

Patch based Semi-supervised Linear Regression for Face Recognition

Yuhua Ding¹, Fan Liu^{2,3*}, Ting Rui⁴, and Zhenmin Tang¹

¹School of Computer Science and Engineering, Nanjing University of Science and Technology
Nanjing, 210094 – CHINA
[e-mail: dingyuhua1210@163.com, tzm.cs@njust.edu.cn]

²College of Computer and Information, Hohai University, Nanjing, 210098 – CHINA

³Nantong Ocean and Coastal Engineering Research Institute, Hohai University, Nanjing, 210098 – CHINA
[e-mail: fanliu@hhu.edu.cn]

⁴Engineering Institute of Engineer Crops, PLA University of Science and Technology, Nanjing, 210007 - CHINA
[e-mail: rtinguu@sohu.com]

*Corresponding author: Fan Liu

*Received May 20, 2017; revised December 25, 2018; accepted February 12, 2019;
published August 31, 2019*

Abstract

To deal with single sample face recognition, this paper presents a patch based semi-supervised linear regression (PSLR) algorithm, which draws facial variation information from unlabeled samples. Each facial image is divided into overlapped patches, and a regression model with mapping matrix will be constructed on each patch. Then, we adjust these matrices by mapping unlabeled patches to $[1, 1, \dots, 1]^T$. The solutions of all the mapping matrices are integrated into an overall objective function, which uses $\ell_{2,1}$ -norm minimization constraints to improve discrimination ability of mapping matrices and reduce the impact of noise. After mapping matrices are computed, we adopt majority-voting strategy to classify the probe samples. To further learn the discrimination information between probe samples and obtain more robust mapping matrices, we also propose a multistage PSLR (MPSLR) algorithm, which iteratively updates the training dataset by adding those reliably labeled probe samples into it. The effectiveness of our approaches is evaluated using three public facial databases. Experimental results prove that our approaches are robust to illumination, expression and occlusion.

Keywords: face recognition, semi-supervised, single sample per person, linear regression

A preliminary conference version of this paper appeared in IEEE BigMM 2017. This is an extended version which has been illustrated in section 1. This work was partially supported by National Natural Science Foundation of China under Grant Nos. 61602150 and 61603190, National major Research Program of China (HGJ Program) under Grant No. 2015ZX01041101, the Fundamental Research Funds for the Central Universities under Grant No. 2018B16214, China Postdoctoral Science Foundation funded project under Grant Nos. 2016M600355 and 2017T100323, Jiangsu Planned Projects for Postdoctoral Research Funds under Grant No. 1601013B, Nantong Science and Technology Project under Grant No. GY12017014.

1. Introduction

Face recognition is a research hotspot because of its potential practical value in the artificial intelligence field. A lot of recognition techniques have been put forward taking into account various situations, in which single sample per person (SSPP) [1] is one of the tough challenging topics. In a large amount of practical applications, such as access control, e-passport and law enforcement, there is usually only one labeled sample available for each individual. In the absence of variation information, it becomes more difficult to fulfill the SSPP face recognition task especially when the probe samples bear dramatic facial variations of illumination, expression, occlusion, etc. Traditional face recognition methods [2-5] usually assume that there are multiple training samples for each subject. When encountering SSPP problem, the performance of these methods will degrade seriously. And what is worse, some methods do not work any more.

This paper presents a patch based semi-supervised linear regression (PSLR) algorithm to deal with SSPP problem. In the proposed method, facial variation information is drawn from large number of unlabeled samples. Borrowing from equidistant prototypes embedding [6], the labels of facial images are treated as coordinate space, and gallery samples are placed at the unit coordinates. In the coordinate space, we term the coordinates that provide no identification information as equidistant points, such as $[0, 0, \dots, 0]^T$ and $[1, 1, \dots, 1]^T$, whose components are equal to each other. In other words, the distances between one equidistant point and each unit coordinate are equal. Different from linear regression analysis with generic learning (LRA-GL) [6] that describes facial variation information by mapping the intra-class differences in the generic dataset to zero vector, we place unlabeled samples at $[1, 1, \dots, 1]^T$ to introduce the variation information, which is illustrated in Fig. 1. Incorporating lots of unlabeled samples into training procedure also helps to avoid overfitting. In order to enhance discrimination ability of mapping matrix and reduce the influence of noise, we adopt $\ell_{2,1}$ -norm for the objective function [7]. To take full advantage of local information from human faces and reduce the dimension of training data, each facial image is divided into a collection of overlapped patches, and a regression model with mapping matrix is constructed for each patch. In order to harvest both local and global benefits, we incorporate the solutions of mapping matrices from all patches into an overall objective function. After mapping matrices are computed, we classify the probe samples by majority voting [8].

Although PSLR can learn face variation information by utilizing the unlabeled samples, the number of the labeled samples is small under SSPP circumstance. Most often, more correctly labeled samples can provide more useful information and lead to higher recognition rate. Therefore, we aim to further take advantages of those probe samples that are regarded to be reliably labeled by PSLR. Based on the observation, we further propose a multistage PSLR (MPSLR) algorithm to incorporate the discrimination information between the probe samples into PSLR. We first classify the probe samples by PSLR. Then, we pick out those reliably labeled probe samples that have more than a certain percentage of votes on their categories. These selected probe samples with their estimated labels are added into the training dataset to learn more reliable mapping matrices. By this way, MPSLR repeatedly updates the training dataset and retrains regression model until it achieves satisfactory result. Experimental evaluation of effectiveness is made on the proposed approaches, which exhibits that our approaches achieve excellent performance.

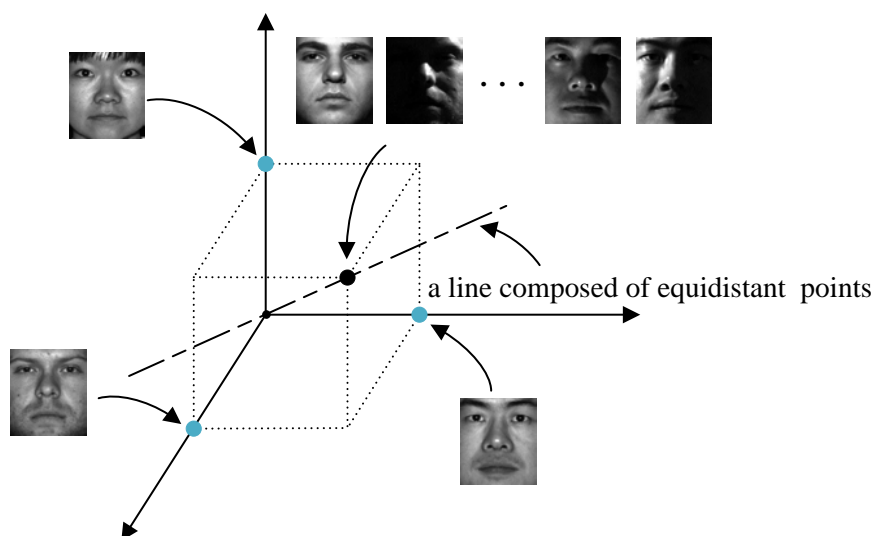


Fig. 1. The illustration of label coordinate space, in which the coordinates of three training images are respectively at $[0,0,1]^T$, $[0,1,0]^T$, $[1,0,0]^T$, and all the unlabeled images are placed at $[1,1,1]^T$. The distances from any certain equidistant point to each unit coordinate are equal, and all the equidistant points lie in one line.

