

Article

Manifold Regularized Principal Component Analysis Method Using $L_{2,p}$ -Norm

Minghua Wan ^{1,2,3}, Xichen Wang ¹ , Hai Tan ^{1,*} and Guowei Yang ^{1,4}¹ School of Information Engineering, Nanjing Audit University, Nanjing 211815, China² Jiangsu Modern Intelligent Audit Integrated Application Technology Engineering Research Center, Nanjing Audit University, Nanjing 211815, China³ Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210014, China⁴ School of Electronic Information, Qingdao University, Qingdao 266071, China

* Correspondence: 270602@nau.edu.cn

Abstract: The main idea of principal component analysis (PCA) is to transform the problem of high-dimensional space into low-dimensional space, and obtain the output sample set after a series of operations on the samples. However, the accuracy of the traditional principal component analysis method in dimension reduction is not very high, and it is very sensitive to outliers. In order to improve the robustness of image recognition to noise and the importance of geometric information in a given data space, this paper proposes a new unsupervised feature extraction model based on $l_{2,p}$ -norm PCA and manifold learning method. To improve robustness, the model method adopts $l_{2,p}$ -norm to reconstruct the distance measure between the error and the original input data. When the image is occluded, the projection direction will not significantly deviate from the expected solution of the model, which can minimize the reconstruction error of the data and improve the recognition accuracy. To verify whether the algorithm proposed by the method is robust, the data sets used in this experiment include ORL database, Yale database, FERET database, and PolyU palmprint database. In the experiments of these four databases, the recognition rate of the proposed method is higher than that of other methods when $p = 0.5$. Finally, the experimental results show that the method proposed in this paper is robust and effective.

Keywords: principal component analysis; manifold learning; features extracting; $l_{2,p}$ -norm; neighborhood preserving embedding

MSC: 68U10

Citation: Wan, M.; Wang, X.; Tan, H.; Yang, G. Manifold Regularized Principal Component Analysis Method Using $L_{2,p}$ -Norm.

Mathematics **2022**, *10*, 4603. <https://doi.org/10.3390/math10234603>

Academic Editors: Florin Leon, Mircea Hulea and Marius Gavrilescu

Received: 1 November 2022

Accepted: 2 December 2022

Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To solve the problem caused by high dimensions, researchers have summarized many dimensionality reduction methods [1–3], including principal component analysis (PCA) [4] that belongs to unsupervised learning and linear discriminant analysis (LDA) [5] that belongs to supervised learning, and these two methods generally project data from high-dimensional space to low dimensional space first. In order to solve the problem of ignoring the structure information embedded in the pixel when converting the two-dimensional image data into one-dimensional image vector [6], 2DPCA [7] was proposed. Inspired by 2DPCA, 2DLDA [8] and multi-directional principal component analysis (MPCA) [9] have also been proposed one after another. These algorithms can extract more effective features from the image itself.

In recent years, l_1 -norm [10] has been greatly developed, and when the image is noisy, the recognition accuracy of the image is still high [11–15]. To further improve the robustness of subspace learning method, l_p -norm is proposed, and because of it, PCA [16] and LDA [17] are further developed. However, the above methods do not have the purpose

of minimizing the reconstruction error. Therefore, Ding et al. [18] proposed a l_1 -norm rotation invariant algorithm of PCA objective function, which is called rotation invariant l_1 -norm PCA (R_1 -PCA). To further improve the performance of PCA algorithm, $l_{2,p}$ -norm [19] is proposed. Bi et al. [20] proposed locally invariant robust principal component analysis (LIRPCA), which uses $l_{2,p}$ -norm to constrain PCA to solve the problem of underwater image recognition [21]. Although LIRPCA solves the problem of PCA in image reconstruction to a certain extent, it also reduces the influence of large distance as much as possible. However, LIRPCA is difficult to capture the nonlinear structure of manifolds, and there are also some limitations, for example, it is unable to generalize new samples, and its training time is too long.

The above methods can only deal with the dimensionality reduction of linear data. Therefore, in order to solve some nonlinear image data dimensionality reduction problems, scholars have proposed many dimensionality reduction methods that can solve nonlinear problems, and manifold learning [22] is one of them. Isometric mapping (Isomap) [23] and laplacian eigenmaps (LE) [24], which belongs to classical manifold learning methods, can learn some nonlinear manifold structures, but these methods lack the ability of generalization, in other words, it means that these algorithms have weak adaptability to new sample databases. Locally linear embedding (LLE) [25] and neighborhood preserving embedding (NPE) [26,27] based on manifold learning [28,29] solve this problem well. As a linear approximation of LLE, NPE has a very good effect on image dimensionality reduction and is easy to process new image samples. A manifold regularization is used to consider non-linearity, so kernel PCA (KPCA) [30], which is another popular extension of PCA that considers non-linearity, is proposed.

As we all know, images will be affected by various interferences in the process of recognition, such as occlusion, blurring, etc. First of all, in order to extract important features of an image, this paper improves the PCA algorithm, and proposes a new principal component analysis method called manifold regularized principal component analysis method using $l_{2,p}$ -norm ($l_{2,p}$ -MRPCA). This method uses $l_{2,p}$ -norm to reconstruct the distance measurement between the error and the original input data. If the noise of the experimental data is relatively large, there is no obvious deviation between the expected projection direction and the desired solution of $l_{2,p}$ -MRPCA, so as to minimize the reconstruction error of the data and improve the recognition accuracy. Secondly, in order to improve the modeling performance, manifold regularization terms are used. Manifold learning shows that observations are always collected from low dimensional manifolds embedded in high-dimensional environment space. $l_{2,p}$ -MRPCA is a generalized robust metric learning method of PCA, and this method not only has strong robustness to outliers, but also maintains the good characteristics of PCA. Finally, the structure of $l_{2,p}$ -MRPCA is relatively simple, belonging to unsupervised subspace learning algorithm, and the ability of model learning task is high. This paper mainly contains the following three contributions:

1. A new algorithm based on PCA is proposed. The model adopts $l_{2,p}$ -norm as the function measure, which is a robust model.
2. This method combines the advantages of regularization and manifold learning, and has higher robustness and recognition effect.
3. In the non greedy iterative algorithm, the weighted covariance matrix is considered to further reduce the reconstruction error.

The following has four sections. Section 2 mainly presents the algorithms which are related to this paper, including PCA, R_1 -PCA, NPE, and LIRPCA. Section 3 mainly presents the objective function, algorithm optimization, and algorithm flow of $l_{2,p}$ -MRPCA. Section 4 analyses experimental comparisons on the ORL, Yale, FERET, and PolyU palm-print databases. Section 5 summarizes the full text.

2. Related Work

The related work includes the definition of the normal form mentioned in the paper and some related algorithms, such as PCA, R_1 -PCA, NPE and LIRPCA.

2.1. Symbols and Definitions

Let the data set $X = (x_1, x_2, \dots, x_n)$ to represent a standardized training sample matrix, which contains n samples, and each sample is an m dimensional column vector. In this paper, l_2 -norm, R_1 -norm and $l_{2,p}$ -norm are adopted. The definition of l_2 -norm is given as follows:

$$\|X\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} \tag{1}$$

R_1 -norm is defined as:

$$\|X\|_{R_1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m |x_{ij}|^2} \tag{2}$$

$l_{2,p}$ -norm is defined as:

$$\|X\|_{2,p} = \sqrt[p]{\sum_{i=1}^n \sqrt[p]{\sum_{j=1}^m |x_{ij}|^2}} \tag{3}$$

2.2. Principal Component Analysis (PCA)

PCA is a common feature extraction algorithm, which is mainly used in image recognition field. Assuming that $U \in R^{m \times q}$ is a projection matrix. This method uses l_2 -norm as constraint, and we can obtain the optimal projection matrix U after finding the solution of the following optimization problem:

$$\min_U \sum_{i=1}^n \|x_i - UU^T x_i\|_2^2 \quad s.t. \quad U^T U = I_q \tag{4}$$

where I_q is a $q \times q$ identity matrix. Through matrix tracing operation, we can convert Equation (4) into:

$$\max_U \sum_{i=1}^n \|UU^T x_i\|_2^2 = \max_U \text{tr}(U^T G_t U) \tag{5}$$

where $G_t = \sum_{i=1}^n x_i(x_i)^T$ is called the image covariance matrix, and the projection matrix U of Equation (4) is composed of G_t eigenvector corresponding to the maximum eigenvalue of q . However, because l_2 -norm is sensitive to noise [31], and its robustness is low, and the iterative process is cumbersome, the traditional PCA method is relatively limited.

