

Article

Two-Dimensional Exponential Sparse Discriminant Local Preserving Projections

Minghua Wan ^{1,2,3}, Yuxi Zhang ¹, Guowei Yang ^{1,4} and Hongjian Guo ^{1,*}¹ School of Computer Science (School of Intelligent Auditing), Nanjing Audit University, Nanjing 211815, China² Jiangsu Key Lab of Image and Video Understanding for Social Security, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China³ Key Laboratory of Intelligent Information Processing, Nanjing Xiaozhuang University, Nanjing 211171, China⁴ School of Electronic Information, Qingdao University, Qingdao 266071, China

* Correspondence: 209432@nau.edu.cn

Abstract: The two-dimensional discriminant locally preserved projections (2DDLPP) algorithm adds a between-class weighted matrix and a within-class weighted matrix into the objective function of the two-dimensional locally preserved projections (2DLPP) algorithm, which overcomes the disadvantage of 2DLPP, i.e., that it cannot use the discrimination information. However, the small sample size (SSS) problem still exists, and 2DDLPP processes the whole original image, which may contain a large amount of redundant information in the retained features. Therefore, we propose a new algorithm, two-dimensional exponential sparse discriminant local preserving projections (2DESDLPP), to address these problems. This integrates 2DDLPP, matrix exponential function and elastic net regression. Firstly, 2DESDLPP introduces the matrix exponential into the objective function of 2DDLPP, making it positive definite. This is an effective method to solve the SSS problem. Moreover, it uses distance diffusion mapping to convert the original image into a new subspace to further expand the margin between labels. Thus more feature information will be retained for classification. In addition, the elastic net regression method is used to find the optimal sparse projection matrix to reduce redundant information. Finally, through high performance experiments with the ORL, Yale and AR databases, it is proven that the 2DESDLPP algorithm is superior to the other seven mainstream feature extraction algorithms. In particular, its accuracy rate is 3.15%, 2.97% and 4.82% higher than that of 2DDLPP in the three databases, respectively.

Keywords: feature extraction; SSS problem; two-dimensional local discriminant preserving projections; matrix exponential; elastic net regression

MSC: 68U10

Citation: Wan, M.; Zhang, Y.; Yang, G.; Guo, H. Two-Dimensional Exponential Sparse Discriminant Local Preserving Projections. *Mathematics* **2023**, *11*, 1722. <https://doi.org/10.3390/math11071722>

Academic Editors: Guangwei Gao, Juncheng Li and Zhi Li

Received: 6 March 2023

Revised: 29 March 2023

Accepted: 30 March 2023

Published: 4 April 2023



Copyright: © 2023 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

Feature extraction is an important part of pattern recognition. Therefore, extracting the required features accurately and effectively has always been a concern of researchers [1–3]. In recent years, researchers have successfully discovered the inherent features of low-dimensional nonlinear manifold structures and put forward some relevant algorithms [4,5]. For example, the classical algorithm locality preserving projections (LPP) [6] is widely applied in feature extraction. However, because LPP is an algorithm based on a one-dimensional vector, it is easy to encounter unusual problems in its solution process. Therefore, scholars have proposed the two-dimensional locally preserving projection (2DLPP) [7] algorithm. 2DLPP directly replaces 1D vectors with 2D image matrices. Although the 2DLPP algorithm can process sample data more effectively than the LPP algorithm, its disadvantage is that it cannot use class label information to classify. The two-dimensional discriminant locally preserving projection (2DDLPP) [8] algorithm proposed by Zhi et al.

adds the inter-class scatter matrix and the intra-class discriminant matrix to the objective function of 2DLPP. Although the 2DDLPP algorithm improves the accuracy of image recognition, it is still restricted by the problem of small sample size when processing high dimensional data. In the practical application of face feature extraction, the facial image has high dimensional characteristics, but due to the influence of computer storage capacity and shooting permission, the number of samples that can be used for classification training is far less than the sample dimension. This is a classic small sample size problem. Limited by the SSS problem, the generalization ability of the 2DDLPP algorithm is not strong, and feature extraction ability is also affected, which leads to certain limitations in its application scope. Therefore, the SSS problem has become urgent in the field of feature extraction of high dimensional data.

In recent years, the matrix exponential has been of interest to researchers and widely used to solve the small sample size problem in scientific computing. Firstly, in order to solve the SSS problem of the linear discriminant analysis (LDA) algorithm [9], researchers pre-processed the sample data with the principal component analysis (PCA) algorithm [10], but some important feature information would be lost in this way. Therefore, the exponential discriminant analysis (EDA) [11] algorithm proposed by Zhang et al. used the matrix exponential to address this problem skillfully. Considering that LPP may encounter the same problem as LDA when the sample size is smaller than the sample dimension, singularity of the matrix will be caused. Therefore, Wang et al. integrated LPP with the matrix exponential, and then proposed the exponential locality preserving projection (ELPP) [12] algorithm, with better performance. Similarly, exponential local discriminant embedding (ELDE) [13] overcomes the SSS problem of local discriminant embedding (LDE) [14] by invoking the matrix exponential. Matrix exponential based discriminant locality preserving projections (MEDLPP) [15] introduce the matrix exponential into discriminant locality preserving projections (DLPP) to solve the small sample size problem, with corresponding improvements to shorten running time. In reference [16], the matrix exponential is added to the accelerating algorithm and incremental algorithm of large-scale semi-supervised discriminant embedding. In view of this, we use the matrix exponential to solve the SSS problem of 2DDLPP.

At present, sparse algorithms are one of the important research directions in feature extraction. In the process of sparse learning, L_1 and L_2 norms are used to optimally represent the relationship between samples and features, drawing on the idea of regression analysis. This process can not only simplify the data but also retain the key information [17–20]. By integrating sparse representation with classical algorithms, many new algorithms have been proposed. Recently, Zhang et al. proposed a new unsupervised feature extraction method, joint sparse representation and local preserving projection (JSRLPP) [21], which integrates a graph structure and projection matrix into a general framework to learn graphs that are more suitable for feature extraction. The performance of traditional L_1 and L_2 norms are reduced by ignoring local geometric structure in addressing the SSS problem. Similarly, Liu et al. not only used L_1 and L_2 norms for joint sparse regression but also used capped the L_2 norm in the loss function of the locality preserving robust regression (LPRR) [22] algorithm to further enhance its robustness, achieving good intelligibility and quality. Therefore, in this paper, we use elastic net regression integrated with other algorithms for sparse feature extraction.

