Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2054546, 14 pages
https://doi.org/10.1155/2022/2054546

Research Article

New Robust Regression Method for Outliers and Heavy Sparse Noise Detection via Affine Transformation for Head Pose Estimation and Image Reconstruction in Highly Complex and Correlated Data: Applications in Signal Processing

Peidong Liang,[1] Habte Tadesse Likassa ,[2] and Chentao Zhang [2,3]

[1]*Fujian (Quanzhou)-HIT Research Institute of Engineering and Technology, Quanzhou 362000, China*
[2]*Department of Statistics, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia*
[3]*Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen, China*

Correspondence should be addressed to Chentao Zhang; 2265275624@qq.com

Received 26 August 2021; Accepted 16 January 2022; Published 18 February 2022

Academic Editor: Fangqing Wen

In this work, we propose a novel method for head pose estimation and face recovery, particularly to solve the potential impacts of noises in signal processing to get an efficient and effective model that is more resilient with annoying effects through adding affine transformation with the low-rank robust subspace regression. Consequently, the corrupted images can be correctly recovered by affine transformations to render more best regression outcomes. Thereby, we need to search so as to get optimal parameters which can be regarded as convex constrained optimization techniques. Afterward, the alternating direction method for multipliers (ADMM) approach is considered and a new set of updated equations is well established so as to update the optimization parameters and affine transformations iteratively in a round-robin manner. Additionally, the convergence of these new updating equations is well scrutinized as well. Thus, the experimental simulations reveal that the proposed method outperforms the state-of-the-art works for head pose estimation and face recovery on some public databases.

1. Introduction

Images, processing particularly for the head pose estimation and image recovery, have been important research potential topics and can have applications in a variety of areas such as surveillance systems [1, 2], signal processing [3, 4], image denoising [5–9] and recovery [10, 11], communications [12], computational imaging [13, 14], and computer vision [15–19]. However, analyzing visual data is a difficult task due to miscellaneous adverse effects such as illuminations, outliers, and sparse noises. It is thus of importance to develop robust face recovery and head pose estimation algorithms, which are resilient to various annoying effects.

After the inception of the pioneering baselines of robust principal component analysis (RPCA) by Candes et al. [20], a myriad of methods has been considered for robust sparse-low-rank image recovery, e.g., [21, 22]. However, these methods do not work well when the outliers and heavy sparse noises are not normally distributed.

To tackle this drawback, Oh et al. [23] proposed a new partial singular value thresholding (PSVT) method, which replaced the nuclear norm in RPCA [8, 20] with the partial sum of singular values to improve the recovery of the low-rank part. Lu et al. [24] proposed a tensor robust principal component (T-RPCA) approach to find the clean tuber low-rank component. However, T-RPCA is not scalable and robust when the number of tensors becomes large. Currently, several algorithms, which combined regression with RPCA [25], were proposed to further enhance the performance. For instance, Ji et al. [26] addressed a regularized sparse regression via combining RPCA [22, 27] with lasso regression [28] to mitigate the influence of outliers in the head pose estimation. But, the performance of this relaxation method degrades when the fitting error is grossly increasing. Wang et al. [29] developed a robust regression method via self-scaled regularization to boost the performance in the presence of gross outliers. Huang et al. [30] proposed a low-rank robust regression (LR-RR) algorithm to clean the outliers and sparse errors from highly contaminated data. Although LR-RR can mitigate the impact of sparse errors inside and outside subspaces, its sensitivity to sparse errors and outliers lying in the disjoint subspaces jeopardizes its performance in some severe scenarios. Yin et al. [31] considered a robust multinomial logistic and binary regression to remove the sparse noise and outliers from the contaminated data. Yang et al. [32] proposed a matrix regression scheme for face image representation based on the nuclear norm. Zhang et al. [33, 34] proposed a low-rank-sparse subspace representation for robust regression (LRS-RR) method to find the clean low-rank part by low-rank subspace recovery along with regression to deal with errors or outliers lying in the corrupted disjoint subspaces. To resolve this, Zeng et al. [35] addressed labeled-robust regression, but its performance is not yet promising to denoise the high dimensional images, particularly in signal processing; to tackle this, Wu et al. [36, 37] addressed the sparse prior information.

This work proposes a new robust method for head pose estimation and image recovery to denoise the potential impacts of outliers and heavy sparse noises in signal processing. To develop a method that is working well with various annoying effects, the new approach incorporates affine transformations taken from [38–40] into the robust regression methods [30, 33] with the robust regression for more faithful low-rank-sparse image representation. Consequently, the corrupted high dimensional images can be recovered correctly by affine transformations to achieve more promising regression outcomes in statistical signal processing. The newly developed algorithm is first cast as a convex optimization programming, in which the affine transformations, low-rank subspace recovery, and regression are carried out simultaneously. Afterward, the alternating direction method for multiplier (ADMM) method is applied and a new set of equations is established to update the optimization variables and affine transformations iteratively in a round-robin manner. Moreover, the convergence of the entire newly developed equations is scrutinized as well. Conducted simulation results reveal that the proposed method excels the state-of-the-art works for head pose estimation and face recovery on some public datasets. The main contributions of this paper include the following:

(1) The affine transformations are incorporated into the low-rank-sparse decomposition to correct the illuminated and highly distorted or misaligned images to attain more precise low-rank image decomposition

(2) The ADMM method is proposed to solve the convex constrained optimization problems and a new set of updating equations is developed to iteratively update the optimization parameters and affine transformations

(3) The convergence of the derived iterative equations that considers more updating parameters is investigated

(4) The proposed method outperforms to the baselines

This work is organized as follows. Section 2 gives an overview of the related works. Section 3 addresses the formulation of the new problem. Section 4 depicts the new set of updating equations to solve the formulated convex optimization problem and Section 5 analyzes its convergence characteristics. Experimental simulation results are provided in Section 6 to verify the proposed method. Section 7 draws some concluding remarks to summarize the paper.

2. Related Works

A number of robust methods have been reported for image recovery [41–45]. For instance, Wei et al. [14, 43, 45] addressed the least trimmed squares to alleviate the gross errors in the regression to explicitly search a data subset that reduces the square of the errors. The discriminatory least square regression [46] and the worst-case linear discriminant analysis [47] were proposed to solve the least square loss function that influences the correlation between the explanatory and response variables. These methods, however, are not that robust for big data. Bunea et al. [48] scrutinized a rank selection criterion to select the best rank estimator of the coefficient regression matrix in the multivariate regression approach. To deal with linearly structured matrices, Zachariah et al. [49, 50] addressed an iterative algorithm via least square estimation for low-rank matrix reconstruction, but it required the prior knowledge of the matrix structure. Chen et al. [51] proposed an iterative reweighted least squared method for sparsity recovery, which incorporated the structure of sparsity along with an orthogonal basis and the total variation. However, it requires conducting matrix inversion at each iteration, leading to high computational complexity. Instead of using the handcrafted least square regressions as in [13, 52–55], some more recent methods [56–61] have improved the visual quality in face reconstruction via the low-rank approximation.

Head pose estimation has also received a considerable amount of research attention. For instance, linear regression methods were considered in [62, 63] for head pose estimation. These methods, however, are sensitive to occlusions and uncontrolled illuminations. Lathuiliere et al. [64] proposed a deep mixture regression approach to replace the supervised manifold learning in [65, 66] to perform head pose estimation. Recently, Sun et al. [61, 67, 68] proposed a probabilistic method for head pose estimation by directly mapping the feature vectors onto the yaw angles. Diaz-Chito et al. [69] addressed an algorithm to narrow down the gap between the head yaw angles and the regression by combining manifold embedding methods with linear regression. Meyer et al. [70] considered a three-dimensional head pose estimation method to handle large pose angles and partial occlusions. However, it cannot prune out outliers from disjoint subspaces in head pose estimation.

3. Problem Formulation

Given n images, $\{\mathbf{I}_i^0\} \in \Re^{w\times h}$, $i = 1, \ldots, n$, where w and h denote respectively the weight and height of the highly illuminated and corrupted images, all of which consists the same objects. In many dilemmas, these images are highly correlated and corrupted by occlusions and outliers. We can stack these images into a matrix: $\mathbf{M} = [\mathrm{vec}(\mathbf{I}_1^0)|vec(\mathbf{I}_2^0)|\ldots|\mathrm{vec}(\mathbf{I}_n^0)] \in \Re^{d_m\times n}$, where $\mathrm{vec}(\cdot)$ denotes the vector stacking operator. In light of the fact that the subspaces may not be independent from each other or the data are illuminated and contaminated by large illuminations and noises and outliers, we decompose the original data matrix into a low-rank component and sparse errors, i.e., $\mathbf{M} = \mathbf{AC} + \mathbf{E}$ [71–73], where $\mathbf{A} \in \Re^{d_m\times n}$ is the low-rank component, $\mathbf{C} \in \Re^{n\times n}$ is a reconstruction coefficient matrix used to represent $\mathbf{M}$, and $\mathbf{E} \in \Re^{d_m\times n}$ denotes a sparse error matrix incurred by some adverse effects.

In reality, $\{\mathbf{I}_i^0\}$ are generally not well corrected and aligned, which makes the issue of the low-rank and sparse separation to be imprecise. To mitigate the issue of the misalignment, inspired by [74, 75], we apply affine transformations τ_i to $\{\mathbf{I}_i^0\}$ to get the transformed images $\mathbf{I}_i = \mathbf{I}_i^0 o\tau_i$, where the operator o indicates the transformation applied to the potentially misaligned input images. After taking the affine transformation on $\mathbf{M}$, we can obtain $\mathbf{M}_{o\tau} = [\mathrm{vec}(\mathbf{I}_1)|\mathrm{vec}(\mathbf{I}_2)|\ldots|\mathrm{vec}(\mathbf{I}_n)] \in \Re^{d_m\times n}$, where $\mathbf{I}_i = \mathbf{I}_i^0 o\tau_i$ is an aligned version of the i^{th} image after the transformation. The aligned images can be treated as samples taken from a union of low-dimensional subspaces, which exhibits a low-rank subpace structure as the rank of the transformed images is as small as possible, up to some outliers and heavy sparse errors. To improve the issue of the nonlinearity in $\mathbf{M}_{\mathbf{o}\tau}$, we can further represent that the change produced by these affine transformations τ is small and an initial affine transformation of τ is known, then we can further linearize it by using the first-order Taylor approximation as $\mathbf{M}_{o(\tau+\Delta\tau)} \approx \mathbf{M}_{\mathbf{o}\tau} + \sum_{i=1}^n \mathbf{J}_i \Delta\tau \mathbf{v}_i \mathbf{v}_i^T$, where $\mathbf{M}_{\mathbf{o}\tau} \in \Re^{d_m\times n}$ is the transformed image, $\Delta\tau \in \Re^{p\times n}$, where p indicates the number of parameters, $\mathbf{J}_i = \partial\mathrm{vec}(\mathbf{I}_i\mathbf{o}\tau_i)/\partial\tau_i \in \Re^{d_m\times p}$ denotes the Jacobian of the i^{th} image with respect to τ_i, and $\mathbf{v}_i$ denotes the standard basis for $\Re^n$.

