2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Facial Expression Recognition under Random Block Occlusion Based on Maximum Likelihood Estimation Sparse Representation

S. S. Liu, Y. Zhang, K. P. Liu

Abstract—**Occlusion is a big challenge for facial expression recognition and previous efforts are largely limited to a few occlusion types without considering the random characteristic of occlusion. Since the original sparse coding model actually assumes that the coding residual follows the Gaussian distribution, which may not be accurate enough to describe the coding errors in practice, so we propose a new scheme by modeling the sparse coding as a sparsely constrained robust regression problem in this paper. Firstly, in order to reduce the influence of occlusion for facial expression, the test facial expression image will be assigned different weights in all pixels. Secondly, because the occluded pixels should have lower weight values, hence, we update the weight through constant iterative until the convergence is achieved. Finally, the final sparse representation of test image can be calculated using the optimal weight matrix. And the class of test image can be determined by the minimal residual which associated with each class of training samples to the test image. The proposed method achieves better performance in JAFFE database and Cohn-Kanade database and experimental results show that it is robust to facial expression recognition under random block occlusion.**

I. Introduction

FACIAL expression recognition technology is a challenging cross-subject of physiology, psychology, image processing, pattern recognition, computer vision and other fields [1]. In order to ensure information integrity, scientists used the facial expression images with no occlusion to conduct experiments and research under controlled laboratory conditions [2]. Unfortunately, face occlusion is very common in real life, sunglasses can occlude the eyes region and a scarf or a medical mask occludes the mouth region [3]. Therefore, facial expression recognition in occluded situations remains one of the most important bottlenecks for practical facial expression recognition systems.

S. S. Liu is with the School of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, CO 130012 China (e-mail: liu-shuaishi@126.com).

Y. Zhang is with the School of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, CO 130012 China (e-mail: 260259423@qq.com).

K. P. Liu is with the School of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, CO 130012 China (corresponding author . phone: 0431-85716441, fax: 0431-85716441; e-mail: liukeping@mail.ccut.edu.cn).

This work was supported by the Project of Science and Technology Foundation of Changchun University of Technology (Grant No: 2010XN07), the Youth Research Fund Project of Science and Technology Development Plan for Jilin Province (Grant No: 20140520065JH) and Natural Science Foundation of Jilin Province (Grant No: 20101505).

In the recent years, a great many of methods have been proposed for the partial occlusion facial expression recognition. Kotsia [4] proposed a novel multiclass classifier inspired by the Fisher's Linear Discriminant Analysis and the Support Vector Machine (SVM) to research the effects of facial expression recognition under different organs occluded. Zhang [5] used a Monte Carlo algorithm to select Gabor templates from the gallery images and template matching over a search area to generate features robust against occlusions.

However, the above methods only work for limited types of occlusion, and are less likely to adapt to random occlusion. The occlusions in facial appearance can be at anywhere and with any size or shape in a given face image and we don't have any prior knowledge about it [6]. So a method should be proposed which can overcome the problem. Wright [7] applied sparse coding to face recognition and proposed the sparse representation of the test image using the available training samples whose classes are known, the impressive facial expression recognition performance was achieved under random block occlusion.

In order to improve the robustness and effectiveness of sparse representation, we propose a maximum likelihood estimation sparse representation (MLESR) model in this paper. Inspired by the robust regression theory [8]which using iterative methods to assign different weights to residuals until the estimation process converges, we design the signal fidelity term as a maximum likelihood estimation (MLE) estimator, which minimizes some function that associated with the distribution of the coding residual. The proposed MLESR scheme utilizes the MLE principle to robustly regress the given signal with sparse regression coefficients, and then the minimization problem is transformed into an iteratively reweighted sparse coding problem. A reasonable weight function is designed for applying MLESR to facial expression recognition. The experiments on the JAFFE database [2] and Cohn-Kanade database [9] demonstrate that the proposed method is effective in dealing with random block occlusion.