According to reference [23], 2D methods are not always superior to 1D methods. Although the 2DDLPP algorithm is simple and efficient, it performs poorly with limited training samples. When the dimension of the samples exceeds the sample size, singularity of the matrix will be generated. It also appears to have limitations in some public databases. In summary, we propose the two-dimension exponential sparse discriminant local preserving projection (2DESDLPP) algorithm in this paper, which adds matrix exponential into the 2DDLPP objective function to improve performance. Based on the property of the matrix function, 2DDLPP is reconstructed using the generalized matrix exponential function. If 2DESDLPP first uses PCA to reduce the dimension of the image data, it will lead to the loss of a large amount of effective feature information and reduce the recognition performance of

the 2DESDLPP algorithm. The matrix exponential function avoids the singular divergence matrix in the generalized eigenvalue problem of the 2DDLPP algorithm by ensuring the orthogonality of the basis vector obtained, so as to solve the SSS problem. Then, for each test sample, Euclidean distance is used to measure the similarity relationship between the test sample and each training sample, and as the weight of the training sample, in order to form a weighted training sample set. This is equivalent to transforming the original image to another new subspace, using distance diffusion mapping that further widens the margins between labels. After solving the small sample problem, each row (or column) of the 2D face image is then treated as a separate vector using the 2D expansion form of elastic net regression, and these vectors are then used as independent model units to perform the corresponding vector-based regression in order to obtain the optimal sparse projection matrix. Finally, experiments were conducted on three face databases, ORL, Yale and AR, to verify the effectiveness of the algorithm by comparing the average recognition rate with two 1D algorithms (ELPP and MEDLPP) and five other 2D algorithms (2DPCA, 2DLDA, 2DLPP, 2DDLPP and 2DEDLPP). The development route for 2DESDLPP is shown in Figure 1 below.

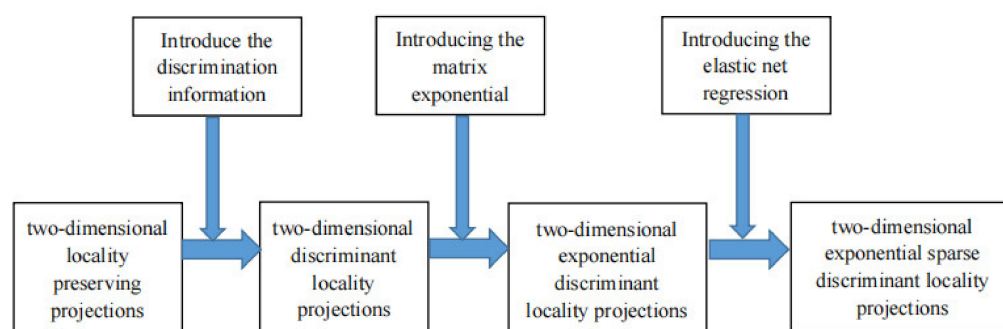


Figure 1. The development route of the 2DESDLPP algorithm.

The main contributions of the algorithm we propose are as follows:

- (1) The 2DESDLPP algorithm not only retains the classification information between samples satisfactorily, but also solves the SSS problem of the 2DDLPP algorithm by integrating matrix exponential.
- (2) We use the idea of sparse feature extraction not only to obtain the optimal projection matrix through the 2D extended form of elastic net regression, but also to reduce the computational complexity of 2DESDLPP.
- (3) The 2DESDLPP algorithm can widen the distance between category labels and discriminate the identification information contained in the zero space, so it has higher recognition accuracy.

The content of this paper is arranged as follows. We briefly review three underlying algorithms, namely 2DDLPP, matrix exponential and elastic net regression, in Section 2. Section 3 describes the proposed 2DESDLPP in detail. Section 4 is composed of seven experiments designed to evaluate our 2DESDLPP algorithm. Finally, Section 5 provides conclusions.

2. Introduction of Underlying Algorithms

Assume that $X = [X_1, X_2, \dots, X_N]$ is a training sample set, among which N is the amount of training sample images whose size is $m \times n$ dimension. We aim to map the initial space sized $m \times n$ into the space sized $m \times d$ by means of a linear transformation, where $d \ll n$. Let $A = [a_1, a_2, \dots, a_d]$ be the matrix sized $n \times d$, where a_i is a cell column vector. The projection applied to each data is as follows:

$$Y_i = X_i A \quad (1)$$

Each X_i , $i \in \{1, 2, \dots, N\}$ whose size is $m \times n$ dimension is mapped through the projection matrix A to obtain matrix Y_i whose size is $m \times d$ dimension.

2.1. 2DDLPP Algorithm

2DDLPP minimizes intra-class distances and maximizes inter-class distances by adding inter-class dispersion constraints and class label discriminant information to the 2DLPP objective function. Suppose $c_i \in \{1, 2, \dots, C\}$ are the class labels of x_i , among which C is the amount of c_i . Y_i^c and Y_j^c represent the c class projection image matrix of the original images. According to reference [9], we know that the 2DDLPP objective function is defined as follows:

$$J(Y) = \frac{\sum_{c=1}^C \sum_{i,j=1}^{n_c} \|Y_i^c - Y_j^c\|^2 B_{ij}^c}{\sum_{i,j=1}^C \|M_i - M_j\|^2 W_{ij}} \quad (2)$$

where B_{ij}^c and W_{ij} are both weight matrices. The sample size of the category c is n_c , and M_i, M_j is the mean value matrix of the projection matrix of category i and j samples, respectively:

$$\begin{cases} M_i = \frac{1}{n_i} \sum_{k=1}^{n_i} Y_k^i \\ M_j = \frac{1}{n_j} \sum_{k=1}^{n_j} Y_k^j \end{cases} \quad (3)$$

Referring to the algorithm idea of 2DDLPP, the 2DESDLPP algorithm proposed in this paper not only ensures that adjacent points maintain a neighborhood relationship after projection to maintain local information for the data, but also makes full use of label information of data features to narrow the distance between data in the same category and to increase the distance between data in different labels, which is more favorable for feature extraction.

2.2. Matrix Exponential

A square matrix $M (M \in R^{m \times n})$ whose size is $m \times n$ dimension is given. Its matrix exponential is shown below:

$$\exp(M) = I + M + \frac{M^2}{2!} + \dots + \frac{M^m}{m!} + \dots \quad (4)$$

I is a unit matrix whose size is $n \times n$ dimension. The properties of an exponential matrix are as follows:

- (1) $\exp(M)$ is the sum of a finite matrix sequence.
- (2) $\exp(M)$ is the full rank matrix.
- (3) Supposing that matrix M and matrix Q are commutative, like $MQ = QM$, then we can have $\exp(M + Q) = \exp(M) \exp(Q)$.
- (4) Assuming that the matrix Q is non-singular, we have $\exp(Q^{-1}AQ) = Q^{-1} \exp(A)Q$.
- (5) For every eigenvector v_1, v_2, \dots, v_n of M that corresponded to eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, with $e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}$ as eigenvalues of $\exp(M)$ having the same eigenvectors, then the matrix is non-singular.

