

RESEARCH ARTICLE

Robust sparse smooth principal component analysis for face reconstruction and recognition

Jing Wang^{1,2*}, Xiao Xie^{1,2}, Li Zhang³, Jian Li^{1,2}, Hao Cai^{1,2}, Yan Feng^{1,2}

1 School of Computer and Information Technology, Xinyang Normal University, Xinyang, Henan, China, **2** Henan Key Laboratory of Analysis and Applications of Education Big Data, Xinyang Normal University, Xinyang, Henan, China, **3** School of Early-Childhood Education, Nanjing Xiaozhuang University, Nanjing, Jiangsu, China

* wangjing@xynu.edu.cn



OPEN ACCESS

Citation: Wang J, Xie X, Zhang L, Li J, Cai H, Feng Y (2025) Robust sparse smooth principal component analysis for face reconstruction and recognition. PLoS One 20(5): e0323281. <https://doi.org/10.1371/journal.pone.0323281>

Editor: Debo Cheng, University of South Australia, AUSTRALIA

Received: January 30, 2024

Accepted: March 27, 2025

Published: May 27, 2025

Copyright: © 2025 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The dataset required to reproduce the study's findings is available at: https://figshare.com/articles/dataset/Benchmark_face_databases_for_face_recognition_and_reconstruction/27643026. The source code for this study can be accessed at: <https://github.com/yuzhounh/RSSPCA>.

Abstract

Existing Robust Sparse Principal Component Analysis (RSPCA) does not incorporate the two-dimensional spatial structure information of images. To address this issue, we introduce a smooth constraint that characterizes the spatial structure information of images into conventional RSPCA, generating a novel algorithm called Robust Sparse Smooth Principal Component Analysis (RSSPCA). The proposed RSSPCA achieves three key objectives simultaneously: robustness through L1-norm optimization, sparsity for feature selection, and smoothness for preserving spatial relationships. Within the Minorization-Maximization (MM) framework, an iterative process is designed to solve the RSSPCA optimization problem, ensuring that a locally optimal solution is achieved. To evaluate the face reconstruction and recognition performance of the proposed algorithm, we conducted comprehensive experiments on six benchmark face databases. Experimental results demonstrate that incorporating robustness and smoothness improves reconstruction performance, while incorporating sparsity and smoothness improves classification performance. Consequently, the proposed RSSPCA algorithm generally outperforms existing algorithms in face reconstruction and recognition. Additionally, visualization of the generalized eigenfaces provides intuitive insights into how sparse and smooth constraints influence the feature extraction process. The data and source code from this study have been made publicly available on the GitHub repository: <https://github.com/yuzhounh/RSSPCA>.

1. Introduction

Principal Component Analysis (PCA) [1,2] has been widely applied in dimensionality reduction, signal reconstruction, and pattern classification [3–5]. However, traditional PCA adopts L2-norm in the objective function, which is easily affected by noise. Applying L1-norm to the objective function of PCA generates the PCA with L1-norm (PCA-L1) [6], which is robust and can effectively reduce the influence of data noise.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 31900710, Grant 31600862, and Grant 62006205; and in part by the Nanhu Scholars Program for Young Scholars of Xinyang Normal University.

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

Sparsity [7,8] is another important property. Sparse modelling can automatically find relevant features from training data while ignoring irrelevant features. It not only improves the generalization ability of an algorithm but also increases the interpretability of the results. Therefore, sparse modelling has been widely applied in signal processing, machine learning, pattern recognition, and many other fields [9,10]. Traditional PCA cannot extract sparse principal components. To address this issue, L1-norm is applied to the constraint function of traditional PCA, generating the Sparse PCA (SPCA) [11]. Due to the sparsity-promoting property of L1-norm, the principal components extracted by SPCA are sparse.

Inspired by PCA-L1 and SPCA, Robust SPCA (RSPCA) [12] applies L1-norm to both the objective and constraint functions of traditional PCA for simultaneously robust and sparse modelling. However, when processing facial images using PCA and its variants, these images must be reshaped into vectors before further processing, which inevitably leads to the loss of inherent two-dimensional spatial structure information. While Two-dimensional PCA (2DPCA) [13] and its improved variants [14–16] attempt to address this limitation by expressing face images as matrices, they still do not fully capture and utilize the rich spatial structure information present in image data.

Various approaches have been developed to effectively preserve and utilize spatial structure information in image processing tasks. These include texture analysis methods that capture local patterns and regularities [17], graph-based techniques that model relationships between image regions [18,19], geometric approaches that analyze shapes and spatial configurations [20], deep learning architectures specifically designed to maintain spatial correlations [21], etc. These approaches leverage the inherent continuity and gradual variations present in natural images to preserve essential spatial relationships. Among these methods, smoothness-based approaches [22] have demonstrated particular effectiveness in handling data with rich spatial and temporal structural characteristics.

The smooth constraint [23] characterizes the interactions among adjacent features and has achieved remarkable success in brain decoding applications, where preserving spatial and temporal relationships is crucial. It was initially introduced in Electroencephalogram (EEG) decoding [23], where it proved essential for capturing the continuity of brain activity patterns. Subsequently, the combination of sparseness and smoothness was fully investigated by [24]. Its effectiveness led to widespread adoption in various neuroimaging applications, including Magnetoencephalography (MEG) [25], functional Magnetic Resonance Imaging (fMRI) [26–28], and Electrocorticographic (ECoG) [29] decoding. Recently, smoothness was also applied in many other fields, including functional connectivity-based brain region parcellation [30], hyperspectral image classification [31], reconstruction of compressively sensed multichannel EEG signals [32], foreground estimation in neuronal images [33], etc. The success in these applications stems from the constraint's ability to model and preserve the intrinsic spatial and temporal relationships within the data, which is particularly crucial when dealing with complex, structured information.

The implementation of smoothness is typically achieved through a graph Laplacian matrix [34], which effectively represents the spatial and temporal structure information of data. This approach is rooted in spectral graph theory [35,36], where Laplacian eigenmaps capture the intrinsic geometry of high-dimensional data [37,38]. While traditional dimensionality reduction techniques like Laplacian eigenmaps [39,40] have been widely used in machine learning and pattern recognition, they primarily focus on general data representation rather than explicitly preserving the spatial and temporal relationships between adjacent features.

Building upon these insights and the success of smooth constraints in brain decoding [23–33], we propose to incorporate this constraint into conventional RSPCA, resulting in a novel approach termed Robust Sparse Smooth PCA (RSSPCA). This integration addresses a major limitation of RSPCA, i.e., its inability to account for spatial structure information in images. The proposed RSSPCA achieves three key objectives simultaneously: robustness through L1-norm optimization, sparsity for feature selection, and smoothness for preserving spatial relationships. These combined properties make RSSPCA particularly well-suited for face image processing tasks, where preserving spatial structure is crucial for accurate reconstruction and recognition.

The synergy between sparsity and smoothness in RSSPCA is theoretically justified by their complementary nature [24,41]. While sparsity helps identify the most relevant features and reduces noise, smoothness ensures that the spatial coherence of these features is maintained, leading to more naturalistic and interpretable results. This combination has proven particularly effective in applications where both feature selection and structural preservation are important, such as in brain decoding and image processing tasks.

To validate the effectiveness of RSSPCA, we conducted comprehensive experiments on six benchmark face databases, comparing its performance with four competing algorithms in terms of face reconstruction and recognition accuracy. The results demonstrate that RSSPCA significantly outperforms existing methods, confirming the advantages of incorporating smooth constraints into robust sparse PCA frameworks.

The remainder of this paper is organized as follows. Section 2 reviews traditional PCA and its robust and sparse variants. Section 3 presents our proposed RSSPCA methodology, including the problem formulation, related techniques, and iterative solutions. Section 4 demonstrates the effectiveness of our approach through comprehensive experiments on face reconstruction and recognition. Section 5 discusses the limitations of our approach and suggests directions for future research. Finally, Section 6 concludes the paper.

2. Related works

In this paper, lowercase letters represent scalars, boldface lowercase letters represent column vectors, boldface uppercase letters represent matrices; $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_F$ represent L1-norm, L2-norm, and Frobenius norm, respectively; $\text{sign}(\cdot)$ represents the sign function; $\text{diag}(\mathbf{a})$ represents a square diagonal matrix with the elements of vector \mathbf{a} on the main diagonal.

This section reviews the traditional PCA and its robust and sparse variants. For the variants of PCA, we emphasize on finding the first projection vector. After that, multiple projection vectors can be extracted by implementing a deflation scheme.

2.1 PCA

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be n training images where each row is a feature and each column is an image. The images are assumed to be mean-centered, i.e., $\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$. PCA [1,2] finds the first principal component $\mathbf{w} \in \mathbb{R}^{d \times 1}$ by solving the following optimization problem:

$$\max_{\mathbf{w}} \left\| \mathbf{X}^T \mathbf{w} \right\|_2^2, \quad \text{s.t.} \quad \|\mathbf{w}\|_2^2 = 1. \quad (1)$$

The projection vector \mathbf{w} can be obtained by conducting eigen decomposition of the image covariance matrix and preserving the eigenvector with the largest eigenvalue.

2.2 PCA-L1

PCA with L1-norm (PCA-L1) [6] is formulated by replacing the L2-norm in the objective function of PCA with L1-norm. That is, PCA-L1 finds the first projection vector by solving the following optimization problem:

$$\max_{\mathbf{w}} \left\| \mathbf{X}^T \mathbf{w} \right\|_1, \quad \text{s.t.} \quad \|\mathbf{w}\|_2^2 = 1. \quad (2)$$

The projection vector \mathbf{w} can be computed by an iterative procedure. Let k be the iteration number, $\mathbf{w}^{(k)}$ be the projection vector at the k th step, then \mathbf{w} is updated by:

$$\mathbf{v}^{(k)} = \mathbf{X} \text{sign} \left(\mathbf{X}^T \mathbf{w}^{(k)} \right), \quad (3)$$

$$\mathbf{w}^{(k+1)} = \frac{\mathbf{v}^{(k)}}{\|\mathbf{v}^{(k)}\|_2}. \quad (4)$$

By incorporating L1-norm into the objective function of PCA, the resulting PCA-L1 algorithm demonstrates enhanced robustness against the impact of data noise.

