# Image Set Classification Algorithm Based on Prototype and Orthogonal Projection Learning (Postprint)

**Authors:** Ren Zhenwen, Wu Mingna

**Date:** 2019-04-01T00:00:00+00:00

## Abstract

To leverage set information within image sets for improving image recognition accuracy and robustness to image variations, thereby substantially reducing the impact of factors such as pose, illumination, occlusion, and misalignment on recognition accuracy, we propose an Image Set Prototype and Projection Learning algorithm (LPSOP) for image set classification. This algorithm simultaneously learns representative points (prototypes) for each image set and an orthogonal global projection matrix, such that each image set in the target subspace can be optimally classified to the nearest prototype set of the same class. Using learned prototypes to represent image sets not only reduces interference from redundant images but also decreases storage and computational overhead, while the learned projection matrix can significantly improve classification accuracy and noise robustness. Experimental results on three datasets—UCSD/Honda, CMU Mobo, and YouTube Celebrities—demonstrate that LPSOP achieves higher recognition accuracy and better robustness compared to currently popular image set classification algorithms.

## Full Text

## Preamble

### Learning of Prototype Set and Orthogonal Projection for Image Set Classification

Ren Zhenwen[1,2], Wu Mingna[1]
(1. School of National Defence Science & Technology, Southwest University of Science & Technology, Mianyang, Sichuan 621010, China;
2. Dept. of Computer Science, Nanjing University of Science & Technology, Nanjing 210094, China)

**Abstract:** To improve identification accuracy and robustness by leveraging collection information from image sets, thereby greatly reducing the influence of factors such as posture, lighting, misalignment, and occlusion, this paper proposes a novel method called Learning Prototype Set and Orthogonal Projection (LPSOP) for image set classification. This algorithm simultaneously learns representative points (prototypes) and an orthogonal global projection matrix for each image set, enabling optimal classification of any image set in the target subspace to its nearest prototype set of the same class. The learned prototypes not only reduce redundant image interference but also decrease storage and computational overhead, while the projection matrix significantly enhances classification accuracy and noise robustness. Experimental results on the UCSD/Honda, CMU Mobo, and YouTube Celebrities datasets demonstrate that LPSOP achieves higher recognition accuracy and better robustness compared to current state-of-the-art image set classification algorithms.

**Keywords:** image set classification; prototype learning; scale learning; face recognition; target recognition; pattern recognition

## 0 Introduction

With advances in computer chip technology and increasingly diverse image acquisition methods, traditional single-image-based recognition methods have exposed limitations in accuracy and robustness, failing to meet the demands of natural image variations. Compared to single-image-based recognition, image set classification studies video frame sequences over time or multiple images collected under different times, scenes, and lighting conditions. By modeling image sets to select appropriate classification criteria, this approach can easily eliminate adverse factors (such as illumination, pose, and occlusion), thereby more effectively enhancing discriminative power and improving recognition rates. Consequently, image set classification algorithms have found increasingly widespread application in machine learning domains including face recognition, object recognition, and security surveillance.

Based on the characteristics of sets themselves, image set classification faces two primary challenges: set representation (modeling) and similarity measurement between sets. According to modeling approaches, related methods can be categorized into four main types: statistical model-based methods such as ProNN [?] and Stiefel and Grassmann Manifolds [?]; linear subspace-based methods such as DCC [?]; nonlinear manifold-based methods such as MMD [?, ?]; and affine subspace-based methods such as ProNN [?] (statistical model + affine subspace), RNP [?], AHISD, and CHISD [?].

Affine hull-based image set classification methods [?, ?] model image sets as affine hulls, where similarity between image sets is defined as the distance between the nearest points in the two affine hulls. This geometric approach makes classification results dependent on point positions in high-dimensional affine space, where outliers in image sets significantly reduce classification accuracy. To address the problem of overly large geometric regions caused by loose affine hull modeling, sparse representation-based nearest point methods [?, ?] represent image sets through sparse linear representation over a dictionary, classifying image sets based on sparse reconstruction residuals. Regularized nearest point methods [?] model each image set as a regularized affine hull. These nearest point-based methods heavily rely on the position of each sample in the image set, and outliers and noise substantially affect model performance.