The introduction of the matrix exponential function not only solves the SSS problem of the 2DDLPP algorithm, but also stretches the distance between different categories of samples, which makes the algorithm achieve better classification ability in the face of recognition tasks.

2.3. Elastic Net Regression

Suppose there are datasets $(X_i, y_i), i = 1, 2, \dots, N$, where N is the sample size. $X_i = (x_{i1}, \dots, x_{ip})^T$ are the independent variables corresponding to the i^{th} observations.

p is the number of columns of X_i . y_i are corresponding variables. Its regression model is defined as:

$$\hat{\beta} = \operatorname{argmin} \left(\sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda_1 \sum_{j=1}^p \beta_j^2 + \lambda_2 \sum_{j=1}^p |\beta_j| \right) \quad (5)$$

If $\varepsilon = \lambda_1 + \lambda_2$, $\theta = \frac{\lambda_2}{\lambda_1 + \lambda_2}$, then

$$\hat{\beta} = \operatorname{argmin} \left(\sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \varepsilon \left(\theta \sum_{j=1}^p |\beta_j| + (1 - \theta) \sum_{j=1}^p \beta_j^2 \right) \right) \quad (6)$$

Here $\lambda_1, \lambda_2 \geq 0$ are the penalty parameters, β_j is the variable coefficient, and β_0 is the constant term, which can be generally ignored in the penalty function, because the constant term does not affect the regression coefficient. Thus, the regular term of Elastic Net regression is a convex linear combination of the regular terms of Lasso regression and Ridge regression. When $\theta = 0$, it is the Ridge regression algorithm. When $\theta = 1$, it is the Lasso regression algorithm. 2DESDLPP uses the elastic net regression method for face recognition and classification. After sparse representation of the elastic net, the problem of feature information redundancy can be solved, which not only reduces the computational complexity of the 2DESDLPP algorithm, but also shortens its running time, and greatly improves the efficiency of face recognition.

3. Two-Dimensional Exponential Sparse Discriminant Local Preserving Projections

The 2DDLPP algorithm uses two-dimensional images to represent data, which is beneficial in maintaining discriminant information about the local manifold structure of the data. Although it has been successfully applied in many fields, matrix-based methods do not always outperform vector-based methods, even with limited training samples. There may be a lot of redundant information in the retained features of the 2DDLPP algorithm, which requires a high computational cost. To address the above problems, we propose a new algorithm, two-dimensional exponential sparse discriminant local preserving projections (2DESDLPP), for feature extraction. This allows for the selection of salient features that fit the current pattern from the abundant features in the original data, in order to find a minimal subset of the original set of features for an optimal representation of the data.

Firstly, according to Equation (1), the molecule of Equation (2) can be simplified as:

$$\begin{aligned} & \sum_{c=1}^C \sum_{i,j=1}^{n_c} \|Y_i^c - Y_j^c\|^2 B_{ij}^c \\ &= \sum_{c=1}^C \sum_{i,j=1}^{n_c} \left(X_i^c A - X_j^c A \right)^T \left(X_i^c A - X_j^c A \right) B_{ij}^c \\ &= 2 \sum_{c=1}^C A^T \left[\sum_{i=1}^{n_c} (X_i^c)^T D_{ii}^c X_i^c - \sum_{i,j=1}^{n_c} (X_i^c)^T B_{ij}^c X_j^c \right] A \\ &= 2 A^T \left[\sum_{c=1}^C X_c^T ((D_c - B_c) \otimes I_n) X_c \right] A \\ &= 2 A^T X^T (L \otimes I_n) X A \end{aligned} \quad (7)$$

Here, B_{ij}^c is the weight matrix between any two samples in the c^{th} class, which is defined as follows:

$$B_{ij}^c = \begin{cases} \exp\left(\|x_i - x_j\|^2 / t\right), & x_i, x_j \in c \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Consisting of the diagonal matrix, D_{ii} is the sum of the rows or columns of B , $D_{ii} = \sum_j B_{ji}$. The symbol \otimes represents the Kronecker product of a matrix. For the

i^{th} point, the larger value of D_{ii} , the more important is x_i , because matrix D provides a natural measure of the data point corresponding to raw images. L is a Laplace matrix, $L = D - B$.

Next, the denominator of Equation (2) is simplified similarly:

$$\begin{aligned} & \sum_{i,j=1}^C \|M_i - M_j\|^2 W_{ij} \\ &= \sum_{i,j=1}^C (M_i - M_j)^T (M_i - M_j) W_{ij} \\ &= \sum_{i,j=1}^C A^T (F_i - F_j)^T (F_i - F_j) A W_{ij} \\ &= 2A^T F^T (H \otimes I_n) F A \end{aligned} \quad (9)$$

F_i is the mean value matrix of the i^{th} sample, $F_i = \frac{1}{n_i} \sum_{k=1}^{n_i} X_k^i$. W_{ij} is the weight of the means of any two kinds of sample, $W_{ij} = \exp\left(\frac{\|F_i - F_j\|^2}{t}\right)$, while t is an adjustable positive parameter. Consisting of the diagonal matrix E_{ii} is the sum of the rows or columns of W , $E_{ii} = \sum_j W_{ji}$. H is a Laplace matrix, $H = E - W$.

Substituting the simplified Equation (7) as well as Equation (8) into Equation (2), the objective function of 2DDLPP is:

$$J(A) = \frac{A^T X^T (L \otimes I_n) X A}{A^T F^T (H \otimes I_n) F A} \quad (10)$$

Minimizing Equation (10), then we can obtain:

$$A = \operatorname{argmin}_A \frac{A^T X^T (L \otimes I_n) X A}{A^T F^T (H \otimes I_n) F A} \quad (11)$$

Next, the matrix exponential and sparse constraint are added to Equation (11) to obtain the objective function of 2DESDLPP as:

$$\begin{cases} A = \operatorname{argmin}_A \frac{A^T \exp(X^T (L \otimes I_n) X) A}{A^T \exp(F^T (H \otimes I_n) F) A} \\ \operatorname{Card}(A) = K \end{cases} \quad (12)$$

The number of non-zero elements in matrix A is K , $K \leq n$. Therefore, $\operatorname{Card}(A)$ is the sparse constraint on matrix A , using the elastic net regression. Equation (12) does not have a solution in a closed form, so it can be expressed more simply:

$$\begin{cases} A = \operatorname{argmin}_A (A^T \exp(F^T (H \otimes I_n) F) A)^{-1} (A^T \exp(X^T (L \otimes I_n) X) A) \\ \operatorname{Card}(A) = K \end{cases} \quad (13)$$

Inspired by the 2D-MELPP algorithm in reference [24], we solve the eigenvalues of Equation (13) to obtain its solution.

$$\begin{cases} \exp(X^T (L \otimes I_n) X) A = \lambda \exp(F^T (H \otimes I_n) F) A \\ \operatorname{Card}(A) = K \end{cases} \quad (14)$$

After orthogonalizing the projection matrix A , we can achieve:

$$\begin{cases} \exp(X^T (L \otimes I_n) X) a_i = \lambda_i \exp(F^T (H \otimes I_n) F) a_i \\ \operatorname{Card}(A) = K \end{cases} \quad (15)$$

where λ_i is the eigenvalue and a_i is the eigenvector corresponding to λ_i . Then, we take the eigenvectors corresponding to the first d minimum non-zero eigenvalues of Equation (15) and combine them to obtain the projection matrix A .