2.3 RSPCA

Robust Sparse PCA (RSPCA) [12] is formulated by incorporating the L1-norm into both the objective and constraint functions of PCA. That is, RSPCA finds the first projection vector by solving the following optimization problem:

$$\max_{\mathbf{w}} \left\| \mathbf{X}^T \mathbf{w} \right\|_1, \quad \text{s.t.} \quad \|\mathbf{w}\|_2^2 = 1, \quad \|\mathbf{w}\|_1 \leq c, \quad (5)$$

where c is a positive constant. RSPCA has two different iterative solutions. The first solution [12] only relaxes the objective function and then solves the relaxed optimization problem by soft-thresholding [42]. That is, the projection vector \mathbf{w} is updated by:

$$\mathbf{v}^{(k)} = \mathbf{X} \text{sign} \left(\mathbf{X}^T \mathbf{w}^{(k)} \right), \quad (6)$$

$$\mathbf{u}^{(k)} = \text{soft} \left(\mathbf{v}^{(k)}, \lambda \right) = \text{sign} \left(\mathbf{v}^{(k)} \right) \left(|\mathbf{v}^{(k)}| - \lambda \right)_+, \quad (7)$$

$$\mathbf{w}^{(k+1)} = \frac{\mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|_2}, \quad (8)$$

where $\text{soft}(\mathbf{v}, \lambda)$ is the soft-thresholding operator, $(x)_+ = \max(x, 0)$, λ is the Lagrange multiplier. The $\text{sign}(\cdot)$ and $(\cdot)_+$ functions can be applied on a vector in an elementwise manner. Inspired by [14,15], RSPCA can also be solved by simultaneously relaxing the objective and constraint functions. That is, the projection vector \mathbf{w} is updated by:

$$\mathbf{v}^{(k)} = \mathbf{X} \text{sign} \left(\mathbf{X}^T \mathbf{w}^{(k)} \right), \quad (9)$$

$$u_i^{(k)} = v_i^{(k)} \frac{|w_i^{(k)}|}{\lambda + |w_i^{(k)}|}, \quad i = 1, 2, \dots, d, \quad (10)$$

$$\mathbf{w}^{(k+1)} = \frac{\mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|_2}. \quad (11)$$

where $w_i^{(k)}$, $v_i^{(k)}$, and $u_i^{(k)}$ are the i th elements of $\mathbf{w}^{(k)}$, $\mathbf{v}^{(k)}$, and $\mathbf{u}^{(k)}$, respectively. The parameter λ is a nonnegative scalar that adjusts the relative weight between the sparse constraint and the L2-norm constraint. When λ is set to zero, the

L1-norm constraint becomes invalid and RSPCA reduces to PCA-L1. When λ increases, the sparsity of the projection vector \mathbf{w} increases. When λ is set to positive infinity, the L2-norm constraint becomes invalid. As a result, the projection vector \mathbf{w} becomes a vector with only one nonzero element that equals one [15,43]. This is the reason why the L2-norm constraint is reserved in RSPCA [12] and its 2D counterpart, i.e., 2DPCA-L1 with Sparsity (2DPCAL1-S) [14]. Due to the L1-norm in the objective and constraint functions, RSPCA achieves robustness and sparseness simultaneously.

After obtaining the first r projection vectors $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r] \in \mathbb{R}^{d \times r}$ for PCA, PCA-L1, or RSPCA, $1 \leq r < d$, the $(r+1)$ th projection vector \mathbf{w}_{r+1} is calculated likewise on the deflated sample [44]

$$\mathbf{x}^{\text{deflated}} = (\mathbf{I} - \mathbf{W}\mathbf{W}^T) \mathbf{x}. \quad (12)$$

By iteratively implementing the deflation procedure, we can extract multiple projection vectors [6,12].

3. Proposed methodology

This section presents our proposed Robust Structured Sparse PCA (RSSPCA) algorithm. We first formulate the optimization problem and define the objective function. Then, we introduce the Laplacian matrix and its associated smoothness penalty term to capture spatial relationships. Next, we elaborate on the Minorization-Maximization (MM) framework that underpins our solution approach, followed by two essential inequalities that facilitate our theoretical derivation. Finally, based on the MM framework and the introduced inequalities, we develop a complete iterative solution to the RSSPCA optimization problem.

3.1 Problem formulation

Based on the robust and sparse PCA algorithms presented above, we propose RSSPCA by incorporating a smooth constraint. That is, RSSPCA finds its first projection vector by solving the following optimization problem:

$$\max_{\mathbf{w}} \|\mathbf{X}^T \mathbf{w}\|_1, \quad \text{s.t.} \quad \|\mathbf{w}\|_2^2 = 1, \quad \|\mathbf{w}\|_1 \leq c_1, \quad \mathbf{w}^T \mathbf{L} \mathbf{w} \leq c_2, \quad (13)$$

where c_1 and c_2 are positive constants, \mathbf{L} is a Laplacian matrix representing the two-dimensional spatial structure information of images.

The optimization problem of RSSPCA has four parts. The first part $\|\mathbf{X}^T \mathbf{w}\|_1$ is the objective function. The L1-norm in the objective function makes RSSPCA robust against noise. The second part $\|\mathbf{w}\|_2^2 = 1$ is the L2-norm constraint. It is reserved for the same reason as in RSPCA [12] and 2DPCAL1-S [14]. The third part $\|\mathbf{w}\|_1 \leq c_1$ is the sparse constraint, which makes the projection vector \mathbf{w} sparse. The fourth part $\mathbf{w}^T \mathbf{L} \mathbf{w} \leq c_2$ is the smooth constraint, which makes the projection vector \mathbf{w} spatially smooth. Therefore, RSSPCA is modelled to be robust, sparse, and smooth at the same time. It is anticipated to offer superior performance in face reconstruction and recognition tasks when compared to PCA and its variants.

3.2 Laplacian matrix

The Laplacian Eigenmaps approach [35,36] was initially proposed to capture the intrinsic geometric structure of low-dimensional manifolds. However, in the optimization problem of RSSPCA, we employ the Laplacian matrix to represent the two-dimensional spatial structure information inherent in images and subsequently construct the smoothness constraint. The definition of the Laplacian matrix is elaborated as follows. Suppose there are two pixels P_i and P_j on an image. Their coordinates on the two-dimensional plane are represented as (x_i, y_i) and (x_j, y_j) , respectively, as illustrated in Fig 1.

The Euclidean distance between the two pixels is

$$d_{ij} = \|(x_i, y_i) - (x_j, y_j)\|_2. \quad (14)$$

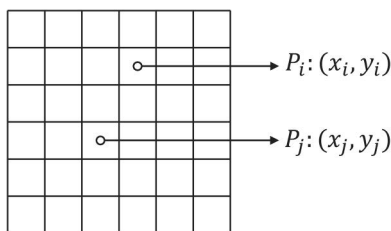


Fig 1. An illustration of the coordinates of two pixels in an image.

<https://doi.org/10.1371/journal.pone.0323281.g001>

Define an adjacency matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ whose elements are

$$A_{ij} = \begin{cases} 1, & d_{ij} = 1, \\ 0, & \text{others}, \end{cases} \quad (15)$$

$i, j = 1, 2, \dots, d$. That is, the corresponding element in the adjacency matrix \mathbf{A} is one if and only if two pixels are adjacent on the two-dimensional plane, otherwise it is 0. The degree matrix \mathbf{D} is defined based on the adjacency matrix as:

$$\mathbf{D} = \text{diag}(\mathbf{1}_d^T \mathbf{A}) \quad , \quad (16)$$

where $\mathbf{1}_d \in \mathbb{R}^{d \times 1}$ represents a d -dimensional column vector with all elements equal to one. The degree matrix \mathbf{D} is a diagonal matrix. Each element on the main diagonal of \mathbf{D} equals the number of pixels adjacent to the current pixel. Then the Laplacian matrix \mathbf{L} is calculated based on the adjacency matrix and the degree matrix as

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad . \quad (17)$$

Fig 2 shows examples of the adjacency matrix, degree matrix, and Laplacian matrix constructed by reshaping a 6×6 image into a vector.

According to the definition of Laplacian matrix, we have

$$\mathbf{w}^T \mathbf{L} \mathbf{w} = \mathbf{w}^T (\mathbf{D} - \mathbf{A}) \mathbf{w} = \mathbf{w}^T (\text{diag}(\mathbf{1}_d^T \mathbf{A}) - \mathbf{A}) \mathbf{w} = \frac{1}{2} \sum_{i,j=1}^d (w_i - w_j)^2 A_{ij} \quad . \quad (18)$$

For two adjacent pixels i and j , A_{ij} equals one, otherwise A_{ij} equals zero. By incorporating this constraint, the difference between w_i and w_j is punished. That is, the weights corresponding to two adjacent pixels will be close. Therefore, this constraint achieves a smoothing effect on the projection vector.

3.3 Minorization-Maximization framework

The optimization problem of RSSPCA is challenging to solve due to the presence of L1-norm. This paper adopts the Minorization-Maximization (MM) framework [45] to address this issue. Suppose $f(\mathbf{w})$ is the objective function to be maximized. Within the MM framework, if there exists a surrogate function $g(\mathbf{w}|\mathbf{w}^{(k)})$ that satisfies the following two key conditions:

$$f(\mathbf{w}^{(k)}) = g(\mathbf{w}^{(k)}|\mathbf{w}^{(k)}) \quad , \quad (19)$$

$$f(\mathbf{w}) \geq g(\mathbf{w}|\mathbf{w}^{(k)}) \quad , \quad \forall \mathbf{w} \quad , \quad (20)$$

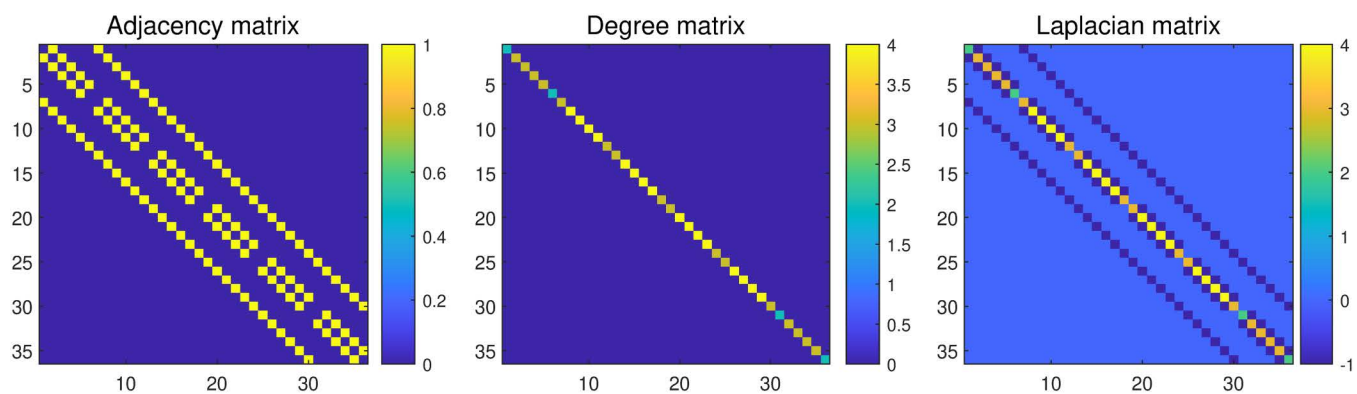


Fig 2. An illustration of the adjacency matrix, degree matrix, and Laplacian matrix.

