Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 3, No. 3, 2023

Non-negative Sparse Discriminative Low Rank Preserving Projection for Robust Face Feature Extraction

Rui Li [*], Haishuang Yin

School of Electrical Engineering, Northeast Petroleum University, Daqing, Heilongjiang, China

[*] **Corresponding author**: Rui Li

Abstract: Face recognition is an important aspect of intelligent security for new energy vehicles. Existing methods extract features without taking low-rank and discriminative similarity relations of data into account and resulting in low quality feature distribution. In addition, the coefficients of the learned representation matrix can be negative and lack interpretability. To address the above issues, a method named non-negative sparse discriminative low rank preserving (NNSDLRPP) which introduces sparsity, non-negativity and block-diagonal regularization is proposed. As a result, NNSDLRPP improves the interpretability of representation features while capturing discriminative information of data. Extensive experiments on two face datasets show that the proposed method outperforms other state-of-the-art methods.

Keywords: Face recognition; Feature extraction; Low rank representation.

1. Introduction

With the rapid development of new energy technology and artificial intelligence, face recognition is of great significance to automotive intelligent security. Feature extraction plays an important role in face recognition. The low-rank-based feature extraction methods have received widespread attention due to their ability to construct robust graphs. In, both the global and local structures are used for feature extraction. We et al. proposed a low-rank preserving projection via graph regularized reconstruction (LRPP_GRR) without increasing complexity. However, the low-rank-based methods only perform well when the data is strictly sampled in its original subspace, which means that the discriminant features are orderly distributed. In other words, a sample should be adaptively represented by its intraclass samples. So the ideal representation of the data is the best to have block-diagonal structure. Otherwise, samples of different classes may be close to each other, resulting in a decrease in classification accuracy. In addition, the learned representation coefficient matrix may be negative and lack interpretability, while non-negativity is more in accordance with practical data representation. Finally, these methods often result in a dense graph, which leads to unnecessary connections.

Inspired by the above insights, we construct a non-negative sparse discriminative low-rank preserving projection (NNSDLRPP) for face classification. This model solves the above limitations by imposing the sparsity, non- negativity and block-diagonal regularization to obtain more discriminative feature representation. With such method, a potential subspace with good discriminative information and strong interpretability can be found and achieving better classification accuracy.

2. Methodology

2.1. Model Formulation

In order to preserve the global and local structure of the data during feature extraction without increasing complexity, we adopt the LRPP_GRR framework. In addition, we introduce three helpful regularization terms which are group-wise constraint, sparse constraint and non-negative constraint.

To remedy the shortcoming of low discrimination, we seek a representation matrix Z with a block-diagonal structure. Specifically, the desired representation Z with c classes is as follows:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & \mathbf{Z}_c \end{bmatrix} \tag{1}$$

To this end, a group-wise regularization $\sum_i^c \|\mathbf{Z}_{-i}\|_F^2$ is introduced. $\mathbf{Z}_{-i}$ means the representation matrix of the i-th class samples subject to X with i-class samples removed. The label information is incorporated into this regularization term, which can promote each sample to be adaptively represented by the corresponding samples and the representation coefficient of interclass samples to be suppressed. So a more ordered and discriminative representation matrix can be obtained. Sparsity helps to reduce unnecessary connections between samples with smaller weights and ensure that the involved neighbors are fewest possible. Nonnegativity ensures that every data vector is in the convex hull of its neighbors. The three constraints introduce richer information on affinity representation matrix Z, which enables the structure of Z to reflect the relationships between samples more accurately. Based on the above analysis, the objective function of NNSDLRPP can be formulated as follows:

$$\begin{gathered}\min_{\mathbf{R},\mathbf{P},\mathbf{Z}} \sum_{i=1}^{n}\sum_{j=1}^{n} \left\|\mathbf{x}_i - \mathbf{R}\mathbf{P}^T\mathbf{X}\mathbf{z}_j\right\|_2^2 \mathbf{W}_{ij} + \alpha\|\mathbf{Z}\|_* + \frac{\beta}{2}\sum_{i=1}^{c}\|\mathbf{Z}_{-i}\|_F^2 + \gamma\|\mathbf{Z}_1\|_1 + \lambda\|\mathbf{P}\|_{2,1} \\ s.t. \mathbf{R}^T\mathbf{R} = \mathbf{I}, \mathbf{Z} \geq \mathbf{0}\end{gathered} \tag{2}$$