According to the properties of the matrix exponential, $\exp(X^T(L \otimes I_n)X)$ and $\exp(F^T(H \otimes I_n)F)$ are both full rank matrices. That is, even in the case of SSS problems, $\exp(X^T(L \otimes I_n)X)$ and $\exp(F^T(H \otimes I_n)F)$ are non-singular, although $X^T(L \otimes I_n)X$ and $F^T(H \otimes I_n)F$ are still singular. Therefore, 2DESDLPP can extract the identification information contained in the $F^T H F$ null space.

However, the resulting projection matrix A is not the sparsest at present. Therefore, we take a 2D extension of elastic net regression by treating each row (or column) of a 2D face image as a separate vector and then use these vectors as independent model units for the corresponding vector-based regression. That is:

$$\begin{cases} A_{\text{Sparse}} = \operatorname{argmin} \left(\sum_{i=1}^N \sum_{k=1}^m (x_i(k, :) \times A - y_i)^2 + \varepsilon \left(\sum_{j=1}^d \theta |a_j| + (1 - \theta) \sum_{j=1}^d a_j^2 \right) \right) \\ \operatorname{Card}(A_{\text{Sparse}}) = K \end{cases} \quad (16)$$

Now, the optimal sparse projection matrix A obtained not only has higher recognition accuracy, because it is based on the image matrix, but also has a reduced computational complexity. Next, we will carry out the specific analysis.

The time complexity of calculating $\exp(X^T(L \otimes I_n)X)$ and $\exp(F^T(H \otimes I_n)F)$ is $O(d^3)$. Therefore, for the eigenvalue problem of Equation (15), the complexity of the operation is also the same. Then, after sparse processing of Equation (16), the calculation complexity is reduced to $O(d^2)$. Finally, the whole time complexity of the 2DESDLPP algorithm is $O(d^2)$, while the calculation complexity of 2DEDLPP is $O(d^3)$.

4. Results and Analysis of the Experiment

The experiments in feature extraction were conducted on the three public face databases, ORL, Yale and AR, using Euclidean distance and nearest neighbor classifiers. Then we compared the performance of several other algorithms (ELPP, MEDLPP, 2DPCA, 2DLDA, 2DLPP, 2DDLPP and 2DEDLPP) to verify the effectiveness of the 2DESDLPP algorithm proposed. In the experiments, we resized the images in each face database to 50×40 to reduce the amount of computer memory used, and randomly selected l ($l = 2, 3, 4, 5, 6$) images for each person as training samples. The rest were used as test samples.

We first set $K = 10$, which is used to count the number of non-zero elements in the projection matrix, and the number of training samples was six. Then, a comparison experiment was conducted with the original images with a size of 92×112 , and the resized images with a size of 50×40 , in the ORL database to test how the accuracy of the 2DESDLPP algorithm depends on the size of the training sample. Finally, the comparison results are recorded in Table 1.

Table 1. Max accuracy (%) of 2DESDLPP experimenting on original images and resized images, and the corresponding number of dimensions in the ORL database.

l	Original Images	Resized Images
2	92.33 (92×85)	91.79 (50×16)
3	94.12 (92×86)	93.61 (50×18)
4	94.57 (92×98)	94.13 (50×26)
5	95.48 (92×95)	95.25 (50×22)
6	98.61 (92×110)	98.50 (50×20)

It can be seen from Table 1 that the size of the training sample has little influence on the accuracy of the 2DESDLPP algorithm, and its variation range is within 0.5, which is acceptable.

4.1. Experiments to Determine the Value of K

First of all, we set K as 10, 20, 30 and 40. Next, six face images of each person in ORL, Yale and AR databases were randomly selected as training samples, and finally their average values were recorded and compared after ten experiments. Figures 2–4 show the average recognition accuracy (%) of the 2DESDLPP algorithm, corresponding to different feature dimensions in ORL, Yale and AR databases with different K values, respectively. It is obvious that the average recognition accuracy reached the maximum in all three face databases when $K = 10$. Therefore, this is selected in subsequent experiments.

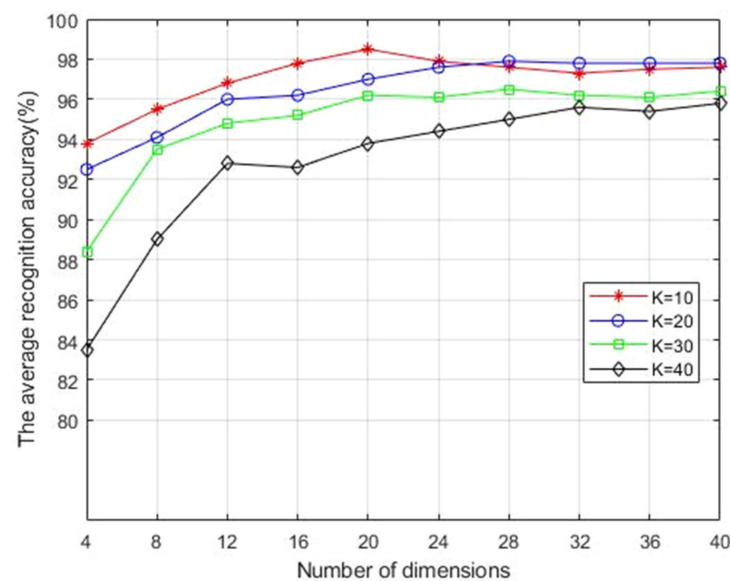


Figure 2. Average recognition accuracy (%) of the 2DESDLPP algorithm corresponding to different K values with feature dimensions when the training sample in the ORL database is 6.