To mitigate these effects, reference [?] utilizes covariance matrices and Riemannian kernel functions to represent image sets, measuring similarity between image sets through mapping from Riemannian manifolds to Euclidean space. Considering correlations between gallery sets, references [?] propose collaborative representation methods for image sets that further improve recognition accuracy. Notably, Wang Wen from the Chinese Academy of Sciences proposed an effective image set classification algorithm PDL [?], which learns a projection matrix to map samples from the original space to a target subspace, then models image sets as affine hulls within the target subspace, updating prototype positions by refining affine coefficients, with the learned projection matrix and prototypes jointly optimizing classification performance.

Although these methods achieve good classification results, several shortcomings remain: (a) they require samples to satisfy certain probability distributions; (b) they involve large computational and storage costs when processing extensive data; and (c) they exhibit low robustness and classification accuracy in the presence of noise or outliers. In recent years, while deep learning has achieved remarkable success in some tasks, its application to image set classification remains limited.

To address these issues, this paper proposes a Learning Prototype Set and Orthogonal Projection (LPSOP) algorithm for image set classification. For each image set, the algorithm first learns a set of representative samples as initial base prototypes through sparse representation and linear discriminant analysis. It then proposes an objective function based on minimizing classification error probability to simultaneously learn optimal prototypes and a linear projection matrix. The final learned representative points (prototypes) can optimally describe the image set while reducing computational time and space overhead, and the learned orthogonal projection matrix significantly enhances the discriminative capability of the prototype set. Furthermore, to further enhance stability and robustness in the target subspace, the projection matrix is constrained to be orthonormal.

To simultaneously optimize prototype and projection learning, this paper develops a gradient descent algorithm that alternately iterates between optimizing the prototype set and projection matrix. Unlike PDL, this paper first utilizes the Fisher discriminant criterion to learn base prototypes rather than modeling image sets as affine hulls, then learns the global projection matrix while updating prototype positions. Comparative experiments on the UCSD/Honda, CMU

Mobo, and YouTube Celebrities datasets against eight popular recent image set classification algorithms confirm that LPSOP achieves higher recognition accuracy and better robustness.

---

## 1.1 Basic Prototype Learning

For a classification scenario with $C$ image sets $\{T_1, T_2, \dots, T_C\}$, where $\mathbf{x}_{c,i} \in \mathbb{R}^d$ is the $d$-dimensional column vector obtained by vectorizing the $i$-th image in the $c$-th class, we learn a representative subset $P_c \subseteq T_c$ based on reference [?] by learning a sparse coefficient matrix $M_c$ (where $M_c$ is the sparse coefficient submatrix corresponding to $T_c$). The learning model is:

$$\min_{M_c} \frac{1}{2}\|T_c - T_c M_c\|_F^2 + \alpha\|M_c\|_{2,1} + \beta f(M_c) \quad (1)$$

where the first term is the reconstruction error, $\|\cdot\|_{2,1}$ denotes the mixed $\ell_{2,1}$ norm, $f(M_c)$ is a discriminant function on coefficient matrix $M_c$, and $\alpha$ and $\beta$ are scalar balancing parameters.

To ensure good discriminative capability between classes, we employ the Fisher discriminant criterion to minimize within-class scatter $S_W(M)$ and maximize between-class scatter $S_B(M)$. The scatter calculations are as follows:

$$S_W(M) = \sum_{c=1}^{C} \sum_{\mathbf{m}_k \in M_c} (\mathbf{m}_k - \mu_c)(\mathbf{m}_k - \mu_c)^T \quad (2)$$

$$S_B(M) = \sum_{c=1}^{C} n_c(\mu_c - \mu)(\mu_c - \mu)^T \quad (3)$$

where $\mathbf{m}_k$ is the $k$-th column of $M$, $\mu_c$ and $\mu$ are the mean vectors of $M_c$ and $M$ respectively, and $n_c$ represents the number of samples in class $c$. Combining equations (2) and (3), the discriminant term $f(M)$ is:

$$f(M) = \mathrm{tr}(S_W(M)) - \eta \cdot \mathrm{tr}(S_B(M)) \quad (4)$$

Substituting equation (4) into equation (1) yields:

$$\min_{M} \frac{1}{2}\|T - TM\|_F^2 + \alpha\|M\|_{2,1} + \beta\left(\mathrm{tr}(S_W(M)) - \eta \cdot \mathrm{tr}(S_B(M))\right) \quad (5)$$

Optimizing objective function (5) yields a reduced representative prototype set $P = TM$, where the representative prototype set $P$ not only optimally describes the training image sets $T$ but also provides good discriminative capability between classes. To address the problem of decreased classification accuracy and stability caused by noise or outliers in practical applications, this paper proposes an algorithm that simultaneously learns prototypes and a projection matrix to further enhance classification accuracy and robustness.