Where α, β, γ and λ are regularization parameters; The first term imposes graph constraints on data reconstruction errors to capture global and local information of data. Redundant information can be removed by introducing the $\ell_{2,1}$-norm into the projection matrix P. Latent subspace derived by the proposed method is more discriminant and interpretable.

2.2. Optimization Strategy

To make (2) separable, we first introduce three auxiliary variables Z=A, Z=J and RTXZ=Y. Then, the augmented Lagrangian function of (3) is constructed as follows:

$$\begin{aligned} L = & tr\left(\mathbf{XDX}^T - 2\mathbf{XWY}^T\mathbf{R}^T\right) + tr(\mathbf{YDY}^T) + \alpha\|\mathbf{A}\|_* + \frac{\beta}{2}\sum_{i=1}^{c}\|\mathbf{Z}_{-i}\|_F^2 + \gamma\|\mathbf{J}\|_1 + \lambda\|\mathbf{P}\|_{2,1} \\ & + \frac{\mu}{2}\left(\|\mathbf{P}^T\mathbf{XZ}-\mathbf{Y}\|_F^2 + \|\mathbf{Z}-\mathbf{A}\|_F^2 + \|\mathbf{Z}-\mathbf{J}\|_F^2\right) + \left\langle \mathbf{C}_1, \mathbf{P}^T\mathbf{XZ}-\mathbf{Y}\right\rangle \\ & + \left\langle \mathbf{C}_2, \mathbf{Z}-\mathbf{A}\right\rangle + \left\langle \mathbf{C}_3, \mathbf{Z}-\mathbf{J}\right\rangle \end{aligned} \quad (3)$$

where $\mathbf{C}_1, \mathbf{C}_2$ and $\mathbf{C}_3$ are the Lagrangian multipliers, μ is a penalty parameter, D is a diagonal matrix with $\mathbf{D}_{ii} = \sum_j w_{ij}$. By alternately updating all variables, we can obtain the solutions of (2).

Step 1: update P with the other variables fixed.

$$L(\mathbf{P}) = \lambda\|\mathbf{P}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{P}^T\mathbf{XZ} - \mathbf{Y} + \mathbf{C}_1/\mu\right\|_F^2 \quad (4)$$

Taking the derivative of $L(\mathbf{P})$ and setting it to zero, then P is obtained:

$$\mathbf{P} = \left(\lambda\mathbf{G} + \mu\mathbf{XZZ}^T\mathbf{X}^T\right)^{-1}\left(\mu\mathbf{XZH}^T\right) \quad (5)$$

where $\mathbf{H} = \mathbf{Y} - \mathbf{C}_1/\mu$, $\mathbf{G}_{ii} = 1/\|\mathbf{P}_{i\cdot}\|_2$, $\mathbf{P}_{i\cdot}$ is the i-th row vector of Q

Step 2: update Z with the other variables fixed.

$$L(\mathbf{Z}) = \frac{\beta}{2}\sum_{i=1}^{c}\|\mathbf{Z}_{-i}\|_F^2 + \frac{\mu}{2}\left(\|\mathbf{P}^T\mathbf{XZ}-\mathbf{Y}+\mathbf{C}_1/\mu\|_F^2 + \|\mathbf{Z}-\mathbf{A}+\mathbf{C}_2/\mu\|_F^2 + \|\mathbf{Z}-\mathbf{J}+\mathbf{C}_3/\mu\|_F^2\right) \quad (6)$$

