Superimposed Sparse Parameter Classifiers for Face Recognition

Qingxiang Feng, Chun Yuan, Jeng-Shyang Pan, *Senior Member, IEEE*, Jar-Ferr Yang, *Fellow, IEEE*, Yang-Ting Chou, Yicong Zhou, *Senior Member, IEEE*, and Weifeng Li

***Abstract*—In this paper, a novel classifier, called superimposed sparse parameter (SSP) classifier is proposed for face recognition. SSP is motivated by two phase test sample sparse representation (TPTSSR) and linear regression classification (LRC), which can be treated as the extended of sparse representation classification (SRC). SRC uses all the train samples to produce the sparse representation vector for classification. The LRC, which can be interpreted as L2-norm sparse representation, uses the distances between the test sample and the class subspaces for classification. TPTSSR is also L2-norm sparse representation and uses two phase to compute the distance for classification. Instead of the distances, the SSP classifier employs the SSPs, which can be expressed as the sum of the linear regression parameters of each class in iterations, is used for face classification. Further, the fast SSP (FSSP) classifier is also suggested to reduce the computation cost. A mass of experiments on Georgia Tech face database, ORL face database, CVL face database, AR face database, and CASIA face database are used to evaluate the proposed algorithms. The experimental results demonstrate that the proposed methods achieve better recognition rate than the LRC, SRC, collaborative representation-based classification, regularized robust coding, relaxed collaborative representation, support vector machine, and TPTSSR for face recognition under various conditions.**

***Index Terms*—Face recognition, linear regression, sparse representation, two phase sparse representation.**

Manuscript received April 26, 2015; revised January 1, 2016; accepted January 5, 2016. Date of publication January 25, 2016; date of current version January 13, 2017. This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/106/2013/A3, in part by the Research Committee at University of Macau under Grant MYRG2014-00003-FST and Grant MRG017/ZYC/2014/FST, in part by the National High Technology Development Plan (863) under Grant 2011AA01A205, and in part by the National Significant Science and Technology Projects of China under Grant 2013ZX01039001-002-003. This paper was recommended by Associate Editor S. Zafeiriou. *(Corresponding author: Yicong Zhou.)*

Q. Feng was with the Graduate School at Shenzhen, Tsinghua University, Shenzhen 518000, China. He is now with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: fengqx1988@gmail.com).

C. Yuan and W. Li are with the Graduate School at Shenzhen, Tsinghua University, Shenzhen, China (e-mail: yuanc@sz.tsinghua.edu.cn; li.weifeng@sz.tsinghua.edu.cn).

J.-S. Pan is with the Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518000, China (e-mail: jspan@cc.kuas.edu.tw).

J.-F. Yang and Y.-T. Chou are with the Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan (e-mail: jfyang@ee.ncku.edu.tw; yangting115@gmail.com).

Y. Zhou is with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yicongzhou@umac.mo).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2016.2516239

I. Introduction

PATTERN recognition systems are relied on classifiers. Nearest neighbor (NN) [1] and nearest subspace (NS) [2] classifiers are the well-known approaches in pattern recognition area. The NN classifies the test sample based on the best representation to a selected single training sample, whereas the NS classifier is based on the best linear representation in terms of all the training samples in each class.

The number of prototype samples is usually very small, which makes the classification of NN be very difficult. So, the nearest feature line (NFL) proposed by Li *et al.* [3]–[7] attempts to enhance the representational capacity of the limited sample set by using the line passing through each pair of the samples in the same class. After the NFL being proposed, Chien and Wu [8] proposed the nearest feature plane (NFP); Zheng *et al.* [9] proposed the NN line and NN plane; and Gao and Wang [10] proposed the center-based NN [10]. Pan *et al.* [42] proposed neighborhood feature line segment, Feng *et al.* [43] proposed center-based NFP.

Since the samples from a specific subject class are laid on a linear subspace [11]–[13], linear regression-based classification (LRC) [14] is proposed for face identification. The LRC formulates the task of face recognition into a linear regression problem. The LRC classifier can be treated as an extension of the NS classifier. For face recognition, the LRC related approaches, including kernel-LRC [15], kernel center-based LRC (KCWLR) [44], linear discriminant analysis-LRC [16], improved-principal components analysis-LRC [17], and unitary-LRC [18], were proposed to further improve the recognition performance under different situations such as variable illumination and facial expressions.

Different to the LRC with the class-model, sparse representation-based classification (SRC) [19], [20] uses the all-class model to classify the test sample. After the SRC classifier, some other improved methods [21]–[34], such as the two-phase test sample sparse representation (TPTSSR) [23], the collaborative representation-based classification (CRC) [27], regularized robust coding (RRC) [28], and relaxed collaborative representation (RCR) [29] classifiers are proposed to achieve higher accuracy under certain conditions.

Actually, the LRC can be treated as an L2-norm minimum problem of the SRC approaches. The LRC and SRC-related approaches adopt the distance measure, i.e., the projection error, between test sample and class subspace for classification, where the regression parameters are estimated from global all-class model. The final classification is achieved by selecting the class, which has the minimum distance away from the test sample. For all-class linear regression problem, the minimum prediction error of the all-class model might not be a good measure for various conditions since the prediction could take the samples from several different classes to get the optimal solution. However, the least square solution of a linear regression problem will give the higher correlation samples with the higher regression parameters. Motivated by the SRC and statistical properties of stable regression parameters, the superimposed sparse parameter (SSP) classifier is proposed for face recognition in this paper. The SSP classifier first uses the test sample vector and whole train space (all the class subspaces) to calculate the global sparse regression parameters. Then, the SSP computes the SSP score, which is the superposition of the sum of the sparse parameters belonging to the same class iteratively. The final largest SSP score of the class will be used to classify the test sample. A mass of experiments on Georgia Tech (GT) face database, ORL face database, CVL face database, AR face database, and CASIA face database are used to evaluate the proposed algorithm. The experimental results show that the SSP methods achieve better recognition rate (RR) than the up-to-date LRC, SRC, CRC, RRC, RCR, support vector machine (SVM), and TPTSSR classifiers.

II. Reviews

Let $\boldsymbol{Y} = \{\boldsymbol{y}_i^c, i = 1, 2, \ldots, N_c, c = 1, 2, \ldots, M\} \in R^{a\times b}$ denote the prototype image set, where $\boldsymbol{y}_i^c$ with $a \times b$ pixels is the ith prototype image of the cth class, M is the number of classes, and N_c is the number of prototype images in the cth class.

A. Linear Regression Classification Algorithm

For linear regression, each image is transformed to the image vector by column concatenation as $\boldsymbol{y}_i^c \in R^{a\times b} \to \boldsymbol{x}_i^c \in R^{q\times 1}$, where $q = a \times b$. By using the concept that the patterns from the same class lie on a linear subspace, the LRC develops a class-specific model $\boldsymbol{X}_c$ by stacking q-dimensional image vectors as

$$\boldsymbol{X}_c = \begin{bmatrix} \boldsymbol{x}_1^c & \boldsymbol{x}_2^c & \cdots & \boldsymbol{x}_{N_c}^c \end{bmatrix} \in R^{q\times N_c}. \tag{1}$$

Let $\boldsymbol{y}$ be an unlabeled test image and the recognition problem is to classify $\boldsymbol{y}$ as one of the classes. We first transform $\boldsymbol{y}$ into $\boldsymbol{x}$ into the vector form. If $\boldsymbol{x}$ belongs to the cth class, it can be represented as a linear combination of the training images from the same class as

$$\boldsymbol{x} \approx \boldsymbol{x}_{\mathrm{LRC}}^c = \boldsymbol{X}_c \boldsymbol{\beta}_c \tag{2}$$

where $\boldsymbol{\beta}_c \in R^{N_c\times 1}$ is the vector of weighting parameters, which can be calculated by the least square error to obtain

$$\boldsymbol{\beta}_c = \left(\boldsymbol{X}_c^T \boldsymbol{X}_c\right)^{-1} \boldsymbol{X}_c^T \boldsymbol{x}. \tag{3}$$

In (2), the predicted vector $\boldsymbol{x}_{\mathrm{LRC}}^c$ can be treated as the projection of $\boldsymbol{x}$ onto the cth class subspace. The LRC calculates the distance measure between the predicted response vector $\boldsymbol{x}^c$ and the original response vector $\boldsymbol{x}$ as

$$d_{\mathrm{LRC}}^c(\boldsymbol{x}) = \left\|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{LRC}}^c\right\| \tag{4}$$

where $||*||$ means L2-norm. The LRC classification rule is in favor of the class with minimum distance

$$\min_{c*} d_{\mathrm{LRC}}^c(\boldsymbol{x}), \ \ c = 1, 2, \ldots, M. \tag{5}$$