This is an extended version of our conference work [9]. The newly incorporated contributions are highlighted as follows:

- 1) We propose MPSLR algorithm to further improve recognition performance.
- 2) We also provide a theoretical explanation for PSLR. The computational complexities and convergences of our algorithms are also analyzed.
- 3) We carry out more experiments to verify the effectiveness of our methods.

The rest of the paper is organized as follows. We introduce the related work in Section 2. Section 3 presents the details of the proposed algorithms. In Section 4, we carry out experiments and analyze the results. We make a conclusion in Section 5.

2. Related Work

As far as SSPP problem is concerned, the estimation of inter-subject facial variations would become inaccurate, and what is worse, it is impossible to estimate the intra-subject facial variations. To alleviate these difficulties and get as much information as possible, some generic learning based methods were proposed, which are based on the intuition that the intra-subject facial variations of different individuals are similar, and thus can be approximated by estimating from sufficient number of generic faces. Therefore, to collect the effective identification information, a generic training dataset is adopted, in which multiple training samples are available for each individual. For example, Su et al. [1] inferred the identification information of the SSPP gallery dataset by learning a prediction model from the generic training dataset. Deng et al. [6] proposed to map gallery samples and intra-subject differences in the generic training dataset to the equidistant locations and zero vectors, respectively. Some other methods [10][11][12] utilize plentiful intra-subject variations in the generic training dataset, together with gallery images, to represent probe images. In this way, the descriptive ability of these methods on unobserved variations can be substantially

strengthened.

All the above methods dealing with SSPP problem belong to global representation based methods since they represent the global face image with a high-dimensional vector form. However, these holistic methods do not pay attention to the distinctiveness of different parts of human faces and thus are prone to be affected by those regions with variations of illumination, expression, occlusion, etc. In view of this, a large number of patch based methods came out, in which two different treatments can be applied. One way is to treat the patches of each subject as different samples of the same class and then perform feature extraction on them by some discriminant learning techniques. For instance, Chen et al. [13] proposed to divide each face image into several patches and apply Fisher linear discriminative analysis (FLDA) [3] to the data set of newly produced samples. In [14], Lu et al. treated the local patches of each individual as a manifold and maximized the manifold margins of different individuals by learning multiple feature spaces. The other way is to represent different patch using one feature vector separately. And then those well-known classification techniques, such as collaborative representation based classification (CRC) [15] and sparse representation based classification (SRC) [16], can be employed to predict the label of each patch. Then the label results of all the patches can be aggregated to make the final decision by majority voting. Although local region partition can significantly improve the robustness and performance [8][17], patch based methods still have no knowledge about facial variations and suffer from indiscriminative regions.

3. Patch based Semi-supervised Linear Regression

3.1 Motivation

As mentioned in related work, the methods for solving SSPP problem in face recognition generally fall into two categories: global methods and local methods, whose advantages are rather complementary. Therefore, we can harvest their advantages to overcome their disadvantages. In other words, we should take into account both global and local information. Furthermore, the facial variations cannot be fully measured by the conventional unsupervised or supervised methods due to the challenge of SSPP problem. Considering that facial variation information can be drawn from lots of unlabeled samples, we should also make full use of facial variation information from unlabeled samples to deal with SSPP problem. Based on the above motivation, we propose an intuitive and effective method, called patch based semi-supervised linear regression (PSLR), which fully utilizes unlabeled samples and integrates the strengths of both global and local representation.

3.2 Patch based Linear Regression

In light of the fact that more identification information can be provided by local region partition, we propose a patch based linear regression (PLR) algorithm. Each face image is divided into a large number of overlapped patches, and the mapping matrices of different patches are formulated into an overall objective function.

For clarity, we give the definition that $T(i) \in \mathcal{R}^K$ is the label indicator vector of the i -th person, i.e. $T(i) = [0, \dots, 1, \dots, 0]^T$, the i -th component of which is 1 and the others are zero. Without loss of generality, we assume that gallery samples are arranged according to which label they belong to. Then the label matrix of gallery samples, i.e. $Y = [T(1), T(2), \dots, T(K)]$, is a $K \times K$ identity matrix.

For any face image, we partition it into S overlapped patches. And the patches at the same

location of all gallery face images are collected to form a patch gallery dataset. In this way, we can acquire S patch gallery datasets $\{X_1, X_2, \dots, X_S\}$, where $X_i \in \mathcal{R}^{d \times K}$ ($i=1, 2, \dots, S$). S patch probe datasets $\{Z_1, Z_2, \dots, Z_S\}$ are also acquired as well. After dividing S local patches, the corresponding mapping matrix W_i ($i=1, 2, \dots, S$) can be obtained by solving the following optimization problem:

$$\min_w \|E\|_{2,1} + \lambda \|W\|_{2,1} \quad s.t. \quad E_i = Y^T - X_i^T W_i, \quad \forall i, \quad (1)$$

where $\|E\|_{2,1} = \sum_k \|E(k, :)\|_2$, $\|W\|_{2,1} = \sum_k \|W(k, :)\|_2$, $E = [E_1^T, \dots, E_S^T]^T$, $W = [W_1^T, \dots, W_S^T]^T$ and λ is the regularized parameter. In order to improve discrimination ability of mapping matrices and reduce the influence of noise, we adopt $\ell_{2,1}$ -norm for the formulations of loss function and regularization respectively. This formulation actually describes a global objective function, which compute the mapping matrix of each patch as a whole.

For the i -th patch of a probe face image z , its label is given by:

$$\text{label}(z_i) = \arg \max_k (y = W_i^T z_i). \quad (2)$$

3.3 Patch based Semi-supervised Linear Regression

If the label space of facial images are treated as coordinate system where the label indicator vectors are unit coordinates, all the unlabeled or labeled samples should have their own locations. Usually, we place gallery samples at unit coordinates $\{T(i)\}_{i=1}^K$ by reference to equidistant prototypes embedding [6]. Nevertheless, where should those unlabeled samples be located in the coordinate system? Those unlabeled samples do not fall within any class in the training stage because of their ‘unlabeled’ status. In view of this, the location of one unlabeled sample should not provide any identification information, which means that the location is equally distant from each unit coordinate. In other words, the probabilities of samples at the location falling within each class are the same. Such locations without identification information are termed as equidistant points. More specifically, we denote the coordinate of one unlabeled sample as $t = [t_1, t_2, \dots, t_K]^T$, then we obtain the following inference:

$$\|t - T(1)\|_2 = \dots = \|t - T(K)\|_2 \Rightarrow t_1 = t_2 = \dots = t_K. \quad (3)$$

