

RESEARCH ARTICLE

Robust Face Recognition via Multi-Scale Patch-Based Matrix Regression

Guangwei Gao^{1*}, Jian Yang², Xiaoyuan Jing³, Pu Huang⁴, Juliang Hua², Dong Yue¹

1 Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China, **2** School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China, **3** School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China, **4** School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China

* csggao@gmail.com



OPEN ACCESS

Citation: Gao G, Yang J, Jing X, Huang P, Hua J, Yue D (2016) Robust Face Recognition via Multi-Scale Patch-Based Matrix Regression. PLoS ONE 11 (8): e0159945. doi:10.1371/journal.pone.0159945

Editor: Zhaohong Deng, Jiangnan University, CHINA

Received: April 6, 2016

Accepted: July 11, 2016

Published: August 15, 2016

Copyright: © 2016 Gao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data have been deposited to Figshare (<http://dx.doi.org/10.6084/m9.figshare.3486989>).

Funding: This work was partially supported by the National Science Fund for Distinguished Young Scholars under Grant Nos. 61125305, 91420201, 61472187, 61233011 and 61373063, the National Natural Science Foundation of China under Grant Nos. 61502245, 61503188 and 61503195, the NUPTSF under Grant Nos. NY214204 and NY214165, the China Postdoctoral Science Foundation No. 2015M571786, the Natural Science Fund for Colleges and Universities in Jiangsu Province No. 15KJB520026, the Natural Science Foundation of Jiangsu Province under Grant Nos.

Abstract

In many real-world applications such as smart card solutions, law enforcement, surveillance and access control, the limited training sample size is the most fundamental problem. By making use of the low-rank structural information of the reconstructed error image, the so-called nuclear norm-based matrix regression has been demonstrated to be effective for robust face recognition with continuous occlusions. However, the recognition performance of nuclear norm-based matrix regression degrades greatly in the face of the small sample size problem. An alternative solution to tackle this problem is performing matrix regression on each patch and then integrating the outputs from all patches. However, it is difficult to set an optimal patch size across different databases. To fully utilize the complementary information from different patch scales for the final decision, we propose a multi-scale patch-based matrix regression scheme based on which the ensemble of multi-scale outputs can be achieved optimally. Extensive experiments on benchmark face databases validate the effectiveness and robustness of our method, which outperforms several state-of-the-art patch-based face recognition algorithms.

Introduction

Object classification is an active topic in the area of pattern recognition [1–8]. Due to the advantages of non-intrusive natural and pronounced uniqueness, face recognition has been an active research topic and has been incorporated into many multimedia applications [9–14], such as surveillance, human machine interaction, access control and photo album management in social networks. Recently, linear regression based face recognition approaches have led to state-of-the-art performance [15–18], with representative examples being sparse representation-based classification (SRC) [15] and linear regression-based classification (LRC) [16]. In SRC, the query sample image is coded as a sparse linear combination of all the training images, and then the classification is made by checking which class yields the least reconstruction error. Many works of SRC have been developed for vision applications, e.g., super-resolution

BK20150849 and BK20150982, the Key Project of Chinese Ministry of Education under Grant No. 313030, the 973 Program No. 2014CB349303, Fundamental Research Funds for the Central Universities No. 30920140121005, and Program for Changjiang Scholars and Innovative Research Team in University No. IRT13072.

Competing Interests: The authors have declared that no competing interests exist.

[19, 20], facial expression recognition [21] and human gait recognition [22, 23]. Alternatively, Naseem et al. [16] proposed LRC for face recognition. Based on the assumption that samples from a specific object class lie on a linear subspace, LRC represents a query image as a linear combination of training images of each class. Yang et al. [24] provided an insight into SRC and sought reasonable supports for its effectiveness. They viewed the L_1 -regularizer as having two properties, sparseness and closeness. Sparseness determines a small number of nonzero representation coefficients, and closeness makes the nonzero representation coefficients concentrating on the training samples having the same class label as the test sample. Zhang et al. [18] discussed the working mechanism of SRC and demonstrated that it is collaborative representation rather than L_1 -norm sparseness that improves the classification performance. In their work, a collaborative representation-based classification (CRC) model was presented with a squared L_2 -regularization, which achieves competitive classification performance but with significantly lower complexity than the sparse representation method.

It is worth noting that the majority of studies assume that the testing images are taken under well-controlled settings (e.g., reasonable illumination, poses and variations, without occlusion or disguise). Their performance is degraded when the testing images are contaminated. By introducing an identity matrix I as a dictionary to code the outliers (e.g., corrupted or occluded pixels), SRC [15] exhibits excellent robustness and promising performance. However, SRC is not robust to contiguous occlusion such as sunglasses or scarves, as the occlusion level exceeds the breakdown point determined by this algorithm. Yang et al. [25] modified the SRC framework for handling outliers such as occlusions in face recognition by modeling the sparse coding as a sparsity-constrained robust regression problem. He et al. [26] unified the algorithms for error correction and detection by using the additive and multiplicative forms, respectively, and established a half-quadratic framework to solve the robust sparse representation problem. From the viewpoint of dictionary learning, Yang et al. [27] constructed a feature pattern dictionary that captures structured information and prior knowledge of image features to represent the unknown feature pattern weight of a query image. Similarly, Ou et al. [28] developed a clear and noise dictionary simultaneously and applied the learned clear dictionary for classification. Observing the distribution of the reconstruction error image, Yang et al. [29–32] used the nuclear norm to characterize the structural information of an error image and proposed a nuclear norm-based matrix regression model that has achieved state-of-the-art performance for face recognition with occlusion and illumination changes.

In spite of aforementioned tremendous achievements, the small sample size (SSS) problem still remains one of the most fundamental and challenging issues in face recognition community. In many real-world applications such as smart card solutions, law enforcement, surveillance and access control, the available training samples per subject may be very limited [33]. Thus, the performance of these regression-based methods is greatly degraded because the query sample cannot be well represented by the few training samples. To tackle the SSS problem, many efforts have been made in the past few decades. Existing methods mainly fall into three categories. The first are patch-based methods, which generally contain steps of local patch representation, local feature extraction and the combination of classification results [34–36]. However, the patch size has a great impact on the output performance in patch-based methods [37, 38]. The second integrate the local and global features for classification [39, 40] because they can provide complementary information for final results. The third employ different feature extractors to extract multiple types of features, and then utilizes decision level fusion scheme for final classification [41, 42]. We mainly focus on patch-based method in the sequel.

To improve the recognition performance of matrix regression in SSS problem and preserve its outstanding ability dealing with occlusion and illumination changes, in this paper, we propose performing matrix regression on patches. The so-called patch-based matrix regression

(PMR) classifies each query matrix patch, and then integrates the recognition outputs of all patches for final decision. Nevertheless, the patch size plays an important role on the final performance in PMR, and the optimal patch size varies greatly across different databases. If the patch size is too small, little information is given, and the method cannot capture the geometric structure of the image; if it is large, the information that can be used is limited. To fully exploit the classification ability and appearance information of different patch sizes, we then devise a multi-scale PMR (MSPMR) scheme by integrating the complementary information from different scales. MSPMR first performs PMR on each scale and then learns optimal scale weights to adaptively fuse multi-scale outputs. To evaluate the performance of the proposed method, we use four databases that involve different recognition tasks: the Extended Yale B, AR and LFW dataset for face recognition without occlusion, the AR database for face recognition with real disguise, and the Extended Yale B dataset for face recognition with block occlusion. The experimental results demonstrate the effectiveness and robustness of the proposed method.

The remainder of the paper is organized as follows. Section 2 briefly reviews two related works. The proposed multi-scale PMR via margin distribution optimization is presented in Section 3. Section 4 conducts extensive experiments, and Section 5 concludes this paper.