B. Collaborative Representation-Based Classification

Suppose that we have M classes of subjects, we can collect the entire class-specific models, $\boldsymbol{X}_c$ for $c = 1, 2, \ldots, M$ to form the global model as

$$\boldsymbol{X}_G = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \cdots & \boldsymbol{X}_M \end{bmatrix} \in R^{q\times MN_c}. \tag{6}$$

If the vector of all-class regression parameters is denoted as $\boldsymbol{\beta} \in R^{MN_c\times 1}$, it can be calculated as

$$\boldsymbol{\beta} = \left(\boldsymbol{X}_G^T \boldsymbol{X}_G\right)^{-1} \boldsymbol{X}_G^T \boldsymbol{x}. \tag{7}$$

The regularized residual of the cth class, r^c is given as

$$r^c = \frac{\left\|\boldsymbol{x} - \boldsymbol{X}^c \boldsymbol{\beta}^c\right\|}{\left\|\boldsymbol{\beta}^c\right\|} \tag{8}$$

where $\boldsymbol{\beta}^c$ corresponding to the coefficient of the sample of class c is the cth sectioned column of $\boldsymbol{\beta}$. The CRC classification rule in favor of the class with the minimum distance becomes

$$\min_{c*} r^c, c = 1, 2, \ldots, M. \tag{9}$$

C. TPTSSR Classifier

The TPTSSR [5] classifier similar to the CRC first computes weighting parameters, $\boldsymbol{\beta} \in R^{MN_c\times 1}$ stated in (7). However, the TPTSSR chooses the first maximum K samples according to the weighted distance with the parameters in $\boldsymbol{\beta}$. Accordingly, the corresponding data vector after selection of K samples, the selected data samples then form a newly-selected K-global data sample model, $\tilde{\boldsymbol{X}}_{G,K}$. In the second phase, the corresponding optimal weighting parameters for the selected data sample becomes

$$\tilde{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{X}}_{G,K}^T \tilde{\boldsymbol{X}}_{G,K}\right)^{-1} \tilde{\boldsymbol{X}}_{G,K}^T \boldsymbol{x}. \tag{10}$$

If the selected K samples from the cth class are $\tilde{\boldsymbol{X}}_s^c, \ldots, \tilde{\boldsymbol{X}}_t^c$. Let $\boldsymbol{x}_{\mathrm{TPTSSR}}^c = \tilde{\boldsymbol{\beta}}_s^c \tilde{\boldsymbol{X}}_s^c + \cdots + \tilde{\boldsymbol{\beta}}_t^c \tilde{\boldsymbol{X}}_t^c$ be the selected projection, we can calculate the deviation of $\boldsymbol{g}^c$ from $\boldsymbol{x}$ by using

$$d_{\mathrm{TPTSSR}}^c(\boldsymbol{x}) = \left\|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{TPTSSR}}^c\right\|. \tag{11}$$

And the TPTSSR selection rule in favor of the class with minimum distance is given by

$$\min_{c*} d_{\mathrm{TPTSSR}}^c(\boldsymbol{x}), \ \ c = 1, 2, \ldots, M. \tag{12}$$

D. Sparse Representation-Based Classification

For the SRC, we first normalize the columns of $\boldsymbol{X}$ stated into (6) to have unit L2-norm, In other words, the SRC classifier revises the original class-specific model, $\boldsymbol{X}_c$ stated in (1) into a normalized class-specific model as

$$\check{\boldsymbol{X}}_c = \begin{bmatrix} \check{\boldsymbol{x}}_1^c \ \check{\boldsymbol{x}}_2^c \ \ldots \ \check{\boldsymbol{x}}_{N_c}^c \end{bmatrix} \in R^{q\times N_c} \tag{13}$$

where the normalized column vectors is defined as

$$\breve{\boldsymbol{x}}_i^c = \frac{\boldsymbol{x}_i^c}{\left\|\boldsymbol{x}_i^c\right\|},\ \ c = 1, \ldots, M;\, i = 1, \ldots, N_c. \quad (14)$$

Then, the normalized global-specific model $\breve{\boldsymbol{X}}_G^0$ by stacking all class-specific models is expressed as

$$\breve{\boldsymbol{X}}_G = \left[\,\breve{\boldsymbol{X}}_1\ \breve{\boldsymbol{X}}_2\ \cdots\ \breve{\boldsymbol{X}}_M\,\right] \in R^{q\times MN_c}. \quad (15)$$

The L1-norm minimization problem is suggested as

$$\boldsymbol{g} = \arg\min_g ||\boldsymbol{g}||_1 \text{ subject to } \breve{\boldsymbol{X}}_G \boldsymbol{g} = \boldsymbol{x}. \quad (16)$$

To compute the regularized residuals r^c as

$$r^c = \left\|\boldsymbol{x} - \breve{\boldsymbol{X}}_c \hat{\boldsymbol{g}}^c\right\| \quad (17)$$

the SRC classification rule in favor of the class with the minimum distance finally can be expressed by

$$\min_{c*} r^c, c = 1, 2, \ldots, M. \quad (18)$$

III. Proposed Method

The SRC related approaches adopt the distance measure between test sample and class subspace for classification. For all-class linear regression problem, the minimum prediction error of the all-class model might not be a good measure for various conditions since the prediction could take the samples from several different classes to achieve the optimal solution. It is noted that the L1-norm minimum solution always gives the higher correlation samples with the higher sparse parameters [19], [33]. That is to say, the sum of parameters of each class can be used for classification: the maximum value represents the most similar class, to the contrary, the minimum value represents the most unlikely class. Motivated by the above situation, we propose an SSP classifier and its fast realization, called the fast SSP (FSSP) classifier in this section.

A. SSP Classifier

Let each training image be an order $a \times b$ pixels and be represented as $\boldsymbol{y}_i^c \in R^{a\times b}$, $c = 1, 2, \ldots, M$, and $i = 1, 2, \ldots, N_c$. Each gallery image is transformed to column vector such that $\boldsymbol{y}_i^c \in R^{a\times b} \rightarrow \boldsymbol{x}_i^c \in R^{q\times 1}$, where $q = ab$. As the SRC method, in this paper, the SSP classifier uses the normalized global-specific model $\breve{\boldsymbol{X}}_G^0 = \breve{\boldsymbol{X}}_G$, which is depicted in (15) in the first iteration. Given a test image $\boldsymbol{x}$, the all-class L2-norm minimum solution of the CRC stated in (7) then becomes

$$\breve{\boldsymbol{\beta}}_0 = \left(\breve{\boldsymbol{X}}_G^{0^T} \breve{\boldsymbol{X}}_G^0\right)^{-1} \breve{\boldsymbol{X}}_G^{0^T} \boldsymbol{x} \quad (19)$$

where $\breve{\boldsymbol{\beta}}_0 = [\breve{\boldsymbol{\beta}}_0^{1^T}, \breve{\boldsymbol{\beta}}_0^{2^T}, \ldots, \breve{\boldsymbol{\beta}}_0^{c^T} \ldots, \breve{\boldsymbol{\beta}}_0^{M^T}]^T \in R^{MN_c\times 1}$ is the vector of all-class sparse parameters and $\breve{\boldsymbol{\beta}}_0^c$ is the sparse parameter of the cth class, i.e., the cth sectioned vector of $\breve{\boldsymbol{\beta}}^c$.

Since the sparse parameters could be negative, this may affect the accuracy of superposition sparse parameters. So, we regulate the sparse parameters into positive ones as

$$\gamma_{0,i}^c = \left(\breve{\beta}_{0,i}^c - \breve{\beta}_{0,\min}^c\right) \Big/ \left(\breve{\beta}_{0,\max}^c - \breve{\beta}_{0,\min}^c\right) \quad (20)$$

where $\breve{\beta}_{0,\max}^c$ and $\breve{\beta}_{0,\min}^c$ denote the maximum and minimum elements of $\breve{\boldsymbol{\beta}}_0^c$, respectively.

After the first iteration, we suggest, r_0^c, the SSP score of the cth class, to be the sum of the regulated sparse parameters as

$$r_0^c = \sum_{i=1}^{N_c} \gamma_{0,i}^c = \sum_{i=1}^{N_c} \left|\gamma_{0,i}^c\right| \ \text{ for } \ c = 1, 2, \ldots, M. \quad (21)$$

If $r_0^{c_m}$ possesses the minimum score among all classes, the c_mth class will be treated as the most unlikely one. Then, the SSP classifier will remove the c_mth-class subspace from $\breve{\boldsymbol{X}}_G$ to obtain the reduced global-model, $\breve{\boldsymbol{X}}_G^1$ after the first iteration.