Since there exist lots of probe samples which have no labels in the training stage, we use them as unlabeled samples, which can be placed at any equidistant point. Unlabeled samples should be far away from gallery samples in the coordinate system since they do not fall into any of the gallery categories during training. To compensate for the deviation between the gallery and probe samples, unlabeled samples should be placed at the same location. To prove this point, we first present the following assumption:

Assumption 1. *The intra-subject variation of any gallery facial images can be approximated by a linear combination of the differences from sufficient number of unlabeled face images.*

The probe sample x of subject k can be regarded as a superposition of the corresponding gallery sample x_k and intra-class variation x_v , i.e., $x = x_k + x_v$. Since the difference of any two unlabeled samples makes one variant base ϕ_j , given the unlabeled samples $\{x_{u_j}\}_{j=1}^N$, there

are m ($m = N(N-1)/2$) variant bases $\{\phi_j\}_{j=1}^m$ ($\{\phi_j | \phi_j = x_{u_p} - x_{u_q}, \forall p, q, p > q\}$). **Assumption 1** ensures that the unobserved variations between the gallery and probe sample of any subject can be approximated by a linear combination of variant bases $\{\phi_j\}_{j=1}^m$. Then, the intraclass variation x_v of subject k can be recovered by

$$x_v \approx \sum_{j=1}^m \beta_j \phi_j. \quad (4)$$

Given the mapping matrix W , the responding label vector y relative to the probe sample x can be calculated as follows:

$$y = W^T x = W^T (x_k + x_v). \quad (5)$$

Ideally, y should be equal to the label vector of the gallery sample x_k , i.e.,

$$y = y_k = W^T x_k. \quad (6)$$

To make (6) true, all the unlabeled samples are mapped to the same location, i.e. for $\forall j$, $W^T \phi_j = W^T (x_{u_p} - x_{u_q}) = \vec{0}$. Therefore, $W^T x_v \approx W^T \sum_{j=1}^m \beta_j \phi_j = \vec{0}$. In other words, if all the unlabeled samples are mapped to the same location, the influence of intra-class variation can be approximately eliminated.

Based on the above analysis, we can make unlabeled samples integrated into the objective function of (1) by mapping them to the equidistant point $t^* = [1, 1, \dots, 1]^T$. Then, the new optimization problem is as follows:

$$\begin{aligned} & \min_W \|E\|_{2,1} + \alpha \|E_u\|_{2,1} + \lambda \|W\|_{2,1} \\ & \text{s.t. } E_i = Y^T - X_i^T W_i, \quad E_{u_i} = Y_u^T - X_{u_i}^T W_i, \quad \forall i, \end{aligned} \quad (7)$$

where $Y_u = [t^*, \dots, t^*] \in \mathcal{R}^{K \times n}$ and X_u denote the labels and data of unlabeled samples, respectively, $X_{u_i} \in \mathcal{R}^{d \times n}$, $E = [E_1^T, \dots, E_S^T]^T$, $W = [W_1^T, \dots, W_S^T]^T$, $E_u = [E_{u_1}^T, \dots, E_{u_S}^T]^T$. And α is a balance factor.

We define that $Y_\alpha = [Y, \alpha Y_u]$ and $X_{\alpha_i} = [X_i, \alpha X_{u_i}]$, and the objective function of (7) can be equivalently reformulated as follows:

$$\min_W \|E_\alpha\|_{2,1} + \lambda \|W\|_{2,1} \quad \text{s.t. } E_{\alpha_i} = Y_\alpha^T - X_{\alpha_i}^T W_i, \quad \forall i. \quad (8)$$

Here $E_\alpha = [E_{\alpha_1}^T, \dots, E_{\alpha_S}^T]^T$ and $W = [W_1^T, \dots, W_S^T]^T$.

3.4 Multistage Patch based Semi-supervised Linear Regression

Though PSLR learns the intra-class variation information by placing the unlabeled samples at equidistant points, it cannot obtain the discrimination information hidden between the probe samples. In order to take advantage of the discrimination information between the probe samples, we further propose a multistage PSLR (MPSLR) algorithm. By using PSLR, we can

obtain the estimated labels of the probe samples. As we adopt majority-voting strategy to classify the probe samples, it is statistically guaranteed that the true label always ranks first in the voting list. In other word, many probe samples receive lots of votes on the category that they actually belong to. Therefore, these probe samples can be regarded as reliable samples with effective identification information. MPCLR aims at repeatedly adding the reliable probe samples into the training dataset and retraining a more reliable regression model.

In order to clearly explain how to pick out the reliable probe samples, we assume that one probe sample is assigned to one category, on which it gets P_C percent of total votes. We set a certain threshold δ , a probe sample will be marked as “reliable” as long as $P_C/100 \geq \delta$. For reducing the influence of probe samples with wrong estimated labels, at most 80 percent of the probe samples can be marked. We denote the reliable probe samples as $\bar{X} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N]$, and their estimated labels are denoted as $\bar{Y} = [\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N]$. To employ the effective identification information from the reliable probe samples, Eq.(7) can be extended as:

$$\begin{aligned} & \min_w \|E\|_{2,1} + \alpha \|E_u\|_{2,1} + \lambda \|W\|_{2,1} + \sum_{i=1}^S \sum_{j=1}^N \|\bar{y}_j^T - \bar{x}_j^T W_i\|_2 \\ & \text{s.t. } E_i = Y^T - X_i^T W_i, \quad E_{u_i} = Y_u^T - X_{u_i}^T W_i, \quad \forall i. \end{aligned} \quad (9)$$

Although more discriminative information is captured, there are two issues to deal with to obtain a more robust mapping matrix W :

- 1) As the training dataset is expanded, the estimated labels of the additional reliable probe samples are not entirely correct.
- 2) The number of the reliable probe samples assigned to each category is different, which means that the sample size of different category is imbalanced. In that case, the learned mapping matrix W will be biased.