<https://doi.org/10.1371/journal.pone.0323281.g002>

the original objective function $f(\mathbf{w})$ can be optimized by iteratively maximizing the surrogate function $g(\mathbf{w}|\mathbf{w}^{(k)})$ as follows:

$$\mathbf{w}^{(k+1)} = \arg \max_{\mathbf{w}} g(\mathbf{w}|\mathbf{w}^{(k)}). \quad (21)$$

Then we have

$$\begin{aligned} f(\mathbf{w}^{(k+1)}) &= f(\mathbf{w}^{(k+1)}) - g(\mathbf{w}^{(k+1)}|\mathbf{w}^{(k)}) + g(\mathbf{w}^{(k+1)}|\mathbf{w}^{(k)}) \\ &\geq f(\mathbf{w}^{(k)}) - g(\mathbf{w}^{(k)}|\mathbf{w}^{(k)}) + g(\mathbf{w}^{(k+1)}|\mathbf{w}^{(k)}) \\ &\geq f(\mathbf{w}^{(k)}) - g(\mathbf{w}^{(k)}|\mathbf{w}^{(k)}) + g(\mathbf{w}^{(k)}|\mathbf{w}^{(k)}) \\ &= f(\mathbf{w}^{(k)}). \end{aligned} \quad (22)$$

The first inequality holds because $f(\mathbf{w}) - g(\mathbf{w}|\mathbf{w}^{(k)})$ reaches the minimum value at $\mathbf{w} = \mathbf{w}^{(k)}$ according to the two key conditions. The second inequality holds because $g(\mathbf{w}|\mathbf{w}^{(k)})$ reaches the maximum value at $\mathbf{w} = \mathbf{w}^{(k+1)}$ according to the update rule. Therefore, the objective function $f(\mathbf{w})$ monotonically increases during the iterative process and will converge to a local optimum. By finding a surrogate function that is easy to handle, we can solve the optimization problem of RSSPCA within the MM framework.

3.4 Inequality

The surrogate function is typically formulated by introducing inequalities. Below are two inequalities that will be used to solve RSSPCA. Let $\mathbf{X} \in \mathbb{R}^{d \times n}$, $\mathbf{w} \in \mathbb{R}^{d \times 1}$, $\mathbf{w}^{(k)} \in \mathbb{R}^{d \times 1}$, the following two inequalities

$$\|\mathbf{X}^T \mathbf{w}\|_1 \geq \text{sign}(\mathbf{X}^T \mathbf{w}^{(k)})^T \mathbf{X}^T \mathbf{w}, \quad (23)$$

$$\|\mathbf{w}\|_1 \leq \frac{1}{2} \mathbf{w}^T \text{diag}(|\mathbf{w}^{(k)}|^{-1}) \mathbf{w} + \frac{1}{2} \|\mathbf{w}^{(k)}\|_1, \quad (24)$$

hold and the inequalities become equalities when $\mathbf{w} = \mathbf{w}^{(k)}$. The proofs for the two inequalities can be found in Wang [15]. Note that Equation (24) requires that $\mathbf{w}^{(k)}$ must not contain zero elements.

3.5 Solution

Based on the MM framework and the introduced inequalities, we can develop an iterative solution to the RSSPCA optimization problem. Maximizing the optimization problem of RSSPCA equals maximizing the Lagrangian as follows:

$$\max_{\mathbf{w}} \left\| \mathbf{X}^T \mathbf{w} \right\|_1 - \lambda_1 \left(\left\| \mathbf{w} \right\|_2^2 - 1 \right) - \lambda_2 \left(\left\| \mathbf{w} \right\|_1 - c_1 \right) - \lambda_3 \left(\mathbf{w}^T \mathbf{L} \mathbf{w} - c_2 \right), \quad (25)$$

where λ_1 , λ_2 , and λ_3 are three Lagrangian multipliers satisfying that $\lambda_1 > 0$, $\lambda_2 \geq 0$, and $\lambda_3 \geq 0$. Denote the Lagrangian as $f(\mathbf{w})$. According to [Equations \(23\)](#) and [\(24\)](#),

$$\begin{aligned} & \left\| \mathbf{X}^T \mathbf{w} \right\|_1 - \lambda_1 \left(\left\| \mathbf{w} \right\|_2^2 - 1 \right) - \lambda_2 \left(\left\| \mathbf{w} \right\|_1 - c_1 \right) - \lambda_3 \left(\mathbf{w}^T \mathbf{L} \mathbf{w} - c_2 \right) \\ & \geq \text{sign} \left(\mathbf{X}^T \mathbf{w}^{(k)} \right)^T \mathbf{X}^T \mathbf{w} - \lambda_1 \left(\left\| \mathbf{w} \right\|_2^2 - 1 \right) \\ & \quad - \lambda_2 \left(\frac{1}{2} \mathbf{w}^T \text{diag} \left(\left| \mathbf{w}^{(k)} \right|^{-1} \right) \mathbf{w} + \frac{1}{2} \left\| \mathbf{w}^{(k)} \right\|_1 - c_1 \right) - \lambda_3 \left(\mathbf{w}^T \mathbf{L} \mathbf{w} - c_2 \right) \end{aligned} \quad (26)$$

holds and the inequality becomes equality when $\mathbf{w} = \mathbf{w}^{(k)}$. Denote the relaxed function as $g(\mathbf{w} | \mathbf{w}^{(k)})$. We have $f(\mathbf{w}^{(k)}) = g(\mathbf{w}^{(k)} | \mathbf{w}^{(k)})$ and $f(\mathbf{w}) \geq g(\mathbf{w} | \mathbf{w}^{(k)})$ for all \mathbf{w} , satisfying the two key conditions of the MM framework. Therefore, $g(\mathbf{w} | \mathbf{w}^{(k)})$ is a feasible surrogate function of the Lagrangian $f(\mathbf{w})$. Within the MM framework, maximizing $f(\mathbf{w})$ can be turned into iteratively maximizing the surrogate function $g(\mathbf{w} | \mathbf{w}^{(k)})$ as follows:

$$\begin{aligned} \mathbf{w}^{(k+1)} &= \arg \max_{\mathbf{w}} g(\mathbf{w} | \mathbf{w}^{(k)}) \\ &= \arg \max_{\mathbf{w}} -\mathbf{w}^T \left(\lambda_1 \mathbf{I} + \frac{\lambda_2}{2} \text{diag} \left(\left| \mathbf{w}^{(k)} \right|^{-1} \right) + \lambda_3 \mathbf{L} \right) \mathbf{w} \\ & \quad + \left(\mathbf{X} \text{sign} \left(\mathbf{X}^T \mathbf{w}^{(k)} \right) \right)^T \mathbf{w} - \frac{\lambda_2}{2} \left\| \mathbf{w}^{(k)} \right\|_1 + \lambda_1 + \lambda_2 c_1 + \lambda_3 c_2. \end{aligned} \quad (27)$$

It is a quadratic optimization problem concerning to \mathbf{w} . Its solution is

$$\mathbf{w}^{(k+1)} = \frac{1}{2} \left(\lambda_1 \mathbf{I} + \frac{\lambda_2}{2} \mathbf{U}^{(k)-1} + \lambda_3 \mathbf{L} \right)^{-1} \mathbf{v}^{(k)}, \quad (28)$$

where $\mathbf{U}^{(k)} = \text{diag} \left(\left| \mathbf{w}^{(k)} \right| \right)$, $\mathbf{v}^{(k)} = \mathbf{X} \text{sign} \left(\mathbf{X}^T \mathbf{w}^{(k)} \right)$. Considering that $\left\| \mathbf{w} \right\|_2^2 = 1$, we have

$$\mathbf{w}^{(k+1)} = \frac{\left(\lambda_1 \mathbf{I} + \frac{\lambda_2}{2} \mathbf{U}^{(k)-1} + \lambda_3 \mathbf{L} \right)^{-1} \mathbf{v}^{(k)}}{\left\| \left(\lambda_1 \mathbf{I} + \frac{\lambda_2}{2} \mathbf{U}^{(k)-1} + \lambda_3 \mathbf{L} \right)^{-1} \mathbf{v}^{(k)} \right\|_2}. \quad (29)$$

Let $\eta_1 = \frac{\lambda_2}{2\lambda_1}$ and $\eta_2 = \frac{\lambda_3}{\lambda_1}$, then $\eta_1 \geq 0$, $\eta_2 \geq 0$. [Equation \(29\)](#) can be rewritten as:

$$\mathbf{w}^{(k+1)} = \frac{\left(\mathbf{I} + \eta_1 \mathbf{U}^{(k)-1} + \eta_2 \mathbf{L} \right)^{-1} \mathbf{v}^{(k)}}{\left\| \left(\mathbf{I} + \eta_1 \mathbf{U}^{(k)-1} + \eta_2 \mathbf{L} \right)^{-1} \mathbf{v}^{(k)} \right\|_2}. \quad (30)$$

Therefore, it is only necessary to tune two parameters, i.e., η_1 and η_2 , in RSSPCA. The parameter η_1 tunes the relative weight between the sparse constraint and the L2-norm constraint. The parameter η_2 tunes the relative weight between the smooth constraint and the L2-norm constraint. In summary, the two parameters adjust the relative weights of the three constraints in the optimization problem of RSSPCA.

Considering that $\mathbf{w}^{(k)}$ might be a sparse vector, calculating $\mathbf{U}^{(k)-1}$ could encounter division by zero errors. To avoid this problem, we rewrite Equation (30) as follows:

$$\mathbf{w}^{(k+1)} = \frac{(\mathbf{U}^{(k)} + \eta_1 \mathbf{I} + \eta_2 \mathbf{U}^{(k)} \mathbf{L})^{-1} \mathbf{U}^{(k)} \mathbf{v}^{(k)}}{\|(\mathbf{U}^{(k)} + \eta_1 \mathbf{I} + \eta_2 \mathbf{U}^{(k)} \mathbf{L})^{-1} \mathbf{U}^{(k)} \mathbf{v}^{(k)}\|_2} . \quad (31)$$

This update rule no longer requires that there are no zero elements in $\mathbf{w}^{(k)}$. Equation (31) can be reformulated in the following two-step form:

$$\mathbf{u}^{(k)} = (\mathbf{U}^{(k)} + \eta_1 \mathbf{I} + \eta_2 \mathbf{U}^{(k)} \mathbf{L})^{-1} \mathbf{U}^{(k)} \mathbf{v}^{(k)} , \quad (32)$$

$$\mathbf{w}^{(k+1)} = \frac{\mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|_2} . \quad (33)$$

Update \mathbf{w} iteratively until convergence, the result is the first projection vector of RSSPCA. Then, multiple projection vectors can be extracted likewise by iteratively implementing the deflation strategy [44]. The algorithm procedure of RSSPCA is listed in Algorithm 1.