Related Works

1. Nuclear norm based matrix regression

By observing the distribution of the reconstruction error image, a nuclear norm-based matrix regression (NMR) [29] model was proposed that uses the nuclear norm to characterize the whole structure of the error image. Here, we define N_i as the number of images from the i -th class and $N = \sum_{i=1}^c N_i$ as the total number of training samples from c classes. Given a set of N training image matrices $A_1, A_2, \dots, A_N \in \mathbb{R}^{row \times col}$ and a query image matrix $B \in \mathbb{R}^{row \times col}$, the NMR model can be represented as

$$\min_x \|B - A(x)\|_* + \frac{\lambda}{2} \|x\|_2^2, \quad (1)$$

where λ is the regularization parameter, x and $A(x) = x_1 A_1 + x_2 A_2 + \dots + x_N A_N$ are the representation coefficient vector and the reconstructed image, respectively. Then, the query image can be classified into the class that yields the minimal reconstruction error, i.e.,

$$\text{Identity}(B) = \arg \min_i \{r_i(B)\}, \quad (2)$$

where $r_i(B) = \|\hat{B} - \hat{B}_i\|_* = \|A(x^*) - A(\delta_i(x^*))\|_*$, where x^* is the optimal solution of Eq (1) and $\delta_i(x)$ is a vector whose only nonzero entries are the entries in x that correspond to Class i . We know that NMR is much more robust and effective for face recognition, particularly with respect to occlusion and illumination changes.

2. Patch-based CRC

Suppose that we have c known pattern classes. Let $X_i = [x_{i1}, x_{i2}, \dots, x_{iN}] \in \mathbb{R}^{M \times N_i}$ be the matrix formed by the training samples of class i , where N_i is the number of training samples of class i . Let $X = [X_1, X_2, \dots, X_c] \in \mathbb{R}^{M \times N}$ be the dataset of all training samples, where $N = \sum_{i=1}^c N_i$. To alleviate the performance degradation of CRC in the small sample size problem, the patch-based CRC (PCRC) [36] model was proposed. For a given query image y , it is first divided into multiple overlapped patches $\{y_1, y_2, \dots, y_p\}$. Then, each patch y_i is tackled by representing it as a

linear combination over a local dictionary D_i . Finally, one can employ the plurality or linear-weighted combination scheme to the recognition outputs for a final decision.

For each patch y_i , its representation weights can be obtained by minimizing the following error:

$$\hat{\beta}_i = \arg \min_{\beta_i} \|y_i - D_i \beta_i\|_2^2 + \eta \|\beta_i\|_2^2. \quad (3)$$

Where $D_i = [D_{i1}, D_{i2}, \dots, D_{ic}]$ denotes the local dictionary located with the same position as y_i , D_{ik} is the sub-dictionary of the k -th class. The recognition result of patch y_i is $\text{Identity}(y_i) = \arg\min_k [43]$, where $r_{ik} = \|y_i - D_{ik} \hat{\beta}_{ik}\|_2 / \|\hat{\beta}_{ik}\|_2$, $\hat{\beta}_{ik}$ is the coefficients associated with the k -th class.

For clarity, four key components (i.e., multi-scale trick, local patch strategy, structure error and pixel error characterization) of several related methods are compared in Table 1.

Multi-Scale Patch-Based Matrix Regression (MSPMR)

1. Motivation

In PCRC, each local patch matrix is first converted to a vector, and then the L_2 -norm is used to characterize the reconstruction error. However, the L_2 -norm (or L_1 -norm) is based on pixel values and thus ignores the structural information of the error image. Fig 1 shows an example where the error between (b) and (a) is shown in (c). By re-arranging the pixels of image (c), we can obtain image (d). The following observations can be made:

1. The nuclear norm can better characterize the structure error than the L_1 or L_2 norm. For example, the L_2 -norm (or L_1 -norm) value of image (c) is the same as that of image (d), so it is difficult to distinguish between them. Fortunately, the nuclear norm values of images (c) and (d) are 47.75 and 58.14, respectively.
2. From the distribution perspective, we can observe that the distribution of the error image does not follow a Laplacian or Gaussian distribution in Fig 1e. Fortunately, it can be seen from Fig 1f that the singular values of the error image (c) fit the Laplacian distribution well. We know that the nuclear norm is the sum of all singular values of a matrix, which can also be considered as the l_1 -norm of the singular value vector. Based on the above example, we believe that the nuclear norm is more suitable to describe the structural error.

2. Patch-based matrix regression (PMR)

To make the model robust and efficient for face recognition with occlusion and illumination changes, matrix regression [29, 30, 32] was proposed using the nuclear norm to characterize

Table 1. Comparisons of related works.

Method	multi-scale	local patch	structure error	pixel error
CRC[18]				✓
SRC[15]				✓
NSC[30]			✓	
PNN[34]		✓		
HQ_A[26]				✓
BlockFLD[38]		✓		
Volterra[35]		✓		
PSRC[15]		✓		✓
PCRC[36]		✓		✓
MSPCRC[36]	✓	✓		✓

doi:10.1371/journal.pone.0159945.t001

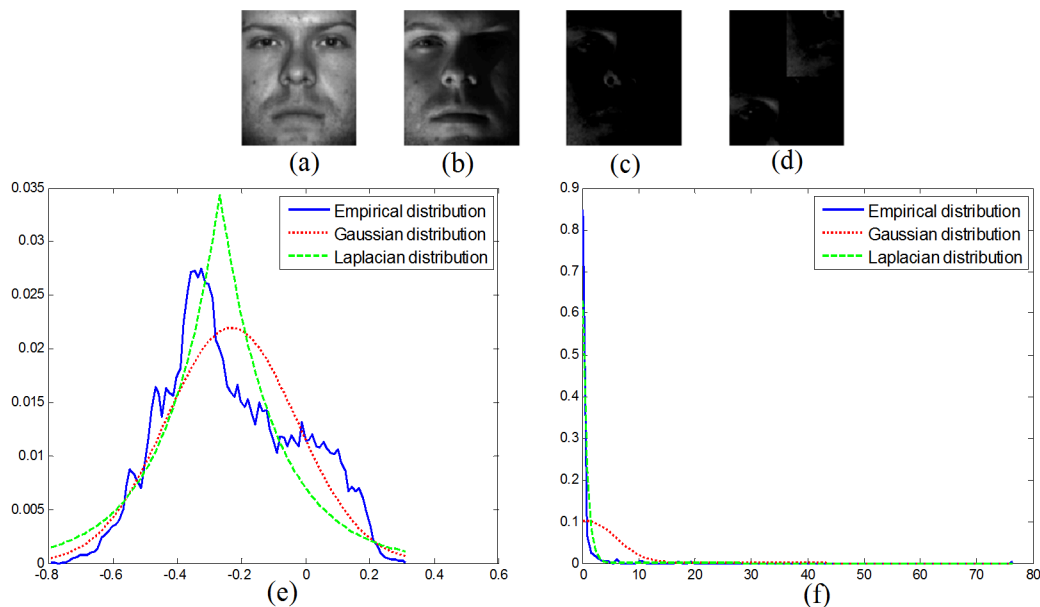


Fig 1. (a) Original image; (b) observed image; (c) error image; (d) rearranged error image; (e) distributions of error image; and (f) distributions of singular values of error image.

doi:10.1371/journal.pone.0159945.g001

the structure of the error image. In our patch-based matrix regression, all local patches are denoted in matrix form. Given a set of N local patches $X_{i1}, X_{i2}, \dots, X_{iN} \in R^{p \times q}$ and a query patch $Y_i \in R^{p \times q}$ located at position i , Y_i can be represented linearly using $X_{i1}, X_{i2}, \dots, X_{iN}$, i.e.,

$$Y_i = F(\alpha_i) + E_i, \quad (4)$$

where $F(\alpha_i) = \alpha_{i1}X_{i1} + \alpha_{i2}X_{i2} + \dots + \alpha_{iN}X_{iN}$, $\alpha_i = (\alpha_{i1}, \dots, \alpha_{iN})^T$ is the representation coefficient vector and E_i is the representation error. Generally, α_i can be determined by the following regularized model

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \|F(\alpha_i) - Y_i\|_* + \lambda \|\alpha_i\|_1, \quad (5)$$

where $\|\cdot\|_*$ denotes the nuclear norm (the sum of the singular values) on $R^{p \times q}$.