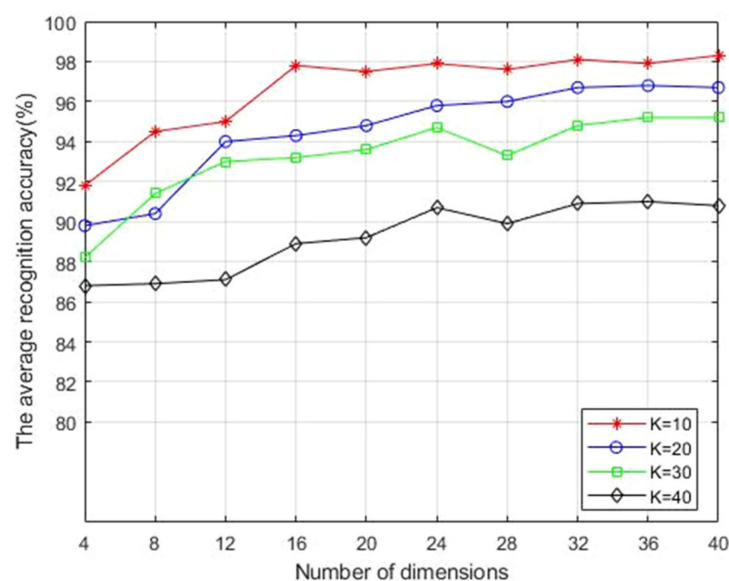


Figure 3. Average recognition accuracy (%) of the 2DESDLPP algorithm corresponding to different K values with feature dimensions when the training sample in the Yale database is 6.

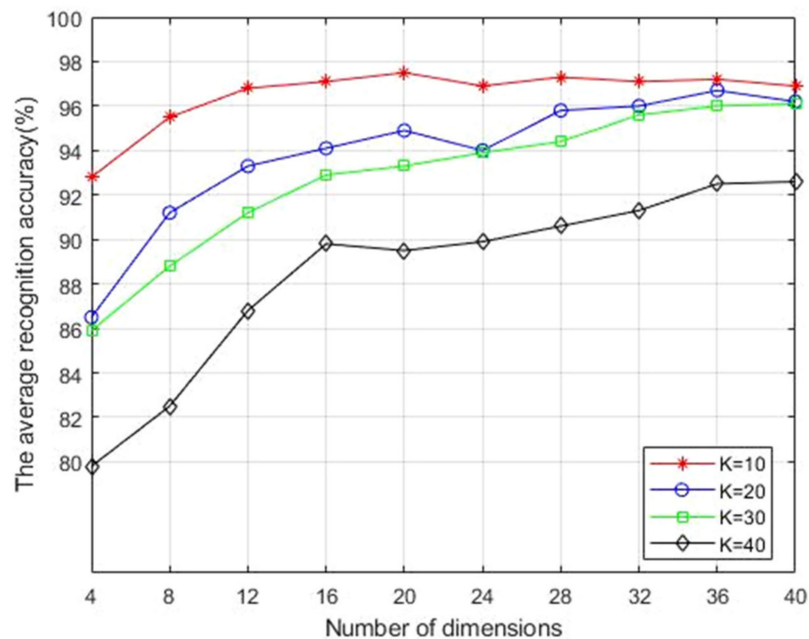


Figure 4. Average recognition accuracy (%) of the 2DESDLPP algorithm corresponding to different K values with feature dimensions when the training sample in the AR database is 6.

4.2. Experiments in ORL Database

There are 400 grayscale faces images from 40 persons in the ORL database and everyone has ten images. All images were acquired under changing external conditions, including lighting intensity, facial angle, posture change and expression changes. All face images in the ORL database were unified into pure black-backed scenery with a gray scale of 92×112 pixels, as shown in Figure 5.



Figure 5. Face image of an individual in ORL face database.

We selected $l(l = 2, 3, 4, 5, 6)$ training sample images randomly from the ORL database and record the maximum average recognition rates and the number of dimensions in Table 2. As can be seen from Table 2, with the increase of the number of training samples, the recognition rate of the eight algorithms has improved. When the number of training samples is two, the recognition rate of all algorithms is not high, but the highest recognition rate of the 2DESDLPP algorithm is still 91.79%, 1.52 times that of the ELPP algorithm, and about 5.54% higher than that of the 2DDLPP algorithm. Even when the number of training samples is much smaller than the feature dimension, the 2DESDLPP algorithm can still achieve better performance, which indicates that the 2DESDLPP algorithm can indeed effectively solve the problem of the small sample size in the 2DDLPP algorithm. When the

training sample size is six, the changes in 2DESDLPP recognition results and those of the other seven comparison algorithms are shown in Figure 6.

Table 2. Maximum average recognition rate (%) of 2DESDLPP and other seven comparison algorithms and the corresponding number of dimensions in the ORL database.

<i>l</i>	ELPP	MEDLPP	2DPCA	2DLDA	2DLPP	2DDLPP	2DEDLPP	2DESDLPP
2	60.36 (50 × 38)	91.20 (50 × 28)	74.85 (50 × 40)	83.75 (50 × 12)	81.40 (50 × 8)	86.25 (50 × 6)	90.22 (50 × 15)	91.79 (50 × 16)
3	75.10 (50 × 40)	92.00 (50 × 32)	85.19 (50 × 40)	88.63 (50 × 2)	86.21 (50 × 17)	90.12 (50 × 4)	91.30 (50 × 16)	93.61 (50 × 18)
4	80.21 (50 × 40)	94.95 (50 × 36)	79.85 (50 × 40)	89.33 (50 × 40)	92.25 (50 × 38)	92.38 (50 × 2)	92.48 (50 × 22)	94.13 (50 × 26)
5	85.96 (50 × 39)	95.18 (50 × 38)	83.76 (50 × 40)	90.96 (50 × 32)	93.49 (50 × 24)	94.18 (50 × 4)	93.90 (50 × 24)	95.25 (50 × 22)
6	90.33 (50 × 40)	95.20 (50 × 40)	92.39 (50 × 40)	95.30 (50 × 40)	94.10 (50 × 40)	94.60 (50 × 8)	96.88 (50 × 28)	98.50 (50 × 20)

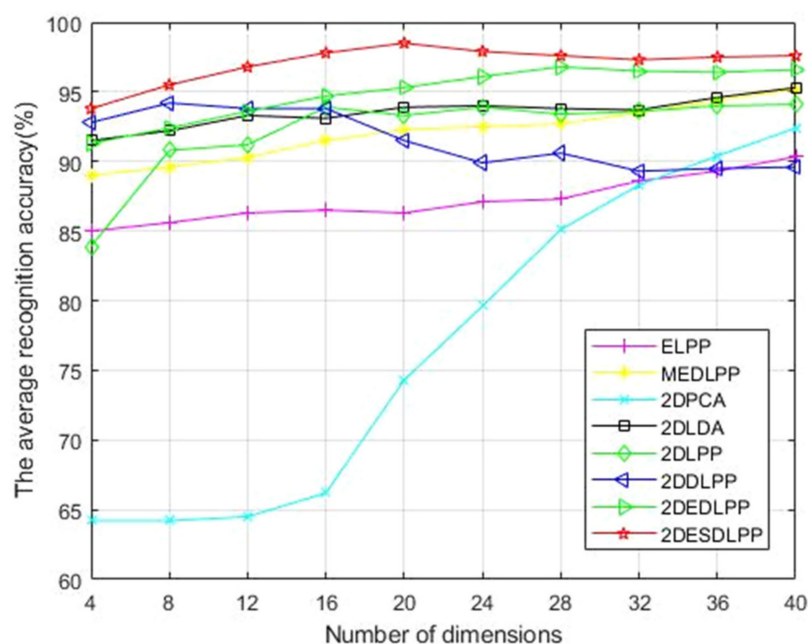


Figure 6. Average recognition rate (%) with feature dimension for different algorithms when the number of training samples in the ORL face database is 6.