After k iterations, we have successively removed k class subspaces to obtain $\breve{\boldsymbol{X}}_G^k$ and its vector of sparse parameters can be calculated as

$$\breve{\boldsymbol{\beta}}_k = \left(\breve{\boldsymbol{X}}_G^{k^T} \breve{\boldsymbol{X}}_G^k\right)^{-1} \breve{\boldsymbol{X}}_G^{k^T} \boldsymbol{x}. \quad (22)$$

By using (20), the normalization sparse parameters γ_k can be obtained from $\breve{\boldsymbol{\beta}}_k$. Thus, the SSP score in the kth iteration of the cth class in term of the sum of the regulated sparse parameters r_k^c is given by

$$r_k^c = \sum_{i=1}^{N_c} \gamma_{k,i}^c. \quad (23)$$

Then, the SSP score by superimposing the previous score of the cth class becomes

$$s_k^c = \begin{cases} s_{k-1}^c + r_k^c & c \in l\left(\breve{\boldsymbol{X}}_G^k\right) \\ s_{k-1}^c & c \notin l\left(\breve{\boldsymbol{X}}_G^k\right) \end{cases} \quad (24)$$

where $l(\breve{\boldsymbol{X}}_G^k)$ denotes the class label of $(M-k)$ class subspaces contained by $\breve{\boldsymbol{X}}_G^k$.

The iterations will be performed until the number of classes is equal to one. And the rule in favor of the class with the maximum SSP score is given as

$$\max_{c*} s_{M-1}^c, \quad c = 1, 2, \ldots, M. \quad (25)$$

It is noted that all the SSP measures stated in (21) and (23) use the L_1-norm sum to compute the SSP measure of each class. According to various applications, of course, the SSP measure in the kth iteration can be modified into the Lp-norm $(p > 0)$ sum as

$$r_k^c = \left(\gamma_{k,1}^c\right)^p + \left(\gamma_{k,2}^c\right)^p + \cdots + \left(\gamma_{k,N_c}^c\right)^p. \quad (26)$$

The evaluation of the appropriately value of p can be found in Section V-F.

The flow chart of the SSP classifier is shown in Fig. 1 and its classification procedure is summarized as follows.

B. Fast SSP

The original SSP classifier removes the most unlikely class one-by-one in each iteration to make the selected classes become more and more discrimination. However, the SSP classifier could face a computation problem if there are a lot of

---

Algorithm 1 SSP

Inputs: M class models $\boldsymbol{X}_c \in R^{q\times N_c}$ for $c = 1, 2, \ldots, M$ and a test image vector $\boldsymbol{x} \in R^{q\times 1}$.

Output: The class index of $\boldsymbol{x}$.

Algorithm procedure:

1. With M class models, we can construct the global-model $\boldsymbol{X}$ as stated in (6).
2. Normalize the columns of $\boldsymbol{X}$ to become unit L2-norm to obtain the normalized global-model $\breve{\boldsymbol{X}}_G$ as stated in (15).
3. Compute the SSP score of each class by (21).
4. For the k^{th} iterations, remove the most unlikely class subspace to obtain the updated global-model $\breve{\boldsymbol{X}}_G^k$ and compute the SSP score of each class by (23).
5. Compute the superimposed SSP score of each class as stated in (24)
6. Repeat Steps 4, and 5 until the number of classes is equal to one.
7. Get the final SSP score vector, s_{M-1} and select the class with the maximum superimposed SSP score by (25).

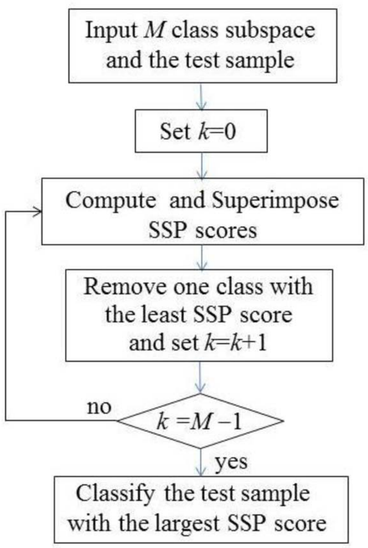

Fig. 1. Flow chart of the SSP classifier.

classes in the database. The SSP classifier might not be practical for the applications with a large number if classes. To solve the computation cost problem, we further propose an FSSP classifier by removing more unlikely classes in each iteration.

From [20], we know that there are only a few classes competing the right class. Since most of classes will have large differences away from the right one, so we can remove $Q(k)$ classes, where $Q(k)$ is a decaying function.

As shown in Fig. 2, the FSSP is similar to the SSP. In the first iteration, the SSP score of the cth class can be computed by (19)–(21). However, the FSSP will remove $Q(k)$ classes. If we plan to reduce the FSSP process from $M-1$ to $\lfloor \log_2(M) \rfloor$ iterations, for example, we can remove about a half of active classes, i.e., $Q(k) \approx M/2^k$. Thus, the FSSP only takes about $\lfloor \log_2(M) \rfloor$ iterations for classification. After k iterations, the

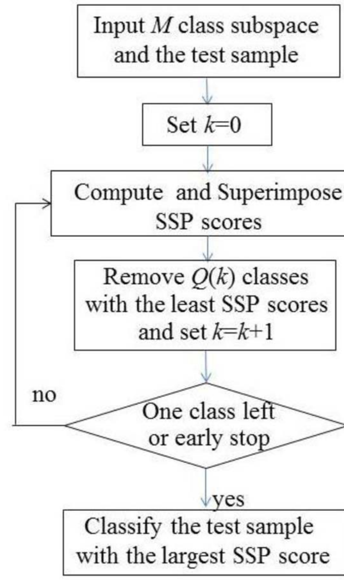

Fig. 2. Flow chart of the FSSP classifier.

final decision rule is in favor of the class with the maximum SSP score as

$$\max_{c*} s_k^c, \quad c = 1, 2, \ldots, M. \tag{27}$$

For 1024 classes, for example, we can reduce the number of iterations from 1024 to 10 with the above half removal strategy. To achieve different computation speed, we can choose other reduction factors or directly-stop strategy to choose the class with the maximum SSP score.

IV. Analyses of the SSP and FSSP Classifiers

For the sparse representation-based approaches, the classification rules are mostly the least regression error for all predictions of all classes. For the linear regression of the all-class model could take the samples from several different classes to achieve the optimization. The least square solution always gives the higher correlation samples with the higher regression parameters. Instead of prediction errors, the regression parameters provide another view for checking the likely and unlikely classes.

A. Features of SSP and FSSP Classifiers

The SSP classifier has two important features: 1) iterative removal of the most unlikely class and 2) successive superposition of score vectors. The removal of the most unlikely class will help to avoid the weakness of linear regression while the successive superposition of score vectors will help to increase the confidence region. If there are ten classes, the sum of linear regression parameter, r_0^c and the score vector $s_0^c (= r_0^c)$ on the original global-model space are shown in Fig. 3. From Fig. 3, we found that the sum of L2-norm sparse parameter of the fourth class is the minimum. So in the first global-model space, the fourth class subspace is removed, that is to say, r_1^4 is set to 0. From Fig. 4, the superimposed score vectors

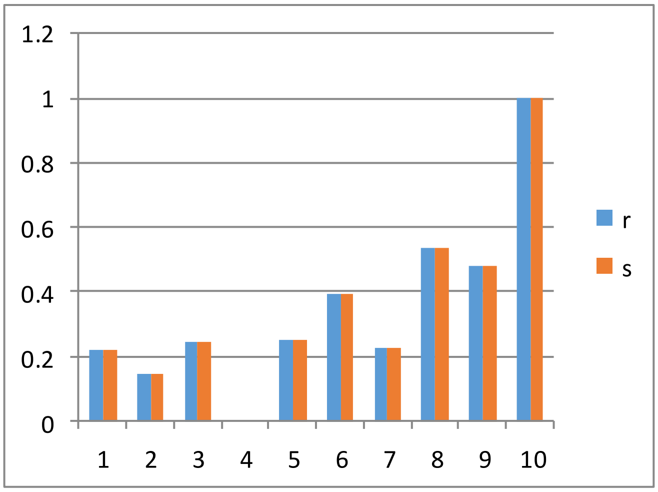

Fig. 3. Sum of sparse parameter r_0^c and the score vector s_0^c in the first iteration.