Algorithm 1. The algorithm procedure of RSSPCA.

Input: training samples $\mathbf{X} \in \mathbb{R}^{d \times n}$, number of projection vectors r , parameters η_1 and η_2 .

Output: projection vector $\mathbf{W} \in \mathbb{R}^{d \times r}$.

```

for  $i = 1 : r$ 
  initialize  $k = 0, \delta = 1$ .
  Initialize  $\mathbf{w}^{(0)}$  by the  $i$ th principal component.
   $f^{(0)} = \|\mathbf{X}^T \mathbf{w}^{(0)}\|_1$ .
  while  $\delta > 10^{-4}$  and  $k < 100$ 
     $\mathbf{U}^{(k)} = \text{diag}(|\mathbf{w}^{(k)}|)$ .
     $\mathbf{v}^{(k)} = \mathbf{X} \text{sign}(\mathbf{X}^T \mathbf{w}^{(k)})$ .
     $\mathbf{u}^{(k)} = (\mathbf{U}^{(k)} + \eta_1 \mathbf{I} + \eta_2 \mathbf{U}^{(k)} \mathbf{L})^{-1} \mathbf{U}^{(k)} \mathbf{v}^{(k)}$ .
     $\mathbf{w}^{(k+1)} = \frac{\mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|_2}$ .
     $f^{(k+1)} = \|\mathbf{X}^T \mathbf{w}^{(k+1)}\|_1$ .
     $\delta = \frac{|f^{(k+1)} - f^{(k)}|}{f^{(k)}}$ .
     $k \leftarrow k + 1$ .
  end while
   $\mathbf{w}_i = \mathbf{w}^{(k+1)}$ .
   $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_i] \in \mathbb{R}^{d \times i}$ .
   $\mathbf{X} \leftarrow (\mathbf{I} - \mathbf{W} \mathbf{W}^T) \mathbf{X}$ .
end for

```

<https://doi.org/10.1371/journal.pone.0323281.t001>

From the optimization problem and the update rule of RSSPCA, it can be inferred that RSSPCA is a generalization of PCA-L1 and RSPCA. When $\eta_1 = 0$ and $\eta_2 = 0$, RSSPCA reduces to PCA-L1. When $\eta_2 = 0$, RSSPCA reduces to RSPCA. When $\eta_1 = 0$, RSSPCA reduces to Robust Smooth PCA (RSMPCA).

4. Experiments

This section presents comprehensive experimental evaluations of our proposed RSSPCA algorithm. We first assessed the reconstruction and recognition performance of RSSPCA against four competing algorithms in face reconstruction and recognition tasks. Subsequently, we visualized the projection vectors computed by different algorithms to provide intuitive insights into their underlying characteristics. Finally, we validated our approach on five additional publicly available benchmark face databases to demonstrate its robustness and generalizability across diverse datasets. The flowchart of the experiments is shown in Fig 3.

Our scripts were written in MATLAB and are publicly available at <https://github.com/yuzhounh/RSSPCA>. The experiments were conducted on four workstations, each equipped with dual 20-core 2.20 GHz Intel(R) Xeon(R) processors and 256 GB of memory. To minimize total computational time, we ran approximately 40 MATLAB sessions in parallel on each workstation. A parallel computing version of the code is available at https://github.com/yuzhounh/RSSPCA_parallel.

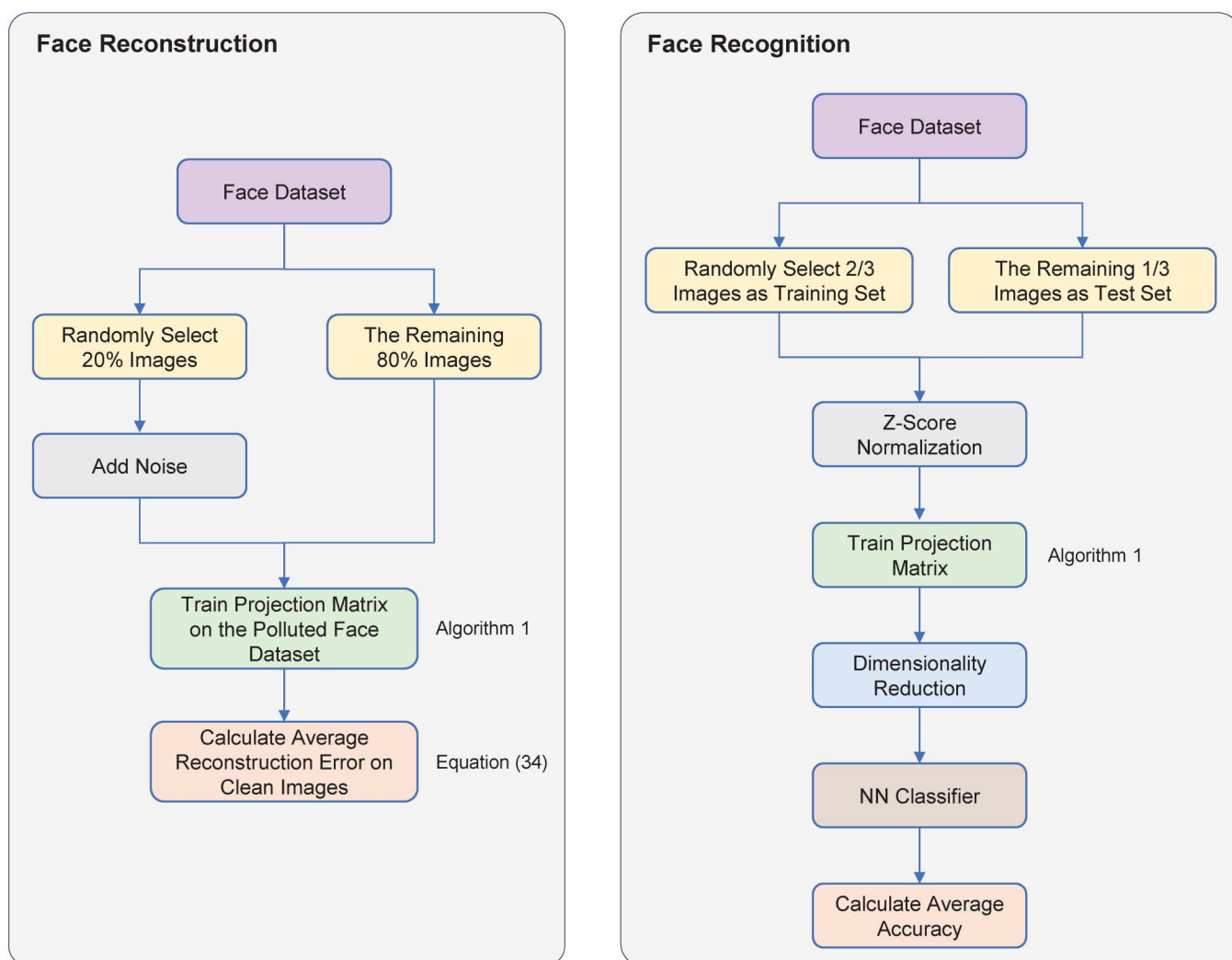


Fig 3. The flowchart of the experiments. The training of the projection matrix can be performed using the proposed RSSPCA method or any other competing algorithms.

<https://doi.org/10.1371/journal.pone.0323281.g003>

4.1 Face reconstruction

We first conducted a face reconstruction experiment to evaluate the reconstruction performance of RSSPCA and the four competing algorithms on the publicly available ORL Face Database [46,47], which was created and distributed by AT&T Laboratories Cambridge for research purposes. Four typical PCA-based algorithms, i.e., PCA [1,2], PCA-L1 [6], RSPCA [12], and RSMPCA, were compared with RSSPCA in the experiment.

The ORL face database contains 400 face images from 40 subjects, with 10 images per subject. The images were captured with different facial expressions, rotations, and slight scale variations. The original image size is 112 by 92. To reduce computational time, we further resized the images to 56 by 46. Among the 400 images in the ORL face database, 80 images were randomly selected and occluded with a rectangular salt-and-pepper noise whose size was not smaller than 10 by 10, located in a random position

Let \mathbf{W} be the projection matrix trained on the polluted ORL face database, which includes 320 clean images and 80 occluded images. Let $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m$ be m clean images that are mean-centered, $m=320$. The average reconstruction error is defined as

$$\frac{1}{m} \sum_{i=1}^m \left\| \mathbf{Z}_i (\mathbf{I} - \mathbf{W}\mathbf{W}^T) \right\|_F. \quad (34)$$

It is used to evaluate the reconstruction performance of the five algorithms.

Fig 4 shows the reconstruction errors of PCA and PCA-L1 with different numbers of projection vectors. In both cases, the reconstruction error monotonically decreases with increasing number of projection vectors. When the number of projection vectors is greater than 7, the reconstruction error of PCA-L1 is lower than that of PCA. By averaging the reconstruction errors with different projection vector numbers that are in the range of [1,30], we obtain the overall reconstruction errors of PCA and PCA-L1. They are 1272.33 and 1203.31, respectively. The results demonstrate that incorporating L1-norm into the objective function of PCA, i.e., incorporating robustness into PCA, can reduce reconstruction errors, consistent with the results in [6,14,15].

RSPCA is a special case of RSSPCA when $\eta_2 = 0$. Therefore, the projection vectors of RSPCA can be calculated by the update rule of RSSPCA. RSPCA only contains a parameter η_1 that tunes the relative weight between the sparse constraint and the L2-norm constraint. The parameters η_1 was selected from $\{10^{-3}, 10^{-2.8}, 10^{-2.6}, \dots, 10^3\}$. That is, $\lg(\eta_1)$ was

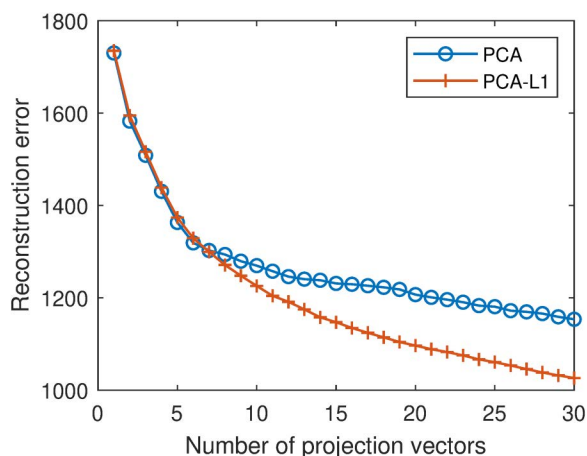


Fig 4. Reconstruction errors of PCA and PCA-L1 with different numbers of projection vectors.