II. Maximum Likelihood Estimation Sparse Representation

In general, the sparse coding problem can be formulated as:

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \quad s.t. \quad \|y - D\alpha\|_2^2 \le \varepsilon \tag{1}$$

Where y is a test facial expression image, D is the dictionary of the training samples, α is the coding vector of y over D, and $\varepsilon > 0$ is a constant. The fidelity term is denoted by $\|y - D\alpha\|_2^2$.

From the viewpoint of maximum likelihood estimation, defining the fidelity term with l_2 norm actually assumes that the coding residual $e = y - D\alpha$ follows the Gaussian distribution. But in practice this assumption may not hold well, especially when occlusions occur in the test facial expression images. So the conventional fidelity term of sparse coding model in (1) may not be robust enough in these cases. In order to construct a more robust model for sparse coding of facial expression images, in this paper, we propose the MLESR to find an MLE solution of the coding coefficients.

A. The Model of Maximum Likelihood Estimation Sparse Representation

The dictionary T of conventional sparse coding is replaced by the $T = [r_1; r_2; \cdots r_n]$, where the row vector r_i is the i^{th} row of D. The coding residual is denoted by $e = y - T\alpha = [e_1; e_2; \cdots; e_n]$. Then each element of e is $e_i = y_i - r_i\alpha, i = 1,2,\cdots,n$. Assume that $e_1, e_2, \cdots, e_n$ are independently and identically distributed according to some probability density function $f_\theta(e_i)$, where θ denotes the parameter set that characterizes the distribution. Without considering the sparsely constraint of α, the likelihood function of coding residual e is as follows:

$$L_\theta(e_1, e_2, \cdots; e_n) = \prod_{i=1}^{n} f_\theta(e_i) \quad (2)$$

The MLE aims to maximize (2) likelihood function or, equivalently, minimize (3) the objective function $F_\theta(e)$:

$$F_\theta(e) = -\ln L_\theta = \sum_{i=1}^{n} \rho_\theta(e_i) \quad (3)$$

Where $\rho_\theta(e_i) = -\ln f_\theta(e_i)$.

With consideration of the sparsely constraint of α and $e = y - T\alpha$, the model of Maximum Likelihood Estimation Sparse Representation can be formulated as the following minimization.

$$\hat{\alpha}_1 = \arg\min \; F_\theta(y - T\alpha) \;\; s.t. \; \|\alpha\|_1 \le \sigma \quad (4)$$

As can be seen from (4), the model is essentially a sparse-constrained Maximum Likelihood Estimation problem. And the least absolute shrinkage and selection operator (LASSO) [10] is mainly used to solve the constrained optimization problem and can be formulated as follows:

$$\hat{\alpha}_1 = \arg\min \|y - T\alpha\|_2^2 \;\; s.t. \; \|\alpha\|_1 \le \sigma \quad (5)$$

Where $\sigma > 0$ is a constant. By comparing (1) with (5), we can find the traditional sparse coding model in (1) is equivalent to the LASSO problem. When $F_\theta(e) = \|e\|_2^2$ in (4), the (5) and (4) have the same solution. $F_\theta(e) = \|e\|_2^2$ means $f_\theta(e_i) = \exp(e_i)$, in other words, coding residual e follows the Gaussian distribution. Hence, the conventional sparse coding model in (5) is a special case of (4). That is to say, the model of Maximum Likelihood Estimation Sparse Representation in (4) is more general than the sparse coding model in (1).

We can get the MLE solution to α with sparsely constraint by solving (4). Obviously, one vital problem is how to determine the distribution f_θ. Specifically, taking f_θ as Gaussian distribution is simple, but not effective enough. In this paper, the distribution f_θ isn't determined directly to solve (4). Instead of the above mentioned general assumptions of f_θ, we transform the minimization problem in (4) into an iteratively reweighted sparse coding problem. And the resulted weights have clear physical meaning, such as the features of occluded area will have low weight values. The weights, however, depend upon the residuals, the residuals depend upon the estimated coefficients, and the estimated coefficients depend upon the weights. An iterative solution is therefore required. By iteratively computing the weights, the solution of Maximum Likelihood Estimation Sparse Representation could be solved efficiently.