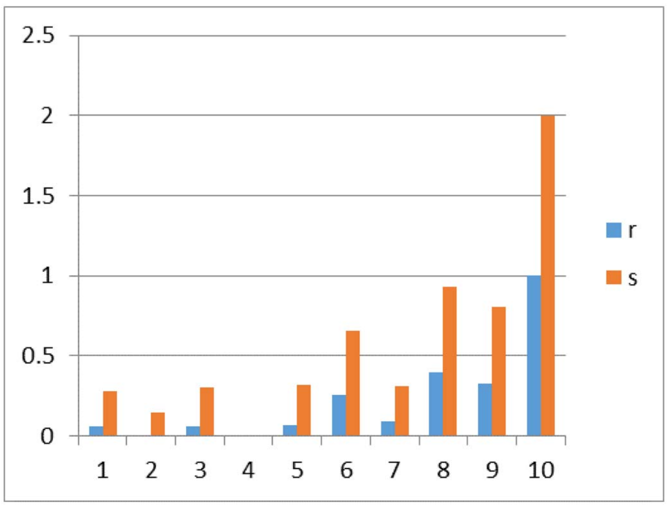

Fig. 4. Sum of sparse parameter r_1^c and the superimposed SSP score vector s_1^c in the second iteration.

obtained in the second iteration becomes more discrimination. For example, the difference value between the maximum value (1) and the second maximum value (0.53) is from 0.47 to the novel difference value (1.09) between the maximum value (2) and the second maximum value (0.91). So, after the superimposed score process, the score vector will become more and more discrimination than the original sparse parameter. The detailed information can be found in Section V-E (Fig. 11). The features of the FSSP are similar to those of the SSP.

B. Relationship With NN, LRC, SRC, CRC, and TPTSSR

The NN classifier uses the distance between the test sample and each prototype sample to classify the test sample. The LRC classifier uses the distance between the test sample and projection samples of each class to classify the test sample. The projection sample is constituted by the L2-norm sparse representation. CRC classifier, TPTSSR, and SRC use the all prototype sample to produce the sparse representation parameter, which is the main difference to the LRC classifier. Besides, SRC uses the L1-norm sparse representation to classify the test sample, which is the second difference to the LRC classifier.

Though the NN classifier, CRC classifier, TPTSSR classifier, and SRC classifier are different, they have one thing in common. It is that they all use the distance between the test sample and an approximated subspace or projection space which is computed from the original prototype samples or the resampled prototype samples, to classify the test sample. However, the SSP classifier, which uses the superposition sparse parameters of the class, uses the different view point of the regression optimization. Since the proposed classifiers is motivated by the TPTSSR, the detailed comparison of the proposed classifiers and TPTSSR is described as follows.

1) Proposed Methods Versus TPTSSR: First, the TPTSSR only iterates two times. The proposed SSP classifiers iterates M times and the FSSP classifiers iterates $\lfloor \log_2(M) \rfloor$ times. The more iteration times may select the better valuable samples, which is helpful for classification.

Second, the TPTSSR does not superimpose results of the first and second steps. The proposed SSP classifier superimposes the results of the M times iterations. The proposed FSSP classifier superimposes the results of the $\lfloor \log_2(M) \rfloor$ times iterations. The superimposition method may make the training set be more and more discrimination, which is also helpful for classification.

Third, the threshold value of the TPTSSR is uncertain. The experimental results in Fig. 13 prove it. The uncertain threshold value will generate a new problem that how to obtain the appropriately threshold value. The problem is a shortcoming of TPTSSR. However, the SSP and the FSSP classifiers only need to check their final SSP scores of the classes. They both have not the problem.

C. Analysis of Computation Cost

Proposition 1: Suppose that $\sum_{c=1}^{k} a_c < b$ and $a_c > 1$ for $c = 1, 2 \ldots, k$. Then $\sum_{c=1}^{k} a_c^p < b^p$ for $p > 1$.

Proof: It is easy to prove the conclusion as follows

$$\sum_{c=1}^{k} a_c^p < \left(\sum_{c=1}^{k} a_c \right)^p < b^p. \qquad \blacksquare$$

Suppose that the computation complexity of the CRC is O_{CRC}. The computation complexity of the first phase of TPTSSR is similar to that of the CRC, the computation complexity of the second phase is according to the number of samples chosen in the first phase. So, O_{TPTSSR} is larger than O_{CRC}. The computation cost of the SRC is much larger than that of the CRC [27], [29]. The computation cost of the first iteration of the SSP and FSSP is also similar to that of the CRC. The whole computation cost of the SSP should be smaller than $M \cdot O_{\mathrm{CRC}}$. O_{CRC} is larger than O(N), where N denotes the number of all the training samples. For FSSP classifier, the sum of the number of the samples in all the iterations (except the first iteration) is smaller than N, according to the

Fig. 5. Some images selected from GT face database.

properties of geometric sequence. Therefore, if half removal strategy is adopted, the computation complexity of the FSSP is less than (close to) $2 \cdot O_{\text{CRC}}$ according to the Proposition 1 and the properties of geometric sequence, which is close to O_{TPTSSR}. Since LRC only uses the class subspace, it is easy to know that O_{LRC} is smaller than O_{CRC}. In summary, the analyzed computation costs of the aforementioned classifiers can be explicated in the order as (28), when the number of classes is large

$$O_{\text{LRC}} < O_{\text{CRC}} < O_{\text{TPTSSR}} < O_{\text{FSSP}} < O_{\text{SRC}} < O_{\text{SSP}}. \quad (28)$$

By simulations, the real run time for each classifier can be found in Table VII in simulation section.

V. Experimental Results

To assess the effectiveness of the proposed methods, the performances of the SSP and FSSP classifiers are compared with those of the SVM [35], [36], RRC, RCR, SRC, LRC, CRC, TPTSSR, and NN classifiers with five famous face databases. In the experiments, SSP(1) and FSSP(1) mean that the SSP score of each class is computed by L1-norm while SSP(2) and FSSP(2) denote that the SSP scores is computed by L2-norm.

"First *N*" scheme is taken for comparisons. The first *N* face images of each class are used as the prototype set. The rest face images of test database are used as test samples. The RR is used to evaluate the performances of all algorithms in each condition, while the average RR (ARR) is exhibited for the average of RRs of all different conditions.

A. Face Recognition With Pose

1) Georgia Tech Face Database [37]: It contains images of 50 people taken in two or three sessions between June 1, 1999 and November 15, 1999 at the Center for Signal and Image Processing at Georgia Institute of Technology. All people in the database are represented by 15 color JPEG images with cluttered background taken at resolution 640×480 pixels. The average size of the faces in these images is 150×150 pixels. The pictures show frontal and/or tilted faces with different facial expressions, lighting conditions, and scale. Each image of the GT face database is manually cropped in 30×40 gray images. Fig. 5 shows some selected face images of GT face database.

In the first experiment, First *N* test scheme is used on the GT face database. The experimental results are exhibited in Table I. The best RR of the TPTSSR is used in Table I. From Table I, we learned that the RRs of the SSP and half-removal FSSP classifiers is superior to those of the other classifiers, when first 3–5 samples are used as the prototype set. The ARR of the SSP(1) classifier outperforms those of the TPTSSR, RCR, RRC, SRC, LRC, CRC, SVM, and NN classifiers with 4.81%, 14.96%, 14.35%, 3.23%, 5.13%, 10.79%, 25.18%, and 8.85% improvements, respectively, when the first 3–5 samples are used as the prototype.

TABLE I
RRs and ARRs of the Classifiers on GT Face Database With 'First *N*' Scheme

Classifier	3	4	5	ARR
NN	0.4983	0.5182	0.5440	0.5202
SVM	0.3067	0.3600	0.4040	0.3569
CRC	0.4600	0.4964	0.5460	0.5008
LRC	0.5183	0.5618	0.5920	0.5574
SRC	0.5317	0.5855	0.6120	0.5764
RRC	0.4483	0.4652	0.4820	0.4652
RCR	0.4267	0.4545	0.4960	0.4591
TPTSSR	0.5317	0.5582	0.5920	0.5606
SSP(1)	0.5717	0.6145	0.6400	0.6087
SSP(2)	0.5717	0.6091	0.6400	0.6069
FSSP(1)	0.5633	0.5945	0.6360	0.5979
FSSP(2)	0.5517	0.6091	0.6300	0.5969

Fig. 6. Some images from CVL face database.