It can be clearly seen from Figure 6 that the recognition rate of supervised algorithms is higher than that of unsupervised algorithms, indicating that the use of label information is indeed helpful in improving the accuracy of the algorithm. Especially when the feature dimension is 20, the recognition rate of the 2DESDLPP algorithm is up to 98.5%, which is about 6.2% higher than that of the 2DDLPP algorithm with the same dimensional number. This is because 2DESDLPP converts the original image into a new subspace using distance diffusion mapping, further expanding the boundaries between labels. This allows more characteristic information to be retained for classification. In addition, the variation trend for algorithm accuracy of 2DESDLPP is relatively flat, indicating that the algorithm has strong robustness. Table 3 shows the average running time of each algorithm in the ORL database when the number of training samples is six. It is not difficult to see that the running time of the 2DESDLPP algorithm after sparse representation is shortened, and is within the acceptable range.

Table 3. The runtime (s) of eight different algorithms when the number of training samples in the ORL database is 6.

Algorithm	ELPP	MEDLPP	2DPCA	2DLDA	2DLPP	2DDLPP	2DEDLPP	2DESDLPP
Runtime	1.633	2.649	2.343	2.494	2.314	3.375	3.890	3.039

4.3. Experiment in Yale Database

In the Yale database, there are 15 different subjects and each has 11 images taken under different external conditions, with a total of 165 grayscale images. Each image was created under different conditions of expression, lighting, etc. Most of them were used to verify the illumination robustness of the algorithm except for the face recognition rate, as shown in Figure 7.



Figure 7. Face image of an individual in Yale face database.

The experimental performance of the algorithm will be affected by external factors such as the facial expression changes of each subject and the different lighting conditions during shooting. Therefore, we chose to test whether 2DESDLPP is susceptible to random factors in the Yale database in order to evaluate its accuracy and robustness on small sample size data. In the experiments, $l(l = 2, 3, 4, 5, 6)$ images of every person were selected as training sample sets randomly, and the rest were used as the test. The recognition performances of 2DESDLPP and the other seven comparison algorithms are shown in Figure 8 when the number of training samples is six. As can be seen, the average recognition rate of 2DESDLPP is highest when the feature dimension is 40, at about 98.33%. Moreover, when the feature dimension is greater than 16, the recognition rate of 2DESDLPP fluctuates little and is relatively stable.

For convenience of comparison, the specific values of the maximum average recognition rates of eight algorithms and the corresponding projected dimension numbers (in brackets) in the Yale database when the training sample sizes are 2, 3, 4, 5 and 6 are shown in Table 4 respectively. When the number of training samples is 2 or 3, the performance of 2DESDLPP is better than the two one-dimensional algorithms and five two-dimensional algorithms used for comparison. After this improvement, the recognition rate of 2DESDLPP is about 3.7% higher than that of 2DDLPP. This indicates that, even when tested on small sample size data affected by external factors such as illumination changes, 2DESDLPP can obtain the best results compared with other algorithms. Table 5 shows the average running time of each algorithm in the Yale database when the number of training samples is six. Compared with 2DDLPP, the computation speed of 2DESDLPP is improved. Although it is not the fastest algorithm, it has the best comprehensive performance when combined with the recognition rate.

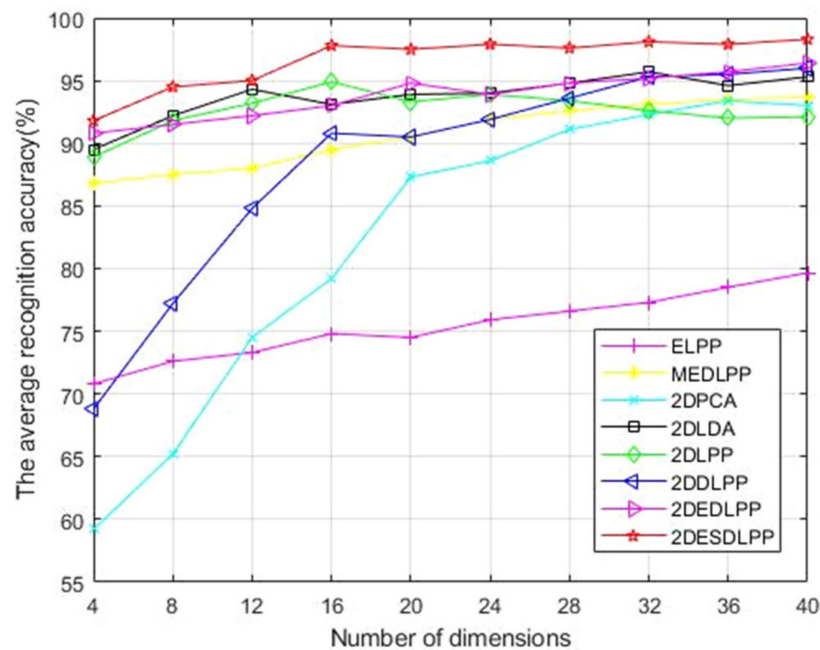


Figure 8. Average recognition rate (%) with feature dimension for different algorithms when the number of training samples in the Yale face database is 6.

Table 4. Maximum average recognition rate (%) of 2DESDLPP and other seven comparison algorithms and the corresponding number of dimensions in the Yale database.

<i>l</i>	ELPP	MEDLPP	2DPCA	2DLDA	2DLPP	2DDLPP	2DEDLPP	2DESDLPP
2	73.22 (50 × 36)	85.92 (50 × 36)	88.67 (50 × 36)	89.72 (50 × 16)	83.93 (50 × 6)	89.61 (50 × 8)	90.45 (50 × 36)	92.96 (50 × 38)
3	74.59 (50 × 38)	86.33 (50 × 38)	90.58 (50 × 36)	91.57 (50 × 18)	86.00 (50 × 4)	90.08 (50 × 3)	91.69 (50 × 36)	94.20 (50 × 39)
4	77.03 (50 × 38)	87.62 (50 × 40)	91.05 (50 × 36)	92.14 (50 × 20)	94.1 (50 × 4)	93.20 (50 × 38)	93.35 (50 × 38)	95.48 (50 × 38)
5	78.48 (50 × 39)	91.80 (50 × 40)	92.00 (50 × 36)	93.22 (50 × 20)	93.11 (50 × 12)	94.08 (50 × 38)	95.77 (50 × 38)	97.08 (50 × 37)
6	79.65 (50 × 40)	93.65 (50 × 40)	93.43 (50 × 36)	95.73 (50 × 32)	94.87 (50 × 16)	96.23 (50 × 40)	96.41 (50 × 40)	98.33 (50 × 40)

Table 5. The runtime(s) of eight different algorithms when the number of training samples in the Yale database is 6.