---

## 1.2 Objective Function Design

After obtaining the corresponding prototype set $P$ from the original image set $T$ through the prototype learning algorithm in Section 1.1, LPSOP optimizes not only the position of each prototype point $\mathbf{p}_c$ in $P$ but also simultaneously learns a linear projection matrix $W$ that projects samples from the original space to a low-dimensional target space, reducing redundant features and improving classification discriminative capability. Figure 1 [Figure 1: see original paper] illustrates the block diagram of the proposed algorithm.

For any sample image $\mathbf{x}$, its projection mapping is represented as $\mathbf{y} = W^T\mathbf{x}$. This paper designs an objective function $J(P, W)$ based on nearest neighbor (NN) classification error estimation to simultaneously learn an optimized prototype set $P$ and linear projection matrix $W$. The objective function is defined as:

$$J(P, W) = \sum_{c=1}^{C} \sum_{\mathbf{x} \in T_c} S_\beta(Q(\mathbf{x})) \quad (6)$$

where $S_\beta(\cdot)$ is a step function: $S_\beta(z) = 0$ when $z \leq 0$, and $S_\beta(z) = 1$ otherwise. From a classification perspective, $J(P, W)$ can be viewed as the sum of classification errors. Therefore, we aim to solve $\min_{P,W} J(P, W)$. However, this function is non-differentiable. To enable gradient descent for simultaneously learning the optimal prototype set $P$ and projection matrix $W$, this paper uses a Sigmoid function with slope $\beta$ and center at 0 to approximate the step function:

$$S_\beta(z) = \frac{1}{1 + e^{-\beta z}} \quad (7)$$

When $\beta$ is large, $S_\beta(z)$ approximates a "window" function that reaches its maximum when $z \geq 0$ and vanishes when $z < 0$. If $\beta$ is large, $S'_\beta(z)$ approximates a Dirac delta function, whereas if $\beta$ is small, $S'_\beta(z)$ is approximately constant within the feasible range of $z$. Furthermore, this paper defines the discriminant function $Q(\mathbf{x})$ as:

$$Q(\mathbf{x}) = \|\mathbf{y} - W^T\mathbf{p}_c^{wn}\|^2 - \|\mathbf{y} - W^T\mathbf{p}_c^{bn}\|^2 \quad (8)$$

where $\mathbf{p}_c^{wn}$ and $\mathbf{p}_c^{bn}$ are the nearest neighbor prototypes to $\mathbf{x}$ in the projected target subspace that belong to the same class and different classes, respectively. They are defined as:

$$\mathbf{p}_c^{wn} = \arg \min_{\mathbf{p}_a \in P, \text{Class}(\mathbf{p}_a)=\text{Class}(\mathbf{x})} \|W^T\mathbf{x} - W^T\mathbf{p}_a\|^2 \quad (9)$$

$$\mathbf{p}_c^{bn} = \arg \min_{\mathbf{p}_b \in P, \text{Class}(\mathbf{p}_b)\neq\text{Class}(\mathbf{x})} \|W^T\mathbf{x} - W^T\mathbf{p}_b\|^2 \quad (10)$$

In the original space, samples belonging to the same class as $\mathbf{x}$ and those belonging to different classes are projected into the target subspace via projection matrix $W$ to obtain two sets. The nearest prototype points are selected from these two sets respectively, yielding prototype points $\mathbf{p}_c^{wn}$ and $\mathbf{p}_c^{bn}$. Here, $\mathbf{x}_c^{wn}$ and $\mathbf{x}_c^{bn}$ are the corresponding samples in the original space for the nearest prototype points $\mathbf{p}_c^{wn}$ and $\mathbf{p}_c^{bn}$. Moreover, for all $\mathbf{x}$, if $Q(\mathbf{x}) > 0$, the function value $S_\beta(Q(\mathbf{x}))$ approximates zero, indicating that $\mathbf{x}$ is correctly classified; conversely, it is considered misclassified.