B. Iteratively Reweighted Sparse Coding

In general, we assume that the probability density function $f_\theta(e_i)$ is symmetric, and $f_\theta(e_i) < f_\theta(e_j)$ if $|e_i| > |e_j|$. So $\rho_\theta(e_i)$ has the following properties:

1) $\rho_\theta(0)$ is the global minimal of $\rho_\theta(e_i)$, namely $\rho'_\theta(0) = 0$.

2) $\rho_\theta(e_i) = \rho_\theta(-e_i)$.

3) $\rho_\theta(e_i) < \rho_\theta(-e_i)$ if $|e_i| > |e_j|$.

We can approximate $F_\theta(e)$ by its first order Taylor expansion in the neighborhood of e_0 as follows:

$$F_\theta(e) = F_\theta(e_0) + (e - e_0)^T F'_\theta(e_0) + R_1(e) \quad (6)$$

Where $F'_\theta(e)$ is the derivative of $F_\theta(e)$ and $R_1(e)$ is the high order residual term. Denote by ρ'_θ the derivative of ρ_θ, and then $F'_\theta(e_0) = [\rho'_\theta(e_{0,1}); \rho'_\theta(e_{0,2}); \cdots; \rho'_\theta(e_{0,n})]$, where $e_{0,i}$ is the i^{th} element of e_0. In sparse coding, it is usually expected that the fidelity term is strictly convex, so we approximate the high order residual term as $R_1(e) = \frac{1}{2}(e - e_0)^T W (e - e_0)$, where W is a diagonal matrix.

In order to get the minimal value of $F_\theta(e)$ at $e = 0$, we calculate the derivative of (6). The result is as follows:

$$F'_\theta(e) = F'_\theta(e_0) + R'_1(e) \quad (7)$$

Where $R'_1(e) = [W_{1,1}(e_{i,1} - e_{0,1}); \cdots; W_{n,n}(e_{i,m} - e_{0,m})]$.

Letting $F'_\theta(0) = 0$, we have the diagonal element of W as

$$W_{i,i} = \rho'_\theta(e_{0,i}) / e_{0,i} \quad (8)$$

According to the properties of $\rho_\theta(e_i)$, $\rho'_\theta(e_i)$ will have the same sign as e_i. So each $W_{i,i}$ is a non-negative scalar. Then $F_\theta(e)$ can be written as $F_\theta(e) = \frac{1}{2}\| W^{1/2} e \|_2^2 + b$, where b is a scalar value determined by e_0. Therefore the model in (4) can be approximated as follows:

$$\hat{\alpha} = \operatorname{argmin} \| W^{1/2}(y - T\alpha) \|_2^2 \quad s.t. \ \|\alpha\|_1 \le \sigma \tag{9}$$

By comparing (9) with (5), we can find that the model of Maximum Likelihood Estimation Sparse Representation is to transform the traditional sparse coding mode into the weighted sparse coding mode. The weight matrix W has a clear physical meaning, $W_{i,i}$ is the weight assigned to each pixel of the test facial expression image y. Intuitively, the occluded pixels should have lower weight values. Considering the logistic function has properties similar to the hinge loss function in SVM [11], we choose it as the weight function. The initialization is vital for the facial expression recognition and we are able to have a very reasonable initialization to achieve good performance. When a testing facial expression image y comes, in order to initialize the weight, we should firstly estimate the coding residual e of y. We can initialize e as $e = y - y_{ini}$, where y_{ini} is some initial estimation of the true facial expression from y_{ini} can be set as the mean facial expression image of all training expression images. In the paper, we simply compute y_{ini} as $y_{ini} = m_t$, where m_t is the mean image of all training samples.

We update W through constant iterative optimization, until the convergence is achieved, namely, the difference of the weight between adjacent iterations as shown in Fig. 2 (a) is small enough. Specifically, we stop the iteration if the following holds:

$$\left\| W^{(t)} - W^{(t-1)} \right\|_2 \Big/ \left\| W^{(t-1)} \right\|_2 < \gamma \tag{10}$$

Where γ is a small positive scalar. In this paper, the convergence can be achieved in 15 iterations.