The main goal is to learn a regression model, denoted as the regression matrix $\mathbf{F} \in \Re^{(d_m)\times(d_m+1)}$, which maps $\mathbf{M}_{\mathbf{o}\tau}$ to the regression output $\mathbf{Y} \in \Re^{(d_y)\times n}$ by minimizing the fitting error $\|\mathbf{W}(\mathbf{Y} - \mathbf{F}\beta)\|_F^2$ [30, 33], where $\mathbf{W} \in \Re^{d_y\times d_y}$ is the diagonal regression matrix that adjusts the regression output dimension and $\beta = [\mathbf{AC}; 1^T] \in \Re^{(d_m+1)\times n}$ denotes the augmented noise-free data matrix with the extra dimension accounting for the regression bias and d_m denoting the dimensional samples. The main objective of this work is to reduce the reconstruction error through extracting the low-rank component from complex highly correlated data in statistical signal processing. The overall problem can thus be posted as the following indicating the constrained convex optimization problem:

$$\begin{aligned} &\min_{\mathbf{A},\mathbf{F},\boldsymbol{\beta},\mathbf{E},\mathbf{C},\mathbf{Q},\Delta\tau} \frac{\gamma}{2}\|\mathbf{U}\|_F^2 + \|\mathbf{A}\|_* + \|\mathbf{C}\|_* + \lambda_1\|\mathbf{Q}\|_1 + \lambda_2\|\mathbf{E}\|_1 \\ &\text{s.t}\, \mathbf{M}_{\mathbf{o\tau}} + \sum_{i=1}^n \mathbf{J}_i\Delta\boldsymbol{\tau}\mathbf{v}_i\mathbf{v}_i^T = \mathbf{AC} + \mathbf{E}, \boldsymbol{\beta} = \left[\mathbf{AC}; 1^T\right], \mathbf{C} = \mathbf{Q}, \mathbf{Q} \succcurlyeq 0 \end{aligned}, \quad (1)$$

where $\mathbf{U} = \mathbf{W}(\mathbf{Y} - \mathbf{F}\beta)$, $1_n \in \Re^n$ indicates a vector of dimension n with all one entries, $\|\mathbf{A}\|_* = \sum_{i=1}^{\min(d_m,n)} \sigma_i(\mathbf{A})$ is the nuclear norm of $\mathbf{A}$, in which $\sigma_i(\mathbf{A})$ indicates the singular values of $\mathbf{A}$, $\|\mathbf{U}\|_F^2 = \mathrm{Trace}(\mathbf{U}^T\mathbf{U})$, λ_1, λ_2, and γ are the regularization parameters, $\langle \mathbf{X}, \mathbf{Y}\rangle = \mathrm{Trace}(\mathbf{X}^T\mathbf{Y})$, and $\|\mathbf{Q}\|_1 = \max_{1\le j\le n} \sum_{i=1}^n |\mathbf{Q}_{ij}|$.

The second term $\|\mathbf{A}\|_*$ in (1) represents the low-rank component of $\mathbf{A}$. The second and the third terms, $\|\mathbf{C}\|_*$ and $\|\mathbf{Q}\|_1$, are to constrain the low-rank and sparse representation, respectively. The last term $\|\mathbf{E}\|_1$ constrains and regularizes the outliers and heavy sparse noises modelled by $\mathbf{E}$ to be sparse. In the constraints of (1), the affine transformations are used to alleviate the impact of outliers and heavy sparse noises, and $\mathbf{Q}$ is constrained to be positive semidefinite to bound the regression errors.

4. Proposed Method

To get the optimal solution of the convex constrained optimization problem in (1), take into consideration the augmented Lagrangian function given by

$$\begin{aligned} \mathscr{L}(\mathbf{F}, \boldsymbol{\beta}, \mathbf{A}, \mathbf{C}, \mathbf{E}, \mathbf{Q}, \Delta\boldsymbol{\tau}) = {}& \frac{\gamma}{2}\|\mathbf{U}\|_F^2 + \|\mathbf{A}\|_* + \|\mathbf{C}\|_* + \lambda_1\|\mathbf{Q}\|_1 + \lambda_2\|\mathbf{E}\|_1 \\ & + \langle \mathbf{Z}_1, \mathbf{M}_{\mathbf{o\tau}} + \sum_{i=1}^n \mathbf{J}_i\Delta\boldsymbol{\tau}\mathbf{v}_i\mathbf{v}_i^T - \mathbf{AC} - \mathbf{E}\rangle + \frac{\mu_1}{2}\left\|\mathbf{M}_{\mathbf{o\tau}} + \sum_{i=1}^n \mathbf{J}_i\Delta\boldsymbol{\tau}\mathbf{v}_i\mathbf{v}_i^T - \mathbf{AC} - \mathbf{E}\right\|_F^2 \\ & + \langle \mathbf{Z}_2, \boldsymbol{\beta} - \left[\mathbf{AC}; 1^T\right]\rangle + \frac{\mu_2}{2}\left\|\boldsymbol{\beta} - \left(\mathbf{AC}; 1^T\right)\right\|_F^2 + \langle \mathbf{Z}_3, \mathbf{C} - \mathbf{Q}\rangle + \frac{\mu_3}{2}\|\mathbf{C} - \mathbf{Q}\|_F^2, \end{aligned} \quad (2)$$

where $\mathbf{Z}_1 \in \mathfrak{R}^{d_m \times n}$, $\mathbf{Z}_2 \in \mathfrak{R}^{(d_m+1)\times n}$, and $\mathbf{Z}_3 \in \mathfrak{R}^{n\times n}$ are the Lagrangian multipliers, and μ_1, μ_2, and μ_3 are the penalty parameters. Considering the linearized alternating direction method with an adaptive penalty (LADMAP) [33, 76], (2) can be rewritten as

$$\begin{aligned}\mathscr{L}(\mathbf{F}, \boldsymbol{\beta}, \mathbf{A}, \mathbf{C}, \mathbf{E}, \mathbf{Q}, \Delta\boldsymbol{\tau}) = {} & \frac{\gamma}{2}\|\mathbf{U}\|_F^2 + \|\mathbf{A}\|_* + \|\mathbf{C}\|_* + \lambda_2\|\mathbf{E}\|_1 \\ & + \lambda_1\|\mathbf{Q}\|_1 + \frac{\mu_1}{2}\left\|\mathbf{M}_{\mathbf{o}\boldsymbol{\tau}} + \sum_{i=1}^{n}\mathbf{J}_i\Delta\boldsymbol{\tau}\mathbf{v}_i\mathbf{v}_i^T - \mathbf{AC} - \mathbf{E} + \frac{\mathbf{Z}_1}{\mu_1}\right\|_F^2 \\ & + \frac{\mu_2}{2}\left\|\boldsymbol{\beta} - \left(\mathbf{AC}; 1^T\right) + \frac{\mathbf{Z}_2}{\mu_2}\right\|_F^2 + \frac{\mu_3}{2}\left\|\mathbf{C} - \mathbf{Q} + \frac{\mathbf{Z}_3}{\mu_3}\right\|_F^2.\end{aligned} \tag{3}$$

Solving (3) directly is its computational load and complexity is very expensive, thereby we consider to iteratively updating the optimization parameters and affine transformations via ADMM [77], which decomposes the minimization into several subproblems. To make the mathematical formulation easy, we additionally assume $\mathbf{W} = \mathbf{I}$ in the following derivations.

Firstly, we need to get the optimal updates of $\mathbf{F}$, we fix $\mathbf{A}$, $\mathbf{E}$, $\Delta\tau$, β, $\mathbf{C}$, and $\mathbf{Q}$ as constants in (3), and $\mathbf{F}^{(k+1)}$ can be determined by

$$\mathbf{F}^{(k+1)} = \underset{\mathbf{F}}{\operatorname{argmin}}\, \mathscr{L}\left(\mathbf{F}, \boldsymbol{\beta}^{(k)}, \mathbf{A}^{(k)}, \mathbf{E}^{(k)}, \mathbf{C}^{(k)}, \mathbf{Q}^{(k)}, \Delta\boldsymbol{\tau}^{(k)}\right), \tag{4}$$

where k is the iteration index and ignoring all the irrelevant terms of $\mathbf{F}$ in (4), equation (4) can be reexpressed as

$$\mathbf{F}^{(k+1)} = \underset{\mathbf{F}}{\operatorname{argmin}}\, \frac{\gamma}{2}\left\|\mathbf{W}\left(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}^{(k)}\right)\right\|_F^2, \tag{5}$$

this is exactly the same as a standard least square regression. Consequently, we can obtain

$$\mathbf{F}^{(k+1)} = \left(\boldsymbol{\beta}^{(k)}\left(\boldsymbol{\beta}^{(k)}\right)^T + \gamma\mathbf{I}_{(d_m+1)}\right)^{-1}\mathbf{Y}\left(\boldsymbol{\beta}^{(k)}\right)^T, \tag{6}$$

where $\mathbf{I}_{d_m+1}$ is a $(d_m + 1)\times(d_m + 1)$ identity matrix.

Secondly, to find the update of β, we fix $\mathbf{A}$, $\mathbf{E}$, $\mathbf{F}$, $\mathbf{C}$, $\mathbf{Q}$, and $\Delta\tau$ and decide $\beta^{(k+1)}$ by

$$\boldsymbol{\beta}^{(k+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \mathscr{L}\mathbf{F}^{(k+1)}, \boldsymbol{\beta}, \mathbf{A}^{(k)}, \mathbf{E}^{(k)}, \mathbf{C}^{(k)}, \mathbf{Q}^{(k)}, \Delta\boldsymbol{\tau}^{(k)}. \tag{7}$$

Ignoring all of the irrelevant terms of β in (7), we can get

$$\boldsymbol{\beta}^{(k+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\left\{\frac{\gamma}{2}\left\|\mathbf{U}^{(k)}\right\|_F^2 + \frac{\mu_2^{(k)}}{2}\left\|\boldsymbol{\beta} - \left(\mathbf{A}^{(k)}\mathbf{C}^{(k)}; 1^T\right) + \frac{\mathbf{Z}_2^{(k)}}{\mu_2^{(k)}}\right\|_F^2\right\}, \tag{8}$$

where $\mathbf{U}^{(k)} = \mathbf{W}(\mathbf{Y} - \mathbf{F}^{(k)}\beta^{(k)})$. Thereby, $\beta^{(k+1)}$ can be determined by

$$\begin{aligned}\boldsymbol{\beta}^{(k+1)} = {} & \left[\gamma\left(\mathbf{F}^{k}\right)^T\mathbf{W}^T\mathbf{W}\mathbf{F}^{(k)} + \mu_2^{(k)}\mathbf{I}_{d_{m+1}}\right]^{-1} \\ & \left[\gamma\left(\mathbf{F}^{(k)}\right)^T\mathbf{W}^T\mathbf{W}\mathbf{Y} - \mathbf{Z}_2^{(k)} + \mu_2^{(k)}\left(\mathbf{A}^{(k)}\mathbf{C}^{(k)}; 1^T\right)\right].\end{aligned} \tag{9}$$