To further enhance stability and robustness in the target subspace, this paper imposes an orthogonal constraint on the projection matrix $W$, i.e., $W^TW = I$ (where $I$ is the identity matrix).

---

## 1.3 Algorithm Optimization

To obtain the optimal prototype set $P$ and linear projection matrix $W$, this paper develops an alternating gradient descent method to minimize the objective function $J(P, W)$: we compute the partial derivatives of $J$ with respect to $W$ and $P$, and perform alternating iterative updates. Considering that the definition of $Q(\mathbf{x})$ involves $\mathbf{p}_c^{wn}$ and $\mathbf{p}_c^{bn}$, and that for some $\mathbf{x}$, when the positions of $W$ and $P$ change, these nearest prototypes also change. It is not difficult to see that the search process for nearest prototypes depends on $W$ and $P$. Since when the positions of $W$ and $P$ change, although $\mathbf{x}$ remains unchanged, $\mathbf{p}_c^{wn}$ and $\mathbf{p}_c^{bn}$ will also change. This involves discontinuities and implicit dependencies. Therefore, this paper assumes that when one of the projection or prototype positions changes slightly, the other remains unchanged. That is, when solving the optimization problem, we can fix $P$ to solve $W$, or fix $W$ to solve $P$.

Under this approximation, the partial derivative of $J$ with respect to $W$ is:

$$\frac{\partial J}{\partial W} \approx \sum_{c=1}^{C} \sum_{\mathbf{x} \in T_c} S'_\beta(Q(\mathbf{x})) \cdot \left((\mathbf{x} - \mathbf{x}_c^{wn})\mathbf{y}^T - (\mathbf{x} - \mathbf{x}_c^{bn})\mathbf{y}^T\right) \quad (12)$$

In equation (12), $\mathbf{y}_k$ is the $k$-th column of $W$, and $y_k$ is the $k$-th element of vector $\mathbf{y}$. For the $c$-th class image set $T_c$, the partial derivative of the optimal prototype set $P_c = \{\mathbf{p}_{c,1}, \mathbf{p}_{c,2}, \dots, \mathbf{p}_{c,m_c}\}$ with respect to $P$ is:

$$\frac{\partial J}{\partial \mathbf{p}_{c,i}} \approx \sum_{\mathbf{x}\in T_c} S'_\beta(Q(\mathbf{x}))\cdot\big(W^T(\mathbf{y}-W^T\mathbf{p}_c^{wn})\cdot \mathbb{I}(\mathbf{p}_{c,i}=\mathbf{p}_c^{wn}) - W^T(\mathbf{y}-W^T\mathbf{p}_c^{bn})\cdot \mathbb{I}(\mathbf{p}_{c,i}=\mathbf{p}_c^{bn})\big) \quad (13)$$

where $\mathbb{I}(\cdot)$ is the indicator function. Using equations (12) and (13), the gradient descent update equations are (where $\gamma$ and $\nu$ are learning step factors respectively):

$$W^{(t+1)} = W^{(t)} - \gamma\frac{\partial J}{\partial W^{(t)}} \quad (14)$$

$$P_c^{(t+1)} = P_c^{(t)} - \nu\frac{\partial J}{\partial P_c^{(t)}} \quad (15)$$

---

## 1.4 Classification

First, using the optimization method from Section 1.3 with the gallery sets as input, we learn the optimal prototype set $P^*$ and linear projection matrix $W^*$. Next, both the test set and prototype set $P^*$ are transformed into the target space using the linear projection matrix $W^*$. Then, the distance between the test image set and prototype set in the target space is computed, specifically the shortest Euclidean distance between test set samples and the corresponding prototypes in the gallery set. Finally, the test set is assigned to the class of the image set containing the nearest prototype, completing the image set classification.

---

# 2 Experiments

To validate the effectiveness of the proposed LPSOP algorithm for image set classification, this paper compares and analyzes eight popular recent image set classification algorithms on three public databases, recording their respective classification accuracies and standard deviations.

### 2.1 Image Sets and Sample Settings

This paper uses three datasets for algorithm evaluation: UCSD/Honda [?], CMU Mobo [?], and YouTube Celebrities (YTC) [?].