The problem is equivalent to

$$\min_{\alpha_i, u_i, E_i} \|E_i\|_* + \lambda \|u_i\|_1, \quad s.t. \quad F(\alpha_i) - Y_i = E_i, \quad \alpha_i = u_i. \quad (6)$$

Problem (6) can be solved by the alternating direction method of multipliers (ADMM), which minimizes the following augmented Lagrangian function:

$$L_\mu = \|E_i\|_* + \lambda \|u_i\|_1 + \text{tr}(y^T(\alpha_i - u_i)) + \text{tr}(\Lambda^T(F(\alpha_i) - Y_i - E_i)) + \frac{\mu}{2} (\|F(\alpha_i) - Y_i - E_i\|_F^2 + \|\alpha_i - u_i\|_F^2). \quad (7)$$

That is,

$$L_{\mu} = \|E_i\|_* + \lambda \|u_i\|_1 + \frac{\mu}{2} \left(\left\| F(\alpha_i) - Y_i - E_i + \frac{1}{\mu} \Lambda \right\|_F^2 + \left\| \alpha_i - u_i + \frac{1}{\mu} y \right\|_F^2 \right) - \frac{1}{2\mu} \|\Lambda\|_F^2 - \frac{1}{2\mu} \|y\|_2^2. \quad (8)$$

The entire algorithm is briefly summarized in **Algorithm 1**, which mainly consists of two steps: a soft-thresholding operator [44] and a singular value thresholding operator [45].

Based on the optimal solution α_i^* , we can obtain the reconstructed image of Y as $\hat{Y}_i = F(\alpha_i^*)$. Let $\delta_k: R^n \rightarrow R^n$ be the characteristic function that selects the coefficients associated with the k -th class. For $\alpha \in R^n$, $\delta_k(\alpha)$ is a vector whose nonzero entries are the entries in α that are associated with Class k . Using the coefficients associated with the k -th class, one can obtain the reconstruction of Y_i in Class k as $\hat{Y}_{ik} = F(\delta_k(\alpha_i^*))$.

Algorithm 1. Solving problem (6) via ADMM

Input: A set of N patches $X_{i1}, X_{i2}, \dots, X_{iN} \in R^{P \times Q}$ and a query patch $Y_i \in R^{P \times Q}$, parameters λ and μ , the termination condition parameter ε .

1: Fix the others and update u_i^{k+1} by

$$u_i^{k+1} = \arg \min_{u_i} \frac{\lambda}{\mu} \|u_i\|_1 + \frac{1}{2} \left\| u_i - \left(\alpha_i^k + \frac{1}{\mu} y^k \right) \right\|_F^2;$$

2: Fix the others and update α_i^{k+1} by

$$\alpha_i^{k+1} = \arg \min_{\alpha_i} \frac{\mu}{2} \left(\left\| \alpha_i - u_i^{k+1} + \frac{1}{\mu} y^k \right\|_F^2 + \left\| F(\alpha_i) - Y_i - E_i^k + \frac{1}{\mu} \Lambda^k \right\|_F^2 \right);$$

3: Fix the others and update E_i^{k+1} by

$$E_i^{k+1} = \arg \min_{E_i} \frac{1}{\mu} \|E_i\|_* + \frac{1}{2} \left\| E_i - \left(F(\alpha_i^{k+1}) - Y_i + \frac{1}{\mu} \Lambda^k \right) \right\|_F^2;$$

4: Update the multipliers

$$y^{k+1} = y^k + \mu(\alpha_i^{k+1} - u_i^{k+1}), \Lambda^{k+1} = \Lambda^k + \mu(F(\alpha_i^{k+1}) - Y_i - E_i^{k+1});$$

5. If the termination condition is satisfied, go to 6; otherwise go to 1.

6. **Output:** Optimal coding vector x^{k+1} .

The corresponding class reconstruction error is defined as

$$e_{ik}(Y_i) = \|\hat{Y}_i - \hat{Y}_{ik}\|_* = \|F(\alpha_i^*) - F(\delta_k(\alpha_i^*))\|_*. \quad (9)$$

The recognition output z_i of query patch Y_i is denoted as $\text{Identity}(Y_i) = \arg \min_k \{e_{ik}(Y_i)\}$.

Then we can combine the classification outputs of all patches by linear weighted combination [37], probabilistic model [40], kernel plurality [34] or majority voting [35]. In this paper, we use majority voting in the final decision making. Fig 2 shows the main diagram of the patch-based matrix regression for face recognition.

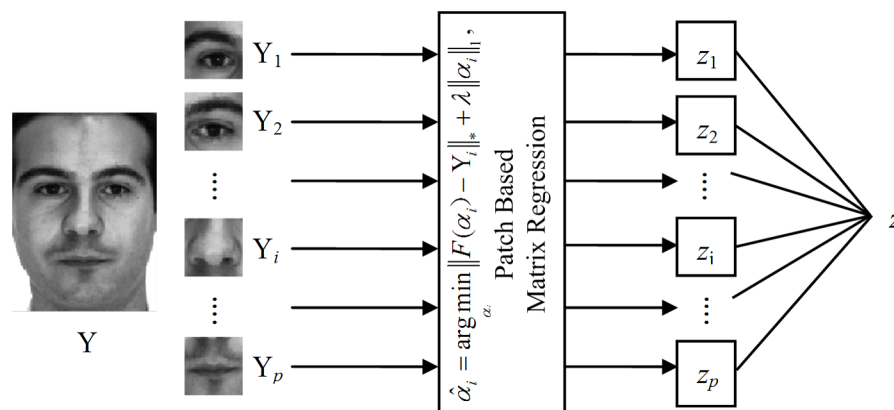


Fig 2. Diagram of patch-based matrix regression for face recognition.

doi:10.1371/journal.pone.0159945.g002

3. Multi-scale ensemble

From the previous introduction of PMR, we can find that the patch size plays an important role on the final output performance. In addition, how to set an optimal scale in advance for various databases remains unclear. Fig 3 exhibits the recognition rates curves versus different training sample sizes and patch sizes on the LFW and Extended Yale B databases, respectively. From Fig 3, the following observations can be made. First, the optimal patch size varies greatly between different databases. Second, the optimal patch size also varies a lot under the variation of the training sample size per person. To tackle the aforementioned difficulties, the recognition outputs of multi-scale PMR can be fused optimally; thus, the complementary information from different scales can be fully applied to further enhance the recognition performance. Motivated by [36], we also incorporate the ensemble learning scheme into our method to integrate multi-scale outputs.

The diagram of the proposed multi-scale PMR is shown in Fig 4. In the following text, we first formulate the multi-scale ensemble problem, and then introduce a margin distribution optimization to obtain the optimal solution.

Problem formulation. Suppose that we have two scales and two sample classes labeled +1 and -1. For any query sample, we can obtain its classification result +1 or -1 on each scale.

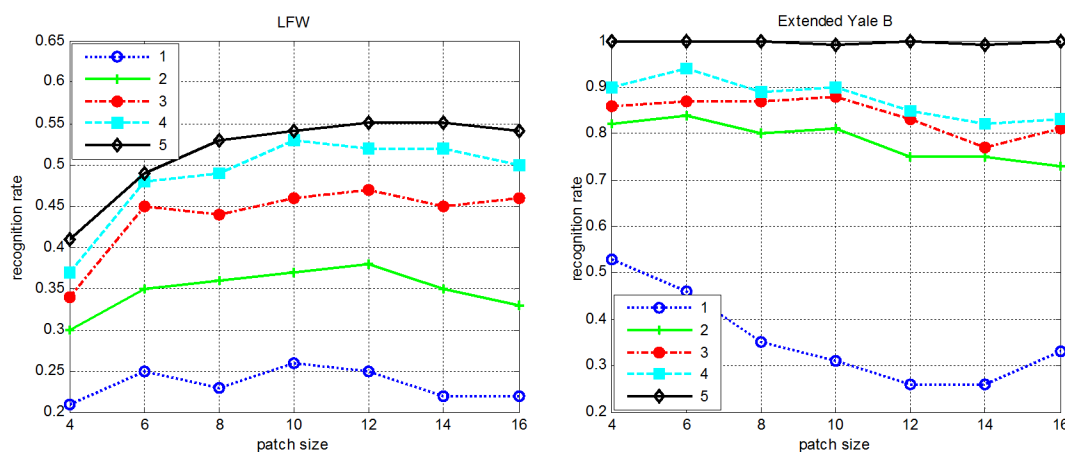


Fig 3. Impact of patch size on PMR (1–5 denote the training sample size per subject).