Similarly, to update $\mathbf{A}$, we fix $\mathbf{E}$, $\mathbf{F}$, β, $\mathbf{Q}$, $\mathbf{C}$, and $\Delta\tau$ and decide $\mathbf{A}^{(k+1)}$ by

$$\mathbf{A}^{(k+1)} = \underset{\mathbf{A}}{\operatorname{argmin}}\, \mathscr{L}\left(\mathbf{F}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{A}, \mathbf{E}^{(k)}, \mathbf{C}^{(k)}, \mathbf{Q}^{(k)}, \Delta\boldsymbol{\tau}^{(k)}\right). \tag{10}$$

By ignoring all of the irrelevant terms of $\mathbf{A}$, (10) can be simplified as

$$\mathbf{A}^{(k+1)} = \underset{\mathbf{A}}{\operatorname{argmin}}\left\{\begin{aligned} & \|\mathbf{A}\|_* + \frac{\mu_1^{(k)}}{2}\left\|\mathbf{M}_{\mathbf{o}\boldsymbol{\tau}} + \sum_{i=1}^{n}\mathbf{J}_i\Delta\boldsymbol{\tau}^{(k)}\mathbf{v}_i\mathbf{v}_i^T - \mathbf{AC}^{(k)} - \mathbf{E}^{(k)} + \frac{\mathbf{Z}_1^{(k)}}{\mu_1^{(k)}}\right\|_F^2 \\ & + \frac{\mu_2}{2}\left\|\boldsymbol{\beta}^{(k+1)} - \left(\mathbf{AC}^{(k)}; 1^T\right) + \frac{\mathbf{Z}_2^{(k)}}{\mu_2^{(k)}}\right\|_F^2\end{aligned}\right\} \tag{11}$$

Considering the linear augmented direction method and the singular value threshold operator [72, 78–80], we can obtain an update of $\mathbf{A}^{(k+1)}$ given by

$$\mathbf{A}^{(k+1)} = \Omega_{1/\kappa_{\mathbf{A}}}\left(\mathbf{A}^{(k)} - \frac{\mathbf{P}_{\mathbf{A}}^{(k)}}{\kappa_{\mathbf{A}}} \right), \tag{12}$$

where Ω is the singular value thresholding operator [79], $\mathbf{P}_{\mathbf{A}}^{(k)} = (\mu_1^{(k)} + \mu_2^{(k)})(\mathbf{A}^{(k)})^T\mathbf{C}^{(k)} - \mu_1^{(k)}(\mathbf{M}_{o\tau} + \sum_i = 1^n \mathbf{J}_i\Delta\tau^{(k)} \mathbf{v}_i\mathbf{v}_i^T - \mathbf{E}^{(k)}) - (\mathbf{Z}_1^{(k)}) - (\mu_2^{(k)}\beta^{(k)} + \mathbf{Z}_2^{(k)})(\mathbf{C}^{(k)})^T$, $\kappa_{\mathbf{A}} = (\mu_1^{(k)} + \mu_2^{(k)})\tau_{\mathbf{A}}/2$, and $\tau_{\mathbf{A}} > \sigma(\mathbf{C}^T\mathbf{C})$ is the proximal parameter, in which $\sigma(\mathbf{C}^T\mathbf{C})$ denotes the spectral radius of $\mathbf{C}^T\mathbf{C}$.

To obtain an update of $\mathbf{E}$, by keeping $\mathbf{A}$, $\mathbf{F}$, β, $\mathbf{Q}$, $\mathbf{C}$, and $\Delta\tau$ as constants, then $\mathbf{E}^{(k+1)}$ is can be determined by

$$\mathbf{E}^{(k+1)} = \underset{\mathbf{E}}{\operatorname{argmin}}\, \mathscr{L}\left(\mathbf{F}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{A}^{(k+1)}, \mathbf{E}, \mathbf{C}^{(k)}, \mathbf{Q}^{(k)}, \Delta\boldsymbol{\tau}^{(k)} \right). \tag{13}$$

Again, by ignoring all of the irrelevant terms of $\mathbf{E}$, (13) can be simplified as

$$\mathbf{E}^{(k+1)} = \underset{\mathbf{E}}{\operatorname{argmin}} \left\{ \lambda_2 \|\mathbf{E}\|_1 + \frac{\mu_1^{(k)}}{2} \left\| \mathbf{M}_{\mathbf{o}\boldsymbol{\tau}} + \sum_{i=1}^{n} \mathbf{J}_i \Delta\boldsymbol{\tau}^{(k)} \mathbf{v}_i\mathbf{v}_i^T - \mathbf{A}^{(k+1)}\mathbf{C}^{(k)} - \mathbf{E} + \frac{\mathbf{Z}_1^{(k)}}{\mu_1^{(k)}} \right\|_F^2 \right\}. \tag{14}$$

By employing the linearized alternating direction method, $\mathbf{E}^{(k+1)}$ is updated by

$$\mathbf{E}^{(k+1)} = \Gamma_{\lambda_1/\mu_1}\left(\mathbf{M}_{\mathbf{o}\boldsymbol{\tau}} + \sum_{i=1}^{n} \mathbf{J}_i \Delta\boldsymbol{\tau}^{(k)} \mathbf{v}_i\mathbf{v}_i^T - \mathbf{A}^{(k+1)}\mathbf{C}^{(k)} + \frac{\mathbf{Z}_1^{(k)}}{\mu_1^{(k)}} \right), \tag{15}$$

where $\Gamma_{\lambda_1/\mu_1}(\Theta) = \operatorname{sgn}(\Theta)\max(|\Theta| - \lambda_1/\mu_1, 0)$ is the soft shrinkage thresholding operator [78, 79], in which $\operatorname{sgn}(\Theta)$ denotes the sign function.

Next, to get an update of $\mathbf{C}$, we again keep $\mathbf{A}$, $\mathbf{E}$, $\mathbf{F}$, β, $\mathbf{Q}$, and $\Delta\tau$ as constants and $\mathbf{C}^{(k+1)}$ can then be determined by

$$\mathbf{C}^{(k+1)} = \underset{\mathbf{C}}{\operatorname{argmin}}\, \mathscr{L}\left(\mathbf{F}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{A}^{(k+1)}, \mathbf{E}^{(k+1)}, \mathbf{C}, \mathbf{Q}^{(k)}, \Delta\boldsymbol{\tau}^{(k)} \right). \tag{16}$$

By ignoring all of the irrelevant terms of $\mathbf{C}$, (16) can be simplified as

$$\mathbf{C}^{(k+1)} = \underset{\mathbf{C}}{\operatorname{argmin}} \left\{ \begin{array}{l} \|\mathbf{C}\|_* + \dfrac{\mu_1^{(k)}}{2} \left\| \mathbf{M}_{\mathbf{o}\boldsymbol{\tau}} + \displaystyle\sum_{i=1}^{n} \mathbf{J}_i \Delta\boldsymbol{\tau}^{(k)} \mathbf{v}_i\mathbf{v}_i^T - \mathbf{A}^{(k+1)}\mathbf{C} - \mathbf{E}^{(k+1)} + \dfrac{\mathbf{Z}_1^{(k)}}{\mu_1^{(k)}} \right\|_F^2 \\ + \dfrac{\mu_2^{(k)}}{2} \left\| \boldsymbol{\beta}^{(k+1)} - \left(\mathbf{A}^{(k+1)}\mathbf{C}; 1^T\right) + \dfrac{\mathbf{Z}_2^{(k)}}{\mu_2^{(k)}} \right\|_F^2 + \dfrac{\mu_3^{(k)}}{2} \left\| \mathbf{C} - \mathbf{Q}^{(k)} + \dfrac{\mathbf{Z}_3^{(k)}}{\mu_3^{(k)}} \right\|_F^2 \end{array} \right\}. \tag{17}$$

To solve this subproblem, we can apply a linearized augmented Lagrangian multiplier with the singular value threshold operator and update $\mathbf{C}^{(k+1)}$ by

$$\mathbf{C}^{(k+1)} = \Omega_{2/\kappa_c}\left(\mathbf{C}^{(k)} - \frac{\mathbf{P}_{\mathbf{C}}^{(k)}}{\kappa_{\mathbf{C}}} \right), \tag{18}$$

where $\kappa_{\mathbf{C}} = (\mu_1^{(k)} + \mu_2^{(k)} + \mu_3^{(k)})\tau_C/2$ with $\tau_{\mathbf{C}} > \sigma(\mathbf{A}\mathbf{A}^T)$ and $\mathbf{P}_{\mathbf{C}}^{(k)} = \mathbf{A}^{(k)}(\mu_1^{(k)} + \mu_2^{(k)})\mathbf{A}^{(k)} - \mu_1^{(k)}(\mathbf{M}_{o\tau} + \sum_{i=1}^{n} \mathbf{J}_i\Delta\tau^{(k)}\mathbf{v}_i\mathbf{v}_i^T - \mathbf{E}^{(k)}) - \mathbf{Z}_1^{(k)} - (\mu_2^{(k)}\beta^{(k)} + \mathbf{Z}_2^{(k)}) + \mu_3^{(k)}(\mathbf{C}^{(k)} - \mathbf{Q}^{(k)}) + \mathbf{Z}_3^{(k)}$.

To get an update of $\mathbf{Q}$, we keep $\mathbf{A}$, $\mathbf{E}$, $\mathbf{F}$, β, $\mathbf{C}$, and $\Delta\tau$ as constants and determine $\mathbf{Q}^{(k+1)}$ by

$$\mathbf{Q}^{(k+1)} = \underset{\mathbf{Q}}{\operatorname{argmin}}\, \mathscr{L}\left(\mathbf{F}^{(k+1)}, \boldsymbol{\beta}^{(k+1)}, \mathbf{A}^{(k+1)}, \mathbf{E}^{(k+1)}, \mathbf{C}^{(k+1)}, \mathbf{Q}, \Delta\boldsymbol{\tau}^{(k)} \right). \tag{19}$$

By removing all of the irrelevant terms of $\mathbf{Q}$, (19) is reduced to

$$\mathbf{Q}^{(k+1)} = \underset{\mathbf{Q}}{\operatorname{argmin}} \left\{ \lambda_1 \|\mathbf{Q}\|_1 + \frac{\mu_3^{(k)}}{2} \left\| \mathbf{C}^{(k)} - \mathbf{Q} + \frac{\mathbf{Z}_3^{(k)}}{\mu_3^{(k)}} \right\|_F^2 \right\}. \tag{20}$$

Similarly, employing the soft threshold operator and the augmented Lagrangian multiplier, we can update $\mathbf{Q}^{(k+1)}$ by

$$\mathbf{Q}^{(k+1)} = \Gamma_{\lambda_2/\mu_3}\left(\mathbf{C}^{(k+1)} + \frac{\mathbf{Z}_3^{(k)}}{\mu_3^{(k)}} \right). \tag{21}$$