2) CVL Face Database [38]: It contains image of 114 persons with seven images for each person, which is done with Sony Digital Mavica under uniform illumination, no flash, and with projection screen in the background. The age of persons are mostly around 18 (pupils and some professors) and mostly male (around 90%). All people in the database are represented by color JPEG images with resolution 640×480 pixels. In our experiments, the subset of CVL face database contains 770 images of 110 people with seven face images. Each image of the subset is manually cropped in 30×40 gray images. Fig. 6 shows some selected face images of CVL face database.

In the second experiment, First *N* test scheme is used on the CVL face database. The experimental results are exhibited in Table II. The best RR of the TPTSSR is used in Table II. From Table II, we learned that the RR of the SSP and half-removal FSSP classifiers are superior to the RRs of the other classifiers, when the first 4–6 samples are used as the prototype set. The ARR of the FSSP(1) classifier outperforms those of the TPTSSR, RCR, RRC, SRC, LRC, CRC, SVM, and NN classifiers with 4.22%, 11.22%, 10.11%, 5.89%, 7.67%, 7.78%, 3.33%, and 8.78% improvements, respectively, when the first 3–5 samples are used as the prototype.

TABLE II
RRs AND ARRs OF THE CLASSIFIERS ON CVL FACE DATABASE WITH 'FIRST *N*' SCHEME

Classifier	4	5	6	ARR
NN	0.3467	0.4000	0.6400	0.4622
SVM	0.3800	0.4900	0.6800	0.5167
CRC	0.3867	0.4300	0.6000	0.4722
LRC	0.3800	0.4400	0.6000	0.4733
SRC	0.3933	0.4600	0.6200	0.4911
RRC	0.3667	0.4000	0.5800	0.4489
RCR	0.3533	0.4200	0.5400	0.4378
TPTSSR	0.3933	0.4700	0.6600	0.5078
SSP(1)	0.4267	0.5300	0.6800	0.5456
SSP(2)	0.4467	0.5100	0.6800	0.5456
FSSP(1)	0.4200	0.5300	0.7000	0.5500
FSSP(2)	0.4133	0.5300	0.7000	0.5478

Fig. 7. Some face images of the subset of AR database. The first, second, third, and fourth columns are natural, smile, anger, and scream face images, respectively.

B. Face Recognition With Expressions

1) AR Database [39]: It contains over 4000 face images of 126 subjects (70 men and 56 women). To reduce the computational complexity, the subset of AR database includes 800 face images of 100 individuals with eight face images of different expressions (smile, anger, and scream) per subject, and all images in AR database were manually cropped into 50×40 pixels. Some face images of the subset of AR database are shown in Fig. 7. In the third experiments, we first test the RR of smile expression. When the smile face images are used as the test set, the rest face images (natural, anger, and scream) are used as the train set. The evaluation methods of the other expressions for anger and scream are performed in similar manners.

In the experiment, we test the performance of expressions on the AR face database. The experimental results are exhibited in Table III. The best RR of the TPTSSR is used in Table III. From Table III, we learned that the RRs of the SSP and half-removal FSSP classifiers is superior to the RRs of the other classifiers for expression recognition. The ARR of the SSP(2) classifier outperforms those of the TPTSSR, RCR, RRC, SRC, LRC, CRC, SVM, and NN classifiers with 4.50%, 4.33%, 0.16%, 2.33%, 3.98%, 4.00%, 11.00%, and 5.33% improvements, respectively.

2) Cambridge ORL Database [40]: It contains 40 distinct persons, each person having ten different images, taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/no-smiling), and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background and the persons are in upright, frontal position (with tolerance for some side movement). Each image of the ORL face database is manually cropped in 28×23 gray images. Fig. 8 shows some selected face images of ORL face database.

In the fourth experiment, we test the performance of expressions on the ORL face database. The experimental results are exhibited in Table IV. The best RR of the TPTSSR is used in Table IV. From Table IV, we learned that the RRs of the SSP and half-removal FSSP classifiers are superior to the RRs of the other classifiers, when the first 3–5 samples are used as the prototype set. The ARR of the SSP(1) classifier outperforms those of the TPTSSR, RCR, RRC, SRC, LRC, CRC, SVM, and NN classifiers with 2.38%, 11.94%,

TABLE III
RR AND ARRS OF THE CLASSIFIER ON AR FACE DATABASE WITH EXPRESSIONS

Classifier	smile	anger	scream	ARR
NN	0.9900	0.9450	0.7650	0.9000
SVM	0.9600	0.9500	0.6200	0.8433
CRC	1	0.9950	0.7450	0.9133
LRC	0.9950	0.9754	0.7700	0.9135
SRC	1	0.9900	0.8000	0.9300
RRC	1	1	0.8550	0.9517
RCR	1	0.9900	0.7400	0.9100
TPTSSR	1	0.9950	0.7300	0.9083
SSP(1)	1	0.9950	0.8600	0.9517
SSP(2)	1	0.9950	0.8650	0.9533
FSSP(1)	1	0.9900	0.8550	0.9483
FSSP(2)	1	0.9950	0.8600	0.9517

Fig. 8. Some selected images from ORL face database.

TABLE IV
RR AND ARR OF SEVERAL CLASSIFIER ON ORL FACE DATABASE WITH 'FIRST *N*' SCHEME

Classifier	3	4	5	ARR
NN	0.8536	0.8875	0.9100	0.8837
SVM	0.6643	0.7417	0.7900	0.7320
CRC	0.8607	0.9000	0.9050	0.8886
LRC	0.8214	0.8500	0.8950	0.8555
SRC	0.8714	0.9083	0.9300	0.9032
RRC	0.8000	0.8375	0.8300	0.8225
RCR	0.8071	0.8083	0.8350	0.8168
TPTSSR	0.8929	0.9292	0.9150	0.9124
SSP(1)	0.9036	0.9500	0.9550	0.9362
SSP(2)	0.9000	0.9292	0.9450	0.9247
FSSP(1)	0.9000	0.9417	0.9450	0.9289
FSSP(2)	0.9000	0.9292	0.9350	0.9214

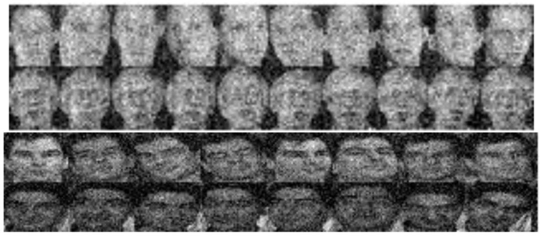

Fig. 9. Noisy test face images of ORL and GT face database (first two rows are from ORL face database; the last two rows are from GT face database).

TABLE V
RRs AND ARRs OF THE CLASSIFIERS ON NOISY ORL DATABASE WITH 'FIRST N' SCHEME

Classifier	3	4	5	ARR
NN	0.8357	0.8792	0.8900	0.8683
SVM	0.6750	0.7458	0.7500	0.7236
CRC	0.8429	0.8750	0.8900	0.8693
LRC	0.8250	0.8500	0.8850	0.8533
SRC	0.8679	0.9042	0.9150	0.8957
RRC	0.6857	0.7167	0.6700	0.6908
RCR	0.7250	0.7208	0.7400	0.7286
TPTSSR	0.8607	0.9048	0.9050	0.8902
SSP(1)	0.9036	0.9375	0.9400	0.9270
SSP(2)	0.8929	0.9375	0.9450	0.9251
FSSP(1)	0.8857	0.9167	0.9200	0.9075
FSSP(2)	0.8821	0.9250	0.9450	0.9174

TABLE VI
RRs AND ARRs OF THE CLASSIFIERS ON NOISY GT FACE DATABASE WITH 'FIRST N' SCHEME

Classifier	3	4	5	ARR
NN	0.4867	0.5109	0.5340	0.5105
SVM	0.2867	0.3345	0.3920	0.3377
CRC	0.4167	0.4709	0.4680	0.4519
LRC	0.5133	0.5600	0.5800	0.5511
SRC	0.5167	0.5436	0.5860	0.5488
RRC	0.3150	0.2836	0.2520	0.2835
RCR	0.3567	0.3455	0.3420	0.3481
TPTSSR	0.5100	0.5182	0.5520	0.5267
SSP(1)	0.5417	0.6000	0.6180	0.5866
SSP(2)	0.5517	0.6091	0.6400	0.6003
FSSP(1)	0.5383	0.5764	0.6080	0.5742
FSSP(2)	0.5317	0.5855	0.6080	0.5751

Fig. 10. Some selected face images from CASIA face database.

11.37%, 3.30%, 8.07%, 4.76%, 20.42%, and 5.25% improvements, respectively, when the first 3–5 face images are used as the prototype.