And then the final coefficient of sparse representation $\hat{\alpha}$ can be got through (9). Finally, we calculate the residual of training sample in each class approximates the test facial expression image y which can be calculated as follows:

$$r_i(y) = \left\| y - T\delta_i(\hat{\alpha}) \right\|_2, i = 1, \cdots, k \tag{11}$$

Where $\delta_i(\hat{\alpha})$ is the coefficient vector in $\hat{\alpha}$ of the i^{th} class training sample space.

According to the rule of minimal approximation residual as in (12), the test facial expression image y will be classified in the class that its approximation residual is smallest.

$$identity(y) = \arg \min_i (r_i(y)) \tag{12}$$

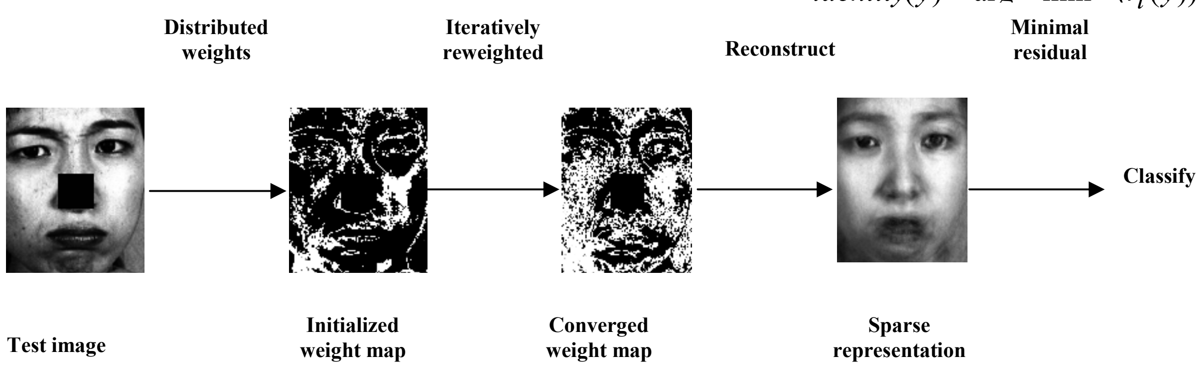


Fig. 1 Procedure of the facial expression recognition under random occlusion based on the proposed method

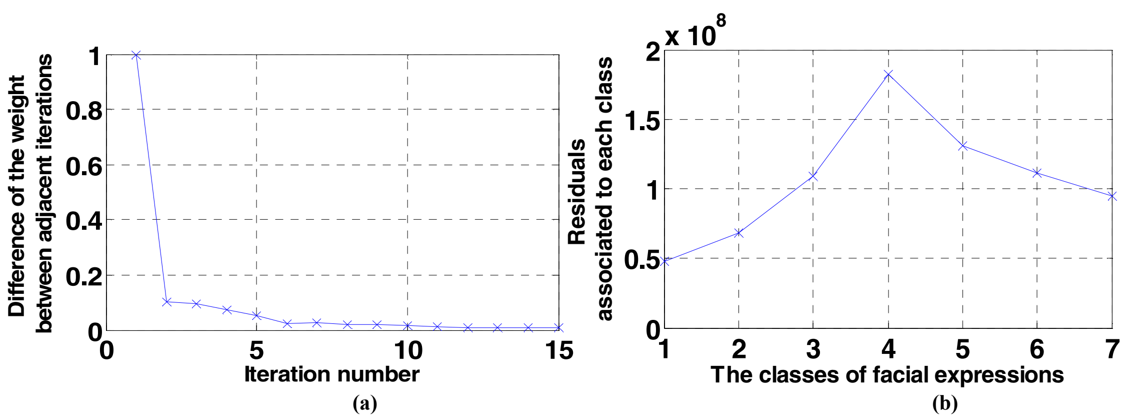


Fig. 2 The convergence curve of the weight (a) and the residual of each training class approximates the test image (b)