To reduce the influence of the reliable probe samples with wrong estimated labels and suppress the imbalance of different category, we introduce into Eq.(9) a weight vector $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_N]^T \in \mathcal{R}^N$, where $\gamma_i (i=1, 2, \dots, N)$ is the weight of error item associated with the i -th reliable probe sample \bar{x}_i . About the value of γ_i , we give the following definition. Assume that the i -th reliable probe sample \bar{x}_i belongs to C -th subject and there are M reliable probe samples that belong to C -th class, the value of γ_i is $1/M$. It is formulated as follows:

$$\gamma_i = 1 / \sum_{j=1}^N \text{ind}(\bar{y}_i == \bar{y}_j), \quad (10)$$

where $\text{ind}(x) = \begin{cases} 1, & \text{if } x \text{ true} \\ 0, & \text{if } x \text{ false} \end{cases}$. That is to say that the sum of the weights that associated with

the same subject is 1, which makes the amount of information introduced into each class equal. Then, the new optimization problem is formulated as follows:

$$\begin{aligned} & \min_w \|E\|_{2,1} + \alpha \|E_u\|_{2,1} + \lambda \|W\|_{2,1} + \sum_{i=1}^S \sum_{j=1}^N \gamma_j \|\bar{y}_j^T - \bar{x}_j^T W_i\|_2 \\ & \text{s.t. } E_i = Y^T - X_i^T W_i, \quad E_{u_i} = Y_u^T - X_{u_i}^T W_i, \quad \forall i. \end{aligned} \quad (11)$$

Eq.(11) can be expressed briefly as follows:

$$\begin{aligned} & \min_w \|E\|_{2,1} + \alpha \|E_u\|_{2,1} + \|E_\gamma\|_{2,1} + \lambda \|W\|_{2,1} \\ & s.t. E_i = Y^T - X_i^T W_i, E_{u_i} = Y_u^T - X_{u_i}^T W_i, E_{\gamma_i} = Y_\gamma^T - X_{\gamma_i}^T W_i, \forall i, \end{aligned} \quad (12)$$

where $E_\gamma = [E_{\gamma_1}^T, \dots, E_{\gamma_s}^T]^T$, $Y_\gamma = [\gamma, \gamma, \dots, \gamma]^T \odot \bar{Y}$, $X_{\gamma_i} = [\gamma, \gamma, \dots, \gamma]^T \odot \bar{X}_i$, and \odot denotes the Hadamard product of two matrices.

We define that $Y_\varphi = [Y, \alpha Y_u, Y_\gamma]$ and $X_{\varphi_i} = [X_i, \alpha X_{u_i}, X_{\gamma_i}]$, then the optimization problem from (12) can be equivalently reformulated as follows:

$$\min_w \|E_\varphi\|_{2,1} + \lambda \|W\|_{2,1} \quad s.t. E_{\varphi_i} = Y_\varphi^T - X_{\varphi_i}^T W_i, \forall i, \quad (13)$$

where $E_\varphi = [E_{\varphi_1}^T, \dots, E_{\varphi_s}^T]^T$ and $W = [W_1^T, \dots, W_s^T]^T$. To prove the effectiveness of multistage training strategy, we also propose multistage PLR (MPLR) algorithm which is a special case of MPSLR algorithm with $\alpha = 0$.

MPSLR gets more information than PSLR by incorporating the reliable probe samples into the retraining process. After retraining, the mapping matrix W is updated, and the estimated labels of the probe samples are also updated correspondingly. Then, the reliable probe samples are picked out anew. We gather the reliable probe samples \bar{X} and the training dataset Φ ($\Phi = \{X, X_u\}$) to generate the new training dataset Ω ($\Omega = \{X, X_u, \bar{X}\}$). In each loop iteration, Ω is used to update W . In MPSLR, this process is repeatedly conducted until the terminal condition is satisfied. As shown in **Algorithm 1**, the whole procedure of MPSLR algorithm is listed.

Algorithm 1 MPSLR algorithm

- Step 1.** Get the mapping matrix W by solving Eq.(8).
 - repeat**
 - Step 2.** Classify the probe samples by majority voting.
 - Step 3.** Pick out the reliable probe samples \bar{X} and generate Ω .
 - Step 4.** Calculate the weight vector γ .
 - Step 5.** Use Ω to update W by solving Eq.(13).
 - until** the iteration number is reached or 80% of the probe samples are added into the training dataset.
 - Step 6.** Classify the probe samples and output the final result.
-

3.5 Optimization via Inexact ALM

In this section, we will develop the optimization algorithm to solve Eq.(8) and Eq.(13) using the inexact Augmented Lagrange Multiplier (ALM) method [18]. Obviously, both Eq.(8) and Eq.(13) are in the form

$$\min_w \|E\|_{2,1} + \lambda \|W\|_{2,1} \quad s.t. E_i = Y^T - X_i^T W_i, \forall i, \quad (14)$$

where $E = [E_1^T, \dots, E_s^T]^T$ and $W = [W_1^T, \dots, W_s^T]^T$.

For the sake of simplicity, we will give the optimization algorithm to solve Eq.(14). To deal with our problem, an assistant variable J is introduced to make the objective function separable and the model in Eq.(14) is rewritten as:

$$\min_W \|E\|_{2,1} + \lambda \|J\|_{2,1} \quad s.t. \quad E_i = Y^T - X_i^T W_i, \quad J = W, \quad \forall i. \quad (15)$$

The augmented Lagrange function L of (15) is:

$$\begin{aligned} L(E, J, W, G, H, \mu) &= \|E\|_{2,1} + \lambda \|J\|_{2,1} + Tr(G^T (J - W)) + \sum_{i=1}^S Tr(H_i^T (Y^T - X_i^T W_i - E_i)) + \\ &\quad \frac{\mu}{2} (\|J - W\|_F^2 + \sum_{i=1}^S \|Y^T - X_i^T W_i - E_i\|_F^2) \\ &= \sum_{i=1}^S \|E_i\|_{2,1} + \lambda \|J_i\|_{2,1} + \frac{\mu}{2} \left\| Y^T - X_i^T W_i - E_i + \frac{H_i}{\mu} \right\|_F^2 + \\ &\quad \frac{\mu}{2} \left\| J_i - W_i + \frac{G_i}{\mu} \right\|_F^2 - \frac{1}{2\mu} \|H_i\|_F^2 - \frac{1}{2\mu} \|G_i\|_F^2 \end{aligned} \quad (16)$$

where $Tr(\cdot)$ is the trace of a matrix, $G = [G_1^T, \dots, G_S^T]^T$, $H = [H_1^T, \dots, H_S^T]^T$ are the Lagrange multipliers, and $\mu > 0$ is a penalty parameter. Then, we update E, J, W by applying the idea of alternating minimization, i.e. update one variable with the others fixed.

Provided the current point E^k, J^k, G^k, H^k , we update W^{k+1} by minimizing L with respect to W , i.e.

$$W^{k+1} = \arg \min_W \sum_{i=1}^S \frac{\mu}{2} \left\| Y^T - X_i^T W_i - E_i^k + \frac{H_i^k}{\mu} \right\|_F^2 + \frac{\mu}{2} \left\| J_i^k - W_i + \frac{G_i^k}{\mu} \right\|_F^2, \quad (17)$$

which produces the optimal updation as

$$W_i^{k+1} = (X_i X_i^T + I)^{-1} (X_i Y^T - X_i E_i^k + J_i^k + \frac{X_i H_i^k + G_i^k}{\mu}), \quad \forall i, \quad (18)$$

where I is the identity matrix.