<https://doi.org/10.1371/journal.pone.0323281.g004>

selected from -3 – 3 with a step of 0.2 . For each η_1 value, we averaged the reconstruction errors with different numbers of projection vectors to obtain the overall reconstruction errors, as shown in Fig 5. The overall reconstruction error generally increases with increasing $\lg(\eta_1)$ value. The lowest reconstruction error is 1205.41 , obtained when $\lg(\eta_1) = -3.0$, very close to the reconstruction error of PCA-L1, which is 1203.31 . From this result and the trend in Fig 5, we can infer that the lowest reconstruction error of RSPCA is achieved when $\eta_1 = 0$. In other words, RSPCA achieves the lowest reconstruction error when it reduces to PCA-L1. The results demonstrate that incorporating sparsity has little effect on reconstruction, consistent with the results in [14,15].

RSMPCA is a special case of RSSPCA when $\eta_1 = 0$. Similarly, the projection vectors of RSMPCA can be calculated by the update rule of RSSPCA. RSMPCA contains a parameter η_2 that tunes the relative weight between the smooth constraint and the L2-norm constraint. The parameter η_2 was selected from $\{10^{-3}, 10^{-2.8}, 10^{-2.6}, \dots, 10^3\}$. That is, $\lg(\eta_2)$ was selected from -3 – 3 with a step of 0.2 . For each η_2 value, we averaged the reconstruction errors with different projection

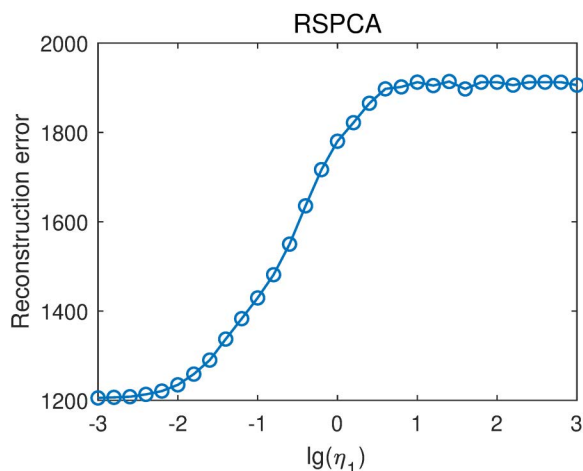


Fig 5. Reconstruction errors of RSPCA with different $\lg(\eta_1)$ values.

<https://doi.org/10.1371/journal.pone.0323281.g005>

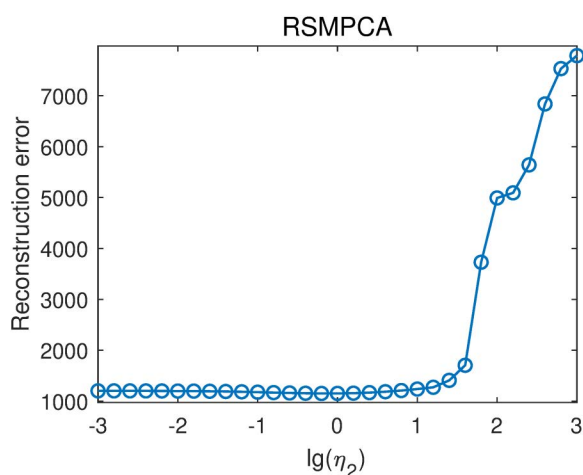


Fig 6. Reconstruction errors of RSMPCA with different $\lg(\eta_2)$ values.

<https://doi.org/10.1371/journal.pone.0323281.g006>

vector numbers to obtain the overall reconstruction errors, as shown in Fig 6. When $\lg(\eta_2) \leq 1$, the reconstruction error was barely affected by $\lg(\eta_2)$. When $\lg(\eta_2) > 1$, the reconstruction error significantly increases with increasing $\lg(\eta_2)$ value. The lowest reconstruction error is 1152.25, obtained when $\lg(\eta_2) = -0.2$. This result is lower than the reconstruction error of PCA-L1, indicating that incorporating smoothness improves reconstruction performance.

RSSPCA contains two parameters η_1 and η_2 , among which η_1 tunes the relative weight between the sparse constraint and the L2-norm constraint, and η_2 tunes the relative weight between the smooth constraint and the L2-norm constraint. For RSSPCA, both $\lg(\eta_1)$ and $\lg(\eta_2)$ were selected from -3–3 with a step of 0.2. The average reconstruction errors of RSSPCA under different parameters are shown in Fig 7. The lowest reconstruction error is 1153.35, obtained when $\lg(\eta_1) = -3.0$ and $\lg(\eta_2) = 0$.

The lowest reconstruction errors and corresponding optimal parameters of the five algorithms are shown in Table 1. Fig 8 shows the reconstruction errors of the five algorithms with different number of projection vectors when the optimal parameters are selected.

Based on the above results, it is evident that the reconstruction errors of PCA are the largest. PCA-L1 incorporates L1-norm into the objective function of PCA and achieves lower reconstruction errors than PCA, indicating that incorporating robustness is beneficial for reconstruction. RSPCA incorporates the sparse constraint into PCA-L1. The reconstruction errors of PCA-L1 and RSPCA are very close, indicating that incorporating sparsity has little effect on reconstruction. These results are consistent with the results in [6,14,15]. RSMPCA incorporates the smooth constraint into PCA-L1 and achieves lower reconstruction errors than PCA-L1, indicating that incorporating smoothness is beneficial for

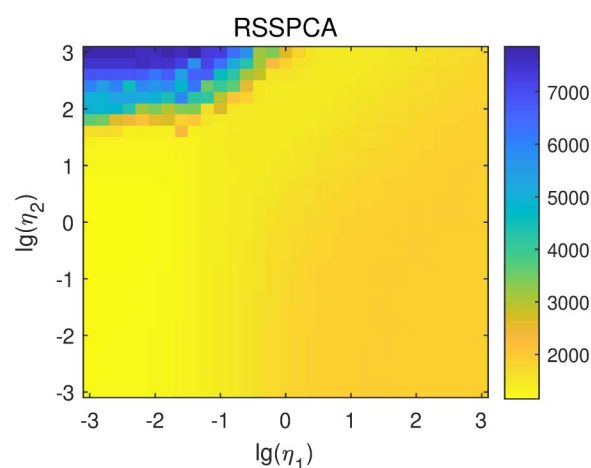


Fig 7. Reconstruction errors of RSSPCA with different $\lg(\eta_1)$ and $\lg(\eta_2)$ values.

<https://doi.org/10.1371/journal.pone.0323281.g007>

Table 1. The lowest reconstruction errors and optimal parameters of the five algorithms.

Algorithm	Optimal parameters		Reconstruction error
	$\lg(\eta_1)$	$\lg(\eta_2)$	
PCA	/	/	1272.33
PCA-L1	/	/	1203.31
RSPCA	-3.0	/	1205.41
RSMPCA	/	-0.2	1152.25
RSSPCA	-3.0	0.0	1153.35

<https://doi.org/10.1371/journal.pone.0323281.t002>

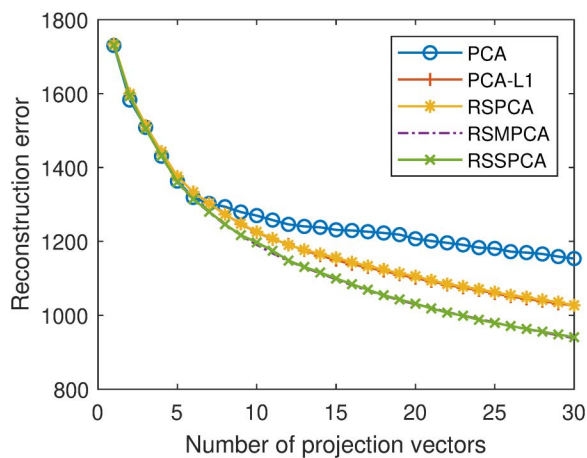


Fig 8. Reconstruction errors of the five algorithms with different numbers of projection vectors. The parameters are set to the optimal ones for each algorithm.

<https://doi.org/10.1371/journal.pone.0323281.g008>

reconstruction. RSSPCA simultaneously incorporates the sparse constraint and the smooth constraint into PCA-L1. It is not surprising that RSSPCA achieves lower reconstruction errors than PCA-L1.

In addition, RSSPCA can be generated by incorporating the smooth constraint into RSPCA. With this adaptation, RSSPCA achieves lower reconstruction errors than RSPCA. It further demonstrates that incorporating smoothness is beneficial for reconstruction. From another perspective, RSSPCA can be generated by incorporating the sparse constraint into RSMPCA. But the reconstruction errors of RSSPCA and RSMPCA are very close. It further demonstrates that incorporating sparsity has little effect on reconstruction.

In summary, incorporating robustness and smoothness reduces reconstruction errors, while incorporating sparsity has little effect on reconstruction. These results demonstrate the superiority of RSSPCA over PCA, PCA-L1, and RSPCA in face reconstruction. When comparing RSSPCA and RSMPCA, they have similar reconstruction performance.

4.2 Face recognition

Next, we conducted a face recognition experiment to evaluate the classification performance of RSSPCA and the four competing algorithms. For each subject in the ORL face database, we randomly selected seven images for training and used the remaining three images for testing. The training images were normalized by z-score so that each feature was centered to have a mean of zero and scaled to have a standard deviation of one. The testing images were then normalized by applying the same parameters. After that, we applied the five algorithms to extract the first 30 projection vectors and applied these projection vectors to reduce the dimension of the normalized data. Finally, Nearest Neighbor (NN) classifier was applied to perform classification. The above procedure was repeated three times, and the average classification accuracy was calculated to evaluate the classification performance of the five algorithms.

Figs 9 and 10 show the average classification accuracies of PCA and PCA-L1 with different numbers of projection vectors. When the number of projection vectors increases, the classification accuracy generally increases. When the number of projection vectors is greater than 10, the classification accuracy increases slowly. By averaging the classification accuracies with different projection vector numbers that are in the range of [1,30], we obtain the overall classification accuracies of PCA and PCA-L1. They are 0.8556 and 0.8552, respectively. The two results are very close, which means that the classification performances of PCA and PCA-L1 are very close. It indicates that incorporating L1-norm into the objective function of PCA, i.e., incorporating robustness into PCA, has little effect on the classification performance, consistent with the results in [14,15].