C. Face Recognition on Noise Face

In this section, we use the GT and ORL face database to test the robustness of the classifier again the noise. The training samples sets are same as that of Sections V-A and V-B. The test samples set is obtained by MATLAB function "imnoise" to insert the Gaussian white noise of zero mean and variance of 0.01. It is noted that we only insert the Gaussian white noise into the test face images, the training face images are used the original images of the database. Some test face images with Gaussian white noise are shown in Fig. 9.

In the fifth experiment, we test the recognition performance of noisy test images on the ORL face database. The experimental results are exhibited in Table V. The best RR of the TPTSSR is used in Table V. From Table V, we learned that the RRs of the SSP and half-removal FSSP classifiers are superior to the RRs of the other classifiers, when the first 3–5 samples are used as the prototype set. The ARR of the SSP(1) classifier outperforms those of the TPTSSR, RCR, RRC, SRC, LRC, CRC, SVM, and NN classifiers with 3.87%, 20.03%, 23.81%, 3.32%, 7.56%, 5.96%, 20.53%, and 6.06% improvements, respectively, when the first 3–5 face images are used as the prototype.

In the sixth experiment, we test the recognition performance of noisy test image on the GT face database. The experimental results are exhibited in Table VI. The best RR of the TPTSSR is used in Table VI. From Table VI, we learned that the RRs of the SSP and half-removal FSSP classifiers are superior to the RRs of the other classifiers, when the first 3–5 samples are used as the prototype set. The ARR of the SSP(1) classifier outperforms those of the TPTSSR, RCR, RRC, SRC, LRC, CRC, SVM, and NN classifiers with 7.36%, 25.22%, 31.68%, 5.15%, 4.92%, 14.84%, 26.26%, and 8.98% improvements, respectively, when the first 3–5 face images are used as the prototype.

D. Evaluation of Computation Cost by Experiments With Large Number of Classes

CASIA [41] Face Image Database Version 5.0 (or CASIA-FaceV5) contains 2500 color facial images of 500 subjects. The face images of CASIA-FaceV5 are captured using Logitech USB camera in one session. The volunteers of CASIA-FaceV5 include graduate students, workers, waiters, etc. All face images are 16 bit color BMP files and the image resolution is 640 × 480. Typical intraclass variations include illumination, pose, expression, eye-glasses, imaging distance, etc. In our experiments, all images are cropped into 40 × 40 gray image. Some face images of CASIA database are shown in Fig. 10.

In Section IV-C, the computation cost of the SSP and the half-removal FSSP as well as the famous classifiers are discussed. If the number of classes is large, the computation cost of the FSSP is much less than that of the SSP. To evaluate the computation cost, we show the run time of each classifier on CASIA face database which contains 500 classes. This experiment is executed on MATLAB 2013a in notebook with Windows 8, Intel Core (i5-3210M) 2.50 GHZ, 3.86 GB RAM.

TABLE VII
RR AND RUN TIME OF SEVERAL CLASSIFIER ON CASIA FACE DATABASE WITH 'FIRST 2' SCHEME

Classifier	Rate	Time (second)
NN	0.1647	0.0322
SVM	0.1413	0.2500
CRC	0.3087	1.4427
LRC	0.1820	0.0855
SRC	0.3300	8.8269
RRC	0.2613	1.0017
RCR	0.2100	0.1795
TPTSSR	0.2853	2.0198
FSSP(1)	0.3387	2.2083
FSSP(2)	0.3647	2.2084

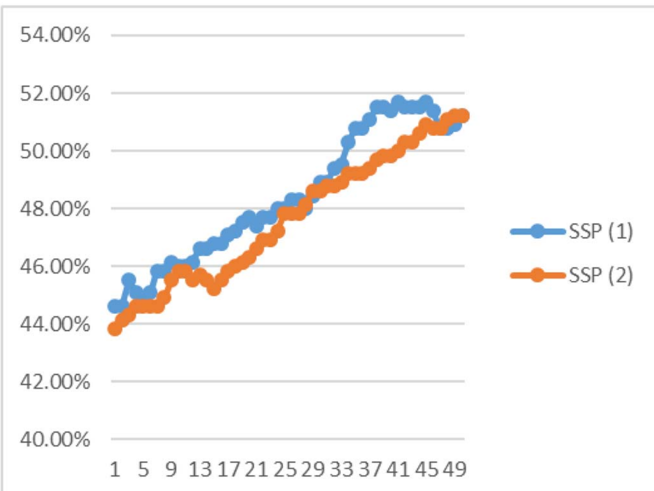

Fig. 11. ARRs versus the number of iterations for the SSP(1) and SSP(2) classifiers on the GT face database.

In the seventh experiment, we test the performance of computation cost on the CASIA face database. The first two samples of each subject are used as the training set, the rest face images are used as the test set. The experimental results in RR and run time are exhibited in Table VII. From Table VII, we learned that the run time of the FSSP classifier is about 25% that of the SRC classifier. Due to a large number of classes (500) and less (2) training images, all the classifiers have low RRs. However, the FSSP(1) and FSSP(2) achieve the best performance with reasonable computational cost among them.

E. Evaluation of the Performance of SSP Iterations

In the experiments 1–6, we gave the RR of the proposed methods for pose, expressions, and noisy variations. In the experiment 7, we provided the run time of the FSSP to show that the proposed FSSP is acceptable for real applications even for very large numbers of classes. In this section, we further show the performance of the proposed method can achieve better to explain "why" the superposition of the SSP score in iterations is needed. To demonstrate the necessity of the iterations, the SSP(1) and SSP(2) classifiers are conducted while we set the early-stop strategy at iteration indices,

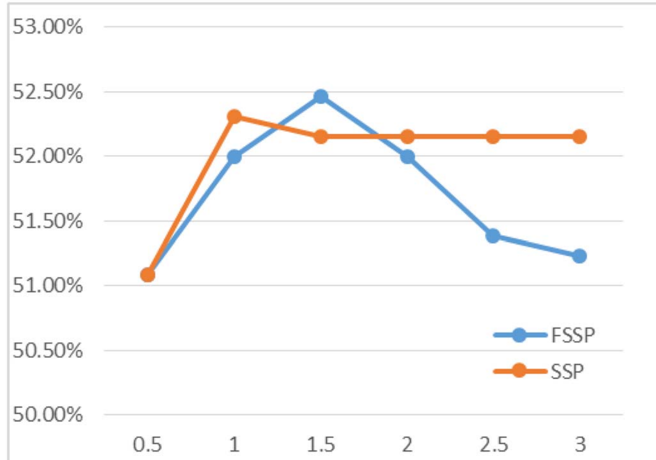

Fig. 12. Performance of the proposed method with the various rule of computing the score. The horizontal axis denotes the value of p.

TABLE VIII
RRS OF THE CLASSIFIERS ON GT FACE DATABASE WITH FIP FEATURE

Deep Learning Feature +Classifiers	RR
FIP feature +LRC	0.4750
FIP feature +CRC	0.4800
FIP feature +SRC	0.4817
FIP feature +SSP(1)	0.5250
FIP feature +SSP(2)	0.5367
FIP feature +FSSP(1)	0.5267
FIP feature +FSSP(2)	0.5317

$k = 1, 2, \ldots, 50$ on the GT face database. Fig. 11 shows ARRs while the horizontal axis denotes the number of iterations conducted. From Fig. 11, we know that the most of iterations can help the SSP classifier to improve the recognition performance.

F. Evaluation With the Various Rule of Computing the Score

In (26), we denote that the appropriately Lp-norm may be helpful for the proposed classifiers. Therefore, in this section, we further show the performance of the proposed method with the various rule of computing the score. In this experiment, the GT face database is used. Two face images of each person are used as training set, the rest face images are used as test set. Fig. 12 shows best RRs while the horizontal axis denotes the value of p. From Fig. 12, we know that the proposed classifier obtains the best performance when the p belongs to the interval of 1–2.

It is noted that we only use the $p = 1$ and $p = 2$ for the proposed classifiers in the above experiments. If we try to find the most appropriately value of p, the proposed classifiers can obtain the better performance.