2.3. Rotation Invariant L1-PCA (R1-PCA)

In R_1 -norm, we use l_2 -norm to measure spatial dimension and l_1 -norm to calculate the sum of different data points. R_1 -PCA is not sensitive to noise [15], so it is easier to process some blurred images. Here is the specific definition of R_1 -PCA:

$$\min_U \sum_{i=1}^n \|x_i - UU^T x_i\|_{R_1} \quad s.t. \quad U^T U = I_q \tag{6}$$

After a series of optimization iterative algorithms, we can obtain the optimal projection matrix U . However, R_1 -PCA uses l_2 -norm to centralize the training samples, so it can not guarantee that the final calculated mean is optimal, so there is still room for improvement.

2.4. Neighborhood Preserving Embedding (NPE)

The idea of NPE is the same as LLE, which is to keep the local linear structure of manifold unchanged in the process of dimensionality reduction, so as to extract useful information from data. The local linear structure is represented by the reconstruction of the

weight matrix, which is the coefficient matrix of the linear reconstruction of the neighbors to the nodes in the neighborhood.

Similar to other classical manifold learning algorithms, NPE has three steps:

1. Constructing Neighborhood Graph;
2. calculating Weight Matrix;
3. and computational mapping.

In conclusion, we can obtain the objective function of NPE in low dimensional space as follows:

$$\min_U \sum_{i=1}^n \left\| U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right\|_2^2 \quad s.t. \quad U^T X X^T U = I_q \tag{7}$$

where the weight matrix W_{ij} mentioned in Formula (7) can be defined as:

$$\sum_{j=1}^m W_{ij} = 1, i = 1, 2, \dots, n \tag{8}$$

where W_{ij} represents the weight value of the edge from node i to node j . If there is no such edge, the value of W_{ij} is 0.

2.5. Locally Invariant Robust Principal Component Analysis (LIRPCA)

LIRPCA hopes to minimize the deviation between the reconstructed image and the original image of each projection data and further enhance the robustness of the model, so as to ensure that the extracted features can well reflect the main information of the original data space. Therefore, LIRPCA uses $l_{2,p}$ -norm to constrain PCA. In order to recover low-dimensional information from high-dimensional environment space, we hope to find a U that ensures that Ux_k and Ux_j are adjacent. Based on the above objectives, LIRPCA is specifically defined as follows:

$$\min_U \sum_{i=1}^n \left(\frac{\|x_i - UU^T x_i\|_2^p}{\|x_i\|_2^p} + \frac{1}{2} \Psi \sum_{j=1}^m \left\| U^T (x_i - x_j) \right\|_2^2 W_{ij} \right) \tag{9}$$

where $\Psi > 0$, and W_{ij} is a weight matrix which can be defined as:

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right), & \text{if } x_i \in M_h(x_j), \\ \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right), & \text{if } x_j \in M_h(x_i), \\ 0 & , \text{ otherwise,} \end{cases} \tag{10}$$

where $\sigma > 0$, and $M_h(x_j)$ is the set of k nearest data of x_i , $M_h(x_i)$ is the set of k nearest data of x_j and W_{ij} represents the i -th, and the j -th column of the matrix W .

3. Manifold Regularized PCA Method Using $l_{2,p}$ -norm ($l_{2,p}$ -MRPCA)

This chapter mainly includes the definition of $l_{2,p}$ -MRPCA and its algorithm optimization process and convergence analysis.

3.1. Motivation and Objective Function

In order to reduce the influence of large distance as a measure and minimize the reconstruction error, combining with the LIRPCA mentioned above, we use $l_{2,p}$ -norm instead of l_2 -norm, and propose $l_{2,p}$ -PCA as follows:

$$\min_U \sum_{i=1}^n \frac{\|x_i - UU^T x_i\|_2^p}{\|x_i\|_2^p} \quad \text{s.t. } U^T U = I_q \tag{11}$$

where p is $0 < p < 2$. By solving this constrained optimization problem, the optimal projection matrix U will be obtained.

However, considering the importance of considering the internal geometric information of data space to improve the performance of the algorithm and ensuring the rotation invariance of the data of the algorithm, popular learning, such as NPE, can be applied to this method. The specific formula of NPE is shown in Formula (7) mentioned above.

To sum up, combining Equations (4) and (11), we can obtain the following objective function:

$$\min_U \sum_{i=1}^n \frac{\|x_i - UU^T x_i\|_2^p}{\|x_i\|_2^p} + \phi \sum_{i=1}^n \left\| U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right\|_2^2 \quad \text{s.t. } U^T U = I_q \tag{12}$$

where $\phi > 0$.

3.2. Optimization

Formula (12) is divided into two parts: $\sum_{i=1}^n \frac{\|x_i - UU^T x_i\|_2^p}{\|x_i\|_2^p}$ and $\sum_{i=1}^n \left\| U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right\|_2^2$.

First, we simplify the $\sum_{i=1}^n \frac{\|x_i - UU^T x_i\|_2^p}{\|x_i\|_2^p}$ part.

$$\begin{aligned} & \sum_{i=1}^n \frac{\|x_i - UU^T x_i\|_2^p}{\|x_i\|_2^p} \\ &= \sum_{i=1}^n \frac{\|x_i - UU^T x_i\|_2^2 \|x_i - UU^T x_i\|_2^{p-2}}{\|x_i\|_2^p} \\ &= \sum_{i=1}^n \text{tr} \left[(x_i - UU^T x_i)^T (x_i - UU^T x_i) \right] q_i \\ &= \sum_{i=1}^n \text{tr} \left[\left((x_i)^T - (x_i)^T UU^T \right) (x_i - UU^T x_i) \right] q_i \\ &= \sum_{i=1}^n \text{tr} \left[(x_i)^T x_i - (x_i)^T UU^T x_i \right] q_i \\ &= \text{tr}(XDX^T) - \text{tr}(U^T XDX^T U) \end{aligned} \tag{13}$$

where $q_i = \frac{\|x_i - UU^T x_i\|_2^{p-2}}{\|x_i\|_2^p}$ and D is a diagonal matrix whose elements on diagonal are q_i .

Then, we simplify the $\sum_{i=1}^n \left\| U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right\|_2^2$ part.

$$\begin{aligned} & \sum_{i=1}^n \left\| U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right\|_2^2 \\ &= \sum_{i=1}^n \left(U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right)^T \cdot \left(U^T x_i - \sum_{j=1}^m W_{ij} \cdot U^T x_j \right) \\ &= \text{tr} \left(U^T X(I - W)^T (I - W) X^T U \right) \\ &= \text{tr} \left(U^T X M X^T U \right) \end{aligned} \tag{14}$$

where I is a $q \times q$ identity matrix.

Finally, Equations (13) and (14) are combined and we obtain the equation:

$$\min_U \sum_{i=1}^n \text{tr}(XDX^T) - \text{tr}(U^T XDX^T U) + \lambda \text{tr}(U^T XMX^T U) \tag{15}$$

where λ is a regularization parameter which should be set to a small real value.