To update the affine transformations, by keeping all other variables as constants, we can get

$$\Delta\boldsymbol{\tau}^{(k+1)} = \underset{\Delta\boldsymbol{\tau}}{\operatorname{argmin}} \left\{ \frac{\mu_1^{(k)}}{2} \left\| \mathbf{M}_{\mathbf{o\tau}} + \sum_{i=1}^{n} \mathbf{J}_i \Delta\boldsymbol{\tau} \mathbf{v}_i \mathbf{v}_i^T - \mathbf{A}^{(k+1)}\mathbf{C}^{(k+1)} - \mathbf{E}^{(k+1)} + \frac{\mathbf{Z}_1^{(k)}}{\mu_1^{(k)}} \right\|_F^2 \right\}, \tag{22}$$

Solving (22) with the threshold operators [74], we can get an update of $\Delta\tau^{(k+1)}$ as

$$\Delta\boldsymbol{\tau}^{(k+1)} = \sum_{i=1}^{n} \mathbf{J}_\mathbf{i}^{+} \left(\mathbf{A}^{(\mathbf{k+1})}\mathbf{C}^{(\mathbf{k+1})} + \mathbf{E}^{(\mathbf{k+1})} - \mathbf{M}_{\mathbf{o\tau}} - \frac{\mathbf{Z}_1^{(k)}}{\boldsymbol{\mu}_1^{(k)}} \right) \mathbf{v_i}\mathbf{v_i}^\mathbf{T}, \tag{23}$$

where $\mathbf{J}_i^{+}$ denotes the Moore–Penrose pseudoinverse of $\mathbf{J}_i$ [81], in which $\mathbf{J}_i$ denotes the Jacobian of the i^{th} image with respect to τ_i as defined in Section 3.

Finally, following the same steps as above, the Lagrangian multipliers $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ are updated by

$$\mathbf{Z}_1^{(k+1)} = \mathbf{Z}_1^{(k)} + \mu_1^{(k)} \left(\mathbf{M}_{\mathbf{o\tau}} + \sum_{i=1}^{n} \mathbf{J}_i \Delta\boldsymbol{\tau}^{(k+1)} \mathbf{v}_i \mathbf{v}_i^T - \mathbf{A}^{(k+1)}\mathbf{C}^{(k+1)} - \mathbf{E}^{(k+1)} \right). \tag{24}$$

$$\mathbf{Z}_2^{(k+1)} = \mathbf{Z}_2^{(k)} + \mu_2^{(k)} \left(\boldsymbol{\beta}^{(k+1)} - \left(\mathbf{A}^{(k+1)}\mathbf{C}^{(k+1)}; 1^T \right) \right). \tag{25}$$

$$\mathbf{Z}_3^{(k+1)} = \mathbf{Z}_3^{(k)} + \mu_3^{(k)} \left(\mathbf{C}^{(k+1)} - \mathbf{Q}^{(k+1)} \right). \tag{26}$$

Likewise, the regularization parameters μ_1, μ_2, and μ_3 are updated respectively by

$$\mu_1^{(k+1)} = \min\left(\mu_{\max}, \rho\mu_1^{(k)} \right). \tag{27}$$

$$\mu_2^{(k+1)} = \min\left(\mu_{\max}, \rho\mu_2^{(k)} \right). \tag{28}$$

$$\mu_3^{(k+1)} = \min\left(\mu_{\max}, \rho\mu_3^{(k)} \right), \tag{29}$$

where ρ and $\mu_{\max}$ are appropriately chosen parameters adjusting the convergence speed of the new approach.

The overall updating equations of our proposed approach can be summarized as follows. First, the regression matrix $\mathbf{F}$ and the regression coefficients β are updated by (6) and (9), respectively. Next, $\mathbf{A}$, $\mathbf{E}$, $\mathbf{C}$, $\mathbf{Q}$, and the affine transformation, $\Delta\tau$, are updated by (12), (15), (18), (21), and (23), respectively. Finally, the Lagrangian multipliers $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ and the regularization parameters μ_1, μ_2, and μ_3 are updated by (24) and (29). The above updating equations proceed in a round-robin manner until convergence. For an easy understanding of the manuscript, the summarized algorithm of this work is given in Algorithm 1.

5. Convergence Analysis

In this section, we consider the convergence behavior of the updating equations, by addressing two theorems related to the convergence of ADMM. We first consider the following two propositions:

Proposition 1. *If $\{\mu_k\}$ is nondecreasing and upper bounded by $\tau_\mathbf{A} > \sigma(\mathbf{CC}^T)$ and $\tau_\mathbf{C} > \sigma(\mathbf{AA}^T)\|$, the subgradients are then defined as*

(a) $-\mu_k\tau_\mathbf{A}(\mathbf{A}^{(k+1)} - \mathbf{A}^{(k)}) - \alpha^*(\mathbf{Z}_1^{(k+1)}) \in \partial f(\mathbf{A}^{(k+1)})$

(b) $-\mu_k\eta_\mathbf{E}(\mathbf{E}^{(k+1)} - \mathbf{E}^{(k)}) - \Psi^*(\mathbf{Z}_2^{(k+1)}) \in \partial g(\mathbf{E}^{(k+1)})$

(c) $-\mu_k\tau_\mathbf{C}(\mathbf{C}^{(k+1)} - \mathbf{C}^{(k)}) - \omega^*(\mathbf{Z}_3^{(k+1)}) \in \partial h(\mathbf{C}^{(k+1)})$

(d) $-\mu_k\eta_\mathbf{Q}(\mathbf{Q}^{(k+1)} - \mathbf{Q}^{(k)}) - \varphi^*(\mathbf{Z}_3^{(k+1)}) \in \partial p(\mathbf{Q}^{(k+1)})$

(e) $-\mu_k\eta_{\Delta\tau}(\Delta\tau^{(k+1)} - \Delta\tau^{(k)}) - \delta^*(\mathbf{Z}_3^{(k+1)}) \in \partial q(\Delta\tau^{(k+1)})$,

Here, ∂f, ∂g, ∂h, ∂p, and ∂q are the subgradients of $\mathbf{A}$, $\mathbf{E}$, $\mathbf{C}$, $\mathbf{Q}$, and $\Delta\tau$ respectively, and α^*, Ψ^*, ω^*, φ^*, and δ^* are the adjoints of the linear identity mapping operators corresponding to $\mathbf{A}$, $\mathbf{E}$, $\mathbf{C}$, $\mathbf{Q}$, and $\Delta\tau$, respectively.

The proof is similar to [76].

Proposition 2. *If $\{\mu_k\}$ is nondecreasing and upper bounded by $\tau_\mathbf{A} > \sigma(\mathbf{CC}^T)$ and $\tau_\mathbf{C} > \sigma(\mathbf{AA}^T)$, then $\{\mathbf{A}^*, \mathbf{E}^*, \mathbf{Z}_1^*, \mathbf{Z}_2^*\}$ and $\{\mathbf{C}^*, \mathbf{Q}^*, \Delta\tau^*, \mathbf{Z}_3^*\}$ are any Karush–Kuhn–Tucker (KKT) point of problem (1). Also,*

(a) $\tau_\mathbf{A}\|\mathbf{A}^{(k)} - \mathbf{A}^*\|^2 - \|\alpha(\mathbf{A}^{(k)}) - \mathbf{A}^*\|^2 + \eta_\mathbf{E}\|\mathbf{E}^{(k)} - \mathbf{E}^*\|^2 - \|\Psi(\mathbf{E}^{(k)}) - \mathbf{E}^*\|^2 + \mu_k^{-2}\|\mathbf{Z}_1^{(k)} - \mathbf{Z}_1^*\|^2 + \mu_k^{-2}\|\mathbf{Z}_2^{(k)} - \mathbf{Z}_2^*\|^2$ *is nonincreasing.*

(b) $\|\mathbf{A}^{(k+1)} - \mathbf{A}^{(k)}\| \longrightarrow 0$, $\|\mathbf{E}^{(k+1)} - \mathbf{E}^{(k)}\| \longrightarrow 0$, $\|\mathbf{Z}_1^{(k+1)} - \mathbf{Z}_1^{(k)}\| \longrightarrow 0$, $\|\mathbf{Z}_2^{(k+1)} - \mathbf{Z}_2^{(k)}\| \longrightarrow 0$.

(c) $\tau_\mathbf{C}\|\mathbf{C}^{(k)} - \mathbf{C}^*\|^2 - \|\omega(\mathbf{C}^{(k)}) - \mathbf{C}^*\|^2 + \eta_\mathbf{Q}\|\mathbf{Q}^{(k)} - \mathbf{Q}^*\|^2 - \|\varphi(\mathbf{Q}^{(k)}) - \mathbf{Q}^*\|^2 + \eta_{\Delta\tau}\|\Delta\tau^{(k)} - \Delta\tau^*\|^2 - \|\delta(\Delta\tau^{(k)}) - \Delta\tau^*\|^2 + \mu_k^{-2}\|\mathbf{Z}_3^{(k)} - \mathbf{Z}_3^*\|^2$ *is nonincreasing.*

Input Data matrix $\mathbf{M} \in \mathfrak{R}^{\mathbf{m}\times\mathbf{n}}$, $\mathbf{A}^0 \in \mathfrak{R}^{m\times n}$, $\mathbf{E}^0 \in \mathfrak{R}^{m\times n}$, $\Delta\tau^0 \in \mathfrak{R}^{p\times n}$, $\mathbf{C}^0 \in \mathfrak{R}^{n\times n}$, $\mathbf{Q}^0 \in \mathfrak{R}^{n\times n}$, λ_1, λ_2, ρ
While not converged do
(1) Update: $\mathbf{A}^{(\mathbf{k}+1)}$ by (12)
(2) Update: $\mathbf{E}^{(k+1)}$ by (15)
(3) Update: $\mathbf{C}^{(k+1)}$ by (18)
(4) Update: $\mathbf{Q}^{\mathbf{k}+1}$ by (21)
(5) Update: $\Delta\tau^{(k+1)}$ by (23)
(6) Update: $\mathbf{Z}_1^{(k+1)}$ by (24)
(7) Update: $\mathbf{Z}_2^{(k+1)}$ by (25)
(8) Update: $\mathbf{Z}_3^{(k+1)}$ by (26)
(9) Update: $\mu_1^{(k+1)}$ by (27)
(10) Update: $\mu_2^{(k+1)}$ by (28)
(11) Update: $\mu_3^{(k+1)}$ by (29)
End while
Outputs : $\mathbf{A}, \mathbf{E}, \mathbf{C}, \mathbf{Q}, \Delta\tau$

Algorithm 1: ADMM for the proposed algorithm.

(d) $\|\mathbf{C}^{(k+1)} - \mathbf{C}^{(k)}\| \longrightarrow 0$, $\|\mathbf{Q}^{(k+1)} - \mathbf{Q}^{(k)}\| \longrightarrow 0$, $\|\Delta\tau^{(k+1)} - \Delta\tau^{(k)}\| \longrightarrow 0$, $\|\mathbf{Z}_3^{k+1} - \mathbf{Z}_3^{k}\| \longrightarrow 0$.