The UCSD/Honda dataset contains 59 videos from 20 subjects, with each image in the videos undergoing pose, illumination, and expression variations. Each face image in the dataset is resized to 20×20 grayscale images and processed with histogram equalization [?]. All video sequences are divided into two groups, with 20 video sequences randomly selected as the training set and the remaining 39 as the test set.

The CMU Mobo dataset contains 96 video sequences from 24 subjects, with each subject having 4 video sequences, each corresponding to a specific walking pattern. First, faces are extracted using the Viola-Jones face detector framework, then resized to 30×30 grayscale images. Finally, one image set is randomly selected from each subject' s video as the training set, with the remainder used as the test set.

The YTC dataset consists of 1910 videos from 47 celebrities, where each video sequence contains significant pose, illumination, and expression variations. Since most videos in this dataset are low-resolution, with large pose variations and diverse lighting intensities, this dataset is more challenging than the previous two. For this dataset, this paper resizes the extracted facial regions to 20×20 grayscale images. In each video segment, 3 image sets are randomly selected as the training set and 6 image sets as the test set. Figure 2 [Figure 2: see original paper] shows sample images from YTC, where each row of face photos belongs to the same image set (the YTC dataset contains naturally collected images with low quality and noise, making it suitable for testing recognition accuracy and robustness).

### 2.2 Comparison Algorithms and Parameter Settings

To demonstrate the effectiveness of LPSOP, this paper compares it with several popular image set classification algorithms, including DCC [?], MMD [?], AHISD [?], CHISD [?], SANP [?], RNP [?], SSDML [?], and PDL [?]. The source code for these algorithms can be downloaded from the respective authors' homepages. To ensure fair experimental results, the relevant parameters for each algorithm are configured according to the recommended experimental parameters in their respective references. In our experiments, each image set consists of randomly selected 50, 100, or 200 face images, and all experimental results for classification accuracy and standard deviation are obtained by averaging over 10 runs. Additionally, the projection dimension of $W$ is set to 100.

### 2.3 Experimental Results and Analysis

This paper first conducts experiments on the UCSD/Honda database, setting image set sizes to 50, 100, and 200 to test LPSOP' s recognition accuracy and standard deviation. Table 1 compares LPSOP with eight other popular algorithms on the UCSD/Honda dataset. The results show that LPSOP achieves higher recognition accuracy and lower standard deviation than other related algorithms, proving the effectiveness of our method. Notably, RNP, PDL, and LPSOP all achieve 100% recognition rate when the image set size is 200.

**Table 1. Classification accuracy and standard deviation data on UCSD/Honda (%)**

| Algorithm | 50 images | 100 images | 200 images |
|---|---|---|---|
| DCC | 77.1±3.5 | 82.2±2.4 | 84.3±2.0 |
| MMD | 68.9±4.4 | 84.3±2.9 | 91.8±1.9 |
| AHISD | 87.4±2.6 | 83.8±2.0 | 86.3±2.0 |
| CHISD | 82.2±2.4 | 87.2±3.1 | 84.6±3.6 |
| SANP | 84.3±2.9 | 90.1±2.1 | 84.4±1.9 |
| RNP | 83.8±2.0 | 91.8±3.2 | 91.8±2.0 |
| SSDML | 87.2±3.1 | 84.3±2.0 | 88.9±1.9 |
| PDL | 90.1±2.1 | 95.1±3.0 | 91.8±1.8 |
| LPSOP | **95.9±2.3** | **96.1±1.3** | **100.0±0.0** |

Next, experiments are conducted on the CMU Mobo database, again testing LPSOP's recognition accuracy and standard deviation with image set sizes of 50, 100, and 200. Table 2 presents the comparative data between LPSOP and other algorithms. The results show that LPSOP still achieves the best classification performance. Compared to UCSD/Honda, the CMU Mobo database is more challenging, yet LPSOP' s recognition rate can still reach up to 96.73%, which is 1.33% higher on average than the similar method PDL.