To update J , we need to solve

$$J^{k+1} = \arg \min_J \lambda \|J\|_{2,1} + \frac{\mu}{2} \left\| J - W^{k+1} + \frac{G^k}{\mu} \right\|_F^2, \quad (19)$$

which is in the form [19] $\min_S \gamma \|S\|_{2,1} + \frac{1}{2} \|S - T\|_F^2$, and its closed-form solution is

$$S(i,:) = \begin{cases} \frac{\|T(i,:)\|_2 - \gamma}{\|T(i,:)\|_2} T(i,:), & \text{if } \|T(i,:)\|_2 > \gamma; \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Given W and J , E can be updated by solving

$$E^{k+1} = \arg \min_E \sum_{i=1}^S \|E_i\|_{2,1} + \frac{\mu}{2} \left\| Y^T - X_i^T W_i^{k+1} - E_i + \frac{H_i^k}{\mu} \right\|_F^2, \quad (21)$$

which can be divided into S subproblems

$$E_i^{k+1} = \arg \min_{E_i} \|E_i\|_{2,1} + \frac{\mu}{2} \left\| Y^T - X_i^T W_i^{k+1} - E_i + \frac{H_i^k}{\mu} \right\|_F^2, \quad \forall i. \quad (22)$$

It is worth noting that the optimization subproblems from (22) can also be solved in the same way with that from (19).

As a conclusion, the updating scheme is described in [Algorithm 2](#).

Algorithm 2 Inexact ALM algorithm for optimizing Eq. (14)

Input: X, Y and Parameter λ

Output: W

Initialize: $E = H = 0$, $J = W = G = 0$, $\mu = 0.5$, $\mu_{\max} = 10^{10}$,
 $\varepsilon = 10^{-8}$, and $\rho = 1.1$.

while not converged, $k \leq \text{maxIter}$ **do**

1. fix J , E and update W via (18).
2. fix W , E and update J via (19).
3. fix W , J and update E via (22).
4. update the multipliers

$$G^{k+1} = G^k + \mu(J^{k+1} - W^{k+1}).$$

$$H_i^{k+1} = H_i^k + \mu(Y^T - X_i^T W_i^{k+1} - E_i^{k+1}), \quad \forall i.$$

5. update μ by $\mu = \min(\rho\mu, \mu_{\max})$.
6. check the convergence conditions:

$$\|Y^T - X_i^T W_i^{k+1} - E_i^{k+1}\|_{\infty} < \varepsilon, \quad \forall i \quad \text{and} \quad \|J^{k+1} - W^{k+1}\|_{\infty} < \varepsilon.$$

end while

3.6 Computational Complexity Analysis

In this part, we make analysis to the computational complexities of our methods. As with others, we use notation O to express the time complexity [20].

Assume that n is the number of samples, d is the feature dimension of patches, S is the number of patches of each sample and c is the class number. As stated in Section 3.5, the proposed optimization problem is solved iteratively. In each iteration, W is computed with the complexity of $O(S(dnc + d^2n + d^3))$ while J is computed with the complexity of $O(Sdc)$. Then, it costs $O(Sdnc)$ to calculate E . Finally, G and H are got with the cost of $O(Sdc)$ and $O(Sdnc)$, respectively. Thus the total time complexity of PSLR is $O(KS(dnc + d^2n + d^3))$, where K is the iterative number in Algorithm 2. And the whole cost of MPSLR is $O(TKS(dnc + d^2n + d^3))$, where T is the iterative number in Algorithm 1.

4. Experimental Results and Analysis

In this section, we evaluate the proposed methods using Extended Yale B [21], AR [22] and LFW [23] databases. Our methods are compared with some popular approaches for SSPP problem including AGL [1], LRA*-GL [6], ESRC [10], SVDL [11], BlockFLD [13], PSRC [16], PCRC [17] and LGR [24]. The gray image is used for all the methods, and we resize all the facial images to 80×80 . For patch based methods including BlockFLD, PSRC, PCRC, LGR and the proposed methods, the patch size is fixed to 11×11 and the overlap between adjacent patches is 4 pixels. For MPSLR and MPLR, $\delta = 0.1$. We use $\lambda = 0.01$, $\alpha = 0.02$ to get the best result in our experiments.

4.1 Extended Yale B Database

The Extended Yale B database [21] includes 38 subjects under 64 kinds of illumination conditions. It is quite challenging for most methods to achieve satisfactory performance on the database due to its extreme lighting conditions. We show some sample images in Fig. 2.



Fig. 2. Sample images from Extended Yale B database.

To prove the effectiveness of our methods for illumination and discuss the effect of semi-supervised and multistage techniques on recognition performance, all the samples are randomly divided into two groups, each with 32 images. For each subject, we randomly choose 1 to 5 facial images from the first group for training. All the images from the second group are used for testing. The testing is performed 5 times and the average recognition rates are shown in Fig. 3. It can be noticed that MPSLR always achieves the best result, with the number of training samples increasing from 1 to 5. Comparing with PLR and PSLR, MPLR and MPSLR get better performances, which can fully validate the effectiveness of multistage technique. As the training sample size increases, the recognition rates of these four algorithms

are approaching gradually, which shows that semi-supervised and multistage techniques play a very small role on dealing with illumination changes as long as there are enough training samples. In other words, the fewer training samples become, the greater semi-supervised and multistage techniques contribute.

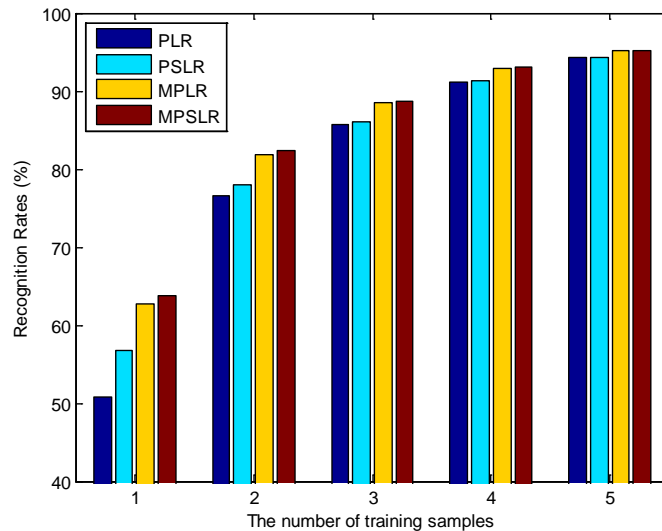


Fig. 3. Recognition rates with different numbers of training samples on Extended Yale B Database.

To further assess the ability of the proposed approaches in dealing with SSPP problem, we compare them with several popular methods designed for SSPP face recognition. The evaluation experiments are carried out with the first 30 subjects, and the other 8 subjects are used as the generic dataset for those generic learning based methods. We use the frontal faces with neutral illumination as gallery samples and the other images as probe samples. We list the experimental results in **Table 1**. Benefiting from local representation, the recognition rate of PSRC, PCRC, LGR and PLR reach 88.47%, 88.10%, 87.51% and 88.99%, respectively. By taking the advantages of both variation information from unlabeled samples and discrimination information from probe samples, MPSLR reaches 96.83% that is the highest recognition rate. The recognition performance of MPLR is similar to that of MPSLR, this is because multistage technique plays a leading role when semi-supervised and multistage techniques are both applied to PLR. In a sense, multistage technique also belongs to semi-supervision, because the labels of reliable probe samples are assigned by the algorithm itself. One can also see that MPSLR outperforms PSLR by 3.23% while MPLR outperforms PLR by 7.52%. And the good performance of our approaches also verifies the robustness of semi-supervised and multistage techniques for illumination.