Therefore, we summarize our method for random occluded facial expression recognition as Fig.1. Firstly, a test facial expression image is distributed different weights. Secondly, the convergent weight can be got through continuous iteration. The convergence curve of the weight is shown in Fig. 2 (a). Thirdly, the final sparse representation of the test image can be got and we calculate the coding residual of the test image. Finally, the approximation residuals between all facial expression classes and the test facial expression image will be calculated and the test image can be classified in the class where its approximation residual is smallest. The residual of each training class approximating the test image is shown in Fig. 2 (b).

III. Experiments and Results

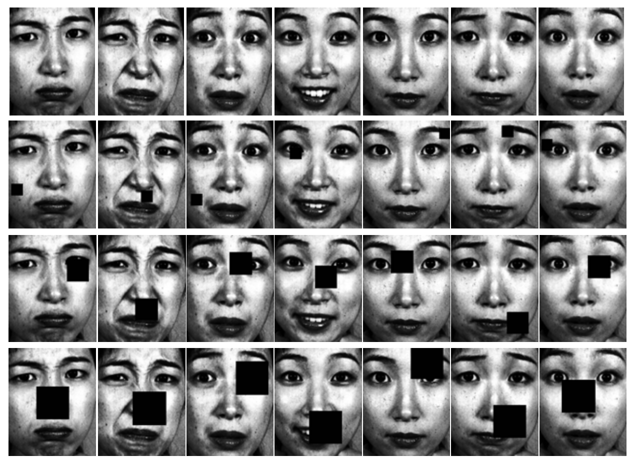
We use the Japanese Female Facial Expression (JAFFE) database and Cohn-Kanade database for the evaluations of the proposed approach. The JAFFE database contains 213 facial expression images of ten persons, everyone has seven kinds of expression and each expression has 3 or 4 samples. 137 images of seven facial expressions (happy-19, surprise-20, sadness-20, fear-20, disgust-18, anger-20 and neutral-20) are chosen as training samples, the remaining 76 images are testing samples. And the Cohn-Kanade database consists of 100 university students were instructed to perform a series of 23 facial displays, seven of which were based on description of basic expressions as JAFFE database. Because the number of facial expression image in JAFFE database is small, so the experiment traverses three conditions to get the average recognition rate. For our experiments, we selected 420 image sequences from 10 persons with seven expressions (happy-6, surprise-6, sadness-6, fear-6, disgust-6, anger-6 and neutral-6). We choose the 210 facial expression images of ten persons as training samples and the rest are testing samples. To test the algorithms' generalization performance on Cohn-Kanade database, we traverse six conditions in our experiments, reported the average recognition results.

A. *The Description of Experiment*

Since the people in the JAFFE database and Cohn-Kanade database posed with slightly different head tilts and have different head sizes, preprocessing is used to eliminate these differences. Our processing is similar to that used in [12]: each image was rotated so that the eyes were horizontally aligned, and the intraocular distance was used to crop a rectangular area from the original image. The size of the original image in the JAFFE database is 256×256 and the original image in the Cohn-Kanade database is 640×490. Then they are normalized to 128×104 and histogram equalization was used to increase the local contrast in some areas of the image.

Unlike other constraints such as pose variations, whose characteristics can be inferred beforehand, facial occlusion is particularly difficult to handle due to its random characteristic, which means that occlusions can occur at random positions and can be arbitrarily large in size. We don't assume any explicit prior knowledge about the location and the size of the occluded regions; the only prior information we have about the occlusion is that the corrupted pixels are likely to be adjacent to each other in the image plane. Fig. 3 shows some expression images under different levels of block occlusion in two databases. And the level means the percentage of occlusion in a total facial expression image. So the level is a positive scalar which locates within the interval (e.g. 0.1 means the occlusion is 10% of a total image).