**Table 2. Classification accuracy and standard deviation data on CMU Mobo (%)**

| Algorithm | 50 images | 100 images | 200 images |
|---|---|---|---|
| DCC | 81.7±3.2 | 91.5±3.4 | 94.8±1.6 |
| MMD | 91.1±2.5 | 91.8±3.0 | 95.4±1.7 |
| AHISD | 90.8±2.7 | 91.2±3.1 | 90.4±2.3 |
| CHISD | 91.5±3.4 | 91.7±2.8 | 95.5±0.9 |
| SANP | 91.8±3.0 | 94.4±2.0 | 91.7±2.6 |
| RNP | 91.2±3.1 | 83.5±2.7 | 97.4±2.2 |
| SSDML | 91.7±2.8 | 93.5±2.2 | 97.2±1.5 |
| PDL | 94.4±2.0 | 93.9±2.1 | 97.4±2.7 |
| LPSOP | **95.1±2.3** | **95.4±1.2** | **98.3±0.9** |

Finally, experiments are conducted on the more challenging YTC database, with image set sizes set to 50, 100, and 200 as in the previous experiments. Table 3 shows the final results. The table demonstrates that LPSOP outperforms other popular algorithms. When image set sizes are 50, 100, and 200, the classification accuracy is 0.5%, 1%, and 0.5% higher than the second-best algorithm, respectively. Particularly when image set sizes are small (50 and 100), LPSOP shows greater advantages.

**Table 3. Classification accuracy and standard deviation data on YouTube (%)**

| Algorithm | 50 images | 100 images | 200 images |
|---|---|---|---|
| DCC | 75.7±2.3 | 78.4±4.3 | 73.5±4.4 |
| MMD | 76.2±4.4 | 72.5±4.2 | 71.7±4.5 |
| AHISD | 68.9±4.4 | 77.7±5.2 | 72.8±7.6 |
| CHISD | 74.6±4.9 | 77.4±3.3 | 73.3±5.3 |
| SANP | 78.4±4.3 | 78.9±2.5 | 74.9±6.0 |
| RNP | 72.5±4.2 | 73.5±4.4 | 68.1±5.3 |
| SSDML | 77.7±5.2 | 71.7±4.5 | 75.3±5.0 |
| PDL | 77.4±3.3 | 73.4±4.1 | 75.2±3.8 |
| LPSOP | **78.9±2.5** | **79.3±3.6** | **76.3±3.6** |

The proposed prototype reduction and orthogonal projection learning method simultaneously learns optimal prototypes and a linear projection matrix. The learned prototype set can effectively describe the structure and features of samples in the image set, while the linear projection matrix can be viewed as a global scale that significantly improves algorithm accuracy and robustness. To demonstrate the learning results of our algorithm, Figure 3 [Figure 3: see original paper] shows images corresponding to some learned prototypes from the YTC database, where each row belongs to the same prototype set. The figure shows that the learned prototypes contain important facial contour and landmark information. Figure 4 [Figure 4: see original paper] displays the projection matrix $W$ (400×100), where the varying element values indicate that this linear projection matrix contributes weights as a global scale.

### 2.4 Algorithm Convergence

Although this paper does not provide a direct theoretical proof of LPSOP’ s convergence properties, experimental analysis demonstrates that the algorithm exhibits fast and smooth convergence characteristics. The convergence curve tested on the YTC database is shown in Figure 5 [Figure 5: see original paper]. As iterations proceed, the loss function value stabilizes after approximately 30 iterations, indicating good convergence performance.

### 2.5 Problems and Future Directions

As shown in Tables 1-3, compared to eight other image set classification methods, LPSOP achieves the highest recognition accuracy and lowest standard deviation on all three datasets. This success primarily stems from two innovations: (a) learning a set of representative prototype points through sparse representation and linear discriminant analysis; and (b) simultaneously learning an optimal prototype set and a global linear projection matrix to enhance the discriminative capability of image sets. However, LPSOP may suffer from overfitting.

Future work will introduce regularization terms to improve LPSOP, such as $\ell_2$ regularization, $\ell_1$ sparse regularization, or structured sparse regularization.

## 3 Conclusion

This paper proposes a Learning Prototype Set and Orthogonal Projection (LPSOP) algorithm for image set classification. The algorithm simultaneously learns a prototype set and a linear projection matrix for image sets. The learned prototype (representative point) set can optimally describe the image set, while the learned orthogonal projection matrix significantly enhances the discriminative capability of the prototype set. Furthermore, to enhance stability and robustness in the target subspace, the projection matrix is constrained to be orthonormal. To simultaneously optimize prototype and projection learning, this paper develops a gradient descent algorithm that alternately iterates between the prototype set and projection matrix. Experiments on three public datasets demonstrate that LPSOP achieves higher recognition accuracy and better robustness compared to other state-of-the-art algorithms for image set classification.