doi:10.1371/journal.pone.0159945.g003

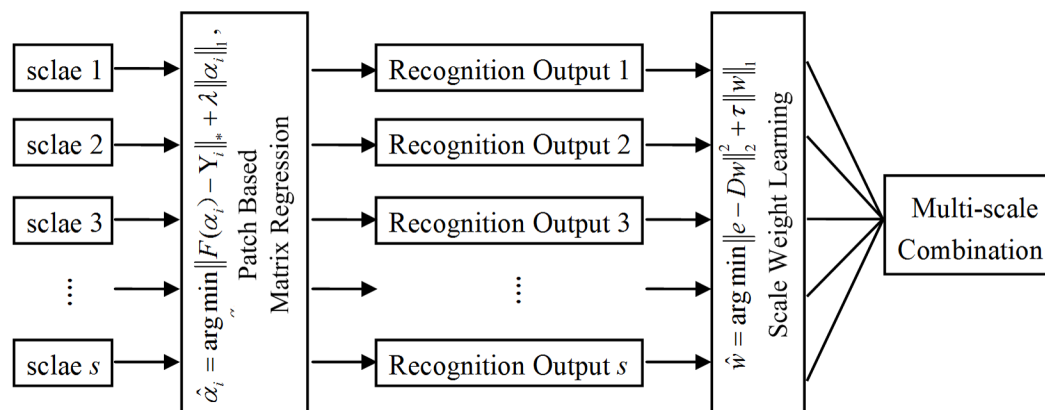


Fig 4. Flow chart of multi-scale learning for PMR.

doi:10.1371/journal.pone.0159945.g004

Thus, each sample will have four possible classification results on these two scales, such as $\{-1, +1\}$, $\{+1, +1\}$, $\{-1, -1\}$ and $\{+1, -1\}$. Given an available training data set, our goal is to learn a classification function f , which can make it possible to classify all the given samples exactly.

Given a sample set $S = \{(x_i, z_i)\}$ ($i = 1, 2, \dots, n$, z_i is the label of x_i) and s scales, the classification results of samples x_i on these s scales can form a space $H \in R^{n \times s}$. Denote by $w = [w_1, w_2, \dots, w_s]$ the scale fusion vector, and $\sum_{j=1}^s w_j = 1$.

Definition 1 [36]: For multi-class classification tasks, the classification results of a given query sample $x_i \in S$ on these s different scales are denoted as [4], $j = 1, 2, \dots, s$. Then the decision matrix $D = \{d_{ij}\}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, s$ can be defined as

$$d_{ij} = f(z_i, h_{ij}) = \begin{cases} +1, & \text{if } z_i = h_{ij} \\ -1, & \text{if } z_i \neq h_{ij} \end{cases}, \quad (10)$$

where z_i is the label of the sample x_i .

Definition 2 [36]: For a query sample $x_i \in S$, [4] ($j = 1, 2, \dots, s$) are the classification results on these s different scales. Then the ensemble margin of x_i is denoted as

$$\mathcal{E}(x_i) = \sum_{j=1}^s w_j d_{ij}. \quad (11)$$

The ensemble loss of x_i can be denoted as [36]

$$l_{x_i} = l(\mathcal{E}(x_i)) = l\left(\sum_{j=1}^s w_j d_{ij}\right), \quad (12)$$

Where $\mathcal{E}(x_i)$ is the ensemble margin of sample x_i . The square loss applied in CRC [18], SRC [15] and least square regression [16] can be used here to evaluate the ensemble loss. For a sample set S , its ensemble square loss can be formulated as

$$\begin{aligned} l(S) &= \sum_{i=1}^n l_{x_i} = \sum_{i=1}^n [1 - \mathcal{E}(x_i)]^2 \\ &= \sum_{i=1}^n \left[1 - \sum_{j=1}^s w_j d_{ij}\right]^2 = \|e_1 - Dw\|_2^2, \end{aligned} \quad (13)$$

where e_1 is a column vector whose entries are 1.

Algorithm of MSPMR. In order to obtain the optimal scale fusion weights, the ensemble square loss in Eq (13) should be minimized. Nevertheless, the solutions may be non-unique for this linear system. Intuitively, we should impose constraint on the objective function in Eq (13)

to make the solution unique and stable. Also, Shawe-Taylor [46] has provided the bound on the generalization error and pointed out that both the norm of w and the ensemble square loss should be optimized simultaneously to enhance the generalization ability.

As in [36], the following constrained l_1 -regularized least square optimization can be used to obtain the optimal scale weights [47]:

$$\begin{aligned} \hat{w} = \arg \min_w & \|e - Dw\|_2^2 + \tau \|w\|_1 \\ \text{s.t.} & \sum_{j=1}^s w_j = 1, w_j > 0 \end{aligned} \quad (14)$$

where τ is a regularization parameter and the regularization term can help to achieve a stable solution. Constraint $\sum_{j=1}^s w_j = 1$ can be converted to $e_2 w = 1$, where $e_2 = [1, 1, \dots, 1]$ is a row vector whose length is s . Then, we have

$$\|e - Dw\|_2^2 = \|e - Dw + 1 - e_2 w\|_2^2 = \|[e; 1] - [D; e_2]w\|_2^2 \quad (15)$$

Denoted by $\hat{e} = [e; 1]$ and $\hat{D} = [D; e_2]$, then we have [36]

$$\hat{w} = \arg \min_w \|\hat{e} - \hat{D}w\|_2^2 + \tau \|w\|_1 \quad \text{s.t.} \quad w_j > 0, j = 1, 2, \dots, s \quad (16)$$

Algorithm 2. Algorithm of multi-scale ensemble learning for PMR

- 1: Choose spatch scale $\delta = [\delta_1, \delta_2, \dots, \delta_s]$;
- 2: Obtain the recognition outputs [4] by PMR;
- 3: Obtain the decision matrix

$$d_{ij} = f(z_i, h_{ij}) = \begin{cases} +1, & \text{if } z_i = h_{ij} \\ -1, & \text{if } z_i \neq h_{ij} \end{cases}$$

- 4: Learn the fusion weights

$$\hat{w} = \arg \min_w \|\hat{e} - \hat{D}w\|_2^2 + \tau \|w\|_1 \quad \text{s.t.} \quad w_j > 0, j = 1, 2, \dots, s.$$

Duo to the fact the decision matrix is usually very small, the scale fusion weights w can be simply obtained by some commonly used l_1 -minimization solvers. In our method, l_1 -ls [48] is employed. Based on the above description, the proposed multi-scale PMR (MSPMR) scheme is summarized in **Algorithm 2**. Once the optimal scale fusion weights are achieved, the recognition output for an arbitrary sample x_i can be represented as $z_i = \arg \max_k \left\{ \sum_j w_j |h_{ij} = k| \right\}$.

4. Computational complexity

In this subsection, we will evaluate the computational complexity of the proposed method. Since the multi-scale fusion weights can be learned off-line, we only discuss the computational complexity of the on-line recognition process involved in the proposed method. As illustrated in Algorithm 2, the proposed face recognition method takes major cost on patch-based matrix regression process. We observe that there are four factors affecting the cost in our method: the training sample size N , the dimension of one patch $m = p \times q$, the number of iterations k in Algorithm 1, the patch number in one image M .

As described in [30], the matrix regression of each patch takes $O(k(m^{1.5} + mN + N^2))$ (in the case that $p = q$) cost. For M image patch, the computational cost is $O(k(m^{1.5} + mN + N^2)M)$. In addition, the scale number s also affect the final running time. Therefore, the computational

cost of the proposed method is about $O(sk(m^{1.5} + mN + N^2)M)$. In Section 4, we will further compare the proposed method with the state-of-the-art approaches in terms of CPU runtime.

Experimental Results and Discussion

In this section, we conduct experiments on the benchmark face databases and the proposed method is compared with state-of-the-art models. For each method, we perform 20 runs of test on each database, and the average recognition rates and the corresponding standard deviations are reported. As in [36], seven scales are adopted in our MSPMR, and the patch sizes are 4×4 , 6×6 , 8×8 , 10×10 , 12×12 , 14×14 and 16×16 . In single scale-based PNN, PSRC, PCRC and PMR, the patch size is 10×10 , and the patches overlap with their neighbors by 5 pixels. For PMR and MSPMR, we choose the optimal $\lambda \in [0.01, 0.1]$. Parameter τ is set as 0.1 for MSPMR. It should be mentioned here that all experiments are done on the original face images, without any feature extraction or image preprocessing step. Some face image datasets were used in this paper to verify the performance of our methods. These face image datasets are publicly available for face recognition research, and the consent was not needed. The face images and the experimental results are reported in this paper without any commercial purpose.