The proof is similar to [76].

Theorem 1. *If $\{\mu_1\}$ and $\{\mu_2\}$ are nondecreasing and upper bounded by $\tau_{\mathbf{A}} > \sigma(\mathbf{C}\mathbf{C}^T)$, then the sequence $\{\mathbf{A}^{(k)}\}$, $\{\mathbf{E}^{(k)}\}$, $\{\mathbf{Z}_1^{(k)}\}$, and $\{\mathbf{Z}_2^{(k)}\}$ generated by ADMM converging to a KKT point of (11) and (14).*

Proof. See Appendix.

Theorem 1 implies that the optimization variables $\{\mathbf{A}^{(k+1)}\}$ and $\{\mathbf{E}^{(k+1)}\}$ are guaranteed to converge to the global optimum with an appropriate choice of $\{\mathbf{Z}_1^{(k)}\}$ and $\{\mathbf{Z}_2^{(k)}\}$ and sufficiently large penalty parameters $\{\mu_1\}$ and $\{\mu_2\}$.

Theorem 2. *If $\{\mu_1\}$, $\{\mu_2\}$, and $\{\mu_3\}$ are nondecreasing and upper bounded by $\tau_{\mathbf{C}} > \sigma(\mathbf{A}\mathbf{A}^T)$, then the sequence $\{\mathbf{C}^{(k)}\}$, $\{\mathbf{Q}^{(k)}\}$, $\{\Delta\tau^{(k)}\}$, and $\{\mathbf{Z}_3^{(k)}\}$ generated by ADMM converges to a KKT point of (17), (20), and (22).*

Proof is similar to that of Theorem 1.

Theorem 2 shows that the updating variables $\{\mathbf{C}^{(k+1)}\}$, $\{\mathbf{Q}^{(k+1)}\}$, and $\{\Delta\tau^{(k+1)}\}$ are guaranteed to converge to the global optimum with an appropriate choice of $\{\mathbf{Z}_3^{(k)}\}$ and sufficiently large penalty parameters $\{\mu_1\}$, $\{\mu_2\}$, and $\{\mu_3\}$.

6. Experimental Results and Discussion

In this section, we first verify the correct recovery guarantee and evaluate the effectiveness of the new method image recovery and head pose estimation based either on synthetic data or on some public databases. Four baseline methods, including T-RPCA [24] +LSR, PSVT [23] +LSR, LR-RR [30], and LRS-RR [33] and the proposed one are conducted for comparison, where T-RPCA + LSR and PSVT + LSR first perform T-RPCA and PSVT on the illuminated and corrupted input data, respectively, and then conduct regression on the error free data using the standard least square regression. Then, first we try to evaluate the effectiveness of the proposed method based on the synthetic datasets. Following this, several public datasets are taken into account to verify the effectiveness of the proposed method.

6.1. Synthetic Data Recovery. First, we further assess the proposed algorithm on the generated synthetic data for prediction.

As [30, 33], we generate 400 three-dimensional samples, in which the first 2 parts of the samples are obtained from a uniform distribution in $[-6, 6]$, while the 3^{rd} one is from two different joint subspaces given by $z = u - v$ and $z = u + v$. In addition, we add zero-mean white Gaussian noise with unit variance into the second dimension, which simulates the in-subspace noise. Similarly, we add zero-mean white Gaussian sparse noise with unit variance in the 3^{rd} subspace to simulate the noise outside subspaces. 200 data samples are randomly chosen for the training and the other 200 samples for the testing. As a fair comparison, we use the relative absolute error (RAE) between the true regression matrix $\mathbf{F}$ and the one learned, $\widehat{F}$, i.e., $RAE_{\mathbf{F}} = \|\widehat{F} - \mathbf{F}\|_F/\|\mathbf{F}\|_F$, and the RAE between the true regression output $\mathbf{Y}$ and the predicted one $\widehat{Y}$, i.e., $RAE_{\mathbf{Y}} = \|\widehat{Y} - \mathbf{Y}\|_F/\|\mathbf{Y}\|_F$, as the performance measure of accuracy for regression.

The comparison of $RAE_{\mathbf{F}}$ and $RAE_{\mathbf{Y}}$ using the aforementioned methods based on the generated synthetic data is shown in Table 1, from which we can see that PSVT + LSR yields better performance than T-RPCA + LSR, as it employs the truncated nuclear norm instead of using more tensors to deal with the outliers and heavy sparse noises.

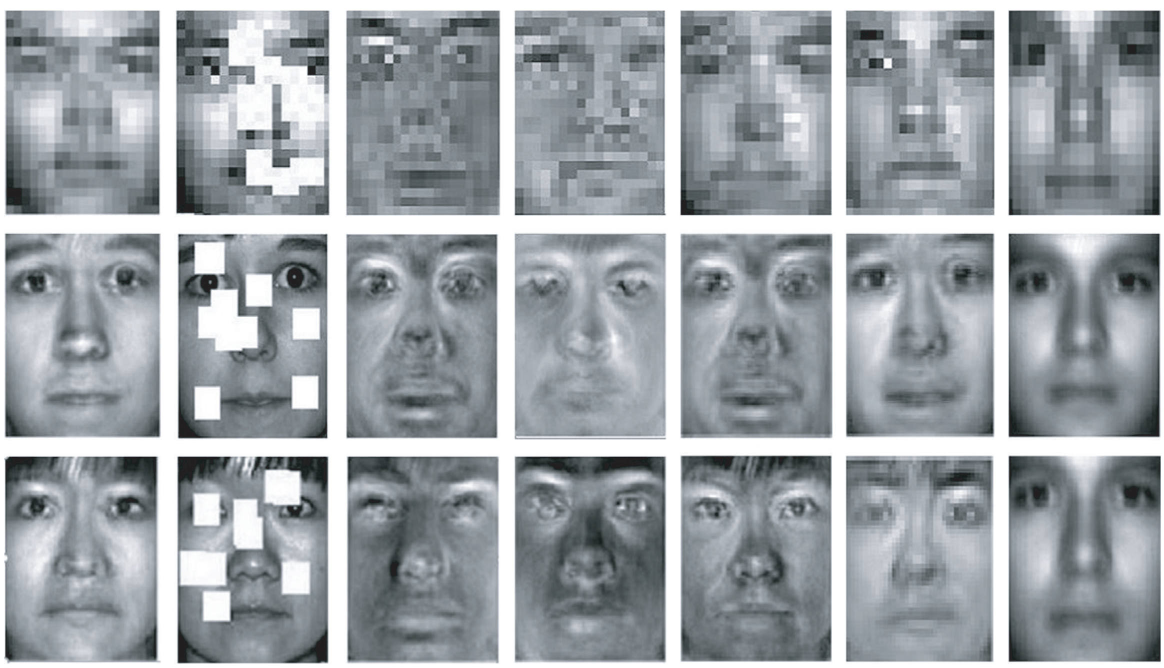
LR-RR is superior to PSVT + LSR and T-RPCA + LSR, as it cleans noises and outliers in and outside subspaces in a supervised manner to yield more precise prediction. LRS-RR provides the second best performance, as it can cope with the outliers coming from inside and outside subspaces, and the disjoint subspaces. We can also see that our method outperforms all of the baselines in both of $RAE_{\mathbf{F}}$ and $RAE_{\mathbf{Y}}$. This is because it incorporates the affine transformations with the robust regression for low-rank subspace recovery, so it can handle the aggregated outliers from various subspaces and heavy sparse noises to produce more precise results.

Table 1: Comparison of relative absolute error and standard deviation for the recovery of synthetic data.

Methods	$RAE_{\mathbf{F}}$	$RAE_{\mathbf{Y}}$
T-RPCA [24] + LSR	0.0740 ± 0.1040	0.0530 ± 0.1650
PSVT [23] + LSR	0.0697 ± 0.0889	0.0150 ± 0.0060
LR-RR [30]	0.0350 ± 0.0150	0.0150 ± 0.0060
LRS-RR [33]	0.0050 ± 0.0005	0.0100 ± 0.0030
Ours	0.0047 ± 0.0001	0.0042 ± 0.000038

Table 2: Comparison of relative absolute error and standard deviation for face recovery on the YaleB database.

Methods	$RAE_{\mathbf{F}}$	$RAE_{\mathbf{Y}}$
T-RPCA [24] + LSR	1.3000 ± 0.0177	0.2274 ± 0.0080
PSVT [23] + LSR	1.2783 ± 0.0197	0.2278 ± 0.0066
LR-RR [30]	1.0715 ± 0.0430	0.1854 ± 0.0056
LRS-RR [33]	1.0457 ± 0.0496	0.1659 ± 0.0061
Ours	0.2265 ± 0.0010	0.1482 ± 0.0055



(a) (b) (c) (d) (e) (f) (g)

Figure 1: Face image recovery: (a) original; (b) corrupted; (c) T-RPCA + LSR; (d) PSVT + LSR; (e) LR-RR; (f) LRS-RR; (g) ours.

6.2. Face Image Recovery. In this subsection, we assess the performance of the proposed method for face image reconstruction in terms of $RAE_{\mathbf{F}}$ and $RAE_{\mathbf{Y}}$ based on the images from the YaleB [82] database.

First, we conduct simulations on the YaleB database [82] for face image recovery. The YaleB database consists of over 2300 frontal face images from 38 subjects with varying illuminations. First, the cropped face images with 196×128 pixels are taken from the first fifteen subjects. Next, we compute the twenty-dimensional eigenfaces based on the training images which are taken as $\mathbf{M}$.

In each test face image, ten blocks of 30×30 of the corrupted pixels are generated randomly and added as a synthetic corrupted data. To assess the accuracy, we decide the true regression model based on the unblocked tested face images via the eigenfaces, then we learn the regression matrix by using various algorithms with eigenimages as the input $\mathbf{M}$ and the blocked face images as $\mathbf{Y}$.

Table 3: Comparison of the average of the yaw angle errors and its standard deviation on the CMU database.