3.3. Algorithm Optimization

Since the unknown variables U and D have a certain relationship with U , it is difficult to directly solve the optimal projection matrix U . However, in this case, we can use non greedy iterative algorithm to solve U and D . The Lagrangian function of Equation (15) is

$$L(U, \xi) = \text{tr}(XDX^T) - \text{tr}(U^T XDX^T U) + \lambda \text{tr}(U^T XMX^T U) + \text{tr}(\xi(U^T U - I)) \tag{16}$$

where $\xi \in R^{d \times d}$ is a symmetric matrix. Then we can apply the Karush–Kuhn–Tucker (KKT) condition to find the projection matrix. We set $\frac{\partial L(U, \xi)}{\partial U} = 0$, then,

$$\begin{aligned} \frac{\partial L(U, \xi)}{\partial U} &= \frac{\partial \text{tr}(XDX^T)}{\partial U} - \frac{\partial \text{tr}(U^T XDX^T U)}{\partial U} + \lambda \frac{\partial \text{tr}(U^T XMX^T U)}{\partial U} \\ &\quad + \frac{\partial \text{tr}(\xi(U^T U - I))}{\partial U} \\ &= 0 - (XDX^T U + (U^T XDX^T)^T) + \lambda (XMX^T U + (U^T XMX^T)^T) \\ &\quad + \xi(U + (U^T)^T) \\ &= -2XDX^T U + 2\lambda XMX^T U + 2U\xi \\ &= 0 \end{aligned} \tag{17}$$

and Equation (17) can be converted into

$$(XDX^T - \lambda XMX^T)U = U\xi \tag{18}$$

We set $\frac{\partial L(U, \xi)}{\partial \xi} = 0$, then,

$$U^T U = I_q \tag{19}$$

We can substitute Equations (18) and (19) into Equation (15), and the projection matrix U satisfies the objective function can be obtained. Algorithm 1 gives the whole flow of U and q_i calculation.

Algorithm 1. $l_{2,p}$ -MRPCA

Input: Training set X , iterations T , parameters $\lambda, p, q, t = 1$

Output: $U^{(t+1)} \in R^{m \times q}$

Compute: $W \in R^{n \times n}, D \in R^{n \times n}$ and $M \in R^{n \times n}$ where $M = I - W$

Initialize: $U^{(t)}$ to a $m \times q$ orthogonal matrix

Repeat:

1. compute the diagonal matrix D by each diagonal element q_i .
2. Compute the weighted covariance matrix $XDX^T - \lambda XMX^T$
3. Update matrix $U^{(t+1)}$ which is called the optimal projection matrix by Equation (14).
4. If $J(U^{(t)}) - J(U^{(t+1)}) \leq \delta$ (δ is a small positive real number, such as 10^{-8}), where $J(U) = \text{tr}(XDX^T) - \text{tr}(U^T XDX^T U) + \lambda \text{tr}(U^T XMX^T U)$
5. $t \leftarrow t + 1$

Output the optimal projection matrix $U^{(t+1)}$, and the Algorithm 1 ends.

Theorem 1. Let any two vectors $e^t \in R^m, e^{t+1} \in R^m$, if $0 < p < 2$, we can obtain the following inequality:

$$\frac{\|e^{(t+1)}\|_2^p}{\|e^{(t)}\|_2^p} - \frac{p}{2} \frac{\|e^{(t+1)}\|_2^2}{\|e^{(t)}\|_2^2} - 1 + \frac{p}{2} \leq 0 \tag{20}$$

where e^t must be a non-zero vector, otherwise the denominator is zero, and the inequality is meaningless.

Proof of Theorem 1. Let $f(y) = y^p - \frac{p}{2}y^2 + \frac{p}{2} - 1$, through simple algebraic calculation, we can obtain:

$$\frac{\partial f}{\partial y} = py(y^{p-2} - 1) \tag{21}$$

It can be seen from Equation (21) that when $y > 0$ and $0 < p < 2$, $y = 1$ is the only extreme optimal solution of function f . In addition, we have $f'(y) > 0$ ($0 < y < 1$) and $f'(y) < 0$ ($1 < y$). So $y = 1$ is the maximum point of function f . Substitute $y = 1$ into the function y to obtain $f = 0$.

Combined with the previous analysis, we obtain that for any $y > 0$, $f(y) \leq 0$, Theorem 1 can be proved by setting $y = \frac{\|e^{t+1}\|_2}{\|e^t\|_2}$. □

Theorem 2. By using the iterative method which is described in Algorithm 1, we can obtain that the value of Equation (12) decreases monotonically in each iteration until it converges to the local optimum.

Proof of Theorem 2. As shown in Algorithm 1, in the $t + 1$ iteration, we have:

$$\begin{aligned} & \sum_{i=1}^n tr\left((x_i)^T x_i q_i^{(t)}\right) - \sum_{i=1}^n tr\left(\left(U^{(t+1)}\right)^T x_i (x_i)^T U^{(t+1)} q_i^{(t)}\right) \\ & + \lambda tr\left(\left(U^{(t+1)}\right)^T XMX^T U^{(t+1)}\right) \\ & \leq \sum_{i=1}^n tr\left((x_i)^T x_i q_i^{(t)}\right) - \sum_{i=1}^n tr\left(\left(U^{(t)}\right)^T x_i (x_i)^T U^{(t)} q_i^{(t)}\right) \\ & + \lambda tr\left(\left(U^{(t)}\right)^T XMX^T U^{(t)}\right) \end{aligned} \tag{22}$$

Equation (22) can be transformed into:

$$\begin{aligned} & \sum_{i=1}^n \left\| x_i - U^{(t+1)} \left(U^{(t+1)} \right)^T x_i \right\|_2^2 q_i^{(t)} + \lambda tr\left(\left(U^{(t+1)}\right)^T XMX^T U^{(t+1)}\right) \\ & \leq \sum_{i=1}^n \left\| x_i - U^{(t)} \left(U^{(t)} \right)^T x_i \right\|_2^2 q_i^{(t)} + \lambda tr\left(\left(U^{(t)}\right)^T XMX^T U^{(t)}\right) \end{aligned} \tag{23}$$

Assuming that $e_i^{(t+1)} = x_i - U^{(t+1)}(U^{(t+1)T})x_i$, $e_i^{(t)} = x_i - U^{(t)}(U^{(t)T})x_i$ and $v^{(t)} = x_i$. As we already know that $q_i = \frac{\|x_i - UU^T\|_2^{p-2}}{\|x_i\|_2^p}$, so Equation (23) can be converted into

$$\begin{aligned} & \sum_{i=1}^n \frac{\|e_i^{(t+1)}\|_2^2}{\|v_i^{(t)}\|_2^p \|e_i^{(t)}\|_2^2} \|e_i^{(t)}\|_2^p + \lambda tr\left(\left(U^{(t+1)}\right)^T XMX^T U^{(t+1)}\right) \\ & \leq \sum_{i=1}^n \frac{\|e_i^{(t)}\|_2^p}{\|v_i^{(t)}\|_2^p} + \lambda tr\left(\left(U^{(t)}\right)^T XMX^T U^{(t+1)}\right) \end{aligned} \tag{24}$$

Equation (24) can be transposition into

$$\begin{aligned} & \sum_{i=1}^n \frac{\|e_i^{(t+1)}\|_2^2}{\|v_i^{(t)}\|_2^p \|e_i^{(t)}\|_2^2} \|e_i^{(t)}\|_2^p \\ & \leq \sum_{i=1}^n \frac{\|e_i^{(t)}\|_2^p}{\|v_i^{(t)}\|_2^p} + \lambda \operatorname{tr} \left(\left(U^{(t)} \right)^T X M X^T U^{(t)} \right) - \lambda \operatorname{tr} \left(\left(U^{(t+1)} \right)^T X M X^T U^{(t+1)} \right) \end{aligned} \tag{25}$$

According to the properties of Theorem 1, we multiply $\frac{1}{\|v_i^{(t)}\|_2^p} > 0$ on both sides of Equation (20) to obtain the inequality of each index i :

$$\sum_{i=1}^n \frac{p}{2} \frac{\|e_i^{(t+1)}\|_2^2}{\|v_i^{(t)}\|_2^p \|e_i^{(t)}\|_2^2} \|e_i^{(t)}\|_2^p \geq \sum_{i=1}^n \frac{\|e_i^{(t+1)}\|_2^p}{\|v_i^{(t)}\|_2^p} - \sum_{i=1}^n \frac{\|e_i^{(t)}\|_2^p}{\|v_i^{(t)}\|_2^p} + \sum_{i=1}^n \frac{p}{2} \frac{\|e_i^{(t)}\|_2^p}{\|v_i^{(t)}\|_2^p} \tag{26}$$

