2014. As a member of the Department of Image and Automatic, Institute of Electronics and Telecommunications of Rennes, his research interests include image analysis, pattern recognition, image segmentation, semantic segmentation, image fusion, and remote sensing.

Weixiao Meng was born in Harbin, China, in 1968. He received his B.Sc. degree in Electronic Instrument and Measurement Technology from Harbin Institute of Technology (HIT), China, in 1990. And then he obtained the M.S. and Ph.D. degree, both in Communication and Information System, HIT, in 1995 and 2000, respectively. Now he is a professor in School of Electronics and Communication Engineering, HIT. Besides, he is a senior member of IEEE, a senior member of China Institute of Electronics, China Institute of Communication and Expert Advisory Group on Harbin E-Government. His research interests mainly focus on adaptive signal processing. In recent years, he has published 1 authored book and more than 100 academic papers on journals and international conferences, more than 60 of which was indexed by SCI, EI and ISTP. Up to now, he has totally completed more than 20 research projects and holds 6 China patents. 1 standard proposal was accepted by IMT-Advanced technical group.