As in [36], to learn the optimal scale weights, the training set is divided into subset1 (one image per person) and subset2 (the remainder of the training set). Then, PMR is used to classify samples from subset1 utilizing subset2 as the gallery set and the optimal weights on the seven scales can be learned. It should be noted that at least two samples per person are required to find the optimal scale fusion weights.

1. Face recognition without occlusion

In this subsection, we test the MSPMR for face recognition without occlusion on four face databases (Extended Yale B [49] and AR [50] in controlled environments together with the LFW [36] in uncontrolled environments). The baseline CRC [18], SRC [15], NSC [30], and patch-based methods including PNN [34], BlockFLD [38], Volterra [35], PCRC and MSPCRC [36] are used for comparison.

Extended Yale B database. The first experiment was conducted on the Extended Yale B database; it includes 38 human objects in 9 poses under 64 illumination changes [49]. 64 images of a person with a particular pose are acquired at a camera frame rate of 30 frames per second, so the variations in the head pose and facial expression are small. All the frontal images marked with P00 are utilized in this experiment, and each is reshaped to 32×32 . Some examples are shown in Fig 5. For each subject, 2~5 samples are randomly selected from the first 32 images for training, and another 5 samples are randomly chosen from the rest 32 images for testing. Table 2 tabulates the experimental results.

It can easily be seen that MSPMR obtains the best recognition performance for all tests. Compared with PCRC and MSPCRC, PMR and MSPMR lead to much better results, thus verifying the effectiveness of characterizing the reconstruction error by the nuclear norm.

AR database. The AR database [50] gathers over 4,000 color face images from 126 subjects, containing frontal facial images with different lighting conditions, facial expressions and occlusions. Pictures of 120 subjects were taken in two sessions (separated by two weeks), and each has 13 color images. As in [18], in this experiments, we choose a subset with only illumination and expression changes, which includes 50 male objects and 50 female objects. Fourteen face images (seven from each session) of these 100 individuals are selected and used. For each object, 2~5 samples from session 1 are randomly chosen for training, and another 3 samples from session 2 are randomly chosen for testing. All the images are manually cropped and then resized to 32×32 pixels. Some sample images of one person are presented in Fig 6.

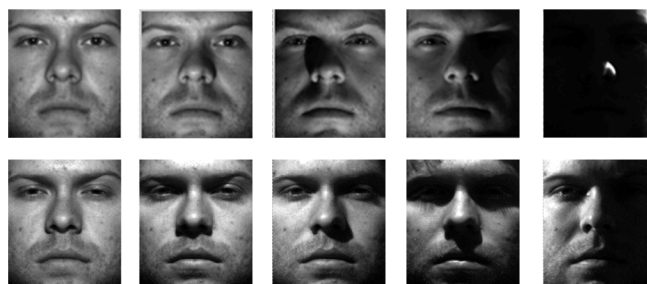


Fig 5. Sample images of a person under various illumination conditions in the Extended Yale B database from different sessions.

doi:10.1371/journal.pone.0159945.g005

The recognition results of different methods are listed in Table 3. The proposed methods always achieve better performance than the other methods. We can observe that in AR database, multi-scale ensemble learning in MSPMR leads to limited improvement over PMR. As described in [36], the reason may be that in this database, the average weight value for the scale 10×10 is approximately 0.9, indicating that patch size 10×10 is a proper choice for PMR in the AR database.

LFW database. Labeled Faces in the Wild (LFW) [43] is a large-scale database of face photographs designed for unconstrained face recognition with variations in pose, illumination, expression, misalignment and occlusion; it contains images of 5,749 subjects. LFW-a is an

Table 2. Recognition rates (%) on the Extended Yale B database.

Method	2	3	4	5
CRC[18]	61.3±16.6	74.0±15.5	81.4±17.6	87.8±13.7
SRC[15]	64.2±17.2	74.2±15.2	82.6±16.8	89.0±12.5
NSC[30]	72.1±15.6	79.4±14.5	87.7±14.0	92.7±11.3
PNN[34]	60.8±14.4	65.6±15.1	73.8±15.8	79.7±14.6
BlockFLD[38]	79.5±8.4	83.8±7.8	88.3±5.4	90.7±5.5
Volterra[35]	69.8±12.9	79.5±12.3	84.0±14.0	86.4±9.6
PCRC[36]	75.7±12.6	82.8±12.4	88.7±8.4	92.0±8.2
MSPCRC[36]	83.0±9.2	88.4±10.1	92.5±6.8	95.0±6.6
PMR	79.4±12.4	84.9±14.1	90.7±10.9	95.0±9.5
MSPMR	85.0±11.9	90.0±12.5	95.0±9.4	97.0±8.6

doi:10.1371/journal.pone.0159945.t002



Fig 6. Sample images of one person from the AR database.

doi:10.1371/journal.pone.0159945.g006

Table 3. Recognition rates (%) on the AR database.

Method	2	3	4	5
CRC[18]	69.9±12.6	80.6±10.4	83.8±9.6	89.1±6.2
SRC[15]	69.7±14.8	79.0±10.6	83.5±8.9	88.2±5.7
NSC[30]	77.6±8.4	84.4±6.7	87.8±5.3	91.1±3.9
PNN[34]	72.7±14.2	82.4±9.3	87.6±8.0	92.2±6.0
BlockFLD[38]	71.5±11.5	78.6±9.8	84.2±8.7	87.6±4.2
Volterra[35]	65.4±12.0	74.9±11.1	79.8±10.5	85.2±6.8
PCRC[36]	82.2±11.3	87.7±9.4	89.9±8.5	92.9±6.7
MSPCRC[36]	82.3±11.5	87.8±10.5	90.2±9.1	93.6±7.6
PMR	83.9±11.2	90.6±7.9	93.2±7.1	95.6±5.8
MSPMR	87.3±9.0	91.5±7.6	93.9±7.2	96.6±5.8

doi:10.1371/journal.pone.0159945.t003

extension of LFW after a commercial face alignment software is applied. As in [36], the objects who have more than ten samples are gathered to form a dataset with 158 objects from LFW-a. All the images are manually cropped and then resized to 32×32 pixels. Fig 7 shows some sample images from this database. For each subject, we randomly choose 2~5 samples for training and another 2 samples for testing.

Table 4 shows the face recognition of each method on the LFW dataset. From Table 4, we can clearly see that the performance of our PMR and MSPMR are superior to that of all other methods. Meanwhile, the recognition performance is greatly improved by MSPMR.



Fig 7. Sample images of the LFW database.

doi:10.1371/journal.pone.0159945.g007

Table 4. Recognition rates (%) on the LFW database.

Method	2	3	4	5
CRC[18]	24.7±2.1	31.9±2.4	37.8±2.6	42.0±3.2
SRC[15]	24.4±2.4	32.7±3.2	38.7±2.4	44.1±2.6
NSC[30]	26.6±2.2	33.7±2.5	40.6±2.1	45.2±2.7
PNN[34]	23.1±2.4	28.1±3.1	33.2±3.1	37.4±2.7
BlockFLD[38]	18.0±2.1	22.3±2.1	26.2±2.6	28.4±2.5
Volterra[35]	26.0±3.0	32.0±3.4	36.4±3.3	40.3±2.7
PCRC[36]	32.0±1.9	37.0±2.8	40.2±2.5	42.9±2.6
MSPCRC[36]	35.0±1.6	41.1±2.8	46.0±3.0	49.0±2.9
PMR	37.3±2.0	45.1±2.6	50.7±2.7	55.6±2.4
MSPMR	40.0±2.5	50.0±2.4	54.2±3.6	58.2±3.3

doi:10.1371/journal.pone.0159945.t004

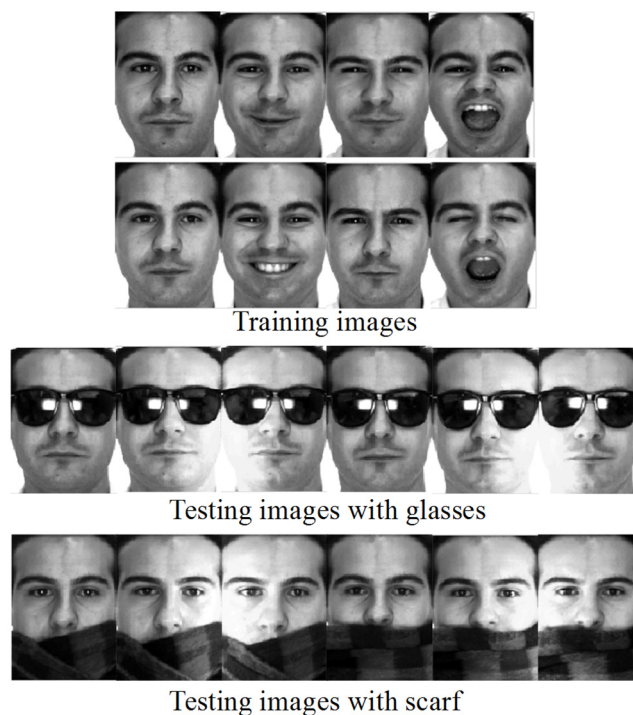


Fig 8. Training and testing images of a person in the AR database.