One of the most important characteristics of sparse coding based facial expression recognition is its robustness to facial occlusion. In order to verify the robustness of the proposed method to random block occlusion, we compare the proposed method with some existing methods (KNN [13], SVM [14], SRC [7] and the recently developed Gabor-SRC (GSRC) [15]) on two databases.



(a)

(b)

Fig. 3 Some samples of occluded facial images in two databases: JAFFE (a), Cohn-Kanade (b). And each one of them from top to bottom: no occlusion, 10%occluded, 20% occluded and 30% occluded.

B. *Results Analysis*

On two facial databases, the expression recognition results of the proposed method and some existing methods under

TABLE I
THE ACCURACIES OF DIFFERENT METHODS ON JAFFE

Occlusion level	Methods				
	The proposed method	KNN	SVM	SRC	GSRC
0	93.42%	85.57%	89.47%	90.79%	92.11%
0.1	92.11%	84.26%	88.16%	90.79%	91.52%
0.2	90.79%	81.58%	86.84%	88.16%	89.47%
0.3	89.67%	80.55%	83.67%	86.84%	88.16%
0.4	88.16%	78.56%	82.43%	84.26%	86.84%
0.5	86.73%	76.49%	80.32%	83.79%	84.26%
0.6	82.89%	75.00%	77.32%	79.78%	81.58%
0.7	78.95%	72.34%	75.00%	76.49%	77.32%
0.8	61.84%	56.47%	59.46%	60.43%	62.45%
0.9	52.63%	47.23%	48.64%	49.42%	51.67%

TABLE II
THE ACCURACIES OF DIFFERENT METHODS ON COHN-KANADE

Occlusion level	Methods				
	The proposed method	KNN	SVM	SRC	GSRC
0	94.29%	86.84%	91.23%	92.79%	93.21%
0.1	92.86%	84.35%	89.32%	90.23%	92.79%
0.2	90.95%	82.12%	85.93%	88.69%	89.23%
0.3	89.04%	79.98%	84.13%	85.97%	88.69%
0.4	86.67%	76.87%	81.87%	83.69%	86.79%
0.5	85.24%	75.21%	79.87%	82.46%	84.27%
0.6	80.00%	73.87%	74.97%	76.78%	78.87%
0.7	75.23%	70.21%	71.34%	72.67%	74.21%
0.8	68.09%	60.57%	62.37%	63.84%	65.43%
0.9	55.71%	49.53%	50.34%	51.24%	52.17%

TABLE III
THE ACCURACIES OF EACH EXPRESSION ON JAFFE IN DIFFERENT LEVELS OF BLOCK OCCLUSIONS

Occlusion level	Expressions						
	An	Di	Fe	Ha	Ne	Sa	Su
0	100%	81.82%	83.33%	100%	100%	100%	90.00%
0.1	100%	81.82%	75.00%	100%	100%	100%	90.00%
0.2	100%	81.82%	75.00%	100%	100%	90.00%	90.00%
0.3	100%	81.82%	75.00%	100%	100%	81.82%	90.00%
0.4	90.00%	81.82%	75.00%	100%	100%	81.82%	90.00%
0.5	90.00%	80.00%	75.00%	91.67%	100%	81.82%	90.00%
0.6	90.00%	72.00%	75.00%	83.33%	100%	70.00%	90.00%
0.7	90.00%	60.83%	70.00%	81.82%	100%	70.00%	80.00%
0.8	90.00	38.46%	58.33%	50.00%	90.00%	63.64%	40.00%
0.9	75.00	27.27%	41.67%	50.00%	90.00%	45.45%	40.00%

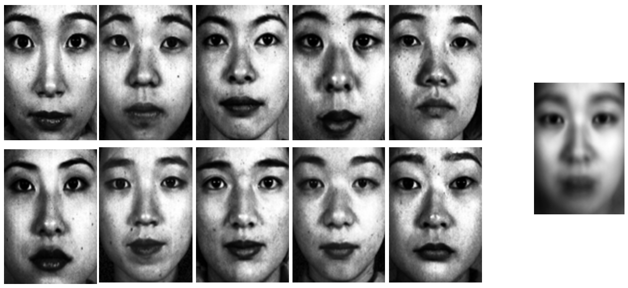
TABLE IV
THE ACCURACIES OF EACH EXPRESSION ON COHN-KANADE IN DIFFERENT LEVELS OF BLOCK OCCLUSIONS