Then, we multiply the whole of Equation (25) by $\frac{p}{2}$ and substitute it into Equation (26), and we obtain

$$\begin{aligned} & \sum_{i=1}^n \frac{\|e_i^{(t+1)}\|_2^p}{\|v_i^{(t)}\|_2^p} - \frac{p}{2} \lambda \operatorname{tr} \left(\left(U^{(t)} \right)^T X M X^T U^{(t)} \right) \\ & \leq \sum_{i=1}^n \frac{\|e_i^{(t)}\|_2^p}{\|v_i^{(t)}\|_2^p} - \frac{p}{2} \lambda \operatorname{tr} \left(\left(U^{(t+1)} \right)^T X M X^T U^{(t+1)} \right) \end{aligned} \tag{27}$$

We substitute $e_i^{(t+1)} = x_i - U^{(t+1)}(U^{(t+1)T})x_i$, $e_i^{(t)} = x_i - U^{(t)}(U^{(t)T})x_i$ and $v^{(t)} = x_i$ into Equation (27), and we can obtain

$$\begin{aligned} & \sum_{i=1}^n \frac{\|x_i - U^{(t+1)}(U^{(t+1)T})x_i\|_2^p}{\|x_i\|_2^p} + \frac{p}{2} \lambda \operatorname{tr} \left(\left(U^{(t+1)} \right)^T X M X^T U^{(t+1)} \right) \\ & \leq \sum_{i=1}^n \frac{\|x_i - U^{(t)}(U^{(t)T})x_i\|_2^p}{\|x_i\|_2^p} + \frac{p}{2} \lambda \operatorname{tr} \left(\left(U^{(t)} \right)^T X M X^T U^{(t)} \right) \end{aligned} \tag{28}$$

Note that $0 < p < 2$, so $\frac{p}{2} \lambda > 0$ is true. Finally, ensuring that $\xi = \frac{p}{2} \lambda$ is established, and combine Equation (28) with Equation (15) to obtain Equation (28):

$$\begin{aligned} & \sum_{i=1}^n \frac{\|x_i - U^{(t+1)}(U^{(t+1)T})x_i\|_2^p}{\|x_i\|_2^p} + \xi \operatorname{tr} \left(\left(U^{(t+1)} \right)^T X M X^T U^{(t+1)} \right) \\ & \leq \sum_{i=1}^n \frac{\|x_i - U^{(t)}(U^{(t)T})x_i\|_2^p}{\|x_i\|_2^p} + \xi \operatorname{tr} \left(\left(U^{(t)} \right)^T X M X^T U^{(t)} \right) \end{aligned} \tag{29}$$

Equation (29) shows that the objective function of Equation (12) decreases monotonically in each iteration. Combining the convergence conditions given by Algorithm 1, it can be determined that the objective function (12) has a lower bound, and finally converges to the local optimal solution, so Theorem 2 is true. □

4. Experiments

The experiment part mainly includes the introduction of several databases, the presentation of the experimental results on each database, and the analysis of the experimental results. The whole experimental analysis is carried out under the windows system which is configured with i5-1035G1 processor, 8G memory, PCI-E 1T solid state disk, and MX250 2G single display. All codes are compiled by using matlab tools.

4.1. Data Sets and Experimental Parameters

In order to verify the effectiveness of $l_{2,p}$ -MRPCA algorithm, this experiment compares $l_{2,p}$ -MRPCA with PCA, R_1 -PCA, KPCA, NPE, and LIRPCA. The databases used in this experiment include ORL face database, YALE face database and FERET face database, and PolyU palmprint database. In order to verify the robustness of the algorithm under

different levels of occlusion, we add 5×5 occlusion block and 10×10 occlusion block to ORL face database and YALE face database respectively, and 5×5 occlusion block to the FERET face database. The original images and continuous occlusion images of the four libraries are shown in Figure 1. The ORL face database randomly selects training samples $n = 3, 4, 5, 6$, YALE face database randomly selects training samples $n = 4, 5$, and FERET face database randomly selects training samples $n = 2, 3, 4, 5$.

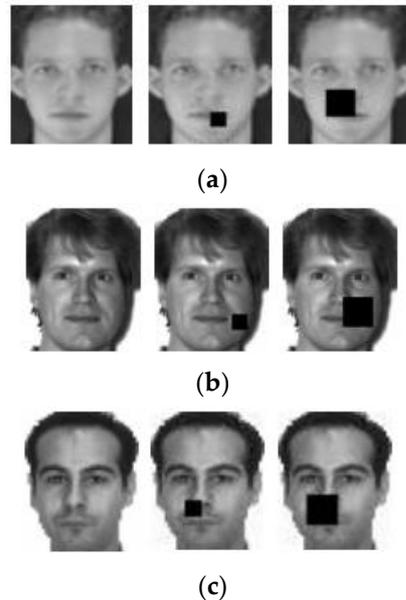


Figure 1. Partial original image and continuous occlusion image on ORL, YALE, and FERET database (a) ORL database (b) YALE database (c) FERET database.

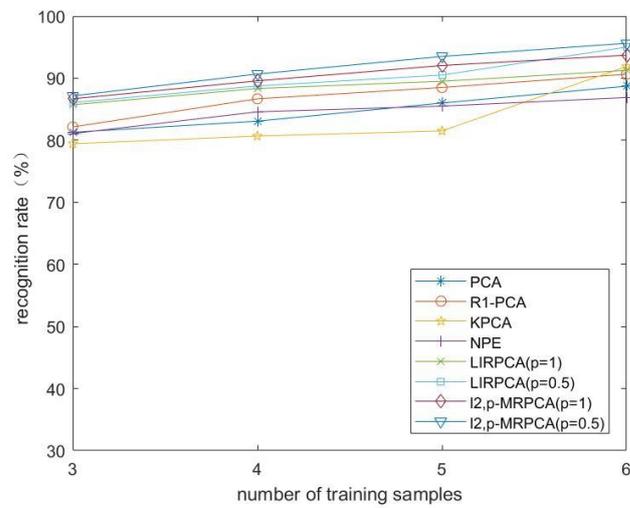
For the parameters mentioned in $l_{2,p}$ -MRPCA algorithm, λ , p , and q are briefly described. We select the optimal parameters of $l_{2,p}$ -MRPCA by crossing validation strategy, and set parameters $\lambda = 0.1$ in the ORL face database, parameters $\lambda = 0.08$ in YALE face database, parameters $\lambda = 0.05$ in FERET face database. Parameter p is chosen as 0.5 and 1, and the two parameters values are substituted into the experiment to obtain the experimental results, so as to select better parameter values. The parameter q represents the number of extracted feature information, which can be determined empirically through Cumulative Percent Variance (CPV), and its formula is as follows:

$$CPV = \left[\frac{\sum_{i=1}^q \lambda_i}{\sum_{j=1}^m \lambda_j} \right] \times 100\% \rightarrow 90\% \tag{30}$$

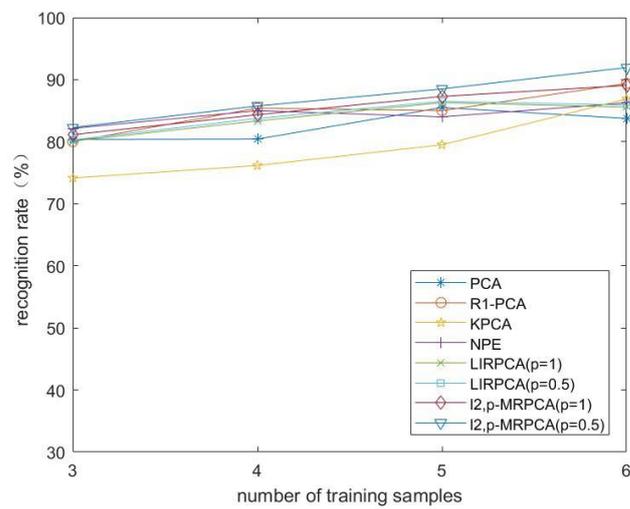
In order to ensure that CPV can reach 90% during the experiment, the corresponding q value is selected. In order to ensure the universality of the experimental results, the experiments in each database are repeated at least 100 times.