Table 1. Recognition Rates (%) on Extended Yale B Database for SSPP Problem.

Method	AGL	LRA*-GL	ESRC	SVDL	BlockFLD	PSRC
Accuracy	60.32	68.20	67.62	66.24	74.55	88.47
Method	PCRC	LGR	PLR	PSLR	MPLR	MPSLR
Accuracy	88.10	87.51	88.99	93.60	96.51	96.83

4.2 AR Database

The AR database [22] consists of over 4,000 color face images of 126 people, which contains frontal faces with different lighting conditions, facial expressions and disguises. 26 pictures are available for each subject, which are taken in two separate sessions. Some facial images are shown in Fig. 4.



Fig. 4. Sample images from AR database.

Following the experimental setup in [10], we collect a subset with 2500 images from 100 individuals, which includes 50 males and 50 females. And we randomly choose 1 to 5 images per subject from Session 1 for training. All of 12 images from Session 2 are used for testing. The evaluation experiment is performed 5 times. Then the average accuracies of our methods are reported, which are shown in Fig. 5. It can be seen that MPLR and MPLR are always superior to PLR and PSLR, no matter how many samples are used for training. The effectiveness of multistage technique is confirmed again. And we can also find that the results from PSLR and MPLR are close to those from PLR and MPLR, respectively. This is mostly because the training samples selected from the AR database may contain expression or occlusion.

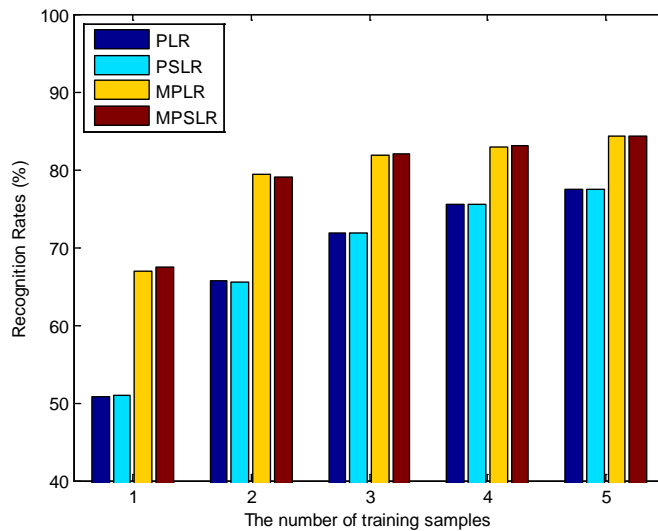


Fig. 5. Recognition rates with different numbers of training samples on AR Database.

We perform experiments to further assess the recognition ability of our approaches for SSPP problem. 80 subjects are used for evaluation, which consists of the first 40 male and the first 40 female subjects. And the other 20 subjects are used as the generic dataset. We use the single face image of each subject with neutral illumination and expression from session 1 as

gallery image. The other images of session 1 and session 2 form the probe dataset. We show the experimental results in **Table 2**. One can observe from the tables that MPSLR obtains the highest average accuracy of 88.91% outperforming PSLR by 4.48% while MPLR works the third best outperforming PLR by 8.33%. Comparing with PLR, PSLR achieves 5% improvement. Although LGR obtains the second highest accuracy being superior to PSLR, the performance of LGR relies greatly on the generic dataset. If we carry out experiments for LGR using PIE-C27 as the generic dataset, PSLR has a better performance compared with LGR. Moreover, the experimental results also demonstrate that the proposed methods are robust to illumination, expression and occlusion.

Table 2. Recognition Rates (%) on AR Database for SSPP Problem.

Method	illumination	expression	disguise	illumination+ disguise	average
AGL	75.63	66.46	55.31	45.63	59.95
LRA*-GL	91.88	75.21	79.38	68.60	77.86
ESRC	89.79	75.42	68.75	57.35	71.88
SVDL	90.63	76.25	71.25	56.88	72.56
BlockFLD	71.25	67.92	68.13	50.32	62.92
PSRC	82.09	76.67	87.19	65.16	75.94
PCRC	88.75	82.50	91.88	73.13	82.50
LGR	89.79	87.82	95.31	84.53	88.48
LGR+PIE	81.25	73.96	89.69	76.41	79.22
PLR	85.63	80.00	88.13	70.00	79.43
PSLR	95.21	80.83	89.06	76.72	84.43
MPLR	96.46	84.58	90.63	82.19	87.76
MPSLR	97.08	86.04	92.50	83.13	88.91

4.3 LFW Database

The LFW database [23] includes images of 5,749 individuals under an unconstrained environment. LFW-a is a aligned version of LFW using commercial software tool [25]. We collected a subset of 158 individuals with 10 or more face images from LFW-a and gathered 10 samples for each person. **Fig. 6** shows several sample images from LFW database.



Fig. 6. Sample images from LFW database.

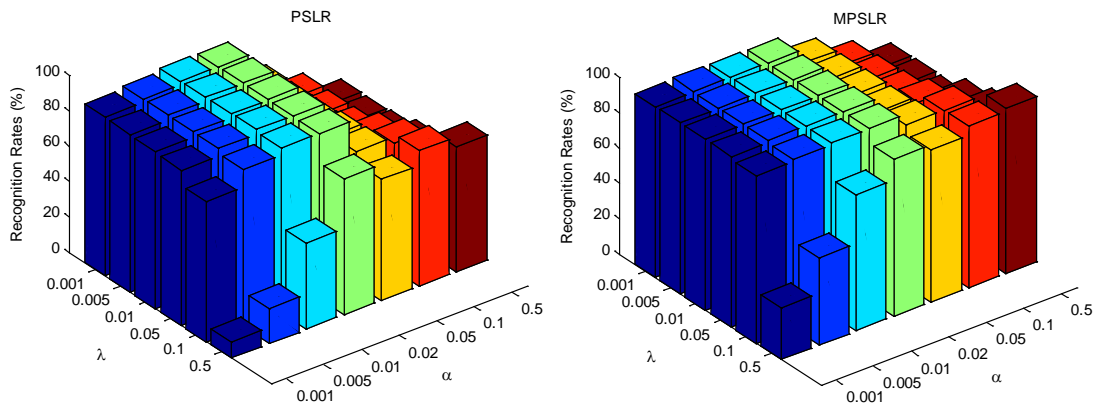
In the experiment of dealing with SSPP problem, we use the first 80 subjects for evaluation and the other subjects as the generic dataset. In our conference paper [9], we randomly select one image as the training sample for each subject. To reproduce the experiment conveniently and evaluate the recognition performance more effectively, we sequentially select one image as the gallery sample and use the other nine images as the probe samples. Thus 10 experiments are carried out. The average results are shown in Table 3 from which one can find that PSLR is still superior to PLR and that MPLSR is slightly inferior to MPLR. This is because there is drastic pose variation in this database, and there are too many types of variations. The number of unlabeled samples is insufficient, which cannot satisfy Assumption 1. In such case, semi-supervised technique does not help. And comparing with PLR and PSLR, MPLR and MPLSR achieve nearly 3% improvement. This suggests that multistage technique can help to improve the robustness for SSPP face recognition in uncontrolled environment.