doi:10.1371/journal.pone.0159945.g008

2. Face recognition with occlusion

In the following experiments, we evaluate the robustness and effectiveness of the proposed method when face images encounter with different occlusions, like real disguise or block occlusion. In this subsection, our method is compared with CRC [18], SRC [15], NSC [30], HQ_A and HO_M [26], PSRC [15], PCRC and MSPCRC [36].

Face recognition with real disguise. As in [29, 32], a subset of the AR face database is applied, containing 50 males and 50 females. Each face image is manually cropped and normalized to a size of 42×30 . Fig 8 shows the sample images for one person from the AR database. In our experiment, for each individual, the first four images (with various facial expressions) from session 1 and session 2 are chosen to form the training set. Two image sets with sunglasses and scarves are used for testing, each of which includes 600 images (three images per session of each individual). For each individual, 2~5 samples are randomly chosen from the training set and another 3 samples from the testing set to evaluate the performance of each method.

The recognition results of each method are shown in Tables 5 and 6, from which we can see that the patch-based method achieves better performance than the corresponding original holistic ones. PMR also gives better results than PCRC or MSPCRC. MSPMR obtains the best performance among all the competing methods when testing images with sunglasses and achieves comparable results when testing images with scarves.

Face recognition with block occlusions. In this subsection, we evaluate the robustness of our method against block occlusions. We adopt Subsets 1 and 2 of the Extended Yale B database for training and Subset 3 for testing. All the face images are normalized to 48×42 pixels. The testing images are corrupted by a randomly located square block of a “baboon” image with an occlusion level of 40%. Fig 9 shows the training and testing sample images for one person from the Extended Yale B database. For each individual, 2~5 samples are randomly chosen

Table 5. Recognition rates (%) on the AR database testing with sunglasses.

Method	2	3	4	5
CRC[18]	23.6±8.3	29.4±7.2	35.4±8.0	40.4±7.8
SRC[15]	30.7±8.0	36.0±7.4	39.7±8.3	41.0±7.9
NSC[30]	58.8±12.9	68.8±9.8	75.2±7.6	77.8±6.3
HQ_A[26]	54.5±10.0	62.0±9.4	67.2±10.5	68.7±9.8
HQ_M[26]	55.7±10.4	63.4±9.7	67.7±10.4	70.8±10.0
PSRC[15]	54.6±14.1	67.2±10.9	74.2±10.9	78.8±8.5
PCRC[36]	63.4±14.2	69.2±11.6	72.8±11.0	74.6±8.7
MSPCRC[36]	68.9±16.6	72.5±13.5	75.1±11.7	76.5±8.5
PMR	65.4±14.7	75.9±10.2	83.6±7.8	86.8±5.5
MSPMR	76.6±10.8	85.6±7.0	88.9±5.8	91.1±5.5

doi:10.1371/journal.pone.0159945.t005

from the training set and another 5 samples from the testing set to evaluate the performance of each method.

The face recognition results of each method are tabulated in Table 7. We can see that by characterizing the reconstruction error with the nuclear norm, NSC overall outperforms CRC, SRC, HQ_A and HQ_M. By virtue of the patch trick, our PMR always outperforms PCRC and PSRC. By incorporating the multi-scale ensemble learning trick, the proposed MSPMR achieves the best performance among all the competitive methods.

Table 6. Recognition rates (%) on the AR database testing with scarves.

Method	2	3	4	5
CRC[18]	31.2±7.4	43.1±9.5	49.6±9.9	53.6±9.3
SRC[15]	29.6±7.3	41.4±8.9	47.4±8.5	52.6±5.5
NSC[30]	45.5±8.9	55.1±7.2	59.8±7.0	63.5±5.5
HQ_A[26]	39.8±7.8	49.0±7.7	55.3±8.1	59.4±7.6
HQ_M[26]	39.4±8.0	48.7±8.3	54.4±8.6	58.6±7.5
PSRC[15]	73.3±9.3	82.3±5.5	86.7±4.5	89.3±3.8
PCRC[36]	75.5±10.3	82.8±5.6	86.2±4.5	87.7±3.5
MSPCRC[36]	77.7±12.4	84.7±5.7	87.8±4.9	89.6±4.6
PMR	82.5±8.7	90.2±3.7	92.4±2.1	93.9±1.6
MSPMR	84.0±8.6	90.9±4.1	93.9±2.5	94.9±3.2

doi:10.1371/journal.pone.0159945.t006

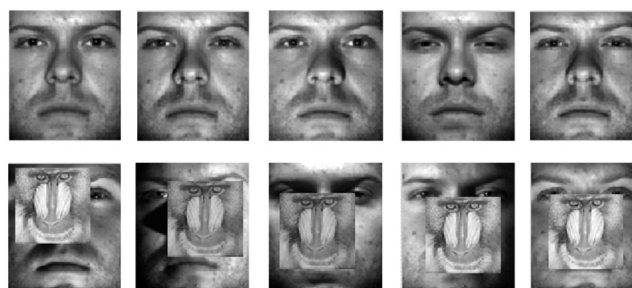


Fig 9. Training (the first row) and testing (the second row) sample images of a person in the Extended Yale B database.

doi:10.1371/journal.pone.0159945.g009

Table 7. Recognition rates (%) on the Extended Yale B database with block occlusion.

Method	2	3	4	5
CRC[18]	40.4±12.5	52.3±12.8	61.1±11.2	67.4±8.9
SRC[15]	50.8±11.5	62.3±10.0	67.5±13.0	71.2±14.5
NSC[30]	67.2±9.5	76.0±8.0	82.6±8.0	85.2±8.7
HQ_A[26]	51.6±12.5	63.5±12.8	68.2±11.1	73.0±14.1
HQ_M[26]	52.0±12.0	66.1±13.8	72.0±12.0	77.1±13.9
PSRC[15]	76.4±12.8	82.3±10.1	88.7±9.3	90.6±7.8
PCRC[36]	74.2±14.2	81.2±11.5	86.1±9.0	90.0±6.4
MSPCRC[36]	81.2±11.7	87.7±9.2	91.6±7.0	93.0±4.8
PMR	78.0±12.2	85.8±10.6	91.1±8.6	93.2±6.6
MSPMR	84.2±10.7	90.2±8.7	93.7±7.3	96.0±5.3