4.2. The ORL Face Database

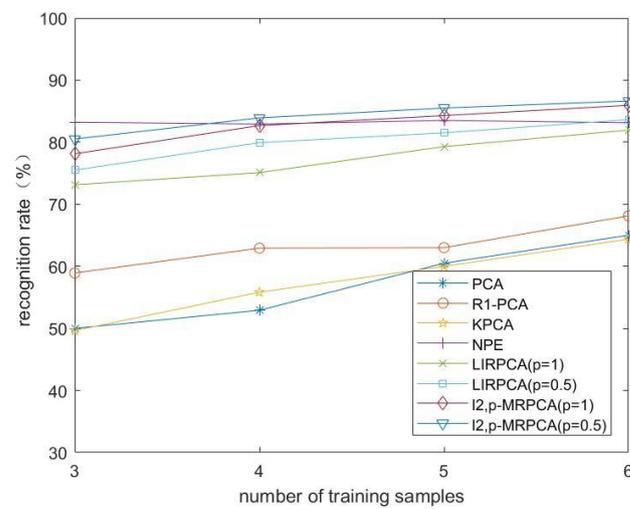
The database has 400 images, including 40 people with 10 images, and each image is 56×46 pixels. The shooting background of these images is relatively dark, which is the front face collected in different time, light, facial expression, and facial detail environment (some images have slight deviation). We obtain the broken line diagram of the recognition rate of ORL database and its occlusion images in PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 1$), LIRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$), and $l_{2,p}$ -MRPCA ($p = 0.5$), as shown in Figure 2.



(a)



(b)



(c)

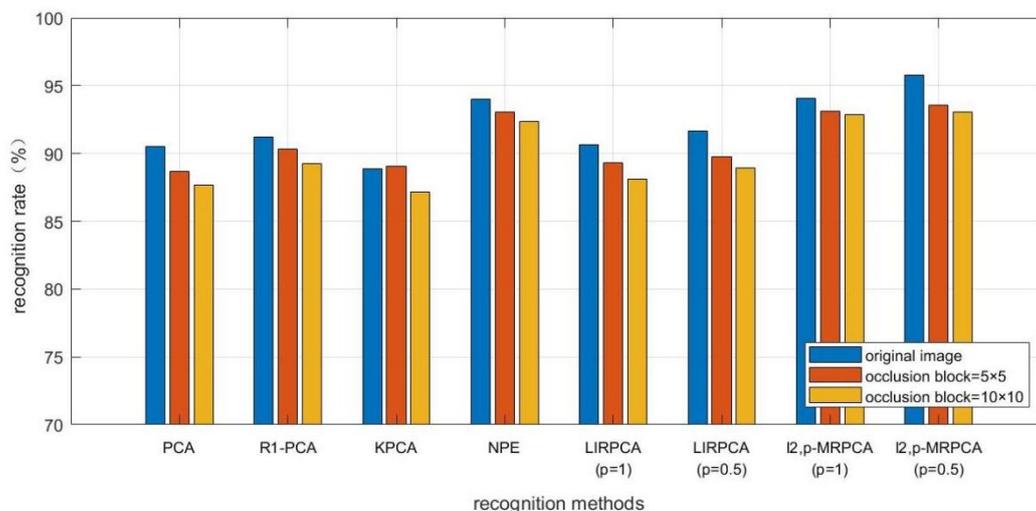
Figure 2. Recognition rate of PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 0.5$), LIRPCA ($p = 1$), $l_{2,p}$ -MRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$) on ORL database (a) original image (b) occlusion block = 5×5 (c) occlusion block = 10×10 .

First of all, it can be seen from Figure 2 that with the increase of the number of training samples, the recognition rates of PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 1$), LIRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$), and $l_{2,p}$ -MRPCA ($p = 0.5$) also increase. Secondly, we compare the robustness of these algorithms. With the increase size of occlusion block, the recognition rate of NPE is improved, which shows that the robustness of NPE algorithm is relatively high, and it is suitable for recognizing occluded images. However, PCA, R_1 -PCA, and KPCA reduce the recognition rate with the increase of occluded block size, which indicates that the two methods are not suitable for recognizing occluded images. Finally, the effect of parameter p on the experiment was observed. It can be seen from Figure 2 that the recognition rate of $l_{2,p}$ -MRPCA is higher than that of LIRPCA when the number of training samples is the same, no matter whether the picture is occluded or not, no matter $p = 0.5$ or $p = 1$. Moreover, the recognition effect of $l_{2,p}$ -MRPCA ($p = 0.5$) is higher than that of $l_{2,p}$ -MRPCA ($p = 1$), which indicates that the value of p also has some influence on the recognition rate.

4.3. The Yale Face Database

The face dataset contains 15 volunteers with 11 images, and each image is 80×100 pixels. The shooting background of these images has more obvious changes in illumination, facial expression, posture, and occlusion than ORL face database. We obtain the histogram of the recognition rate of Yale database and its occlusion images in PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 1$), LIRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$), and $l_{2,p}$ -MRPCA ($p = 0.5$), as shown in Figure 3.

First of all, the pixels of YALE is 80×100 where the pixels of ORL is 56×46 , so YALE has a higher recognition rate than ORL. The reason may be that the shooting background of ORL database is dark, while that of YALE database is bright. It may also be because YALE database has high pixels. Secondly, in this experiment, the recognition rate of LIRPCA is only slightly higher than that of PCA, or even lower than that of NPE. This may be because LIRPCA is not able to capture the linear structure of manifolds. However, the recognition rate of $l_{2,p}$ -MRPCA is still relatively high, which indicates that even if the methods are based on $l_{2,p}$ -PCA, different regularization terms have a greater impact on the experimental results. Finally, in the YALE experiment, the recognition rate of $l_{2,p}$ -MRPCA ($p = 1$) is higher than that of LIRPCA ($p = 0.5$) when the training samples are the same, regardless of whether the pictures are occluded or not. This shows that the robustness and stability of $l_{2,p}$ -MRPCA are higher than that of LIRPCA. Therefore, the introduction of popular regularization in $l_{2,p}$ -MRPCA has certain advantages.



(a)

Figure 3. Cont.