Table 3. Recognition Rates (%) on LFW Database for SSPP Problem.

Method	AGL	LRA*-GL	ESRC	SVDL	BlockFLD	PSRC
Accuracy	32.25	26.90	32.96	33.18	18.01	15.79
Method	PCRC	LGR	PLR	PSLR	MPLR	MPLSR
Accuracy	32.29	29.36	37.00	37.53	40.69	40.56

4.4 Parameter Selection

In this section, we will discuss the impact of α and λ on the classification performances of our approaches by using Extended Yale B and AR databases. We tune α and λ within the range of $\{0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.5\}$ and $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$, respectively. The performances of PSLR and MPLSR under different parameter combinations are presented in Fig. 7. It can be observed that if α and λ are large, the performance tends to deteriorate. And the performance varies slightly with varying α and λ when α and λ are small. Generally speaking, MPLSR yields more satisfactory performances than PSLR.



(a)

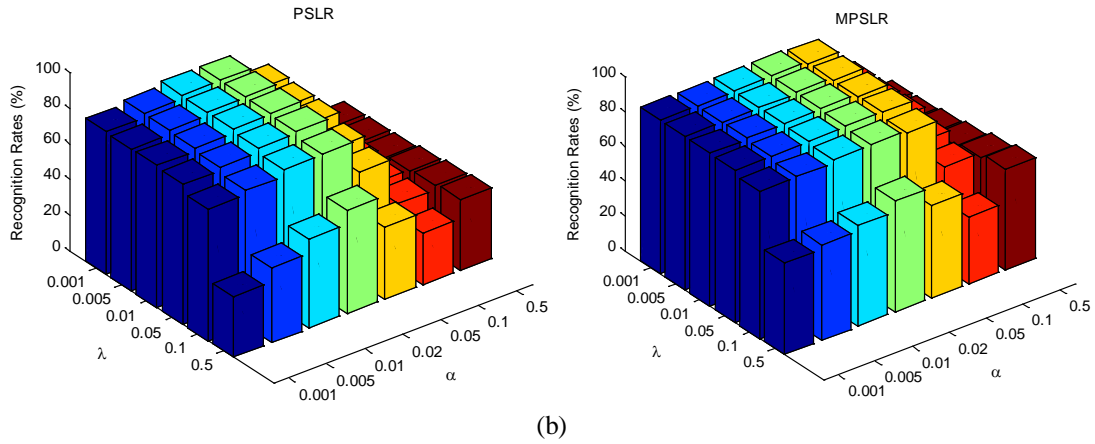
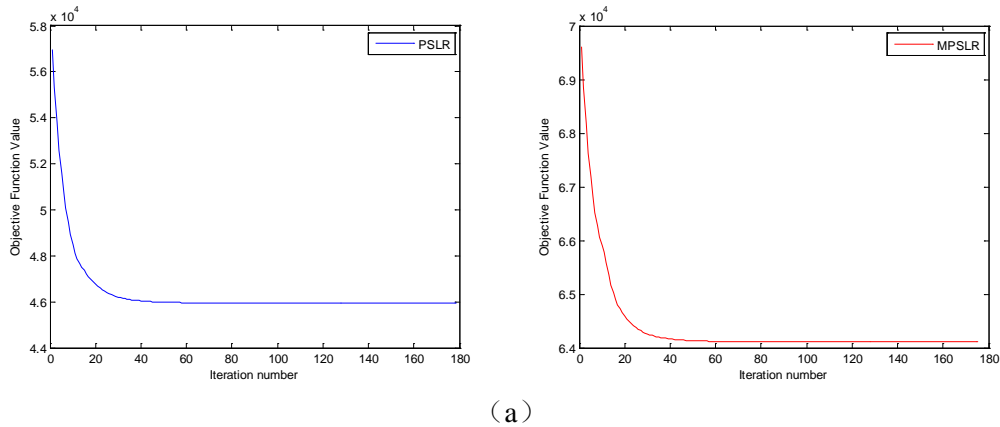


Fig. 7. The classification performances with the varied parameters α and λ on (a) Extended Yale B and (b) AR.

4.5 Convergence Study

Generally, it is hard to guarantee the convergence of inexact ALM when there are more than two blocks [26]. The convergence of **Algorithm 2** would not be proved in theory easily since the objective function of (14) is not smooth and there are three blocks (including W , J and E). Therefore, we verify the convergence of PSLR and MPSLR by experiments. The convergence curves of PSLR and MPSLR on Extended Yale B and AR databases are shown in **Fig. 8**. One can see from these figures that the objective value monotonically decreases and the two methods converge within 60 iterations.



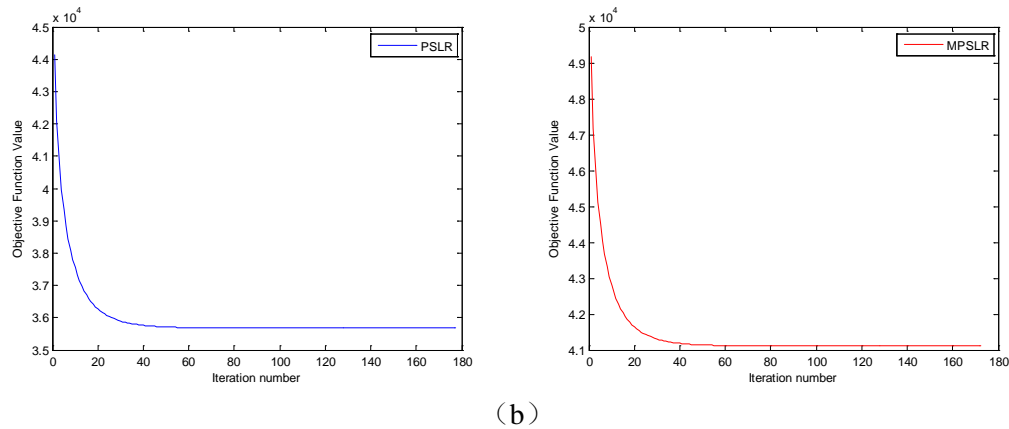


Fig. 8. Convergence curves of PSLR and MPLSR methods. (a) Extended Yale B and (b) AR.

5. Conclusion

In this paper, we propose a patch based semi-supervised linear regression (PSLR) algorithm for the single sample face recognition task, which makes use of unlabeled samples to describe facial variation information and adjust the mapping matrix. The local region partition strategy is adopted, which provides more identification information. We also formulate all regression models relative to each patch into an overall objective function, which harvests both global and local strengths. To further improve the performance of PSLR, we propose multistage PSLR (MPLSR), which adopt multistage strategy to select the reliable probe samples with effective identification information and use them to improve the discriminability of regression model. Experimental results prove that PSLR and MPLSR work well in dealing with SSPP problem.