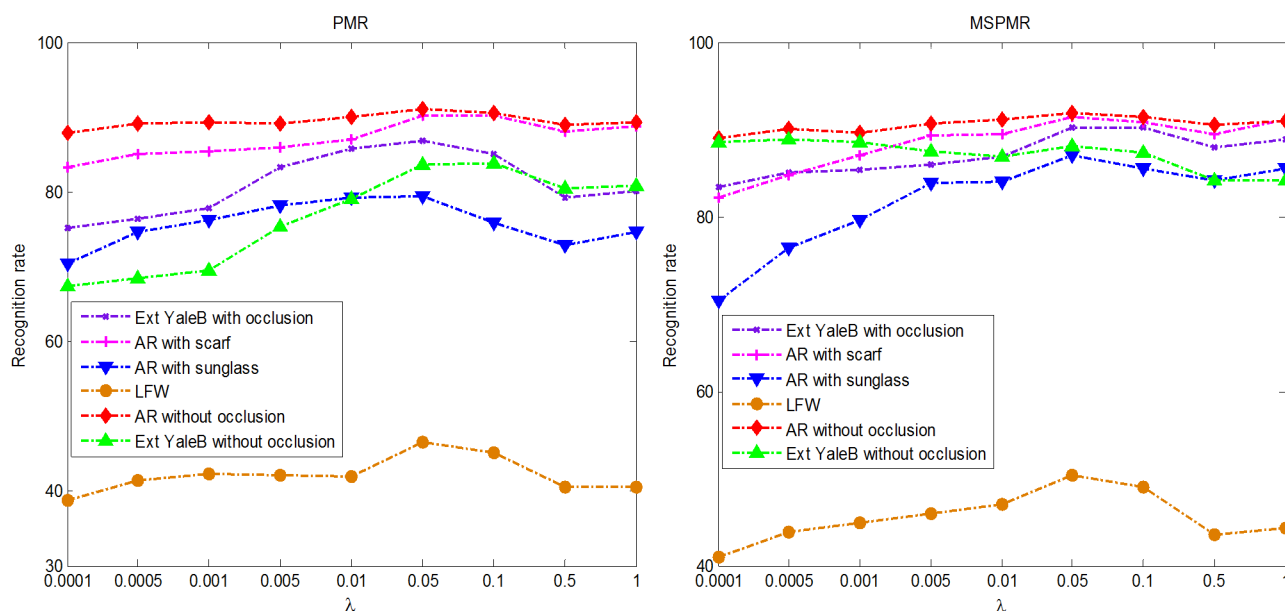
doi:10.1371/journal.pone.0159945.t007

3. Parameter discussion

In this subsection, we mainly discuss how the regularization parameter λ affects the performance of our PMR and MSPMR in different face recognition scenarios. The experimental settings are the same as in the aforementioned experiments in section 4.1 and 4.2 except that the training samples per person are fixed at 3. Fig 10 plots the recognition results of PMR and MSPMR versus the variation in the regularization parameter λ on different face image databases. We can observe that PMR and MSPMR always achieve their optimal or nearly optimal performance in the range of $[0.01, 0.1]$. Thus, we can set the regularization parameter of the proposed method in the above range for real-word scenarios.

4. Running time comparisons

In this subsection, the CPU runtime of the proposed method is compared with the state-of-the-art methods. The compared results on the AR face database testing with scarves are listed

**Fig 10. Recognition rate curves of PMR and MSPMR versus the variations in the regularization parameter in different face recognition scenarios.**

doi:10.1371/journal.pone.0159945.g010

Table 8. Comparisons of CPU time on AR face database testing with scarves.

Method	Time(s)
CRC[18]	0.6480
SRC[15]	46.3025
NSC[30]	67.9925
HQ_A[26]	108.3425
HQ_M[26]	164.7608
PSRC[15]	71.1593
PCRC[36]	9.5541
MSPCRC[36]	173.5952
PMR	256.9996
MSPMR	4767.1186

doi:10.1371/journal.pone.0159945.t008

for demonstration. For each individual, 3 samples are randomly chosen for training and another 3 samples for testing. Table 8 tabulates the CPU time cost on all test images conducted using Matlab R2012b on an Intel Core 8 CPU with 3.6 GHz and 8G memory PC at Windows platform. Since the singular value shrinkage operator in matrix regression, the proposed method consumes much more time compared with other methods. Although the patch-based methods have achieved promising results, they come at the cost of expensive running time. Thanks to the independence of the recognition process of each test image, we can hope to save the cost by parallel computation.

5. Evaluation of the experimental results

The aforementioned experimental results have shown that the proposed method always obtain better performance than some state-of-the-art methods. However, is this superiority statistically significant? In this subsection, we will assess the experimental results by the null hypothesis statistical test [51]. If the evaluated p -value is under the desired significance level (i.e., 0.05), the performance difference between compared approaches is deemed to be statistically significant. The evaluation results are summarized as follows:

1. For face recognition without occlusion, such as on LFW database, MSPMR outperforms MSPCRC significantly for all tests ($p = 0.014, 0.013, 0.016$ and 0.020). On other database, although MSPMR performs better than other state-of-the-art methods, the performance discrepancies between MSPMR and other approaches are not statistically significant.
2. For face recognition with occlusion, MSPMR performs significantly better than other approaches in case of real disguise and block occlusion ($p < 0.001$).

Conclusions and the Future Work

To improve the performance of matrix regression in face of the small sample size problem and preserve the desired performance level in the presence of occlusion and illumination changes, in this paper, we proposed a patch-based matrix regression (PMR) method. PMR first performed matrix regression on each raw patch (without matrix-to-vector conversion), and then combined the recognition outputs of all patches by majority voting. However, it is difficult to pre-define an optimal patch size across different databases. Fortunately, the complementary information across multiple patch scales can be fully utilized to further enhance the recognition performance. To this end, we proposed the multi-scale version of PMR, i.e., MSPMR, to

optimally combine the multi-scale outputs. Our extensive experimental results have demonstrated that the proposed methods are more effective and robust than the state-of-the-art methods.

Although our proposed method has obtained successful performance, there are still many issues to be addressed in future. Generally, two main improvements can be made for our method. (1) With the development of the storage device, many images can be collected from real-world applications. One challenge in our method is how to overcome the expensive computational cost. We will try to design efficient matrix regression algorithm to further improve the robustness and effectiveness of our method. (2) In our method, we have to pre-define several specific scale sizes in advance. However, different database may exhibit scale transformation in real-world applications. We can borrow the idea of scale selective local binary patterns [52] to design adaptive scale selection strategy to further improve the flexibility of our method.

Ethics Statement

Some face image datasets were used in this paper to verify the performance of our methods. These face image datasets are publicly available for face recognition research, and the consent was not needed. The face images and the experimental results are reported in this paper without any commercial purpose.

Author Contributions

Conceived and designed the experiments: GWG JY XYJ.

Performed the experiments: GWG PH JLH.

Analyzed the data: GWG JY XYJ DY.

Wrote the paper: GWG JY XYJ DY.

References

1. Wen X, Shao L, Xue Y, Fang W (2015) A rapid learning algorithm for vehicle classification. *Information Sciences* 295:395–406.
2. Gu B, Sheng VS, Tay KY, Romano W, Li S (2015) Incremental support vector learning for ordinal regression. *IEEE Transactions on Neural Networks and Learning Systems* 26(7):1403–1416. doi: [10.1109/TNNLS.2014.2342533](https://doi.org/10.1109/TNNLS.2014.2342533) PMID: [25134094](https://pubmed.ncbi.nlm.nih.gov/25134094/)
3. Juang C-F, Chiu S-H, Shiu S-J (2007) Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 37(6):1077–1087.
4. Gu B, Sheng VS, Wang Z, Ho D, Osman S, Li S (2015) Incremental learning for ν -support vector regression. *Neural Networks* 67:140–150. doi: [10.1016/j.neunet.2015.03.013](https://doi.org/10.1016/j.neunet.2015.03.013) PMID: [25933108](https://pubmed.ncbi.nlm.nih.gov/25933108/)
5. Deng Z, Choi K-S, Jiang Y, Wang S (2014) Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods. *IEEE Transactions on Cybernetics* 44(12):2585–2599. doi: [10.1109/TCYB.2014.2311014](https://doi.org/10.1109/TCYB.2014.2311014) PMID: [24710838](https://pubmed.ncbi.nlm.nih.gov/24710838/)
6. Gu B, Sheng VS (2016) A Robust Regularization Path Algorithm for ν -Support Vector Classification. *IEEE Transactions on Neural Networks and Learning Systems* doi: [10.1109/TNNLS.2016.2527796](https://doi.org/10.1109/TNNLS.2016.2527796)
7. Jiang Y, Chung F-L, Ishibuchi H, Deng Z, Wang S (2015) Multitask TSK fuzzy system modeling by mining intertask common hidden structure. *IEEE Transactions on Cybernetics* 45(3):534–547.
8. Gu B, Sun X, Sheng VS. Structural Minimax Probability Machine (2016) *IEEE Transactions on Neural Networks and Learning Systems* doi: [10.1109/TNNLS.2016.2544779](https://doi.org/10.1109/TNNLS.2016.2544779)
9. Jain AK, Ross A, Prabhakar S (2004) An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 14(1):4–20.