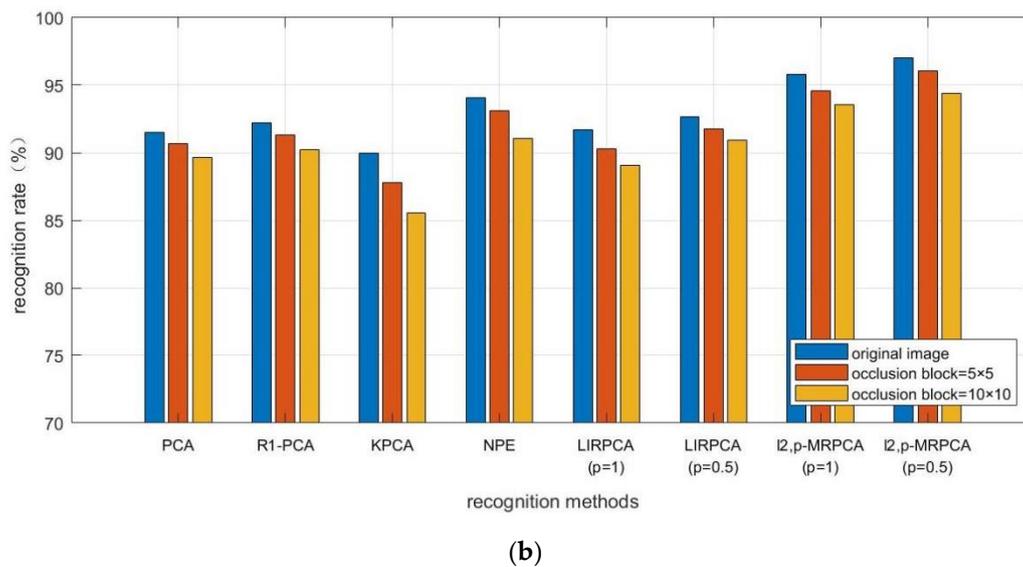
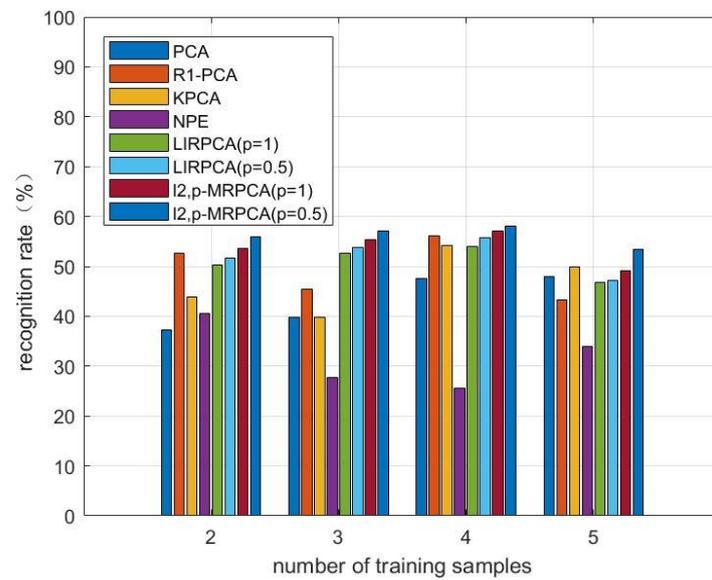


Figure 3. Recognition rate of PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 0.5$), LIRPCA ($p = 1$), $l_{2,p}$ -MRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$) on YALE database (a) Number of training samples $n = 4$ (b) Number of training samples $n = 5$.

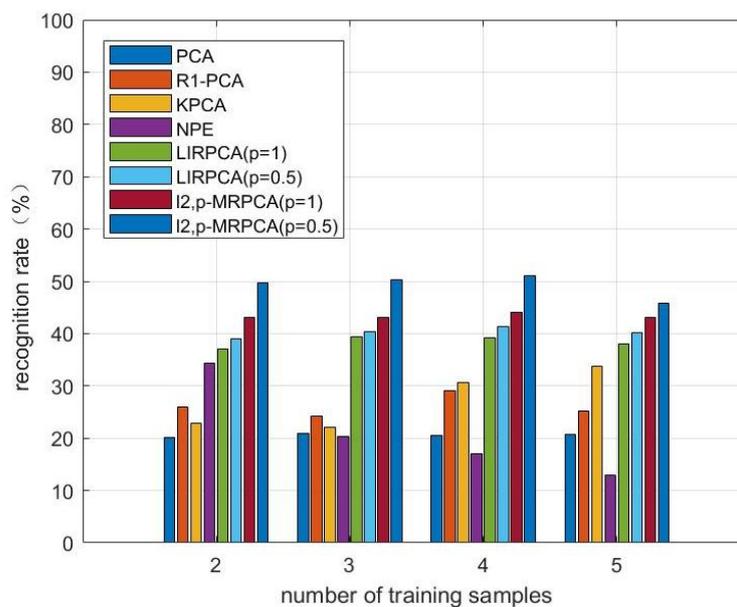
4.4. The FERET Face Database

There are 1400 images in this face dataset, including 200 people, and 7 images for each person, and each image is adjusted to 40×40 pixels. These images are collected under different illumination, facial expression, posture, and age. Most of the subjects are westerners, and the changes of face images contained by each person are relatively single. We obtain the original image of FERET database, and the database when block size is 5×5 . The histogram of the recognition rate of the pictures of occlusion blocks in PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 1$), LIRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$), and $l_{2,p}$ -MRPCA ($p = 0.5$) is shown in Figure 4.

First of all, from the results, the recognition rate of the database is relatively low, which may be because the database has more people but fewer images for each person, insufficient training samples, or the database image pixel is low. Since the recognition rate on the original data is low, the experiment is only carried out on the original database and 5×5 occlusion block. Secondly, on FERET database, the recognition rate of NPE is relatively low, but it is relatively high on ORL database and YALE database. This may be because the stability of NPE is not very strong, so the recognition rate varies greatly on different databases. The recognition rate of R_1 -PCA decreases suddenly when the training sample is 3, and increases suddenly when the training sample is 4. Combined with previous experiments, it may be because of some errors in the experimental process, or because the stability of R_1 -PCA is not very strong. Third, KPCA performs better on FERET database than on ORL database and YALE database, and it is greatly affected by the database and the number of training samples. Finally, the recognition rate of most methods in this experiment is lower when the training sample number is 5 than when the training sample number is 4, which is related to the number of each sample on FERET database. As the results of the previous two experiments, when the number of training samples is the same, the recognition rate of LIRPCA ($p = 0.5$) is higher than that of LIRPCA ($p = 1$), the recognition rate of $l_{2,p}$ -MRPCA ($p = 0.5$) is higher than that of $l_{2,p}$ -MRPCA ($p = 1$), which once again shows that the recognition rate effect of $p = 0.5$ is higher than that of $p = 1$.



(a)



(b)

Figure 4. Recognition rate of PCA, R_1 -PCA, KPCA, NPE, LIRPCA($p = 0.5$), LIRPCA($p = 1$), $l_{2,p}$ -MRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$) on FERET database (a) original image (b) block size = 5×5 .

4.5. The PolyU Palmprint Verification Experiment

There are 600 images in this database, including the palmprint of 100 people. Each person has 6 images, and each image is cut into 50×40 pixels. To better verify the robustness of $l_{2,p}$ -MRPCA, we add 5×5 , 10×10 and 20×20 occlusion blocks to the database as shown in Figure 5, and three palmprint pictures of each person are selected as training samples. Finally, the average recognition accuracy of each algorithm when the number of training samples $n = 3$ can be obtained as shown in Table 1, the training time on PolyU palmprint database is shown in Figure 6, classification recognition rate on PolyU palmprint database is shown in Figure 7.

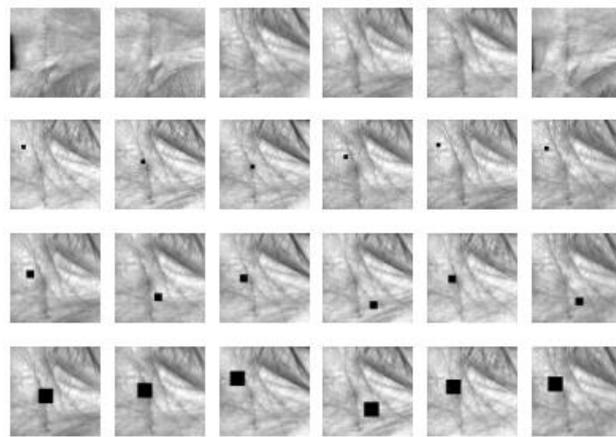


Figure 5. Partial original image and continuous occlusion image on the PolyU palmprint database.

Table 1. Experimental results of recognition rate (standard deviation)(%) on PolyU palmprint database.