Occlusion level	Expressions						
	An	Di	Fe	Ha	Ne	Sa	Su
0	83.33%	90.00%	100%	96.67%	100%	100%	90.00%
0.1	83.33%	83.33%	100%	93.33%	100%	100%	90.00%
0.2	80.00%	80.00%	100%	93.33%	100%	100%	83.33%
0.3	80.00%	73.33%	100%	90.00%	100%	100%	80.00%
0.4	76.67%	73.33%	96.67%	90.00%	100%	93.33%	76.67%
0.5	73.33%	70.00%	93.33%	86.67%	96.67%	90.00%	73.33%
0.6	70.00%	66.67%	90.00%	83.33%	93.33%	86.67%	70.00%
0.7	63.33%	63.33%	86.67%	80.00%	90.00%	80.00%	63.33%
0.8	60.00%	56.67%	73.33%	70.00%	80.00%	76.67%	60.00%
0.9	53.33%	36.67%	56.67%	63.33%	66.67%	60.00%	53.33%

different levels of block occlusion are shown in Table I and Table II. As can be seen in Table I and Table II, the facial expression recognition rates of different methods always decrease with the increase of the level of block occlusion. And all methods achieve good performance when the level of block occlusion increases from 0.1 to 0.5. Since the training and test images in our experiments are represented by their pixel values and no additional feature extraction is required, so the KNN [13] and SVM [14] methods can not get better performance when the level of occlusion is too big, and the two methods are usually used with the additional feature extraction. We can also see that the expression recognition rates of SRC [7] and GSRC [15] are less than the proposed method. By comparing the results of different methods, the performance of the proposed method is better than the others.

Since the proposed method has achieved desirable performance under random occlusion, so the performance obtained by the proposed method alone is used for analyzing the effects of occlusions separately for each facial expression. The effects of different levels of block occlusions on each expression on two databases are shown in Table III and Table IV. The abbreviations An, Di, Fe, Ha, Ne, Sa and Su represent angry, disgust, fear, happy, neutral, sadness and surprise respectively.

As can be seen in Table III and Table IV, increasing size of random occlusions decreases the facial expression recognition performance for all expressions. The angry, happy, neutral and sad achieve ideal recognition rates at the level of block occlusion from 0~0.1, since the level of block occlusion is small and the useful information of these expressions has lost less. All expressions except for angry, happy and neutral are strongly affected by 20% occluded and 30% occluded. When the level of block occlusion is 0.4, the happy and neutral achieve better performance. All expressions except for neutral are strongly affected by the level of block occlusion increases from 0.5~0.7. We can also see that the recognition rates of neutral expression maintain high values under different levels of block occlusion. The recognition rate of neutral expression is 90%, even under 90% occluded in the JAFFE database as shown in Table III. That is due to we choose the mean facial expression image as the initial estimation of the testing image according the definition of LASSO [9] and the neutral expression is similar to the mean facial expression image of all training expression images. Therefore, even under big level occlusion, the neutral expression can be recognized effectively. And the testing neutral expression images and the mean facial expression image of all training expression images are shown in Fig 4. However, the same situation does not occur obviously in Cohn-Kanade database. The recognition rate of neutral expression is 66.67% under 90% occluded in Cohn-Kanade database. Although the accuracy of neutral expression is higher than other expression's accuracies at the occlusion level is 0.9 in Cohn-Kanade database, which is lower than the accuracies in the same condition in JAFFE database. That is due to the facial expression images in JAFFE database are



(a) **(b)**

Fig. 4 Some testing neutral expression images in JAFFE database (a) and the mean facial expression image (b).

female and are the same country, while the images in Cohn-Kanade database are different genders and different counties. Then there is less similar between neural expression and the mean facial expression image of all training expression images as shown in Fig 5. Therefore, the recognition rate of neutral expression in Cohn-Kanade database is lower than in JAFFE database when the level of block occlusion increases from 0.5 to 0.9. Though the recognition rates of neutral expression in two databases differ bigger, the facial expression images in Cohn-Kanade database conform to the realities of the situation due to the images come from different counties and different genders.