Let $\partial L(\mathbf{Z})/\partial\mathbf{Z} = 0$, then we get:

$$\mathbf{Z} = \left(\mu\mathbf{X}^T\mathbf{PP}^T\mathbf{X} + \beta\mathbf{I} + 2\mu\mathbf{I}\right)^{-1}\left(\beta\mathbf{Q} + \mu\mathbf{X}^T\mathbf{PM}_1 + \mu\mathbf{M}_2 + \mu\mathbf{M}_3\right) \quad (7)$$

where $\mathbf{Q}_i = \mathbf{Z}_i$, $\mathbf{M_1} = \mathbf{Y} - \mathbf{C}_1/\mu$, $\mathbf{M_2} = \mathbf{A} - \mathbf{C}_2/\mu$, $\mathbf{M_3} = \mathbf{J} - \mathbf{C}_3/\mu$.

Step 3: update A with the other variables fixed.

$$\min_{\mathbf{A}} \alpha\|\mathbf{A}\|_* + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{A} + \mathbf{C}_2/\mu\|_F^2 \quad (8)$$

The solution of problem (8) can be obtained by singular value thresholding operator [7] S as follows:

$$\mathbf{A} = \mathrm{S}_{\alpha/\mu}\left(\mathbf{Z} + \mathbf{C}_2/\mu\right) \quad (9)$$

Step 4: update J with the other variables fixed.

$$\min_{\mathbf{J}\geq 0} \gamma\|\mathbf{J}\|_1 + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{J} + \mathbf{C}_3/\mu\|_F^2 \quad (10)$$

Then J is obtained by using the soft thresholding operator [8] Ω as follows:

$$\mathbf{J} = \max\left\{\Omega_{\gamma/\mu}\left(\mathbf{Z} + \mathbf{C}_3/\mu\right), 0\right\} \quad (11)$$

Step 5: update Y with the other variables fixed.

$$L(\mathbf{Y}) = tr(\mathbf{YDY}^T - 2\mathbf{XWY}^T\mathbf{R}^T) + \frac{\mu}{2}\|\mathbf{P}^T\mathbf{XZ} - \mathbf{Y} + \mathbf{C}_1/\mu\|_F^2 \quad (12)$$

Let $\mathbf{M_4} = \mathbf{P}^T\mathbf{XZ} - \mathbf{C}_1/\mu$ and $\partial L(\mathbf{Y})/\partial\mathbf{Y} = 0$, then we can get:

$$\mathbf{Y} = \left(2\mathbf{D} + \mu\mathbf{I}\right)^{-1}\left(\mu\mathbf{M}_4 + 2\mathbf{R}^T\mathbf{XW}\right) \quad (13)$$

Step 6: update R with the other variables fixed.

$$\max_{\mathbf{R}^T\mathbf{R}=\mathbf{I}} tr\left(\mathbf{XWY}^T\mathbf{R}^T\right) \quad (14)$$

Problem (15) is an orthogonal Procrustes problem which can be simply solved by SVD, i.e.,

$$\mathrm{SVD}\left(\mathbf{XWY}^T\right) = \mathbf{USV}^T \quad (15)$$

$$\mathbf{R} = \mathbf{UV}^T \quad (16)$$

Step 6: update Lagrangian multipliers and penalty parameter.