However, this paper does not consider the hidden and exploitable features contained in the gallery sets themselves. Future work will employ relevant low-rank algorithms to capture these hidden features and further improve recognition accuracy.

## References

[1] Wang Wen, Wang Ruiping, Shan Shiguang, et al. Probabilistic nearest neighbor search for robust classification of face image sets [C]//Proc of IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Washington DC: IEEE Computer Society, 2015.

[2] Turaga P, Veeraraghavan A, Srivastava A, et al. Statistical computations on Grassmann and stiefel manifolds for image and video-based recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2273-2286.

[3] Kim T K, Kittler J, Cipolla R. Discriminative learning and recognition of image set classes using canonical correlations [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2007, 29(6): 1005-1013.

[4] Du Jixiang, Shao Meiwen, Zhai Chuanmin, et al. Recognition of leaf image set based on manifold–manifold distance [J]. Neurocomputing, 2016, 188(5): 131-138.

[5] Yang Meng, Zhu Pengfei, Van Gool L, et al. Face recognition based on regularized nearest points between image sets [C]//Proc of IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Washington DC: IEEE Computer Society, 2013.

[6] Hu Yiqun, Mian A S, Owens R. Face Recognition using sparse approximated nearest points between image sets [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2012, 34(10): 1992-2004.

[7] Cevikalp H, Triggs B. Face recognition based on image sets [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2011: 2567-2573.

[8] Cheng Gong, Zhou Peicheng, Han Junwei. Duplex metric learning for image set classification [J]. IEEE Trans on Image Processing, 2018, 27(1): 281-292.

[9] Ma Jianhong, Zhang Han, Ji Qiu. Robust face recognition using block based on covariance matrix to represent depth image set [J]. Application Research of Computers, 2016, 33(12): 3847-3851.

[10] Lu Jiwen, Wang Gang, Deng Weihong, et al. Simultaneous feature and dictionary learning for image set based face recognition [J]. IEEE Trans on Image Process, 2017, 26(8): 4042-4054.

[11] Wang Ruiping, Guo Huimin Larry S D, et al. Covariance discriminative learning: a natural and efficient approach to image set classification [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2012.

[12] Cai Sijia, Zhang Lei, Zuo Wangmeng, et al. A probabilistic collaborative representation based approach for pattern classification [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2016.

[13] Li Feng, Zhang Sanyuan. Laplacian sparse coding dictionary for image set based collaborative representation [C]//Proc of the 2nd IEEE International Conference on Computer and Communications. Piscataway, NJ: IEEE Press, 2016: 245-249.

[14] Uzair M, Shafait F, Ghanem B, et al. Representation learning with deep extreme learning machines for efficient image set classification [J]. Neural Computing and Applications, 2018, 30(4): 1211-1223.

[15] Wang Wen, Wang Ruiping, Shan Shiguang, et al. Prototype discriminative learning for face image set classification [C]//Proc of Asian Conference on Computer Vision. Cham: Springer, 2016: 344-360.

[16] Wang Baoxing, Yin Qiyue, Wu Shu, et al. Discriminative Representative Selection via Structure Sparsity [C]//Proc of International Conference on Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 1401-1406.

[17] Lee K C, Ho J, Yang M H, et al. Video-based face recognition using probabilistic appearance manifolds [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2003: 313-320.

[18] Hayat M, Khan S H, Bennamoun M. Empowering simple binary classifiers for image set based face recognition [J]. International Journal of Computer Vision, 2017, 123(3): 479-498.

[19] Kim M, Kumar S, Pavlovic V, et al. Face tracking and recognition with visual constraints in real-world videos [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2008: 1-8.

[20] Liu Yali, Xu Pengfei. Linear filtering multispectral image enhancement algorithm with fuzzy kernel clustering [J]. Application Research of Computers, 2015, 32(5): 1536-1539.

[21] Wang Ruiping, Shan Shiguang, Chen Xilin, et al. Manifold-manifold distance with application to face recognition based on image set [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2008.

[22] Zhu Pengfei, Zhang Lei, Zuo Wangmeng, et al. From Point to Set: Extend the Learning of Distance Metrics [C]//Proc of International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2013: 2664-2671.

*Note: Figure translations are in progress. See original paper for figures.*