G. Evaluation on Various Features

In the experiments 1–8, we test the RR and run time of the proposed methods and several existed classifiers with the original face image vectors. That is to say, all the classifiers (proposed classifier and the existed classifiers) do not need to

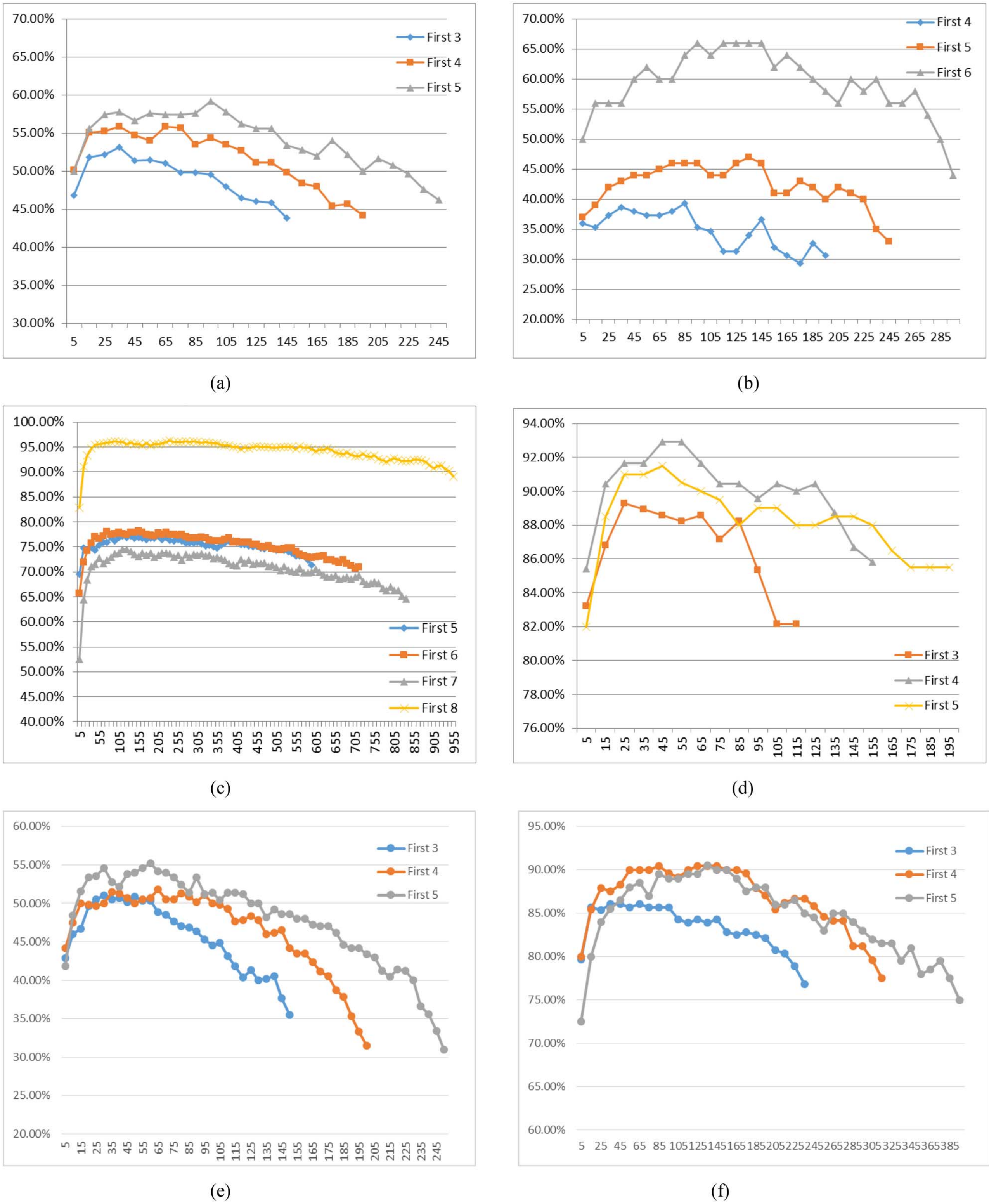

Fig. 13. Performance of the TPTSSR with the various databases. (a) GT face database. (b) CVL face database. (c) AR face database. (d) ORL face database. (e) GT face database with noise. (f) ORL face database with noise. The best RRs of (a)–(f) are shown in Tables I–VI, respectively.

learn the feature before starting the test procedure. However, some researcher pay attention to learn the feature for face recognition [45]–[52]. Therefore, we describe some results of several classifiers on the various features in this section. The used database is the GT face database. Three face images of each person are chosen from the database as the training set, the rest face images are used as the testing set. There are two features are used in this section. They are image

TABLE IX
RRs OF THE CLASSIFIERS ON GT FACE DATABASE WITH IGO FEATURE

Subspace Learning Feature +Classifiers	RR
IGO-PCA +LRC	0.4683
IGO-LDA+LRC	0.4783
IGO-PCA +CRC	0.4667
IGO-LDA+CRC	0.4817
IGO-PCA +SRC	0.4617
IGO-LDA+SRC	0.4733
IGO-PCA +SSP(1)	0.4833
IGO-LDA+ SSP(1)	0.4833
IGO-PCA +SSP(2)	0.4817
IGO-LDA+ SSP(2)	0.4817
IGO-PCA +FSSP(1)	0.4883
IGO-LDA+ FSSP(1)	0.4833
IGO-PCA +FSSP(2)	0.4850
IGO-LDA+ FSSP(2)	0.4817

gradient orientations (IGOs) feature [50], [51] and the face identity-preserving (FIP) feature [52].

IGO Feature: It is based on the subspace learning. This feature is learned from the training set of the GT face database.

FIP Feature: It is based on the deep learning. This feature is shared by Zhu *et al.* [52].

The experimental result is shown in Tables VIII and IX. Analyze the experimental result, we find that the LRC, CRC, and SRC classifiers obtains the completive performance with the IGO feature and FIP feature. However, the proposed classifiers obtain best performance with the FIP feature, which is better than that with the IGO feature.

VI. CONCLUSION

In this paper, a novel SSP classifier and its FSSP are proposed for face recognition. The SSP approaches first adopt the SSP score, which can be expressed in terms of the superposition of the sum of the regression parameters of each class in iterations. The SSP and FSSP classifiers successfully conduct the iterative removal of the most unlikely classes to improve the sparse representation and utilize the successive superposition of score vectors to increase the confidence region. Experimental results show that the proposed SSP and FSSP classifiers achieve better RR than the well-known SRC, RRC, RCR, TPTSSR, LRC, CRC, SVM, and NN classifiers. With comparable computational complexity to the famous classifiers, the FSSP takes the advantage of better recognition performance achieved by the SSP, at the sample time; it uses much less iterations to decrease the computation cost. The analyses and the experimental results on several famous face databases confirm the effectiveness of the proposed algorithms for face recognition.

REFERENCES

[1] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[2] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Madison, WI, USA, 2003, pp. I-11–I-18.

[3] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 439–443, Mar. 1999.

[4] S. Z. Li, K. L. Chan, and C. L. Wang, "Performance evaluation of the nearest feature line method in image classification and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1335–1339, Nov. 2000.

[5] J. Lu and Y.-P. Tan, "Uncorrelated discriminant nearest feature line analysis for face recognition," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 185–188, Feb. 2010.

[6] Q. Feng, J.-S. Pan, and L. Yan, "Restricted nearest feature line with ellipse for face recognition," *J. Inf. Hiding Multimedia Signal Process.*, vol. 3, no. 3, pp. 297–305, Jul. 2012.

[7] Q. Feng, C.-T. Huang, and L. Yan, "Resprentation-based nearest feature plane for pattern recognition," *J. Inf. Hiding Multimedia Signal Process.*, vol. 4, no. 3, pp. 178–191, Jul. 2013.

[8] J.-T. Chien and C.-C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1644–1649, Dec. 2002.

[9] W. Zheng, L. Zhao, and C. Zou, "Locally nearest neighbor classifiers for pattern classification," *Pattern Recognit.*, vol. 37, no. 6, pp. 1307–1309, 2004.

[10] Q.-B. Gao and Z.-Z. Wang, "Center-based nearest neighbor classifier," *Pattern Recognit.*, vol. 40, no. 1, pp. 346–349, Jan. 2007.

[11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[12] R. Barsi and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[13] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.

[14] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.

[15] S.-M. Huang and J.-F. Yang, "Kernel linear regression for low resolution face recognition under variable illumination," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Kyoto, Japan, 2012, pp. 1945–1948.

[16] S.-M. Huang and J.-F. Yang, "Improved principal component regression for face recognition under illumination variations," *IEEE Signal Process. Lett.*, vol. 19, no. 4, pp. 179–182, Apr. 2012.

[17] S.-M. Huang and J.-F. Yang, "Linear discriminant regression classification for face recognition," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 91–94, Jan. 2013.

[18] S.-M. Huang and J.-F. Yang, "Unitary regression classification with total minimum projection error for face recognition," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 443–446, May 2013.