(a) **(b)**

Fig. 5 Some testing neutral expression images in Cohn-Kanade database (a) and the mean facial expression image (b).

IV. CONCLUSION

In this paper, a novel and comprehensive facial expression recognition method under random block occlusion is proposed based on maximum likelihood estimation sparse representation in order to reduce the effect of occlusion for facial expression recognition. This algorithm can achieve good performance by using raw imagery data without dimension reduction, feature selection, synthetic training examples and domain-specific information. The proposed method is compared with the existing methods on the JAFFE database and Cohn-Kanade database under random block occlusion and better recognition rates are achieved. The experimental results clearly demonstrated that the proposed method outperforms significantly previous methods and is robust to random block occlusion.

REFERENCES

[1] M. F. Valstar, M. Mehu, and B. Jiang, “Meta-analysis of the first facial expression recognition challenge,” *IEEE Trans. Sys. Man Cybernetics*, vol. 42, pp. 966-979, April 2012.

[2] L. S. Shi, T. Y. Tao and W. X. Zhu, “Illumination-robust Facial Expression Recognition Based on Symmetric Bilinear Model,” *Acta Automatica Sinica*, vol. 38, pp. 1933-1940, February 2012.

[3] L. S. Shi, T. Y. Tao and W. Chuan, “Facial Expression Recognition Method Based on Gabor Multi-orientation Features Fusion and Block Histogram,” *Acta Automatica Sinica*, vol. 37, pp. 1455-1463, December 2011.

[4] I. Kotsia, I. Pitas and S. Zafeirious, “Novel multiclass classifiers based on the minimization of the within-class variance,” *IEEE Trans. Neural Networks*, vol. 20, pp. 14-34, January 2009.

[5] L. Zhang, D. Tjondronegoro and V. Chandran, “Toward a more robust facial expression recognition in occluded images using randomly sampled Gabor based templates,” in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2011, pp.1-6.

[6] X. Tan, S. Chen and Z. H. Zhou, “Face recognition under occlusions and variant expressions with partial similarity,” *IEEE Trans. Information Forensics and Security*, vol. 4, pp. 217-230, February 2009.

[7] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210-227, February 2009.

[8] R. Andersen. *Modern methods for robust regression, series: Quantitative applications in the social sciences*. SAGE Publications, 2008. pp. 47-67.

[9] T. Kanade, J. Cohn, and Y. Tian, “Comprehensive Database for Facial Expression Analysis,” in *Proc. IEEE Int. Conf. Face and Gesture Recognition*, pp.1-6, 2011.

[10] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Royal Statistical Society*, vol. 58, pp. 267-288, January 1996.

[11] J. Zhang, R. Jin, Y. M. Yang, and A. G. Hauptmann, “Modified logistic regression: An approximation to SVM and its applications in large-scale text categorization,” in *Proc. Int. Conf. Machine Learning*, 2003, pp. 888-895.

[12] S. S. Liu and Y. Zhang, “Facial Expression Recognition under Partial Occlusion Based on Gabor Multi-orientation Feature Fusion and Local Gabor Binary Pattern Histogram Sequence,” in *Proc. IEEE Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing*, 2013, pp. 218-222.

[13] H. Tang, M. H. Johnson and T. Huang, “Non-frontal view facial expression recognition based on ergodic hidden markov model supervectors,” in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2010, pp. 1202-1207.

[14] I. Kotsia and I. Pitas, “Facial expression recognition in image sequences using geometric deformation features and support vector machines,” *IEEE Trans. Image Processing*, vol. 16, pp. 172-187, January 2007.

[15] M. Yang and L. Zhang, “Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary,” *Comput. vison*, vol. 6316, pp. 448-461, September 2010.