Algorithm	ELPP	MEDLPP	2DPCA	2DLDA	2DLPP	2DDLPP	2DEDLPP	2DESDLPP
Runtime	1.332	2.512	2.258	2.397	2.416	3.985	4.026	3.335

4.4. Experiment in AR Database

The AR database includes 126 people (70 male and 56 female) with a total of more than 3000 frontal face images. The photos of each person were taken under different conditions of expression, lighting and occlusion. It is worth noting that the most significant external factors for the AR database focus on expression changes and facial occlusion, so its use consists mainly of face and expression recognition. Everyone has 15 images, as shown in Figure 9.



Figure 9. Facial image of an individual in the AR face database.

We use AR database photos to test whether the 2DESDLPP algorithm is affected by the changes in human expressions and the occlusion of human faces, so as to evaluate its performance on high-dimensional noisy data. In the experiments, we performed a random selection of l ($l = 2, 3, 4, 5, 6$) images among all images of every person as a training sample set, and the rest are used as a test set, and then recorded 2DESDLPP, as well as the other seven comparison methods', maximum average recognition rates and the number of dimensions in Table 6. Although the accuracy of all experimental algorithms in the AR database is lower than the ORL database and the Yale database due to the challenges of the database itself, no matter how small the number of training samples, the accuracy of 2DESDLPP is always slightly higher than that of the other algorithms. When the number of training samples was two, the highest accuracy of 2DESDLPP was 88.24% when the feature dimension was 24. When the number of training samples was three, the highest accuracy of 2DESDLPP was 89.98% when the feature dimension was 26. When the number of training samples was four, the highest accuracy of 2DESDLPP was 90.75% when the feature dimension was 25. When the training sample size is six, the changes are shown in Figure 10. The average highest recognition rate obtained by 2DESDLPP is 97.56% when the number of training samples is six, and its feature dimension is 20 at this time. The highest recognition rate of 2DESDLPP is about 5.06% higher than that of 2DDLPP, and 1.27 times higher than that of ELPP. Table 7 shows the average running time of each algorithm in the AR database when the number of training samples is six. Although facial occlusion in AR database improves the difficulty of face recognition, resulting in an increase in the running time of the eight algorithms, the operation of the 2DESDLPP algorithm does not take too much time.

Table 6. Maximum average recognition rate (%) of 2DESDLPP and other seven comparison algorithms and the corresponding number of dimensions in AR database.

l	ELPP	MEDLPP	2DPCA	2DLDA	2DLPP	2DDLPP	2DEDLPP	2DESDLPP
2	69.82 (50 × 36)	87.08 (50 × 26)	77.18 (50 × 36)	80.75 (50 × 30)	79.00 (50 × 32)	80.08 (50 × 26)	83.77 (50 × 26)	88.24 (50 × 24)
3	72.75 (50 × 38)	87.61 (50 × 25)	80.67 (50 × 34)	84.38 (50 × 38)	82.93 (50 × 28)	84.61 (50 × 40)	86.20 (50 × 28)	89.98 (50 × 26)
4	74.91 (50 × 38)	88.28 (50 × 23)	83.04 (50 × 38)	86.36 (50 × 40)	85.70 (50 × 30)	87.28 (50 × 20)	89.51 (50 × 32)	90.75 (50 × 25)
5	75.33 (50 × 39)	89.32 (50 × 26)	87.63 (50 × 36)	89.98 (50 × 40)	87.75 (50 × 32)	89.76 (50 × 24)	90.33 (50 × 36)	92.11 (50 × 22)
6	76.80 (50 × 40)	91.50 (50 × 32)	90.06 (50 × 38)	92.04 (50 × 40)	89.12 (50 × 40)	92.80 (50 × 40)	94.65 (50 × 40)	97.56 (50 × 20)

Table 7. The runtime (s) of eight different algorithms when the number of training samples in the AR database is 6.

Algorithm	ELPP	MEDLPP	2DPCA	2DLDA	2DLPP	2DDLPP	2DEDLPP	2DESDLPP
Runtime	4.868	5.865	4.201	4.5385	5.251	6.748	7.027	6.199

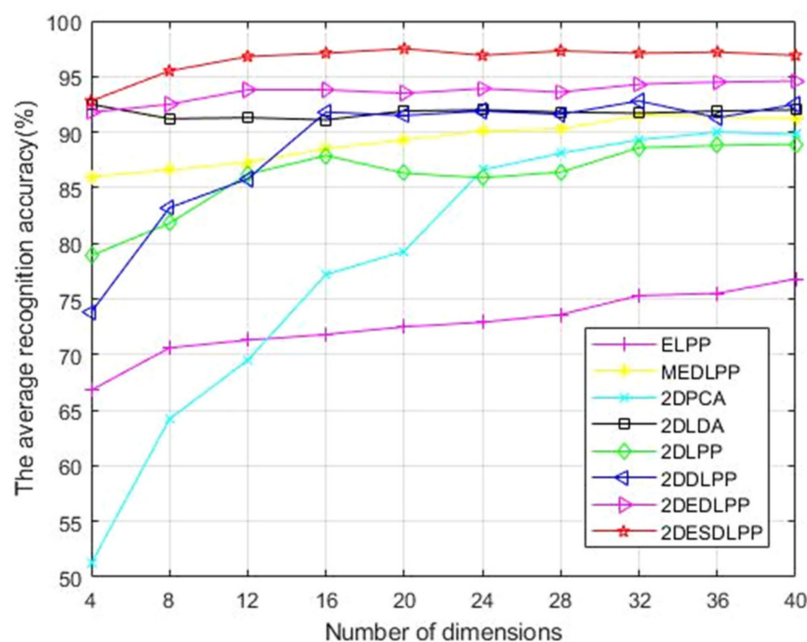


Figure 10. Average recognition rate (%) with feature dimension for different algorithms when the number of training samples in the AR face database is 6.