[19] J. Wright *et al.*, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[20] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[21] Y. Xu, Q. Zhu, Z. Fan, Y. Wang, and J.-S. Pan, "From the idea of 'sparse representation' to a representation-based transformation method for feature extraction," *Neurocomputing*, vol. 113, pp. 168–176, Aug. 2013.

[22] Y. Xu, W. Zuo, and Z. Fan, "Supervised sparse representation method with a heuristic strategy and face recognition experiments," *Neurocomputing*, vol. 79, pp. 125–131, Mar. 2012.

[23] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1255–1262, Sep. 2011.

[24] Y. Xu *et al.*, "Using the idea of the sparse representation to perform coarse to-fine face recognition," *Inf. Sci.*, vol. 238, pp. 138–148, Jul. 2013.

[25] J. Yang, D. Chu, L. Zhang, Y. Xu, and J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1023–1035, Jul. 2013.

[26] Y. Xu *et al.*, "Coarse to fine K nearest neighbor classifier," *Pattern Recognit. Lett.*, vol. 34, no. 9, pp. 980–986, 2013.

[27] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. ICCV*, Barcelona, Spain, 2011, pp. 471–478.

[28] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[29] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. CVPR*, Providence, RI, USA, 2012, pp. 2224–2231.

[30] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *Proc. ICCV*, Portland, OR, USA, Jun. 2013, pp. 399–406.

[31] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, Providence, RI, USA, 2013, pp. 2586–2593.

[32] F. Shen and C. Shen, "Generic image classification approaches excel on face recognition," arXiv preprint arXiv:1309.5594, 2013.

[33] J. Li and C.-Y. Lu, "A new decision rule for sparse representation based classification for face recognition," *Neurocomputing*, vol. 116, pp. 265–271, Sep. 2013.

[34] Y. Xu, X. Li, J. Yang, Z. Lai, and D. Zhang, "Integrating conventional and inverse representation for face recognition," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1738–1746, Oct. 2014.

[35] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, Vancouver, BC, Canada, 2001, pp. 688–694.

[36] D. Roobaert and M. M. van Hulle, "View-based 3D object recognition with support vector machines," in *Proc. IEEE Int. Workshop Neural Netw. Signal Process.*, Madison, WI, USA, 1999, pp. 77–84.

[37] (2013). *Georgia Tech Face Database*. [Online]. Available: http://www.anefian.com/research/face_reco.htm

[38] P. Peer. (2013). *CVL Face Database*. [Online]. Available: http://www.lrv.fri.uni-lj.si/facedb.html

[39] A. M. Martinez and R. Benavente, "The AR face database," Centre de Visio per Computador, Universitat Autonoma de Barcelona, Barcelona, Spain, CVC Tech. Rep. 24, Jun. 1998.

[40] The Olivetti & Oracle Research Laboratory. (1994). *The ORL Database of Faces.* [Online]. Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[41] (2013). *CASIA Face Image Database*. [Online]. Available: http://biometrics.idealtest.org/findTotalDbByMode.do?mode=Face

[42] J.-S. Pan, Q. Feng, L. Yan, and J.-F. Yang, "Neighborhood feature line segment for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 387–398, Mar. 2015.

[43] Q. Feng, J.-S. Pan, and L. Yan, "Two classifiers based on nearest feature plane for recognition," in *Proc. 20th IEEE Int. Conf. Image Process. (ICIP)*, Melbourne, VIC, Australia, 2013, pp. 3216–3219.

[44] Q. Feng, C. Yuan, J. Huang, and W. Li, "Center-based weighted kernel linear regression for image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, 2015, pp. 3630–3634.

[45] Y. Xu *et al.*, "Data uncertainty in face recognition," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1950–1961, Oct. 2014.

[46] Z. Lai, W. K. Wong, Y. Xu, C. Zhao, and M. Sun, "Sparse alignment for robust tensor learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1779–1792, Oct. 2014.

[47] Z. Lai, Y. Xu, Q. Chen, J. Yang, and D. Zhang, "Multilinear sparse principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1942–1950, Oct. 2015.

[48] Q. Feng, J.-S. Pan, and I. Lee, "Face recognition using sparse feature sphere centroid classifier," *Electron. Lett.*, vol. 50, no. 17, pp. 1198–1200, Aug. 2014.

[49] Y. Lu *et al.*, "Low-rank preserving projections," *IEEE Trans. Cybern.*, Doi: 10.1109/TCYB.2015.2457611.

[50] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Subspace learning from image gradient orientations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2454–2466, Dec. 2012.

[51] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Sparse representations of image gradient orientations for visual recognition and tracking," in *Proc. Workshop CVPR Human Behav. Anal.*, Colorado Springs, CO, USA, 2011, pp. 26–33.

[52] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning identity-preserving face space," in *Proc. ICCV*, Sydney, NSW, Australia, 2013, pp. 113–120.

Qingxiang Feng received the B.S. degree in computer science and technology from Henan University, Kaifeng, China, in 2010, and the M.S. degree from Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China, in 2013. He is currently pursuing the Ph.D. degree with the University of Macau, Macau, China.

He has published over 20 scientific papers, which contains 10 papers indexed by Science Citation Index. His current research interests include pattern recognition and machine learning.

Chun Yuan received the M.S. and Ph.D. degrees from the Department of Computer Science and Technology, Tsinghua University, Shenzhen, China, in 1999 and 2002, respectively.

He is currently an Associate Professor with the Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University. He was a Post-Doctoral Research Fellow with INRIA, Paris, France, from 2003 to 2004. In 2002, he was an Intern with Microsoft Research Asia, Beijing, China. His current research interests include computer vision, machine learning, video coding and processing, cryptography, and digital rights management.

Jeng-Shyang Pan (M'01–SM'09) received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan, in 1986, the M.S. degree in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1988, and the Ph.D. degree in electrical engineering from the University of Edinburgh, Edinburgh, U.K., in 1996.

He is currently a Professor with the Innovative Information Industry Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China. He is also invited to be an Adjunct Professor with Flinders University, Adelaide, SA, Australia. He has published over 500 papers in which 170 papers are indexed by Science Citation Index. His current research interests include soft computing, information security, and signal processing.

Dr. Pan is a fellow of the Institution of Engineering and Technology, U.K.

Jar-Ferr Yang (S'84–M'88–SM'98–F'07) was born in Keelung, Taiwan, in 1954. He received the B.S. degree from Chung-Yuan Christian University, Taoyuan, Taiwan, in 1977, the M.S. degree from the National Taiwan University, Taipei, Taiwan, in 1979, and the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA, in 1988, all in electrical engineering.

In 1988, he joined the National Cheng Kung University, Tainan, Taiwan, as an Associate Professor and promoted to a Full Professor in 1994 and a Distinguished Professor in 2004. He has published over 121 journals and 188 conference papers. His current research interests include multimedia processing, coding, and recognition and their applications in smart living services.

Prof. Yang was a recipient of the NSC Excellent Research Award in Taiwan, in 2008. From 2004 to 2005, he was selected into the Distinguished Lecturer Program by the IEEE Circuits and Systems Society. From 2008 to 2009, he was the Chair of the IEEE CAS Multimedia Systems and Applications Technical Committee. He previously served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the *IEEE Circuits and Devices Magazine*. He is currently an Associate Editor of the *EURASIP Journal of Advances in Signal Processing* and an Editorial Board Member of the *IET Signal Processing*.

Yang-Ting Chou received the B.S. degree in electrical engineering from National Chung Hsing University, Taichung, Taiwan, in 2010, and the M.S. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2012, where he is currently pursuing the Ph.D. degree.

His current research interests include face recognition and image processing.

Yicong Zhou (M'07–SM'14) received the B.S. degree from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, in 2008 and 2010, respectively, all in electrical engineering.

He is currently an Assistant Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His current research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning.

Dr. Zhou was a recipient of the Third Prize of the Macau Natural Science Award in 2014. He is a member of the International Society for Photo-Optical Instrumentations Engineers and the Association for Computing Machinery.

Weifeng Li received the M.E. and Ph.D. degrees in information electronics from Nagoya University, Nagoya, Japan, in 2003 and 2006, respectively.

In 2006, he joined the Idiap Research Institute, Martigny, Switzerland. In 2008, he moved to Swiss Federal Institute of Technology in Lausanne, Lausanne, Switzerland, as a Research Scientist. Since 2010, he has been an Associate Professor with the Department of Electronic Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. His current research interests include audio and visual signal processing, biometrics, human–computer interactions, and machine learning techniques.

Dr. Li is a member of the Institute of Electronics, Information and Communication Engineers.