Methods	Yaw angle error
T-RPCA [24] + LSR	$24.1636^o \pm 20.2201^o$
PSVT [23] + LSR	$23.2429^o \pm 19.0697^o$
LR-RR [30]	$1.97^o \pm 5.77^o$
LRS-RR [33]	$1.03^o \pm 5.65^o$
Ours	$0.95^o \pm 5.26^o$

The comparison of $RAE_{\mathbf{F}}$ and $RAE_{\mathbf{Y}}$ based on the proposed method as well as the aforementioned baselines are given in Table 2, from which we can see that LR-RR outperforms PSVT + LSR and T-RPCA + LSR as it cleans the

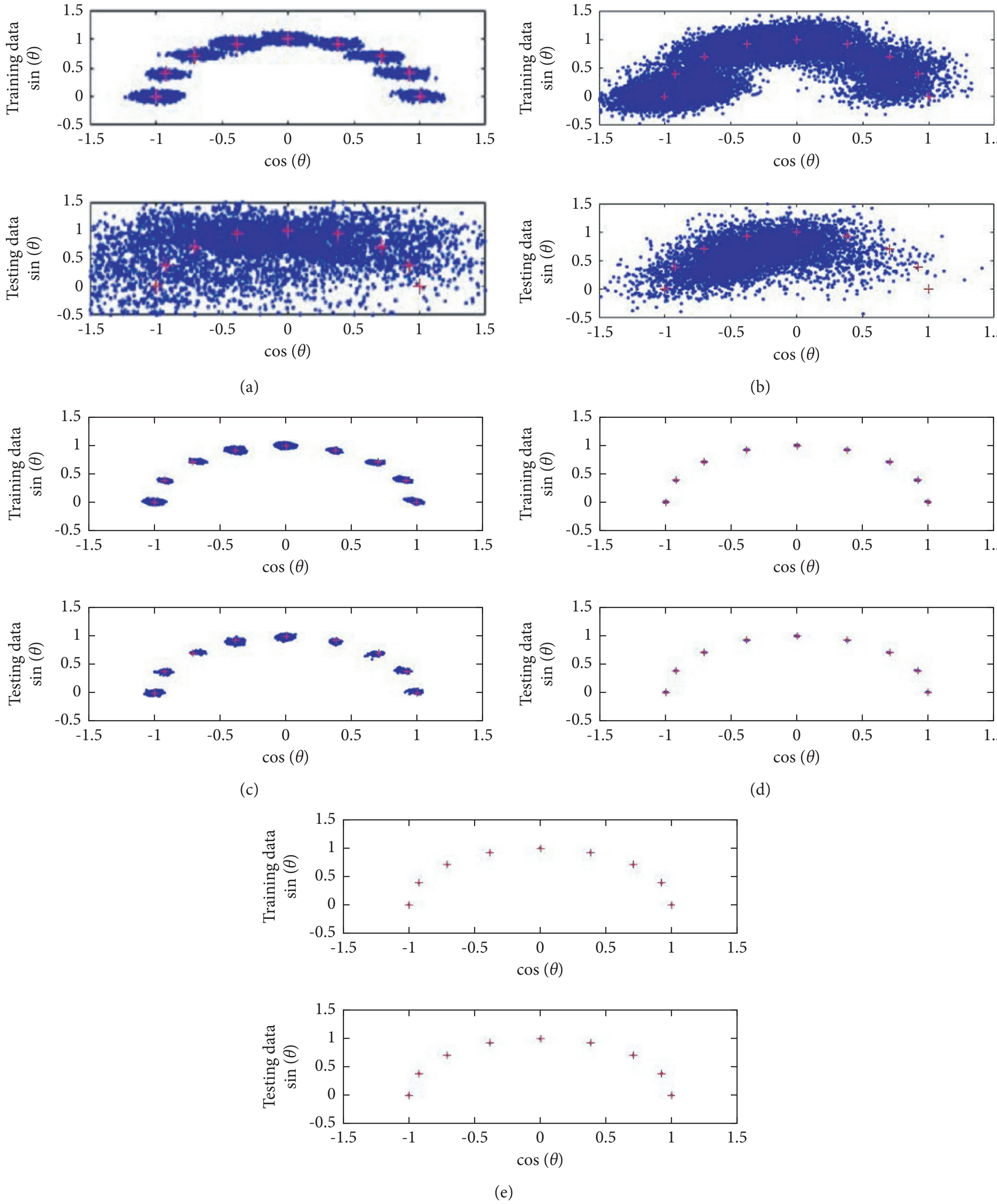


Figure 2: Projection of the head pose images, where (".") denotes the output space and the red ("+") is the ground true location of pose angles. (a) T-RPCA + LSR; (b) PSVT + LSR; (c) LR-RR; (d) LRS-RR; (e) ours.

intrasample in a supervised manner to reduce the overfitting error. LRS-RR is superior to LR-RR, as it considers the outliers and heavy sparse noises coming from disjoint subspaces into consideration, thereby reducing the fitting errors and the model errors. We can also notice that the proposed method provides the best performance. This is because the incorporation of affine transformations with robust regression enables our approach to tackle the impact of outliers and heavy sparse errors better compared with the other methods.

As an illustration, we provide some corrupted images recovered by the aforementioned methods, as depicted in

Figure 1, from which we can observe that the proposed algorithm recovers the corrupted data better compared with the other four baselines. As shown in Figure 1(g), the recovered image provides a clearer visual quality by removing the corruption. This is in agreement with the results in Table 2 and further justifies that the new approach is more resilient to outliers and heavy sparse noises.

6.3. Head Pose Estimation. In this section, we conduct simulations for head pose estimation based on the CMU multi-PIE database [83], which consists of more than 5000 faces taken from 53 different subjects. In this experiment, nine images are chosen with yaw pose angles that vary from $\theta = [-90^o, 90^o]$ with an increment of 22.5°. The face images are cropped around the faces and resized to 48×48. We also readjust the face images after applying linearization to $\mathbf{M}$ and the yaw angles are considered as the output, denoted as $\mathbf{Y} = [\cos(\theta);\ \sin(\theta)]$ [30, 33].

The comparison of the averaged yaw angle errors by the proposed method and the four baselines is given in Table 3, from which we can see that PSVT + LSR yields smaller yaw angle errors than T-RPCA + LSR. This is because the performance of T-RPCA is influenced by the number of tensors, so it cannot work well when there are lots of outliers and heavy sparse noise.

We can also notice that LR-RR outperforms the above two methods, as it is based on a supervised learning to better tackle the impact of outliers and heavy sparse noises. LRS-RR produces even smaller yaw angle errors, as it also takes into account the outliers and sparse errors lying in disjoint subspaces. Our approach is superior to all of the baselines because it combines the affine transformations with more robust low-rank-sparse representation, entailing better resilience to outliers and heavy sparse noises.

To further verify the performance of the proposed method as an illustration, some projections of the head pose images onto the output space $\mathbf{Y}$ are also furnished in Figure 2, from which we can find that the head poses predicted by the new novel method are close to the true ones compared with the other baselines, as shown in Figure 2(e). The superiority of the new approach is to combine the affine transformations with more robust low-rank-sparse representation, so it is more robust against the cumbersome noises, outliers, and heavy sparse noises. This is again in agreement with the results in Table 3.

7. Conclusions

In this work, we considered affine transformation for image recovery and head pose estimation to remove the potential impacts of annoying effects in statistical signal processing. This approach is very useful to correct the distorted or misaligned images. The determination of the affine transformations as well as the optimization parameters are formulated as a convex optimization problem. Thereafter, the ADMM approach is considered and a new set of parameters and equations is derived to update the parameters and affine transformations iteratively in a round-robin manner. Additionally, the convergence of the developed updating equations is addressed as well. The experimental conducted simulations show that the new approach outperforms the state-of-the-art method for head pose estimation and face recovery on some common databases.

Appendix

In this appendix, we will prove Theorem 1 in Section 5. First, let $\Phi = \mathbf{AC} + \mathbf{E}$, $f(\mathbf{A}) = \|\mathbf{A}\|_*$, and $g(\mathbf{E}) = \|\mathbf{E}\|_1$, all of which are convex functions in (1). By Proposition 2(a), $\{\mathbf{A}^{(k)}, \mathbf{E}^{(k)}, \mathbf{Z}_1^{(k)}, \mathbf{Z}_2^{(k)}\}$ are bounded, so the accumulation point of $\{\mathbf{A}^{(kj)}, \mathbf{E}^{(kj)}, \mathbf{Z}_1^{(kj)}, \mathbf{Z}_2^{(kj)}\}$ is $\{\mathbf{A}^\infty, \mathbf{E}^\infty, \mathbf{Z}_1^\infty, \mathbf{Z}_2^\infty\}$. We proceed our proof in two steps.

First, we show that $\{\mathbf{A}^\infty, \mathbf{E}^\infty, \mathbf{Z}_1^\infty, \mathbf{Z}_2^\infty\}$ converge to a KKT point of problem (1). By Proposition 2(b), we have

$$\begin{aligned}\alpha\left(\mathbf{A}^{(k+1)}\right) + \Psi\left(\mathbf{E}^{(k+1)}\right) - \mathbf{\Phi} \\ = \mu_k^{-1}\left(\mathbf{Z}_1^{(k+1)} - \mathbf{Z}_1^{(k)}\right) + \mu_k^{-1}\left(\mathbf{Z}_2^{(k+1)} - \mathbf{Z}_2^{(k)}\right) \longrightarrow 0,\end{aligned} \tag{A.1}$$

which implies the accumulation points of $\{\mathbf{A}^{(k)}, \mathbf{E}^{(k)}\}$ are a feasible solution.

Let $k = k_j - 1$ in Proposition 1, and using the subgradients, we can get

$$\begin{aligned} f\left(\mathbf{A}^{(kj)}\right) + g\left(\mathbf{E}^{(kj)}\right) \le f\left(\mathbf{A}^*\right) + g\left(\mathbf{E}^*\right) \\ + \langle \mathbf{A}^{(kj)} - \mathbf{A}^*, -\mu^{(kj-1)}\boldsymbol{\tau}_{\mathbf{A}}\left(\mathbf{A}^{(kj)} - \mathbf{A}^{(kj-1)}\right) - \alpha^*\left(\mathbf{Z}_1^{(kj)}\right)\rangle \\ + \langle \mathbf{A}^{(kj)} - \mathbf{A}^*, -\mu^{(kj-1)}\boldsymbol{\tau}_{\mathbf{A}}\left(\mathbf{A}^{(kj)} - \mathbf{A}^{(kj-1)}\right) - \alpha^*\left(\mathbf{Z}_2^{(kj)}\right)\rangle \\ + \langle \mathbf{E}^{(kj)} - \mathbf{E}^*, -\mu^{(kj-1)}\eta_{\mathbf{E}}\left(\mathbf{E}^{(kj)} - \mathbf{E}^{(kj-1)}\right) - \Psi^*\left(\mathbf{Z}_1^{(kj)}\right)\rangle \\ + \langle \mathbf{E}^{(kj)} - \mathbf{E}^*, -\mu^{(kj-1)}\eta_{\mathbf{E}}\left(\mathbf{E}^{(kj)} - \mathbf{E}^{(kj-1)}\right) - \Psi^*\left(\mathbf{Z}_2^{(kj)}\right)\rangle. \end{aligned} \tag{A.2}$$