TN	3			
Occlusion Block Size	None	5 × 5	10 × 10	20 × 20
PCA/%	76.32 (0.53)	67.32 (0.51)	61.98 (0.49)	43.12 (0.22)
R1-PCA/%	79.82 (0.42)	70.14 (0.44)	65.75 (0.53)	53.54 (0.35)
KPCA/%	77.23 (0.12)	73.04 (0.50)	58.33 (0.49)	40.98 (0.16)
NPE/%	83.06 (0.42)	71.17 (0.44)	67.38 (0.47)	54.17 (0.54)
LIRPCA ($p = 1$)/%	81.78 (0.27)	72.00 (0.33)	65.47 (0.35)	57.10 (0.41)
LIRPCA ($p = 0.5$)/%	86.53 (0.32)	73.15 (0.35)	66.56 (0.31)	57.64 (0.29)
$l_{2,p}$ -MRPCA ($p = 1$)/%	86.91 (0.30)	73.03 (0.38)	68.96 (0.32)	61.82 (0.07)
$l_{2,p}$ -MRPCA ($p = 0.5$)/%	89.01 (0.33)	74.54 (0.36)	69.34 (0.35)	63.01 (0.43)

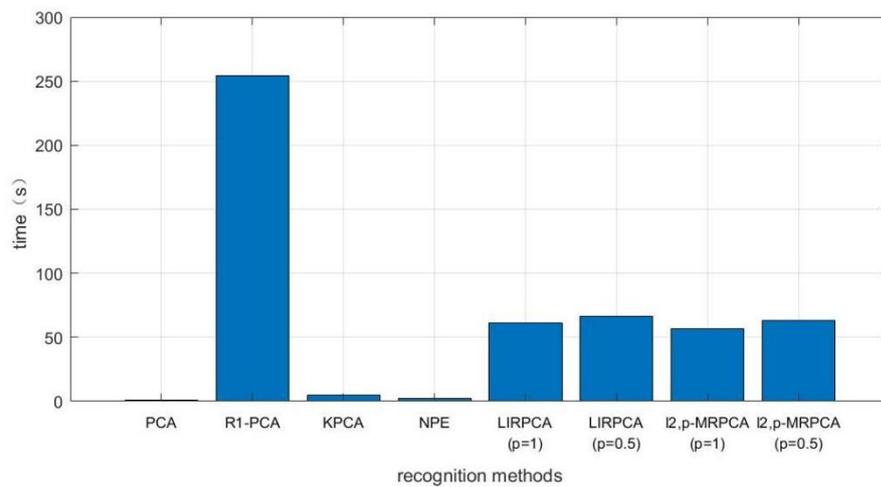


Figure 6. Training time on the PolyU palmprint database.

First, it can be seen from Table 1 that $l_{2,p}$ -MRPCA ($p = 0.5$) has the best recognition effect. With the increase of the number of occluded blocks, the recognition rate of PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 1$), LIRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$), and $l_{2,p}$ -MRPCA ($p = 0.5$) decreases gradually. Secondly, when the picture has no occlusion block, the recognition rate of $l_{2,p}$ -MRPCA is only slightly higher than that of other algorithms. However, as the size of the occlusion block becomes higher, the advantages of $l_{2,p}$ -MRPCA become larger, which shows that the method is robust and suitable for recognizing noisy images. Compared to PCA, R_1 -PCA, and KPCA, LIRPCA has a certain recognition effect, but the recognition rate is always lower than that of $l_{2,p}$ -MRPCA. Although LIRPCA, $l_{2,p}$ -MRPCA all use $l_{2,p}$ -PCA with good robustness, the regularization

term of LIRPCA lacks a certain normalization ability, so the recognition effect is not as good as that of $l_{2,p}$ -MRPCA. Finally, in general, $l_{2,p}$ -MRPCA has good recognition effect and high robustness.

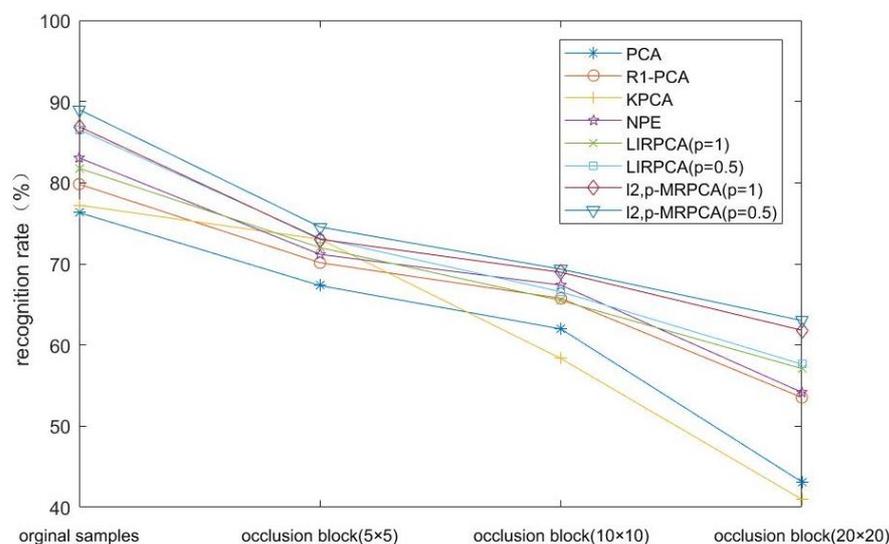


Figure 7. Classification recognition rate of PCA, R_1 -PCA, KPCA, NPE, LIRPCA ($p = 0.5$), LIRPCA ($p = 1$), $l_{2,p}$ -MRPCA ($p = 0.5$), $l_{2,p}$ -MRPCA ($p = 1$) on the PolyU palmprint database.

4.6. Result Analysis

1. From the experimental results, R_1 -PCA, as an improved algorithm of PCA algorithm, has a high recognition rate both in the original image database and in the occluded database.
2. When the experimental data is occluded, NPE, as a manifold learning method, most of the recognition rates are higher than PCA, indicating that the algorithm is less affected by occlusion. When the image is occluded, the recognition effect is better.
3. Compared to LIRPCA, $l_{2,p}$ -MRPCA introduces manifold learning method, so the recognition rate is more significant. In addition, it takes into account the advantages of manifold regularization when the image is occluded, so the recognition effect is better. As the clarity of each database in the experiment is different, the recognition rate made by different databases is relatively different.
4. The training time on $l_{2,p}$ -MRPCA is longer than PCA, KPCA and NPE, and is shorter than R_1 -PCA and LIRPCA. Considering the recognition rate, robustness, and algorithm time of the algorithm, the training time on $l_{2,p}$ -MRPCA is acceptable.
5. The parameter p also has a certain impact on the recognition effect. Whether it is LIRPCA or $l_{2,p}$ -MRPCA, the recognition efficiency is slightly higher when $p = 0.5$ than when $p = 1$.

In order to further verify the stability of $l_{2,p}$ -MRPCA, the algorithm is tested on PolyU palmprint dataset. The results show that even in the case of occlusion, $l_{2,p}$ -MRPCA can still have a high recognition rate, so it shows that the algorithm has good robustness.

5. Conclusions

In this paper, we propose a manifold regularization principal component analysis method by using $l_{2,p}$ -norm constraints. This method effectively combines $l_{2,p}$ -PCA and manifold learning methods. It is not only robust to outliers, but also maintains the rotation invariance of the algorithm, and protects the true geometric information of the original data space. In the non greedy iterative algorithm of the model, the weight covariance matrix is considered to further reduce the reconstruction error. Therefore, the model has good expression ability, and it can effectively extract the algebraic features of images. This method is mainly divided into the following three steps:

1. Optimize the formula of $l_{2,p}$ -MRPCA;
2. the equation of the optimal matrix is obtained by using KKT condition;
3. and according to the algorithm proposed in this paper, the convergence of the objective function is obtained, and the optimal projection matrix is obtained.

The experimental results show that the recognition rate of $l_{2,p}$ -MRPCA algorithm is higher than some of the existing advanced algorithms, and it still has good robustness when there is occlusion. However, since this algorithm specifies many parameters in the implementation, which limits its application in practice, the following research will focus on parameter adjustment.