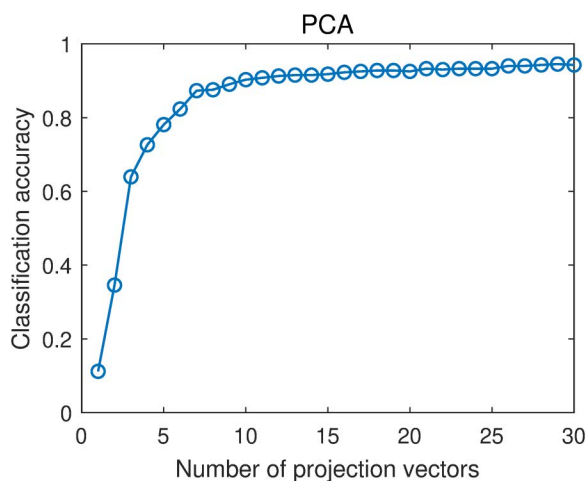


Fig 9. Average classification accuracies of PCA with different numbers of projection vectors.

<https://doi.org/10.1371/journal.pone.0323281.g009>

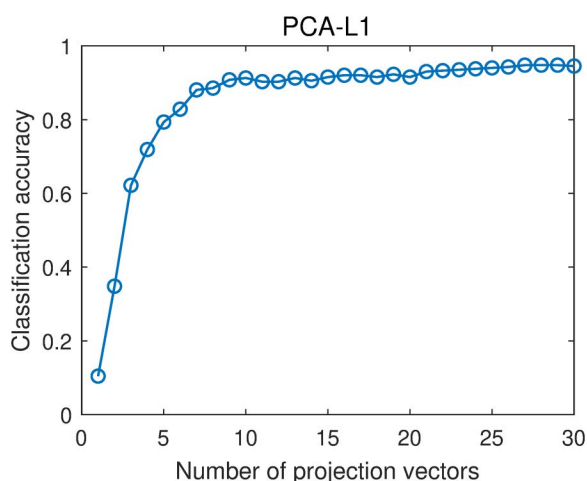


Fig 10. Average classification accuracies of PCA-L1 with different numbers of projection vectors.

<https://doi.org/10.1371/journal.pone.0323281.g010>

For RSPCA, the parameters η_1 was selected from $\{10^{-6}, 10^{-5.9}, 10^{-5.8}, \dots, 10^6\}$. That is, $\lg(\eta_1)$ was selected in the range of -6 – 6 with a step of 0.1 . For each η_1 value, we averaged the classification accuracies with different projection vector numbers to obtain the overall classification accuracies, as shown in Fig 11. When $\lg(\eta_1) < -1$ or $\lg(\eta_1) > 2$, the overall classification accuracies are stable with different values of $\lg(\eta_1)$. When $-1 \leq \lg(\eta_1) \leq 2$, the overall classification accuracy generally decreases with increasing $\lg(\eta_1)$ value. The highest classification accuracy is 0.8695 , obtained when $\lg(\eta_1) = -1.2$. RSPCA outperforms PCA-L1, indicating that incorporating sparsity improves classification performance, consistent with the results in [14,15].

For RSMPCA, the parameter η_2 was selected from $\{10^{-2}, 10^{-1.9}, 10^{-1.8}, \dots, 10^6\}$. That is, $\lg(\eta_2)$ was selected in the range of -2 – 6 with a step of 0.1 . The overall classification accuracies of RSMPCA with different $\lg(\eta_2)$ values are shown in Fig 12. When $\lg(\eta_2) < 1.4$ or $\lg(\eta_2) > 3.3$, the overall classification accuracies are stable with different values of $\lg(\eta_2)$. When $1.4 \leq \lg(\eta_2) \leq 3.3$, the overall classification accuracy generally decreases with increasing $\lg(\eta_2)$ value. The highest

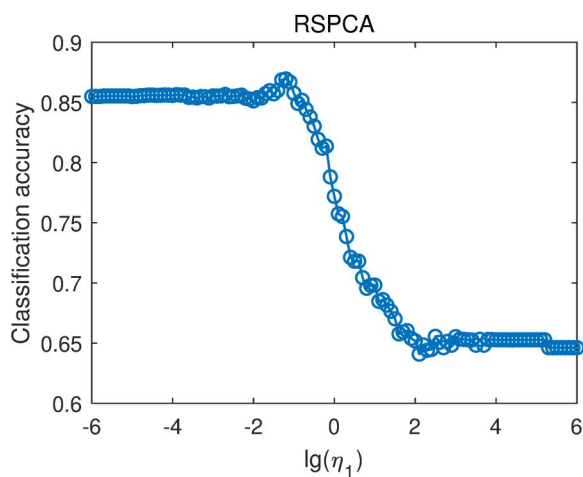


Fig 11. Average classification accuracies of RSPCA with different $\lg(\eta_1)$ values.

<https://doi.org/10.1371/journal.pone.0323281.g011>

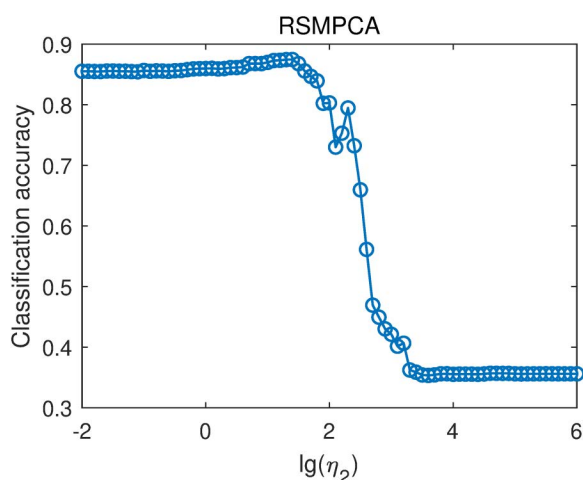


Fig 12. Average classification accuracies of RSMPCA with different $\lg(\eta_2)$ values.

<https://doi.org/10.1371/journal.pone.0323281.g012>

classification accuracy is 0.8747, obtained when $\lg(\eta_2) = 1.4$. RSMPCA outperforms PCA-L1, indicating that incorporating smoothness improves classification performance.

For RSSPCA, two parameters were selected according to the classification results of RSPCA and RSMPCA. Specifically, $\lg(\eta_1)$ was selected in the range of -3–1 with a step of 0.2, and $\lg(\eta_2)$ was selected in the range of -1–3 with a step of 0.2. The purpose of setting the step size to 0.2 is to reduce the number of parameter combinations, thereby shortening the total calculation time. The overall classification accuracies of RSSPCA with different parameters are shown in Fig 13. The highest classification accuracy is 0.8810, obtained when $\lg(\eta_1) = -0.6$ and $\lg(\eta_2) = 2.0$.

The highest classification accuracies and corresponding optimal parameters of the five algorithms are shown in Table 2. The classification accuracies of PCA and PCA-L1 are very close. It indicates that incorporating L1-norm into the objective function of PCA, i.e., incorporating robustness, has little effect on classification performance. RSPCA incorporates L1-norm constraint into PCA-L1 and achieves a higher classification accuracy than PCA-L1, indicating that incorporating

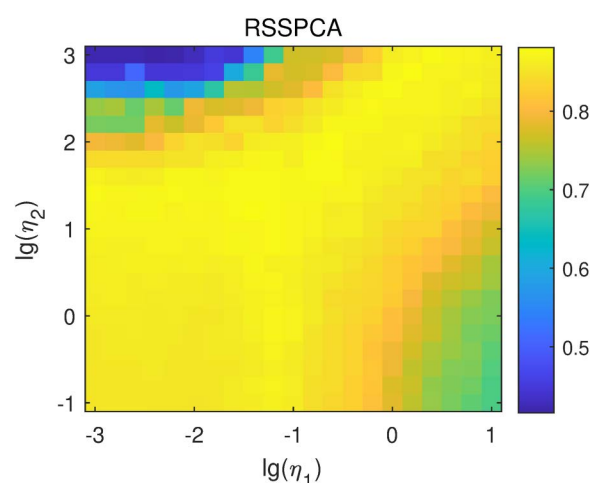


Fig 13. Average classification accuracies of RSSPCA with different $\lg(\eta_1)$ and $\lg(\eta_2)$ values.

<https://doi.org/10.1371/journal.pone.0323281.g013>

Table 2. The highest classification accuracies and optimal parameters of the five algorithms.

Algorithm	Optimal parameters		Classification accuracy
	$\lg(\eta_1)$	$\lg(\eta_2)$	
PCA	/	/	0.8556
PCA-L1	/	/	0.8552
RSPCA	-1.2	/	0.8695
RSMPCA	/	1.4	0.8747
RSSPCA	-0.6	2.0	0.8810

<https://doi.org/10.1371/journal.pone.0323281.t003>

the sparse constraint is beneficial for classification. These results are consistent with the results in [14,15]. RSMPCA incorporates the smooth constraint into PCA-L1 and achieves higher classification accuracy than PCA-L1, indicating that incorporating the smooth constraint is also beneficial for classification. RSSPCA simultaneously incorporates the sparse constraint and the smooth constraint into PCA-L1, and achieves the highest classification accuracy among the five algorithms.

In addition, RSSPCA can be generated by incorporating the sparse constraint into RSMPCA. With this adaptation, RSSPCA achieves a higher classification accuracy than RSMPCA. It further demonstrates that incorporating the sparse constraint is beneficial for classification. In another viewpoint, RSSPCA can be generated by incorporating the smooth constraint into RSPCA. With this adaptation, RSSPCA achieves a higher classification accuracy than RSPCA. It further demonstrates that incorporating the smooth constraint is also beneficial for classification.

In summary, incorporating robustness has little effect on classification performance, while incorporating either sparsity or smoothness improves classification performance. Therefore, RSSPCA has advantages over the other four algorithms in face recognition.

4.3 Experiments on five additional face databases

To further demonstrate the reconstruction and classification performance of the proposed algorithm, we conducted similar experiments on five additional benchmark face databases, i.e., the AR face database [48], the FEI face database [49], the FERET face database [50], the GT face database [51], and the Yale face database [52]. These datasets are publicly

available online and widely distributed for research purposes. The comprehensive details of these face databases, along with their corresponding experimental results, are fully documented in [S1 Text](#) within the Supporting Information section.

[Table 3](#) summarizes the lowest reconstruction error and highest classification accuracy of the five algorithms on these face databases. It also includes the corresponding optimal parameters and running time (in seconds) required to compute the projection matrix for each experiment. Additionally, results of the ORL face database are included to provide a comprehensive comparison.