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{C}_1 + \mu\left(\mathbf{P}^T\mathbf{XZ} - \mathbf{Y}\right) \\ \mathbf{C}_2 &= \mathbf{C}_2 + \mu\left(\mathbf{Z} - \mathbf{A}\right) \\ \mathbf{C}_3 &= \mathbf{C}_3 + \mu\left(\mathbf{Z} - \mathbf{J}\right) \\ \mu &= \min(\rho\mu, \mu_{\max}) \end{aligned} \quad (17)$$

3. Experimental Result

3.1. Experimental Setup

In this section, several experiments are conducted to evaluate the performance of the proposed algorithm. For comparison, some related methods are selected including LPP, NPE, SPP, LSPP, LRPP_GRR. The datasets used in the experiments are ORL and Extended Yale B. The ORL dataset contains 400 images of 40 individuals and the image size is 32×32. Fig. 1(a) show some ORL images with different percentages of pixel corruption. The Extended Yale B contains 2414 images with the size of 32×32. Fig. 1(b) show some EYaleB images with different levels of block occlusion.

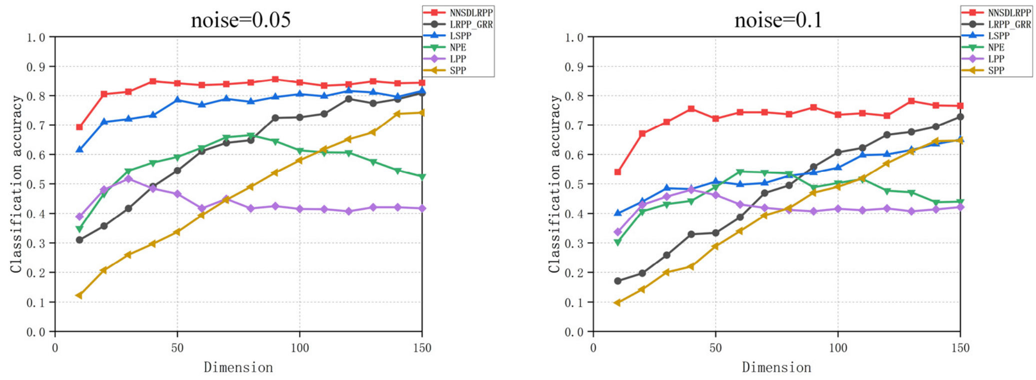
(a) The ORL dataset with random pixel corruptions of none, 0.05,0.1

(b) The EYaleB dataset with block occlusions of none, 6×6, 8×8

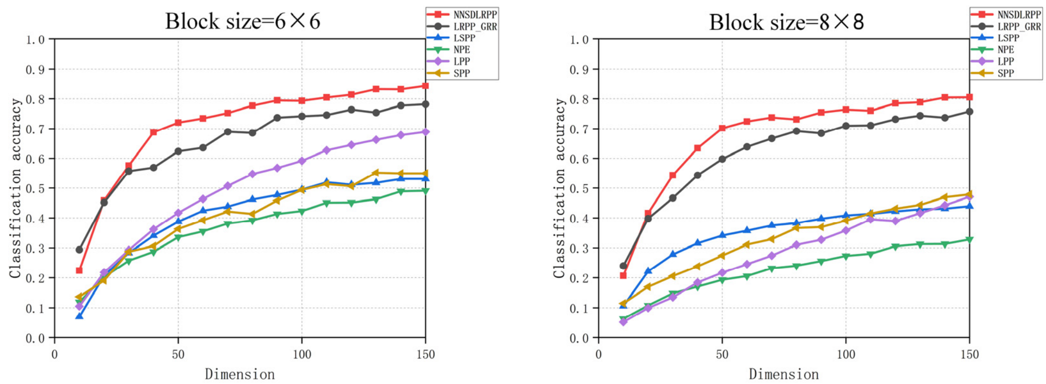
Figure 1. Some corrupted samples from the ORL and EYaleB datasets.

3.2. Experimental Results and Analysis

Figure 2 is the classification accuracy of different methods under different feature dimensions in the ORL dataset and the EYaleB dataset. Other comparison methods use suggested parameters. In random pixel corruptions experiment on the ORL dataset, our algorithm outperforms other algorithms and the highest classification accuracy are 85.6% and 78.1%, respectively. In the block occlusion experiment on the EYaleB dataset, the proposed algorithm achieved the best performance with the highest classification accuracy of 84.3% and 80.5%. In general, the proposed method outperforms other methods with stronger feature extraction capability and robustness.