10. Wong WK, Lai Z, Xu Y, Wen J, Ho CP (2015) Joint Tensor Feature Analysis For Visual Object Recognition. *IEEE Transactions on Cybernetics* 45(11):2425–2436. doi: [10.1109/TCYB.2014.2374452](https://doi.org/10.1109/TCYB.2014.2374452) PMID: [26470058](https://pubmed.ncbi.nlm.nih.gov/26470058/)
11. Yang M, Feng Z, Shiu SC, Zhang L (2014) Fast and robust face recognition via coding residual map learning based adaptive masking. *Pattern Recognition* 47(2):535–543.
12. Gao G, Yang J, Wu S, Jing X, Yue D (2015) Bayesian sample steered discriminative regression for biometric image classification. *Applied Soft Computing* 37:48–59.
13. Chen B, Shu H, Coatrieux G, Chen G, Sun X, Coatrieux JL (2015) Color image analysis by quaternion-type moments. *Journal of mathematical imaging and vision* 51(1):124–144.
14. Yang W, Wang Z, Sun C (2015). A collaborative representation based projections method for feature extraction. *Pattern Recognition* 48(1):20–27.
15. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2):210–227. doi: [10.1109/TPAMI.2008.79](https://doi.org/10.1109/TPAMI.2008.79) PMID: [19110489](https://pubmed.ncbi.nlm.nih.gov/19110489/)
16. Naseem I, Togneri R, Bennamoun M (2010). Linear Regression for Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 32(11):2106–2112. doi: [10.1109/TPAMI.2010.128](https://doi.org/10.1109/TPAMI.2010.128) PMID: [20603520](https://pubmed.ncbi.nlm.nih.gov/20603520/)
17. Wagner A, Wright J, Ganesh A, Zhou Z, Mobahi H, Ma Y (2012) Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(2):372–386. doi: [10.1109/TPAMI.2011.112](https://doi.org/10.1109/TPAMI.2011.112) PMID: [21646680](https://pubmed.ncbi.nlm.nih.gov/21646680/)
18. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: Which helps face recognition? 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, p. 471–478.
19. Gao G, Yang J (2014) A novel sparse representation based framework for face image super-resolution. *Neurocomputing* 134:92–99.
20. Jiang J, Hu R, Wang Z, Han Z (2014) Noise robust face hallucination via locality-constrained representation. *IEEE Transactions on Multimedia* 16(5):1268–1281.
21. Tawari A, Trivedi MM (2013) Face expression recognition by cross modal data association. *IEEE Transactions on Multimedia* 15(7):1543–1552.
22. Lai Z, Xu Y, Jin Z, Zhang D (2014). Human gait recognition via sparse discriminant projection learning. *IEEE Transactions on Circuits and Systems for Video Technology* 24(10):1651–1662.
23. Lai Z, Xu Y, Chen Q, Yang J, Zhang D (2014) Multilinear sparse principal component analysis. *IEEE Transactions on Neural Networks and Learning Systems* 25(10):1942–1950. doi: [10.1109/TNNLS.2013.2297381](https://doi.org/10.1109/TNNLS.2013.2297381) PMID: [25291746](https://pubmed.ncbi.nlm.nih.gov/25291746/)
24. Yang J, Zhang L, Xu Y, Yang J-y (2012) Beyond sparsity: The role of L 1-optimizer in pattern classification. *Pattern Recognition* 45(3):1104–1118.
25. Yang M, Zhang L, Yang J, Zhang D (2011) Robust Sparse Coding for Face Recognition. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp:625–632.
26. He R, Zheng W-S, Tan T, Sun Z (2014) Half-quadratic-based iterative minimization for robust sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(2):261–275. doi: [10.1109/TPAMI.2013.102](https://doi.org/10.1109/TPAMI.2013.102) PMID: [24356348](https://pubmed.ncbi.nlm.nih.gov/24356348/)
27. Yang M, Zhu P, Liu F, Shen L (2015) Joint representation and pattern learning for robust face recognition. *Neurocomputing* 168:70–80.
28. Ou W, You X, Tao D, Zhang P, Tang Y, Zhu Z (2014) Robust face recognition via occlusion dictionary learning. *Pattern Recognition* 47(4):1559–1572.
29. Yang J, Luo L, Qian J, Tai Y, Zhang F, Xu Y (2016) Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence preprints* doi: [10.1109/TPAMI.2016.2535218](https://doi.org/10.1109/TPAMI.2016.2535218)
30. Luo L, Yang J, Qian J, Yang J (2014) Nuclear Norm Regularized Sparse Coding. 2014 22nd International Conference on Pattern Recognition (ICPR), IEEE, pp: 1834–1839.
31. Zhang F, Yang J, Tai Y, Tang J (2015). Double Nuclear Norm-Based Matrix Decomposition for Occluded Image Recovery and Background Modeling. *IEEE Transactions on Image Processing* 24(6):1956–1966. doi: [10.1109/TIP.2015.2400213](https://doi.org/10.1109/TIP.2015.2400213) PMID: [25667350](https://pubmed.ncbi.nlm.nih.gov/25667350/)
32. Qian J, Luo L, Yang J, Zhang F, Lin Z (2015). Robust nuclear norm regularized regression for face recognition with occlusion. *Pattern Recognition* 48(10):3145–3159.
33. Tan X, Chen S, Zhou Z-H, Zhang F (2006) Face recognition from a single image per person: A survey. *Pattern recognition* 39(9):1725–1745.

34. Kumar R, Banerjee A, Vemuri BC, Pfister H (2011) Maximizing all margins: Pushing face recognition with kernel plurality. 2011 IEEE International Conference on Computer Vision (ICCV), IEEE, pp: 2375–2382.
35. Kumar R, Banerjee A, Vemuri BC (2009) Volterrafaces: Discriminant analysis using volterra kernels. IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp: 150–155.
36. Zhu P, Zhang L, Hu Q, Shiu SC (2012) Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. 2012 European Conference on Computer Vision (ECCV), Springer, pp: 822–835.
37. Tan X, Chen S, Zhou Z-H, Zhang F (2005) Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. IEEE Transactions on Neural Networks 16(4):875–886. PMID: [16121729](#)
38. Chen S, Liu J, Zhou Z-H (2004) Making FLDA applicable to face recognition with one sample per person. Pattern recognition 37(7):1553–1555.
39. Su Y, Shan S, Chen X, Gao W (2009) Hierarchical ensemble of global and local classifiers for face recognition. IEEE Transactions on Image Processing 18(8):1885–1896. doi: [10.1109/TIP.2009.2021737](#) PMID: [19556198](#)
40. Lin D, Tang X (2006) Recognize high resolution faces: From macrocosm to microcosm. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, pp. 1355–1362.
41. Wolf L, Hassner T, Taigman Y (2011) Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(10):1978–1990. doi: [10.1109/TPAMI.2010.230](#) PMID: [21173442](#)
42. Guillaumin M, Verbeek J, Schmid C (2009) Is that you? Metric learning approaches for face identification. 2009 IEEE International Conference on Computer Vision (ICCV), IEEE, pp. 498–505.
43. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst.
44. Lin Z, Chen M, Ma Y (2010) The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055.
45. Cai JF, Candes EJ, Shen ZW (2010) A Singular Value Thresholding Algorithm for Matrix Completion. Siam Journal on Optimization 20(4):1956–1982.
46. Shawe-Taylor J, Cristianini N (1999) Robust bounds on generalization from the margin distribution. 4th European Conference on Computational Learning Theory.
47. Shen C, Li H (2010) On the dual formulation of boosting algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(12):2216–2231. doi: [10.1109/TPAMI.2010.47](#) PMID: [20975119](#)
48. Kim S-J, Koh K, Lustig M, Boyd S, Gorinevsky D (2007) An interior-point method for large-scale l1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing 1(4):606–617.
49. Lee KC, Ho J, Kriegman DJ (2005) Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5):684–698. PMID: [15875791](#)
50. Martinez AM, Benavente R (1998) The AR face database. CVC Technical Report 24.
51. Beveridge JR, She K, Draper B, Givens GH (2001) Parametric and nonparametric methods for the statistical evaluation of human id algorithms. 3rd Workshop on the Empirical Evaluation of Computer Vision Systems pp. 1–14.
52. Guo Z, Wang X, Zhou J, You J (2016) Robust Texture Image Representation by Scale Selective Local Binary Patterns. IEEE Transactions on Image Processing 25(2):687–699. doi: [10.1109/TIP.2015.2507408](#) PMID: [26685235](#)