Author Contributions: Conceptualization, M.W. and X.W.; methodology, M.W.; software, X.W.; validation, M.W., X.W. and G.Y.; formal analysis, M.W.; investigation, X.W.; resources, G.Y. and H.T.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W.; visualization, G.Y.; supervision, M.W. and H.T.; project administration, M.W. and H.T.; funding acquisition, M.W. and X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the Postgraduate Research and Practice Innovation Program of Jiangsu Province Nos. SJCX21_0890; the National Science Foundation of China under Grant Nos. 61876213, 61861033, 61991401, 62172229, 61976117, 71972102, 61976118; the KeyR&D Program Science Foundation in Colleges and Universities of Jiangsu Province Grant Nos. 18KJA520005, 19KJA360001, 20KJA520002; the Natural Science Fund of Jiangsu Province under Grants Nos. BK20201397, BK20191409, BK20211295; and the Jiangsu Key Laboratory of Image and Video Understanding for Social Safety of Nanjing University of Science and Technology under Grants J2021-4. The Future Network Scientific Research Fund Project SRFP-2021-YB-25, and China's Jiangxi Province Natural Science Foundation (No. 20202ACBL202007). The Significant Project of Jiangsu College Philosophy and Social Sciences Research "Research on Knowledge Reasoning of Emergency Plan for Emergency Decision" (No: 2021SJZDA153). This work is funded in part by the "Qinglan Project" of Jiangsu Universities under Grant D202062032.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: [http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/att_faces.tar.Z; <http://www.itl.nist.gov/iad/humanid/feret/>; <http://www.comp.polyu.edu.hk/~biometrics>] (all accessed on 1 October 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Wu, X.T.; Yan, Q.D. Analysis and Research on data dimensionality reduction method. *Comput. Appl. Res.* **2009**, *26*, 2832–2835.
2. Yu, X.X.; Zhou, N. Research on dimensionality reduction method of high-dimensional data. *Inf. Sci.* **2007**, *25*, 1248–1251.
3. Wan, M.H.; Lai, Z.H.; Yang, G.W. Local graph embedding based on maximum margin criterion via fuzzy set. *Fuzzy Sets Syst.* **2017**, *2017*, 120–131. [[CrossRef](#)]
4. Yang, J.; Zhang, D.D.; Yang, J.Y. Constructing PCA Baseline Algorithms to Reevaluate ICA-Based Face-Recognition Performance. *IEEE Trans Multimed.* **2007**, *37*, 1015–1021.
5. Zuo, W.; Zhang, D.; Yang, J.; Wang, K. BDPCA plus LDA: a novel fast feature extraction technique for face recognition. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2006**, *36*, 946–953.
6. Kim, Y.G.; Song, Y.J.; Chang, U.D.; Kim, D.W.; Yun, T.S.; Ahn, J.H. Face recognition using a fusion method based on bidirectional 2DPCA. *Appl. Math. Comput.* **2008**, *205*, 601–607. [[CrossRef](#)]
7. Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J.Y. Two dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 131–137. [[CrossRef](#)]
8. Yang, J.; Zhang, D.; Yong, X.; Yang, J.Y. Two dimensional discriminant transform for face recognition. *Pattern Recognit.* **2005**, *38*, 1125–1129. [[CrossRef](#)]
9. Wang, J.; Barreto, A.; Wang, L.; Chen, Y.; Rische, N.; Andrian, J.; Adjouadi, M. Multilinear principal component analysis for face recognition with fewer features. *Neurocomputing* **2010**, *73*, 1550–1555. [[CrossRef](#)]

10. Wan, M.; Yao, Y.; Zhan, T.; Yang, G. Supervised Low-Rank Embedded Regression (SLRER) for Robust Subspace Learning. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 1917–1927. [[CrossRef](#)]
11. Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; Ma, Y. Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices. *Adv. Neural Inf. Process. Syst.* **2009**, *22*, 2080–2088.
12. Wan, M.; Chen, X.; Zhao, C.; Zhan, T.; Yang, G. A new weakly supervised discrete discriminant hashing for robust data representation. *Inf. Sci.* **2022**, *611*, 335–348. [[CrossRef](#)]
13. Ke, Q.F.; Kanade, T. Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming. In Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 739–746.
14. He, R.; Hu, B.G.; Zheng, W.S.; Kong, X.W. Robust Principal Component Analysis Based on Maximum Correntropy Criterion. *IEEE Trans. Image Process.* **2011**, *20*, 1485–1494.
15. Kwak, N. Principal Component Analysis Based on L1-Norm Maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1672–1680. [[CrossRef](#)] [[PubMed](#)]
16. Kwak, N. Principal Component Analysis by L-p-Norm Maximization. *IEEE Trans. Cybern.* **2014**, *44*, 594–609. [[CrossRef](#)] [[PubMed](#)]
17. Ye, Q.; Fu, L.; Zhang, Z.; Zhao, H.; Naiem, M. Lp- and Ls-Norm Distance Based Robust Linear Discriminant Analysis. *Neural Netw.* **2018**, *105*, 393–404. [[CrossRef](#)]
18. Ding, C.; Zhou, D.; He, X.; Zha, H. R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. In Proceedings of the 23rd International Conference on Machine Learning, ACM, New York, NY, USA, 25–29 June 2006; pp. 281–288.
19. Wang, Q.; Gao, Q.; Gao, X.; Nie, F. L2,p-norm based PCA for image recognition. *IEEE Trans. Image Process.* **2008**, *27*, 1336–1346. [[CrossRef](#)]
20. Bi, P.; Du, X. Application of Locally Invariant Robust PCA for Underwater Image Recognition. *IEEE Access* **2021**, *9*, 29470–29481. [[CrossRef](#)]
21. Xu, J.; Bi, P.; Du, X.; Li, J.; Chen, D. Generalized Robust PCA: A New Distance Metric Method for Underwater Target Recognition. *IEEE Access* **2019**, *7*, 51952–51964. [[CrossRef](#)]
22. Wan, M.; Chen, X.; Zhan, T.; Xu, C.; Yang, G.; Zhou, H. Sparse Fuzzy Two-Dimensional Discriminant Local Preserving Projection (SF2DDLPP) for Robust Image Feature Extraction. *Inf. Sci.* **2021**, *563*, 1–15. [[CrossRef](#)]
23. Tasoulis, S.; Pavlidis, N.G.; Roos, T. Nonlinear Dimensionality Reduction for Clustering. *Pattern Recognit.* **2020**, *107*, 107508. [[CrossRef](#)]
24. Luo, W.Q. Face recognition based on Laplacian Eigenmaps. In Proceedings of the International Conference on Computer Science and Service System, Nanjing, China, 27–29 June 2011; pp. 27–29.
25. Roweis, S.; Saul, L. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
26. Hu, K.L.; Yuan, J.Q. Statistical monitoring of fed-batch process using dynamic multiway neighborhood preserving embedding. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 195–203. [[CrossRef](#)]
27. Song, B.; Ma, Y.; Shi, H. Multimode process monitoring using improved dynamic neighborhood preserving embedding. *Chemom. Intell. Lab. Syst.* **2014**, *135*, 17–30. [[CrossRef](#)]
28. Wan, M.; Chen, X.; Zhan, T.; Yang, G.; Tan, H.; Zheng, H. Low-rank 2D Local Discriminant Graph Embedding for Robust Image Feature Extraction. *Pattern Recognit.* **2023**, *133*, 109034. [[CrossRef](#)]
29. Chen, X.; Wan, M.; Zheng, H.; Xu, C.; Sun, C.; Fan, Z. A New Bilinear Supervised Neighborhood Discrete Discriminant Hashing. *Mathematics* **2022**, *10*, 2110. [[CrossRef](#)]
30. Li, W.H.; Gong, W.G.; Cheng, W.M. Method based on wavelet multiresolution analysis and KPCA for face recognition. *Comput. Appl.* **2005**, *25*, 2339–2341.
31. De, F.; Torre, L.; Black, M.J. A Framework for Robust Subspace Learning. *Int. J. Comput. Vis.* **2003**, *54*, 117–142.