References

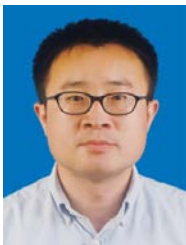
- [1] Y. Su, S. Shan, X. Chen, W. Gao, "Adaptive Generic Learning for Face Recognition from a Single Sample per Person," in *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 2699-2706, June 13-18, 2010. [Article \(CrossRef Link\)](#).
- [2] M. Turk, A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991. [Article \(CrossRef Link\)](#).
- [3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997. [Article \(CrossRef Link\)](#).
- [4] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, December 8-13, 2003. [Article \(CrossRef Link\)](#).
- [5] F. Liu, Z. Tang, and J. Tang, "WLBP: Weber Local Binary Pattern for Local Image Description," *Neurocomputing*, vol. 120, pp. 325-335, 2013. [Article \(CrossRef Link\)](#).
- [6] W. Deng, J. Hu, X. Zhou, and J. Guo, "Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning," *Pattern Recognition*, vol. 47, no. 12, pp. 3738-3749, 2014. [Article \(CrossRef Link\)](#).
- [7] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization," *Advances in Neural Information Processing Systems*, pp. 1813-1821, December 6-9, 2010. [Article \(CrossRef Link\)](#).

- [8] F. Liu, J. Tang, Y. Song, L. Zhang, and Z. Tang, "Local structure-based sparse representation for face recognition," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no. 1, pp. 2:1-2:20, 2015. [Article \(CrossRef Link\)](#).
- [9] F. Liu, Y. Ding, S. Yang, and F. Xu, "Patch based Semi-supervised Linear Regression for Single Sample Face Recognition," in *Proc. of IEEE International Conference on Multimedia Big Data*, April 19-21, 2017. [Article \(CrossRef Link\)](#).
- [10] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled Face Recognition via Intra-class Variant Dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864–1870, 2012. [Article \(CrossRef Link\)](#).
- [11] M. Yang, Luc Van, and L. Zhang, "Sparse variation dictionary learning for face recognition with a single training sample per person," in *Proc. of IEEE International Conference on Computer Vision*, pp. 689-696, December 3-6, 2013. [Article \(CrossRef Link\)](#).
- [12] D. Huang, and Y. Wang, "With one look: robust face recognition using single sample per person," in *Proc. of the 21st ACM international conference on Multimedia*, pp. 601-604, October 21-25, 2013. [Article \(CrossRef Link\)](#).
- [13] S. C. Chen, J. Liu, Z. H. Zhou, "Making FLDA Applicable to Face Recognition with One Sample per Person," *Pattern Recognition*, vol. 37, no. 7, pp. 1553-1555, 2004. [Article \(CrossRef Link\)](#).
- [14] J. Lu, Y.P. Tan, G. Wang, "Discriminative Multi-Manifold Analysis for Face Recognition from a Single Training Sample per Person," in *Proc. of IEEE International Conference on Computer Vision*, pp. 1943-1950, November 6-13, 2011. [Article \(CrossRef Link\)](#).
- [15] L. Zhang, M. Yang and X. Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?" in *Proc. of IEEE International Conference on Computer Vision*, pp. 471-478, November 6-13, 2011. [Article \(CrossRef Link\)](#).
- [16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210–227, 2009. [Article \(CrossRef Link\)](#).
- [17] P. F. Zhu, L. Zhang, Q.H. Hu, and S. C.K. Shiu, "Multi-scale Patch based Collaborative Representation for Face Recognition with Margin Distribution Optimization," in *Proc. of European Conference on Computer Vision*, pp. 822–835, October 7-13, 2012. [Article \(CrossRef Link\)](#).
- [18] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," *UIUC Technical Report UILU-ENG-09-2215, Tech. Rep.*, 2009. [Article \(CrossRef Link\)](#).
- [19] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569-592, 2009. [Article \(CrossRef Link\)](#).
- [20] Z. Li, J.Liu, J. Tang, and H. Lu, "Robust Structured Subspace Learning for Data Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2085-2098, 2015. [Article \(CrossRef Link\)](#).
- [21] A. Georghiades, P. Belhumeur, D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, 2001. [Article \(CrossRef Link\)](#).
- [22] A. M. Martinez, "The AR Face Database," *CVC Technical Report 24*, 1998. [Article \(CrossRef Link\)](#).
- [23] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *Technical Report 07-49, University of Massachusetts, Amherst*, vol. 1, no. 2, 2007. [Article \(CrossRef Link\)](#).
- [24] P. Zhu, et al, "Local Generic Representation for Face Recognition with Single Sample per Person," in *Proc. of Asian Conference on Computer Vision*, pp. 34-50, November 1-5, 2014. [Article \(CrossRef Link\)](#).

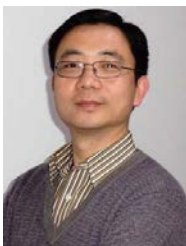
- [25] L. Wolf, T. Hassner, and Y. Taigman, "Similarity scores based on background samples," in *Proc. of Asian Conference on Computer Vision*, pp. 88-97, September 23-27, 2009. [Article \(CrossRef Link\)](#).
- [26] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust Recovery of Subspace Structures by Low-Rank Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171-184, 2013. [Article \(CrossRef Link\)](#).



Yuhua Ding received his bachelor's degree in software engineering from Nanjing University of Science and Technology, Nanjing, China, in 2011, where he is currently working toward his PhD. His current research interests include pattern recognition, image processing, and subspace learning.



Fan Liu is currently a lecturer of HoHai University. He received his B.S. degree in network engineering from Nanjing University of Science and Technology (NUST) in June 2009. From September 2008 to December 2008, he studied in Ajou University of South Korea as an exchange student. He received his Ph.D. degree from Nanjing University of Science and Technology in January 2015. His research interests include computer vision, image processing, pattern recognition and deep learning. Dr. Liu also serves as reviewer of Information Sciences, Neurocomputing, KSII Transaction on Internet and Information Systems and Pattern Analysis and Application.



Ting Rui received the M.S. degree and Ph.D. from PLA University of Science and Technology, Nanjing, China, in 1998 and 2001 respectively. Ting RUI is Professor of Information Technology Department in the PLA University of Science and Technology. He mainly applies computer vision, machine learning, multimedia, and video surveillance. He has authored and co-authored more than 80 scientific articles.



Zhenmin Tang is currently a professor of School of Computer Engineering, Nanjing University of Science and Technology. He received his B.S. degree from Harbin Engineering University (HEU) in January 1982. He received his M.S. and Ph.D. degrees in January 1988 and March 2003 respectively, both from Nanjing University of Science and Technology (NUST). His current research interests are in the area of image processing, pattern recognition, artificial intelligence and expert system. He has over 60 journal and conference papers in these areas.