Overall, the running time of PCA and PCA-L1 is much shorter than that of RSPCA, RSMPCA, and RSSPCA. This is because the projection vector \mathbf{w} in the latter three algorithms is iteratively updated by [Equations \(32\)](#) and [\(33\)](#), which is a time-consuming process.

In the face reconstruction experiment, the results on the AR and FEI face databases differ from those on the other four face databases. On the AR face database, PCA obtains the lowest reconstruction error among the five competing

Table 3. The reconstruction error and classification accuracy of the five algorithms on the six face databases.

Face database	Algorithm	Reconstruction				Classification			
		$\lg(\eta_1)$	$\lg(\eta_2)$	Error	Time(s)	$\lg(\eta_1)$	$\lg(\eta_2)$	Accuracy	Time(s)
AR	PCA	/	/	1145.12	3.58	/	/	0.5091	3.12
	PCA-L1	/	/	1151.14	11.81	/	/	0.5104	8.98
	RSPCA	-3.0	/	1151.97	57.21	0.1	/	0.6438	482.04
	RSMPCA	/	-1.2	1149.62	62.21	/	-0.1	0.5133	70.95
	RSSPCA	-3.0	-1.0	1150.19	69.18	1.0	0.8	0.6646	629.08
FEI	PCA	/	/	513.13	0.45	/	/	0.6245	0.28
	PCA-L1	/	/	510.90	2.13	/	/	0.6270	1.45
	RSPCA	-3.0	/	513.64	7.04	-1.5	/	0.6300	13.83
	RSMPCA	/	-1.0	510.49	9.02	/	0.6	0.6482	13.26
	RSSPCA	-3.0	-1.0	510.77	8.85	-1.6	0.8	0.6541	17.52
FERET	PCA	/	/	502.88	1.00	/	/	0.2545	0.30
	PCA-L1	/	/	474.41	4.00	/	/	0.2518	2.51
	RSPCA	-3.0	/	474.55	31.98	-0.2	/	0.3142	247.58
	RSMPCA	/	-0.6	461.52	48.35	/	1.5	0.3619	32.78
	RSSPCA	-3.0	-0.6	461.86	49.61	-2.2	1.6	0.3659	56.06
GT	PCA	/	/	784.46	0.19	/	/	0.6972	0.06
	PCA-L1	/	/	745.06	1.84	/	/	0.6920	1.15
	RSPCA	-3.0	/	746.66	22.90	-1.3	/	0.7025	62.10
	RSMPCA	/	-0.2	720.18	27.37	/	1.1	0.7365	38.32
	RSSPCA	-2.8	-0.2	720.23	34.54	-3.0	1.2	0.7362	42.85
ORL	PCA	/	/	1272.33	0.06	/	/	0.8556	0.04
	PCA-L1	/	/	1201.79	3.27	/	/	0.8552	2.76
	RSPCA	-3.0	/	1205.41	104.32	-1.2	/	0.8695	428.65
	RSMPCA	/	-0.2	1152.25	157.35	/	1.4	0.8747	136.01
	RSSPCA	-3.0	0.0	1153.35	167.41	-0.6	2.0	0.8810	513.86
Yale	PCA	/	/	1344.22	0.01	/	/	0.6804	0.01
	PCA-L1	/	/	1231.39	1.95	/	/	0.6822	1.74
	RSPCA	-3.0	/	1237.73	73.22	-2.7	/	0.6838	57.90
	RSMPCA	/	-0.2	1161.28	86.59	/	0.5	0.6887	63.40
	RSSPCA	-2.8	-0.2	1152.87	118.02	-2.6	0.4	0.6877	95.82

<https://doi.org/10.1371/journal.pone.0323281.t004>

Table 4. The impact of the three factors on reconstruction and classification performance.

	Reconstruction	Classification
Robustness	√	
Sparsity		√
Smoothness	√	√

<https://doi.org/10.1371/journal.pone.0323281.t005>

algorithms. However, on the other five face databases, the lowest reconstruction error is obtained by RSMPCA or RSSPCA. This discrepancy may be attributed to the intrinsic properties of the AR face database. On the FEI face database, the reconstruction errors obtained by PCA-L1, RSMPCA, and RSSPCA are lower than those obtained by PCA and RSPCA. This result is slightly different from those obtained on the other face databases.

On the FERET, GT, and Yale face databases, RSMPCA or RSSPCA obtains the lowest reconstruction errors, while PCA obtains the highest reconstruction errors. Furthermore, the reconstruction errors of PCA-L1 and RSPCA are close, as are those of RSMPCA and RSSPCA. These findings are consistent with those obtained on the ORL face database. In summary, except for the results on the AR and FEI face databases, the results on the other four face databases demonstrate that incorporating robustness and smoothness improves reconstruction performance, while incorporating sparsity has little effect on reconstruction performance.

In the face recognition experiment, except for the results on the AR face database, the lowest classification accuracies are obtained by PCA or PCA-L1, while the highest classification accuracies are obtained by RSMPCA or RSSPCA. On the AR face database, the classification accuracy of RSMPCA is much lower than that of RSSPCA. Again, this may be attributed to the intrinsic property of the AR face database. In general, incorporating sparsity and smoothness improves classification performance, while incorporating robustness has little effect on classification performance.

In summary, the incorporation of robustness and smoothness enhances reconstruction performance, and the incorporation of sparsity and smoothness enhances classification performance. Table 4 summarizes the impact of the three factors, i.e., robustness, sparsity, and smoothness, on reconstruction and classification performance.

5. Discussion

This study has three major limitations, as detailed below.

First, identifying the optimal parameters for RSSPCA remains an unresolved issue. This challenge is not unique to RSSPCA but is also present in other parameter-dependent algorithms, such as RSPCA [12,14] and RSMPCA. An exception exists when RSPCA is used for face reconstruction. As indicated in Table 3, RSPCA achieves the lowest reconstruction error with $\lg(\eta_1) = -3.0$ across all face databases. In this case, RSPCA approximates PCA-L1, which aligns with the findings in [14,15], demonstrating that incorporating sparsity does not affect reconstruction. However, outside of this specific case, determining the optimal parameters for RSPCA, RSMPCA, and RSSPCA is challenging because their results vary across different face databases.

Second, while some general conclusions can be drawn from the experimental results, they do not fully align with our expectations. Notably, on the AR face database, PCA achieves the lowest reconstruction error, and the classification accuracy of RSMPCA is much lower than that of RSSPCA. These outcomes are not observed in the other five databases. This discrepancy suggests that the AR face database has intrinsic properties influencing the results.

Third, the running time of RSPCA, RSMPCA, and RSSPCA is much longer than that of PCA and PCA-L1. This is due to the time-consuming update rule used by the first three algorithms, as specified in Equations (32) and (33). The running time of RSPCA can be reduced by using either of the two update rules in Equations (6)–(11). However, for RSMPCA and RSSPCA, no acceleration strategies have been identified yet due to the incorporation of the smooth constraint.

Potential improvements for the current study are outlined as follows.

First, the smooth constraint in this paper only considers the relationships between spatially adjacent pixels. Extending it to a more general form [41,53] that additionally considers the relations between spatially distant pixels can make full use of the two-dimensional spatial structure information of images.

Second, the robustness and sparsity in this paper are achieved by incorporating L1-norm into the objective function and the constraint function, respectively, in the optimization problem of RSSPCA. They can be further enhanced by replacing L1-norm with an arbitrary norm, i.e., Lp-norm [15,43].

Third, this study only compares five algorithms, i.e., PCA, PCA-L1, RSPCA, RSMPCA, and RSSPCA. However, by incorporating sparsity or smoothness into traditional PCA, we can construct three additional algorithms: Sparse PCA [54], Smooth PCA, and Sparse Smooth PCA. These algorithms can be solved similarly within the MM framework. This paper focuses on improving RSPCA by incorporating the smooth constraint. Therefore, the three non-robust algorithms are not investigated.

As for application of RSSPCA, while this paper focuses exclusively on face reconstruction and recognition, the algorithm is versatile and can be extended to analyze various types of data. For example, RSSPCA can be employed to analyze one-dimensional time series data [55], such as stock prices, heart rate recordings, daily electricity usage, hourly traffic volume, etc. In these cases, the smooth constraint can capture the one-dimensional temporal structure information of the data. For two-dimensional images other than face images, RSSPCA is undoubtedly applicable. Additionally, for high-dimensional data like EEG, MEG, and fMRI [23,25–29], the smooth constraint in RSSPCA can capture both spatial and temporal structure information within these datasets. In conclusion, RSSPCA shows great promise for reconstructing or classifying data that contains spatial or temporal structural information.

6. Conclusion

This paper proposes a new algorithm, termed RSSPCA (Robust Sparse Smooth Principal Component Analysis), which enhances RSPCA by incorporating a smooth constraint that captures the two-dimensional spatial structure information of images. An iterative optimization procedure is designed within the Majorization-Minimization (MM) framework to solve the RSSPCA optimization problem.

We evaluate RSSPCA's performance against four existing algorithms: PCA, PCA-L1, RSPCA, and RSMPCA. Experimental results reveal distinct effects of different constraints: incorporating sparsity has minimal impact on reconstruction performance, while robustness and smoothness contribute to improved reconstruction accuracy. Conversely, robustness shows limited influence in classification tasks, while sparsity and smoothness significantly enhance classification performance.

The proposed RSSPCA algorithm demonstrates clear advantages over existing methods in both face reconstruction and recognition by simultaneously integrating three key properties: robustness, sparsity, and smoothness. Visualization of the generalized eigenfaces clearly illustrates how these constraints influence feature extraction: sparsity enables selective feature identification, while smoothness preserves spatial relationships among facial components. Given these promising results, RSSPCA shows great potential for analyzing data with rich spatial or temporal structural information.

Supporting information

S1 Text. Experiments on five additional face databases.
(PDF)

Author contributions

Conceptualization: Jing Wang, Li Zhang, Jian Li, Yan Feng.

Formal analysis: Jing Wang, Li Zhang, Hao Cai.

Funding acquisition: Jing Wang, Li Zhang, Jian Li.

Methodology: Jing Wang, Xiao Xie, Jian Li, Hao Cai, Yan Feng.

Visualization: Xiao Xie, Li Zhang, Hao Cai.

Writing – original draft: Jing Wang, Xiao Xie, Jian Li, Hao Cai.

Writing – review & editing: Jing Wang, Li Zhang, Jian Li, Yan Feng.