If we let $j \longrightarrow +\infty$, by Proposition 2(b), we have

$$\begin{aligned} f(\mathbf{A}^{\infty}) + g(\mathbf{E}^{\infty}) \leq\, & f(\mathbf{A}^{*}) + g(\mathbf{E}^{*}) + \langle \mathbf{A}^{\infty} - \mathbf{A}^{*}, -\alpha(\mathbf{Z}_1^{\infty})\rangle + \langle \mathbf{A}^{\infty} - \mathbf{A}^{*}, -\alpha(\mathbf{Z}_2^{\infty})\rangle \\ & + \langle \mathbf{E}^{\infty} - \mathbf{E}^{*}, -\Psi^{*}(\mathbf{Z}_1^{\infty})\rangle + \langle \mathbf{E}^{\infty} - \mathbf{E}^{*}, -\Psi^{*}(\mathbf{Z}_2^{\infty})\rangle \\ =\, & f(\mathbf{A}^{*}) + g(\mathbf{E}^{*}) - \langle \alpha(\mathbf{A}^{\infty} - \mathbf{A}^{*}), \mathbf{Z}_1^{\infty}\rangle - \langle \alpha(\mathbf{A}^{\infty} - \mathbf{A}^{*}), \mathbf{Z}_2^{\infty}\rangle \\ & - \langle \Psi(\mathbf{E}^{\infty} - \mathbf{E}^{*}), \mathbf{Z}_1^{\infty}\rangle - \langle \Psi(\mathbf{S}^{\infty} - \mathbf{E}^{*}), \mathbf{Z}_2^{\infty}\rangle \\ =\, & f(\mathbf{A}^{*}) + g(\mathbf{E}^{*}) - \langle \alpha(\mathbf{A}^{\infty}) + \mathbf{\Psi}(\mathbf{E}^{\infty}) - \alpha(\mathbf{A}^{*}) - \Psi(\mathbf{E}^{*}), \mathbf{Z}_1^{\infty}\rangle - \alpha(\mathbf{A}^{*}) \\ & + \langle \Psi(\mathbf{E}^{*}), \mathbf{Z}_2^{\infty}\rangle \\ =\, & f(\mathbf{A}^{*}) + g(\mathbf{E}^{*}). \end{aligned} \tag{A.3}$$

Thus, we can see that both $\{\mathbf{A}^{\infty}, \mathbf{E}^{\infty}\}$ and $\{\mathbf{A}^{*}, \mathbf{E}^{*}\}$ are feasible solutions. Therefore, $\{\mathbf{A}^{\infty}, \mathbf{S}^{\infty}\}$ is an optimal solution of (1).

With the definition of the subgradients and letting $k = k_j - 1$, we can get

$$\begin{aligned} f(\mathbf{A}) \geq\, & f\left(\mathbf{A}^{(kj)}\right) + \langle \mathbf{A} - \mathbf{A}^{(kj)}, -\mu^{(kj)}\boldsymbol{\tau}_{\mathbf{A}}\left(\mathbf{A}^{(kj)} - \mathbf{A}^{(kj-1)}\right) - \Psi^{*}\left(\mathbf{Z}_1^{(kj)}\right)\rangle \\ & + \langle \mathbf{A} - \mathbf{A}^{(kj)}, -\mu^{(kj)}\boldsymbol{\tau}_{\mathbf{A}}\left(\mathbf{A}^{(kj)} - \mathbf{A}^{(kj-1)}\right) - \Psi^{*}\left(\mathbf{Z}_1^{(kj)}\right)\rangle. \end{aligned} \tag{A.4}$$

Fixing $\mathbf{A}$ and letting $j \longrightarrow \infty$, we can obtain

$$\begin{aligned} f(\mathbf{A}) \geq\, & f(\mathbf{A}^{\infty}) + \langle \mathbf{A} - \mathbf{A}^{\infty}, -\Psi^{*}(\mathbf{Z}_1^{\infty})\rangle \\ & + \langle \mathbf{A} - \mathbf{A}^{\infty}, -\mathbf{\Psi}^{*}(\mathbf{Z}_2^{\infty})\rangle, \forall \mathbf{A}. \end{aligned} \tag{A.5}$$

Thus, $-\alpha^{*}(\mathbf{Z}_1^{\infty}) \in \partial f(\mathbf{A}^{\infty})$ and $-\alpha^{*}(\mathbf{Z}_2^{\infty}) \in \partial f(\mathbf{A}^{\infty})$. Similarly, $-\Psi^{*}(\mathbf{Z}_1^{\infty}) \in \partial g(\mathbf{E}^{\infty})$ and $-\Psi^{*}(\mathbf{Z}_2^{\infty}) \in \partial g(\mathbf{E}^{\infty})$. Therefore, we can conclude that $\{\mathbf{A}^{\infty}, \mathbf{E}^{\infty}, \mathbf{Z}_1^{\infty}, \mathbf{Z}_2^{\infty}\}$ converges to a KKT point of problem (1).

Next, we prove that the sequence $\left\{\mathbf{A}^{(k)}, \mathbf{E}^{(k)}, \mathbf{Z}_1^{(k)}, \mathbf{Z}_2^{(k)}\right\}$ will converge to a KKT point of problem (1). By choosing $\{\mathbf{A}^{*}, \mathbf{E}^{*}, \mathbf{Z}_1^{*}, \mathbf{Z}_2^{*}\}$ as $\{\mathbf{A}^{\infty}, \mathbf{E}^{\infty}, \mathbf{Z}_1^{\infty}, \mathbf{Z}_2^{\infty}\}$ in Proposition 2, we have

$$\begin{aligned} & \boldsymbol{\tau}_{\mathbf{A}}\left\|\mathbf{A}^{(kj)} - \mathbf{A}^{\infty}\right\|^2 - \left\|\alpha\left(\mathbf{A}^{(kj)}\right) - \mathbf{A}^{\infty}\right\|^2 + \eta_{\mathbf{E}}\left\|\mathbf{E}^{(kj)} - \mathbf{E}^{\infty}\right\|^2 - \left\|\Psi\left(\mathbf{E}^{(kj)}\right) - \mathbf{E}^{\infty}\right\|^2 \\ & + \mu^{(kj)}\left\|\mathbf{Z}_1^{(kj)} - \mathbf{Z}_1^{\infty}\right\|^2 + \mu^{(kj)}\left\|\mathbf{Z}_2^{(kj)} - \mathbf{Z}_2^{\infty}\right\|^2 \longrightarrow 0. \end{aligned} \tag{A.6}$$

By Proposition 2(a), we can obtain

$$\begin{aligned} & \boldsymbol{\tau}_{\mathbf{A}}\left\|(\mathbf{A}) - \mathbf{A}^{\infty}\right\|^2 - \left\|\alpha\left(\mathbf{A}^{(k)} - \mathbf{A}^{\infty}\right)\right\|^2 + \eta_{\mathbf{E}}\left\|\mathbf{E}^{(k)} - \mathbf{E}^{\infty}\right\|^2 - \left\|\Psi\left(\mathbf{E}^{(k)}\right) - \mathbf{E}^{\infty}\right\|^2 \\ & + \mu_k^2\left\|\mathbf{Z}_2^{(k)} - \mathbf{Z}_2^{\infty}\right\|^2 \longrightarrow 0. \end{aligned} \tag{A.7}$$

Therefore, $\left\{\mathbf{A}^{(k)}, \mathbf{E}^{(k)}, \mathbf{Z}_1^{(k)}, \mathbf{Z}_2^{(k)}\right\} \longrightarrow \{\mathbf{A}^{\infty}, \mathbf{E}^{\infty}, \mathbf{Z}_1^{\infty}, \mathbf{Z}_2^{\infty}\}$ and we can conclude that $\left\{\mathbf{A}^{(k)}, \mathbf{E}^{(k)}, \mathbf{Z}_1^{(k)}, \mathbf{Z}_2^{(k)}\right\}$ converges to a KKT point of problem (1). It thus completes the proof.

Data Availability

The data used in this article are freely available for the readers.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant no. 2018YFB1305700 and Quanzhou Scientific and Technological Planning Projects under Grant no. 2019CT009 and Department of Statistics in Addis Ababa University for their contribution in providing materials to successfully publish this latest work. The authors' final warmest appreciation goes to Prof. Wen-Hsien Fang and Prof. Jen-Shiou Leu.

References

[1] H. Van Luong, B. Joukovsky, Y. C. Eldar, and N. Deligiannis, "A deep-unfolded reference-based rpca network for video foreground-background separation," in *Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 1432–1436, IEEE, Amsterdam, Netherlands, January 2021.

[2] X. Wei, Y. Li, H. Shen, F. Chen, M. Kleinsteuber, and Z. Wang, "Dynamical textures modeling via joint video dictionary learning," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2929–2943, 2017.

[3] G. Zheng, Y. Song, and C. Chen, "Height measurement with meter wave polarimetric mimo radar: signal model and music-like algorithm," *Signal Processing*, vol. 190, Article ID 108344, 2021.

[4] H. T. Likassa, W.-H. Fang, and Y.-A. Chuang, "Modified robust image alignment by sparse and low rank decomposition for highly linearly correlated data," in *Proceedings of the 2018 3rd International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, pp. 1–4, IEEE, Yilan, Taiwan, April 2018.

[5] Q. Song, G. Yan, G. Tang, and F. Ansari, "Robust principal component analysis and support vector machine for detection of microcracks with distributed optical fiber sensors," *Mechanical Systems and Signal Processing*, vol. 146, Article ID 107019, 2021.

[6] X. Wei, H. Shen, and M. Kleinsteuber, "Trace quotient meets sparsity: a method for learning low dimensional image representations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5268–5277, Las Vegas, NV, USA, June 2016.

[7] Y. Wang, X. Wei, H. Shen, L. Ding, and J. Wan, "Robust fusion for rgb-d tracking using cnn features," *Applied Soft Computing*, vol. 92, Article ID 106302, 2020.

[8] X. Wei, H. Shen, and M. Kleinsteuber, "Trace quotient with sparsity priors for learning low dimensional image representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 12, pp. 3119–3135, 2019.

[9] H. Zeng, X. Xie, and J. Ning, "Hyperspectral image denoising via global spatial-spectral total variation regularized nonconvex local low-rank tensor approximation," *Signal Processing*, vol. 178, Article ID 107805, 2021.

[10] H. T. Likassa, W. Xian, and X. Tang, "New robust regularized shrinkage regression for high-dimensional image recovery and alignment via affine transformation and tikhonov regularization," *International Journal of Mathematics and Mathematical Sciences*, vol. 2020, 2020.

[11] Z. Tang, T.-H. Chang, X. Ye, and H. Zha, *Low-rank Matrix Recovery with Unknown Correspondence*, 2021, https://arxiv.org/abs/2110.07959.

[12] L. Zhao, Y. Liu, J. Zhao et al., "Robust pca-deep belief network surrogate model for distribution system topology identification with ders," *International Journal of Electrical Power & Energy Systems*, vol. 125, Article ID 106441, 2021.

[13] S. Han, E.-S. Cho, I. Park, K. Shin, and Y.-G. Yoon, "Efficient neural network approximation of robust pca for automated analysis of calcium imaging data," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 595–604, Springer, Strasbourg, France, September 2021.

[14] X. Wei, H. Shen, Y. Li et al., "Reconstructible nonlinear dimensionality reduction via joint dictionary learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 175–189, 2018.

[15] X. Xiang and T. D. Tran, "Linear disentangled representation learning for facial actions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 12, pp. 3539–3544, 2018.

[16] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1380–1410, 2018.

[17] D. Huang, M. Storer, F. De la Torre, and H. Bischof, "Supervised local subspace learning for continuous head pose estimation," in *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2928, IEEE, Hannover, Germany, June 2011.