(a) Classification accuracy in ORL dataset



(b) Classification accuracy in EYaleB dataset

Figure 2. The relationship between features and classification accuracy in two datasets.

3.3. Parameter Sensitiveness

There are 4 parameters to be tuned in our proposed methods, including α,β,γ and λ. In our experiments, we choose them from {10-4,10-3,10-2,10-1,100,101,102} by a grid search manner. The performance variance of each combination of parameter value on the ORL dataset is shown in Fig. 3. For ORL dataset, the optimal parameters are α=10-3, β=10-4, γ =10-3 and λ =10-1. For EYaleB dataset, the optimal parameters are α=10-2, β=10-4, γ=10-4 and λ=10-4.

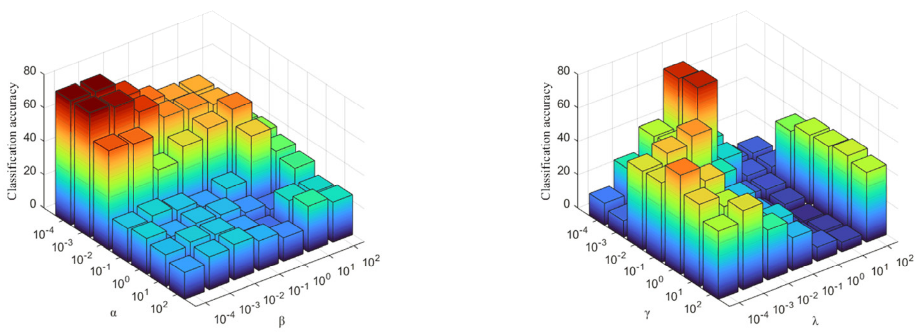


Figure 3. Classification accuracy under different parameter combinations

4. Conclusion

In this paper, we develop a NNSDLRPP model for face feature extraction and classification. In order to get a more discriminant feature distribution, a group-wise regularization is introduced. The non-negative constraint improves the interpretability of representation while the sparse constraint help capture the local information. We develop an alternate algorithm to solve the model. Extensive experiments on two well-known databases have demonstrated the superiority and effectiveness of our method.

References

[1] Lei Chen,Jianjun Xu,Shuang Chen,Hui Yang,Linhu Liu. Power Quality Disturbance Signal Denoising Based on Overcomplete Representation. IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING.2022; 17: 544–555.

[2] Xu Jianjun, Huang Lida, Yan Limei, Yi Na. Insulator Self-Explosion Defect Detection Based on Hierarchical Multi-Task

Deep Learning. Transactions of China Electrotechnical Society, 2021,36(07):1407-1415.

[3] Yan Limei,LIU Yongqiang,XU Jianjun,et al.Broken string diagnosis of composite insulator based on Grabcut segmentation and filler area discrimination.Power System Protection and Control,2021,49(22):114-119.

[4] Yi, Q. Wang, L. Yan, et al., A multi-stage game model for the false data injection attack from attacker's perspective. Sustainable Energy Grids & Networks 28 (2021).

[5] Na Yi,Jianjun Xu,Limei Yan,Lin Huang. Task Optimization and Scheduling of Distributed Cyber-physical System Based on Improved Ant Colony Algorithm. Future Generation Computer Systems, 109(Aug. 2020),134-148.

[6] Yang Zhao, Jianjun Xu, Jingchun Wu. A New Method for Bad Data Identification of Oilfield Power System Based on Enhanced Gravitational Search-Fuzzy C-Means Algorithm. IEEE Transactions on Industrial Informatics. VOL. 15, NO. 11, NOVEMBER 2019 5963-5970.

[7] Cai, J. F., Candès, E. J., & Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. SIAM Journal on optimization, 20(4), 1956-1982.

[8] 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.