4.5. Summary of Experimental Results

The following conclusions are drawn by analyzing the results of the experiments on the three public face databases mentioned above:

- (1) From Figures 2–4, we can see that, when $K = 10$, 2DESDLPP achieves the maximum recognition rate in three face databases, ORL, Yale and AR, indicating that 2DESDLPP has the best feature extraction ability at this time.
- (2) As can be seen from the data in Tables 2, 4 and 6, with the increase in training sample size, the maximum average recognition accuracy increases to some extent for most experiments. As can be seen from the data in Tables 3, 5 and 7, the running time of the 2DESDLPP algorithm after sparse representation is shortened, and is within the acceptable range.
- (3) It is not difficult to find that the recognition accuracy of 2DESDLPP outperforms the other 1D algorithms (ELPP and MEDLPP) and 2D algorithms (2DPCA, 2DLDA, 2DLPP, 2DDLPP and 2DEDLPP) for the same training sample size from Figures 8–10.

5. Conclusions

The 2DESDLPP algorithm we propose is an image-based method that uses sample label discrimination information to satisfy the “minimum within-class distance” and “maximum between-class distance” characteristics without destroying the local structural features of the face, while overcoming the SSS problem by using matrix exponential. Then, elastic net regression is used to remove a large amount of redundant information in the face images, and an optimally sparse result is obtained, further mining the features that are more critical for recognition and classification, making the subspace obtained by the algorithm more discriminative than that of other algorithms. In the final analysis, the results of the comparison experiments in ORL, Yale and AR databases show that the 2DESDLPP algorithm has better feature extraction ability.

Author Contributions: Software, Y.Z.; Investigation, M.W.; Writing—original draft, Y.Z.; Writing—review & editing, G.Y.; Project administration, H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by Postgraduate Research & Practice Innovation Program of Jiangsu Province No. SJCX22_0997, the National Science Foundation of China under Grant Nos. 61876213, 61991401, 62172229, 61976118, the Key R&D Program Science Foundation in Colleges and Universities of Jiangsu Province Grant Nos. 20KJA520002, the Natural Science Fund of Jiangsu Province under Grants Nos. BK20201397, BK20191409, BK20211295. Jiangsu Key Laboratory of Image and Video Understanding for Social Safety of Nanjing University of Science and Technology under Grants J2021-4. Funded by the Future Network Scientific Research Fund Project SRFP-2021-YB-25. 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).

Data Availability Statement: Data available on request due to restrictions e.g., privacy or ethical The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: There is no conflict of interest. All the data set in the manuscript come from open source data. There are no copyright issues or ethical issues.

References

1. Barış, B.; Çiğdem, E.E.; Tanju, E. More learning with less labeling for face recognition. *Digit. Signal Process.* **2023**, *136*, 288.
2. Shi, K.; Liu, Z.; Lu, W.; Ou, W.; Yang, C. Unsupervised domain adaptation based on adaptive local manifold learning. *Comput. Electr. Eng.* **2022**, *100*, 107941. [\[CrossRef\]](#)
3. Abdulhussain Sadiq, H.; Mahmmoud Basheera, M.; AlGhadhban, A.; Flusser, J. Face Recognition Algorithm Based on Fast Computation of Orthogonal Moments. *Mathematics* **2022**, *10*, 2721. [\[CrossRef\]](#)
4. 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. [\[CrossRef\]](#)
5. Ishibashi, H.; Higa, K.; Furukawa, T. Multi-task manifold learning for small sample size datasets. *Neurocomputing* **2022**, *473*, 138–157. [\[CrossRef\]](#)
6. He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H.-J. Face recognition using laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 328–340.
7. Chen, S.; Zhao, H.; Kong, M.; Luo, B. 2D-LPP: A two-dimensional extension of locality preserving projections. *Neurocomputing* **2007**, *70*, 912–921. [\[CrossRef\]](#)
8. Zhi, R.; Ruan, Q. Facial expression recognition based on two-dimensional discriminant locality preserving projections. *Neurocomputing* **2008**, *71*, 1730–1734. [\[CrossRef\]](#)
9. Izenman, A.J. Linear Discriminant Analysis. In *Modern Multivariate Statistical Techniques*; Springer: New York, NY, USA, 2013; pp. 237–280.
10. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [\[CrossRef\]](#)
11. Adil, M.; Abid, M.; Khan, A.Q.; Mustafa, G.; Ahmed, N. Exponential discriminant analysis for fault diagnosis. *Neurocomputing* **2016**, *171*, 1344–1353. [\[CrossRef\]](#)
12. Wang, S.J.; Chen, H.L.; Peng, X.-J.; Zhou, C.-G. Exponential locality preserving projections for small sample size problem. *Neurocomputing* **2011**, *74*, 3654–3662. [\[CrossRef\]](#)
13. Dornaika, F.; Bosaghzadeh, A. Exponential local discriminant embedding and its application to face recognition. *IEEE Trans. Cybern.* **2013**, *43*, 921–934. [\[CrossRef\]](#)
14. Chen, H.T.; Chang, H.W.; Liu, T.L. Local discriminant embedding and its variants. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE, Beijing, China, 20–25 June 2005; Volume 2, pp. 846–853.
15. Lu, G.F.; Wang, Y.; Zou, J.; Wang, Z. Matrix exponential based discriminant locality preserving projections for feature extraction. *Neural Netw.* **2018**, *97*, 127–136. [\[CrossRef\]](#)
16. Dornaika, F.; Baradaaji, A.; El Traboulsi, Y. Semi-supervised classification via simultaneous label and discriminant embedding estimation. *Inf. Sci.* **2021**, *546*, 146–165. [\[CrossRef\]](#)
17. 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. [\[CrossRef\]](#)
18. Yan, L. Application of face expression recognition technology in skilled unsupervised course based on ultra-wide regression network. *J. Intell. Fuzzy Syst.* **2020**, *38*, 7167–7177. [\[CrossRef\]](#)
19. 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. [\[CrossRef\]](#)
20. Ankita, W.; Megha, A. Robust pattern for face recognition using combined Weber and pentagonal-triangle graph structure pattern. *Optik* **2022**, *259*, 10282.
21. Zhang, W.; Kang, P.; Fang, X.; Teng, L.; Han, N. Joint sparse representation and locality preserving projection for feature extraction. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 1731–1745. [\[CrossRef\]](#)

22. Liu, N.; Lai, Z.; Li, X.; Chen, Y.; Mo, D.; Kong, H.; Shen, L. Locality preserving robust regression for jointly sparse subspace learning. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 2274–2287. [[CrossRef](#)]
23. Luciano, D.; Steven, Z.; Ian, W. Iterated Adaptive Neighborhoods for Manifold Learning and Dimensionality Estimation. *Neural Comput.* **2023**, *35*, 2982.
24. Xiong, Z.; Wan, M.; Xue, R.; Yang, G. 2D-MELPP: A two dimensional matrix exponential based extension of locality preserving projections for dimensional reduction. *KSII Trans. Internet Inf. Syst.* **2022**, *16*, 2991–3007.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.