[18] V. Drouard, S. Ba, G. Evangelidis, A. Deleforge, and R. Horaud, "Head pose estimation via probabilistic high-dimensional regression," in *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4624–4628, IEEE, Quebec City, QC, Canada, September 2015.

[19] Y. Xu, S. Chen, J. Li, L. Luo, and J. Yang, "Learnable low-rank latent dictionary for subspace clustering," *Pattern Recognition*, vol. 120, Article ID 108142, 2021.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, p. 11, 2011.

[21] S. Ebadi and E. Izquierdo, "Approximated rpca for fast and efficient recovery of corrupted and linearly correlated images and video frames," in *Proceedings of the Systems, Signals and Image Processing (IWSSIP)*, pp. 49–52, IEEE, London, UK, September 2015.

[22] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization," in *Proceedings of the Advances in neural information processing systems*, pp. 2080–2088, Vancouver, British Columbia, Canada, December 2009.

[23] T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim, and I. S. Kweon, "Partial sum minimization of singular values in robust pca: algorithm and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 744–758, 2016.

[24] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, 2018.

[25] M. Rahmani and P. Li, "Closed-form, provable, and robust pca via leverage statistics and innovation search," *IEEE Transactions on Signal Processing*, vol. 69, 2021.

[26] H. Ji, R. Liu, F. Su, Z. Su, and Y. Tian, "Robust head pose estimation via convex regularized sparse regression," in *Proceedings of the 2011 18th IEEE International Conference on Image Processing*, pp. 3617–3620, IEEE, Brussels, Belgium, September 2011.
[27] Y. Gao, T. Lin, Y. Zhang, S. Luo, and F. Nie, "Robust principal component analysis based on discriminant information," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
[28] H. Xu, C. Caramanis, and S. Mannor, "Robust regression and lasso," *Advances in Neural Information Processing Systems*, vol. 56, no. 7, pp. 1801–1808, 2009.
[29] Y. Wang, C. Dicle, M. Sznaier, and O. Camps, "Self scaled regularized robust regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3261–3269, Boston, MA, USA, June 2015.
[30] D. Huang, R. Cabral, and F. D. Torre, "Robust regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 363–375, 2016.
[31] M. Yin, D. Zeng, J. Gao, Z. Wu, and S. Xie, "Robust multinomial logistic regression based on rpca," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1144–1154, 2018.
[32] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 156–171, 2017.
[33] Y. Zhang, D. Shi, J. Gao, and D. Cheng, "Low-rank-sparse subspace representation for robust regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7445–7454, Honolulu, HI, USA, July 2017.
[34] J. Chen, S. Yang, H. Mao, and C. Fahy, "Multiview subspace clustering using low-rank representation," *IEEE Transactions on Cybernetics*, 2021.
[35] D. Zeng, Z. Wu, C. Ding, Z. Ren, Q. Yang, and S. Xie, "Labeled-robust regression: simultaneous data recovery and classification," *IEEE Transactions on Cybernetics*, 2020.
[36] Z. Wu, C. Su, M. Yin, Z. Ren, and S. Xie, "Subspace clustering via stacked independent subspace analysis networks with sparse prior information," *Pattern Recognition Letters*, vol. 146, pp. 165–171, 2021.
[37] J. Yang, J. Qu, Q. Mi, and Q. Li, "A cnn-lstm model for tailings dam risk prediction," *IEEE Access*, vol. 8, pp. 206491–206502, 2020.
[38] H. T. Likassa, W.-H. Fang, and J.-S. Leu, "Robust image recovery via affine transformation and $L_{2,1}$ norm," *IEEE Access*, vol. 7, pp. 125011–125021, 2019.
[39] H. T. Likassa, "New robust principal component analysis for joint image alignment and recovery via affine transformations, frobenius and norms," *International Journal of Mathematics and Mathematical Sciences*, vol. 2020, 2020.
[40] Y. Wang, X. Luo, L. Ding, S. Fu, and X. Wei, "Detection based visual tracking with convolutional neural network," *Knowledge-Based Systems*, vol. 175, pp. 62–71, 2019.
[41] G. R. Obozinski, M. J. Wainwright, and M. I. Jordan, "High-dimensional support union recovery in multivariate regression," in *Proceedings of the Advances in neural information processing systems*, pp. 1217–1224, Vancouver, British Columbia, Canada, December 2009.
[42] X. Wei, *Learning Image and Video Representations Based on Sparsity Priors*, Technische Universität München, PhD thesis, 2017.
[43] Y. Jin and B. D. Rao, "Algorithms for robust linear regression by exploiting the connection to sparse signal recovery," in *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3830–3833, IEEE, Dallas, TX, USA, MArch 2010.
[44] R. Calabrese, "Regression for recovery rates with both continuous and discrete characteristics," in *Proceedings of the Italian Statistical Society Conference*, Padua, June 2010.
[45] X.-Y. Zhang, L. Wang, S. Xiang, and C.-L. Liu, "Retargeted least squares regression algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 9, pp. 2206–2213, 2015.
[46] S. Shiming Xiang, F. Feiping Nie, G. Gaofeng Meng, C. Chunhong Pan, and C. Changshui Zhang, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1738–1754, 2012.
[47] Y. Zhang and D.-Y. Yeung, "Worst-case linear discriminant analysis," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2568–2576, Vancouver, British Columbia, Canada, December 2010.
[48] F. Bunea, Y. She, and M. H. Wegkamp, "Optimal selection of reduced rank estimators of high-dimensional matrices," *Annals of Statistics*, vol. 39, no. 2, pp. 1282–1309, 2011.
[49] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee, "Alternating least-squares for low-rank matrix reconstruction," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 231–234, 2012.
[50] A. Liu, G. Shi, S.-J. Chung, A. Anandkumar, and Y. Yue, "Robust regression for safe exploration in control," in *Learning for Dynamics and Control*vol. 120, , pp. 608–619, PMLR, 2020.
[51] C. Chen, J. Huang, L. He, and H. Li, *Fast Iteratively Reweighted Least Squares Algorithms for Analysis-Based Sparsity Reconstruction*, 2014, https://arxiv.org/abs/1411.5057.
[52] H. Barreto and D. Maharry, "Least median of squares and regression through the origin," *Computational Statistics & Data Analysis*, vol. 50, no. 6, pp. 1391–1397, 2006.
[53] X. Wei, Y. Huang, Y. Xu et al., *Learning Robust and Lightweight Model through Separable Structured Transformations*, 2021, https://arxiv.org/abs/2112.13551.
[54] X. Cai, C. Ding, F. Nie, and H. Huang, "On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1124–1132, ACM, Chicago Illinois USA, August 2013.
[55] H. Abdi, "Partial least squares regression and projection on latent structure regression (pls regression)," *WIREs Computational Statistics*, vol. 2, no. 1, pp. 97–106, 2010.
[56] E. Richardson, M. Sela, R. Or-El, and R. Kimmel, "Learning detailed face reconstruction from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1259–1268, Honolulu, HI, USA, July 2017.
[57] R. Zhang, F. Nie, M. Guo, X. Wei, and X. Li, "Joint learning of fuzzy k-means and nonnegative spectral clustering with side information," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2152–2162, 2018.
[58] M. Shakeri and H. Zhang, "Corola: a sequential solution to moving object detection using low-rank approximation," *Computer Vision and Image Understanding*, vol. 146, pp. 27–39, 2016.
[59] G. Papageorgiou, P. Bouboulis, and S. Theodoridis, "Robust nonlinear regression: a greedy approach employing kernels

with application to image denoising," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4309–4323, 2017.

[60] Z. Wang, D. Hu, X. Luo, W. Wang, J. Wang, and W. Chen, "Performance guarantees of transformed schatten-1 regularization for exact low-rank matrix recovery," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 12, pp. 3379–3395, 2021.

[61] J. Sun, W. Wang, X. Wei et al., "Deep clustering with intraclass distance constraint for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4135–4149, 2020.

[62] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Head pose estimation: classification or regression?" in *Proceedings of the 2008 19th International Conference on Pattern Recognition*, pp. 1–4, IEEE, Tampa, FL, USA, December 2008.

[63] H. Yu and H. Liu, "Linear regression for head pose analysis," in *Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 987–992, IEEE, Beijing, China, July 2014.

[64] S. Lathuilière, R. Juge, P. Mesejo, R. Munoz-Salinas, and R. Horaud, "Deep mixture of linear inverse regressions applied to head-pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4817–4825, Honolulu, HI, USA, July 2017.

[65] C. BenAbdelkader, "Robust head pose estimation using supervised manifold learning," in *Proceedings of the European Conference on Computer Vision*, pp. 518–531, Springer, Heraklion, Crete, Greece, September 2010.

[66] K. Sundararajan and D. L. Woodard, "Head pose estimation in the wild using approximate view manifolds," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 50–58, Boston, MA, USA, June 2015.

[67] M. Demirkus, D. Precup, J. J. Clark, and T. Arbel, "Probabilistic temporal head pose estimation using a hierarchical graphical model," in *Proceedings of the European conference on computer vision*, pp. 328–344, Springer, Zurich, Switzerland, September 2014.

[68] V. Drouard, R. Horaud, A. Deleforge, S. Ba, and G. Evangelidis, "Robust head-pose estimation based on partially-latent mixture of linear regressions," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1428–1440, 2017.

[69] K. Diaz-Chito, A. Hernández-Sabaté, and A. M. López, "A reduced feature set for driver head pose estimation," *Applied Soft Computing*, vol. 45, pp. 98–107, 2016.

[70] G. P. Meyer, S. Gupta, I. Frosio, D. Reddy, and J. Kautz, "Robust model-based 3d head pose estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3649–3657, Santiago, Chile, December 2015.

[71] Z. Lin, M. Chen, and Y. Ma, *The Augmented lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices*, 2010, https://arxiv.org/abs/1009.5055.

[72] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th international conference on machine learning*, pp. 663–670, ICML-10), Haifa Israel, June 2010.

[73] J. Liu, Y. Cheng, X. Wang, and S. Ge, "One-step robust low-rank subspace segmentation for tumor sample clustering," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[74] Y. Yigang Peng, A. Ganesh, J. Wright, W. Wenli Xu, and Y. Yi Ma, "RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2233–2246, 2012.

[75] C. Zhang, H. T. Likassa, P. Liang, and J. Guo, "New robust part-based model with affine transformations for facial landmark localization and detection in big data," *Modelling and Simulation in Engineering*, vol. 2021, 2021.

[76] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proceedings of the Advances in neural information processing systems*, pp. 612–620, Vancouver, British Columbia, Canada, December 2011.

[77] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[78] L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang, and N. Yu, "Non-negative low rank and sparse graph for semi-supervised learning," in *Proceedings of the2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2328–2335, IEEE, Providence, RI, USA, June 2012.

[79] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[80] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[81] P. Courrieu, *Fast Computation of moore-penrose Inverse Matrices*, 2008, https://arxiv.org/abs/0804.4809.

[82] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[83] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.