References

1. Jolliffe I. A 50-year personal journey through time with principal component analysis. *J Multivar Anal.* 2022;188:104820. <https://doi.org/10.1016/j.jmva.2021.104820>
2. Bharadiya JP. A tutorial on principal component analysis for dimensionality reduction in machine learning. *Int J Innov Sci Res Technol.* 2023;8(5):2028–32.
3. Wu RMX, Zhang Z, Yan W, Fan J, Gou J, Liu B, et al. A comparative analysis of the principal component analysis and entropy weight methods to establish the indexing measurement. *PLoS One.* 2022;17(1):e0262261. <https://doi.org/10.1371/journal.pone.0262261> PMID: 35085274
4. Torell F. Evaluation of stretch reflex synergies in the upper limb using principal component analysis (PCA). *PLoS One.* 2023;18(10):e0292807. <https://doi.org/10.1371/journal.pone.0292807> PMID: 37824570
5. Ahmed W, Ansari S, Hanif M, Khalil A. PCA driven mixed filter pruning for efficient convNets. *PLoS One.* 2022;17(1):e0262386. <https://doi.org/10.1371/journal.pone.0262386> PMID: 35073373
6. Kwak N. Principal component analysis based on l1-norm maximization. *IEEE Trans Pattern Anal Mach Intell.* 2008;30(9):1672–80. <https://doi.org/10.1109/TPAMI.2008.114> PMID: 18617723
7. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B: Stat Methodol.* 1996;58(1):267–88. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
8. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell.* 2009;31(2):210–27. <https://doi.org/10.1109/TPAMI.2008.79> PMID: 19110489
9. Zhang Z, Xu Y, Yang J, Li X, Zhang D. A survey of sparse representation: algorithms and applications. *IEEE Access.* 2015;3:490–530. <https://doi.org/10.1109/access.2015.2430359>
10. Crespo Marques E, Maciel N, Naviner L, Cai H, Yang J. A review of sparse recovery algorithms. *IEEE Access.* 2019;7:1300–22. <https://doi.org/10.1109/access.2018.2886471>
11. Zou H, Hastie T, Tibshirani R. Sparse principal component analysis. *J Comput Graph Stat.* 2006;15(2):265–86. <https://doi.org/10.1198/106186006x113430>
12. Meng D, Zhao Q, Xu Z. Improve robustness of sparse PCA by L1-norm maximization. *Pattern Recognit.* 2012;45(1):487–97. <https://doi.org/10.1016/j.patcog.2011.07.009>
13. Yang J, Zhang D, Frangi AF, Yang J. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell.* 2004;26(1):131–7. <https://doi.org/10.1109/tpami.2004.1261097> PMID: 15382693
14. Wang H, Wang J. 2DPCA with L1-norm for simultaneously robust and sparse modelling. *Neural Netw.* 2013;46:190–8. <https://doi.org/10.1016/j.neunet.2013.06.002> PMID: 23800536
15. Wang J. Generalized 2-D principal component analysis by Lp-norm for image analysis. *IEEE Trans Cybern.* 2016;46(3):792–803. <https://doi.org/10.1109/TCYB.2015.2416274> PMID: 25898326
16. Wang J, Zhao M, Xie X, Zhang L, Zhu W. Fusion of bilateral 2DPCA information for image reconstruction and recognition. *Appl Sci.* 2022;12(24):12913. <https://doi.org/10.3390/app122412913>
17. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst, Man, Cybern.* 1973;SMC-3(6):610–21. <https://doi.org/10.1109/tsmc.1973.4309314>
18. Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Trans Pattern Anal Machine Intell.* 2000;22(8):888–905. <https://doi.org/10.1109/34.868688>
19. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell.* 2012;34(11):2274–82. <https://doi.org/10.1109/TPAMI.2012.120> PMID: 22641706
20. Chan TF, Vese LA. Active contours without edges. *IEEE Trans Image Process.* 2001;10(2):266–77. <https://doi.org/10.1109/83.902291> PMID: 18249617
21. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1989;1(4):541–51. <https://doi.org/10.1162/neco.1989.1.4.541>
22. Rasmussen CE. Gaussian processes in machine learning. *Summer school on machine learning.* Springer; 2003. p. 63–71.

23. Dyrholm M, Parra LC, editors. Smooth bilinear classification of EEG. International Conference of the IEEE Engineering in Medicine and Biology Society; 2006.
24. Heibiri M, vandeGeer S. The smooth-lasso and other ℓ_1 ℓ_2 -penalized methods. *Electron J Stat*. 2011;5:1184–226.
25. de Brecht M, Yamagishi N. Combining sparseness and smoothness improves classification accuracy and interpretability. *Neuroimage*. 2012;60(2):1550–61. <https://doi.org/10.1016/j.neuroimage.2011.12.085> PMID: [22261376](#)
26. Grosenick L, Klingenberg B, Katovich K, Knutson B, Taylor JE. Interpretable whole-brain prediction analysis with GraphNet. *Neuroimage*. 2013;72:304–21. <https://doi.org/10.1016/j.neuroimage.2012.12.062> PMID: [23298747](#)
27. Sun Z, Qiao Y, Lelieveldt BPF, Staring M, Alzheimer's Disease Neuroimaging Initiative. Integrating spatial-anatomical regularization and structure sparsity into SVM: improving interpretation of Alzheimer's disease classification. *Neuroimage*. 2018;178:445–60. <https://doi.org/10.1016/j.neuroimage.2018.05.051> PMID: [29802968](#)
28. Watanabe T, Kessler D, Scott C, Angstadt M, Sripada C. Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine. *Neuroimage*. 2014;96:183–202. <https://doi.org/10.1016/j.neuroimage.2014.03.067> PMID: [24704268](#)
29. Eliseyev A, Aksenova T. Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ECoG) recording. *PLoS One*. 2016;11(5):e0154878. <https://doi.org/10.1371/journal.pone.0154878> PMID: [27196417](#)
30. Ling Q, Liu A, Li Y, Fu X, Chen X, McKeown MJ, et al. A joint constrained CCA model for network-dependent brain subregion parcellation. *IEEE J Biomed Health Inform*. 2022;26(11):5641–52. <https://doi.org/10.1109/JBHI.2022.3196689> PMID: [35930507](#)
31. Chen E, Chang R, Guo K, Miao F, Shi K, Ye A, et al. Hyperspectral image spectral-spatial classification via weighted Laplacian smoothing constraint-based sparse representation. *PLoS One*. 2021;16(7):e0254362. <https://doi.org/10.1371/journal.pone.0254362> PMID: [34255786](#)
32. Tayyib M, Amir M, Javed U, Akram MW, Yousufi M, Qureshi IM, et al. Accelerated sparsity based reconstruction of compressively sensed multi-channel EEG signals. *PLoS One*. 2020;15(1):e0225397. <https://doi.org/10.1371/journal.pone.0225397> PMID: [31910204](#)
33. Liu S, Huang Q, Quan T, Zeng S, Li H. Foreground estimation in neuronal images with a sparse-smooth model for robust quantification. *Front Neuroanat*. 2021;15:716718. <https://doi.org/10.3389/fnana.2021.716718> PMID: [34764857](#)
34. Merris R. Laplacian matrices of graphs: a survey. *Linear Algebra Appl*. 1994;197–198:143–76. [https://doi.org/10.1016/0024-3795\(94\)90486-3](https://doi.org/10.1016/0024-3795(94)90486-3)
35. Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv Neural Inf Process Syst*. 2001;14.
36. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput*. 2003;15(6):1373–96. <https://doi.org/10.1162/089976603321780317>
37. Zhang S, Li X, Zong M, Zhu X, Wang R. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans Neural Netw Learn Syst*. 2018;29(5):1774–85. <https://doi.org/10.1109/TNNLS.2017.2673241> PMID: [28422666](#)
38. Zhang S, Li X, Zong M, Zhu X, Cheng D. Learning k for kNN classification. *ACM Trans Intell Syst Technol*. 2017;8(3):1–19. <https://doi.org/10.1145/2990508>
39. Cai D, He X, Hu Y, Han J, Huang T, editors. Learning a spatially smooth subspace for face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*; 2007.
40. He X, Yan S, Hu Y, Niyogi P, Zhang H-J. Face recognition using laplacianfaces. *IEEE Trans Pattern Anal Mach Intell*. 2005;27(3):328–40. <https://doi.org/10.1109/TPAMI.2005.55> PMID: [15747789](#)
41. Wang J, Xie X, Wang P, Sun J, Liu Y, Zhang L. Incorporating symmetric smooth regularizations into sparse logistic regression for classification and feature extraction. *Symmetry*. 2025;17(2):151. <https://doi.org/10.3390/sym17020151>
42. Donoho DL. De-noising by soft-thresholding. *IEEE Trans Inform Theory*. 1995;41(3):613–27. <https://doi.org/10.1109/18.382009>
43. Liang Z, Xia S, Zhou Y, Zhang L, Li Y. Feature extraction based on Lp-norm generalized principal component analysis. *Pattern Recognit Lett*. 2013;34(9):1037–45. <https://doi.org/10.1016/j.patrec.2013.01.030>
44. Mackey L. Deflation methods for sparse PCA. *Adv Neural Inf Process Syst*. 2008;21.
45. Hunter DR, Lange K. A tutorial on MM algorithms. *Am Stat*. 2004;58(1):30–7. <https://doi.org/10.1198/0003130042836>
46. Samaria FS, Harter AC, editors. Parameterisation of a stochastic model for human face identification. *IEEE Workshop on Applications of Computer Vision*; 1994.
47. Samaria FS. Face recognition using hidden Markov models. Cambridge, UK: University of Cambridge; 1994.
48. Martinez A, Benavente R. The AR face database. *CVC Tech Rep*. 1998;24.
49. Thomaz CE, Giraldo GA. A new ranking method for principal components analysis and its application to face image analysis. *Image Vis Comput*. 2010;28(6):902–13. <https://doi.org/10.1016/j.imavis.2009.11.005>
50. Phillips PJ, Martin A, Wilson CL, Przybocki M. An introduction evaluating biometric systems. *Computer*. 2000;33(2):56–63. <https://doi.org/10.1109/2.820040>
51. Nefian AV, Hayes MH, editors. Hidden Markov models for face recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*; 1998.
52. Georgiades AS, Belhumeur PN, Kriegman DJ. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans Pattern Anal Machine Intell*. 2001;23(6):643–60. <https://doi.org/10.1109/34.927464>

53. Wang J, Zhang B, Xie X, Li J, Zhang L, Guo H. Whole-brain classification based on generalized sparse logistic regression. *J Xinyang Norm Univ (Nat Sci Ed)*. 2022;35(3):488–93.
54. Jolliffe IT, Trendafilov NT, Uddin M. A modified principal component technique based on the LASSO. *J Comput Graph Stat*. 2003;12(3):531–47.
<https://doi.org/10.1198/1061860032148>
55. Yu H-F, Rao N, Dhillon IS. Temporal regularized matrix factorization for high-dimensional time series prediction. *Adv Neural Inf Process Syst*. 2016;29.