IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 16, 2019, accepted July 25, 2019, date of publication August 1, 2019, date of current version September 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932470

Robust Image Recovery via Affine Transformation and $L_{2,1}$ Norm

HABTE TADESSE LIKASSA, WEN-HSIEN FANG, AND JENQ-SHIOU LEU
National Taiwan University of Science and Technology, Taipei 105-95, Taiwan

Corresponding author: Habte Tadesse Likassa (d10502801@mail.ntust.edu.tw)

This work was supported by the Ministry of Science and Technology (MOST) under Contract MOST 107-2221-E-011-078-MY2.

ABSTRACT In this paper, we propose a novel robust algorithm for image recovery via affine transformations and the $L_{2,1}$ norm. To be robust against miscellaneous adverse effects such as occlusions, outliers, and heavy sparse noise, the new algorithm integrates affine transformations with low-rank plus sparse decomposition, where the low-rank component lies in a union of disjoint subspaces, so the distorted or misaligned images can be rectified to render more faithful image representation. In addition, the $L_{2,1}$ norm is employed to remove the correlated samples across the images, enabling the new approach to be more resilient to outliers and large variations in the images. The determination of the variables involved and the affine transformations is cast as a convex optimization problem. To alleviate the computational complexity, the Alternating Direction Method of Multipliers (ADMM) method is utilized to derive a new set of recursive equations to update the optimization variables and the affine transformations iteratively in a round-robin manner. The convergence of the developed updating equations is addressed and experimentally validated as well. The conducted simulations demonstrate that the new algorithm is superior to the state-of-the-art works in terms of accuracy on some public databases.

INDEX TERMS Affine transformation, $L_{2,1}$ norm, robust image recovery, low-rank plus sparse representation.

I. INTRODUCTION

Image recovery has found applications in a variety of areas such as medical imaging, wireless sensor networks, surveillance, batch image denoising, computational imaging, *etc.* [1]–[6]. Image recovery can also be used in background extraction, where the low-rank component corresponds to the background and the sparse component captures the foreground [7]–[9]. However, this problem faces some severe challenges such as illumination variation, occlusion, outliers and heavy sparse noise. It is thus important to develop robust image recovery algorithms to tackle the above adverse effects.

A number of robust algorithms have been addressed for the image alignment or image recovery problems. For instance, De la Torre *et al.* [10] proposed a parameterized component analysis algorithm to find the low-rank component, and used a robust fitting function to reduce the impact of outliers. However, it is non-convex and thus lacks a polynomial-time algorithm to solve the problem. To tackle this dilemma, Ebadi and Izquierdo [11] considered an approximated robust principal component analysis for the recovery of corrupted and linearly correlated images. Chen *et al.* [12] addressed a Non-convex plus Quadratic penalized a Low-rank and Sparse Decomposition (NQLSD) algorithm to clean the outliers and sparse errors. However, it was unable to recover and align images when the outliers and noise follow statistical distributions other than the mixture of Gaussian distributions. Peng *et al.* [13] considered a Robust Alignment for Sparse and Low-rank decomposition (RASL) algorithm by affine transformations to recover the low-rank component. However, it only considers the outliers from a single subspace. Oh *et al.* [14] proposed a new method, Partial Singular Value Thresholding (PSVT), which used the partial sum of the singular values to replace the nuclear norm in the Robust Principal Component Analysis (RPCA) algorithm [15] to find the low-rank component. It is, however, not robust against outliers and sparse errors when there are lots of data samples. He *et al.* [16] considered a similar efficient convex relaxation algorithm. However, it cannot deal with outliers lying outside the main subspaces. To overcome this setback, [17] proposed an Image Alignment Robust Principal Component Analysis via stochastic gradient

The associate editor coordinating the review of this manuscript and approving it for publication was Md Asaduzzaman.

descent (IA-RPCA) by combining geometric transformations with the online RPCA. However, it fails in aligning images when there are severe occlusions, distortions and outliers. To resolve this drawback, Zheng *et al.* [18] considered an optimization problem for batch image alignment via Image Gradient Orientations for Online Robust Image Alignment (IGO-RIA) to minimize the impact of sparse errors.

In this paper, we propose a novel robust algorithm for image recovery via affine transformations and the $L_{2,1}$ norm. To be robust against miscellaneous adverse effects such as occlusions, outliers and heavy sparse noise, the new algorithm integrates affine transformations with low-rank plus sparse decomposition, where the low-rank component lies in a union of disjoint subspaces, so the distorted or misaligned images can be rectified to render more faithful image representation. In addition, the $L_{2,1}$ norm, which enjoys the advantages of the L_1 and L_2 norms, is employed to remove the correlated samples across the images, enabling the new approach to be more resilient to outliers and large variations in the images. The determination of the variables involved and the affine transformations is cast as a convex optimization problem. To alleviate the computational complexity, the Alternating Direction Method of Multipliers (ADMM) method is utilized to derive a new set of recursive equations to update the optimization variables and the affine transformations iteratively in a round-robin manner. The convergence of the developed updating equations is addressed as well. Conducted simulations demonstrate that the new algorithm is superior to the state-of-the-art works in terms of accuracy on some public databases.

The major contributions of this paper include:

1) The affine transformations are aggregated with the low-rank plus sparse representation, where the low-rank component lies in a union of subspaces instead of one single subspace. These transformations can fix the distortion or misalignment in a batch of corrupted images to render more faithful image decomposition, thereby being more robust against heavy sparse errors and outliers.

2) Since large errors in images may happen, which will impact the accuracy of image recovery, the $L_{2,1}$ norm is utilized. This new norm, when combined with the affine transformations, can further enhance the performance.

3) The ADMM method is employed to solve the new convex optimization problem and a new set of updating equations is derived to iteratively update the optimization variables and the affine transformations.

4) The convergence of the developed iterative equations is addressed and experimentally justified.

The remainder of this paper is organized as follows. Sec. II gives an overview of the related works. Sec. III describes the formulation of the problem. Secs. IV and V develop updating equations to solve the proposed problem and analyze its convergence characteristics, respectively. In Sec. VI, experimental results are conducted to justify the proposed method. Sec. VII provides some concluding remarks to summarize the paper.

II. RELATED WORKS

There has been a considerable amount of research in image alignment and image recovery. For instance, Vedaldi *et al.* [27] proposed to minimize a log-determinant cost function to tackle the potential impact of sparse errors in image alignment. Lia and Fang [28] addressed an image alignment algorithm by explicitly modeling the spatially varying illumination multiplication and bias factors by low-order polynomials. However, it does not work well when there are severe outliers and sparse errors in the data. To circumvent this setback, [29], [30] considered a RPCA algorithms, which decomposed the corrupted data into a summation of a low-rank component and sparse errors. Podosinnikova *et al.* [31] developed a robust PCA to minimize the reconstruction error. Shahid *et al.* [32] incorporated the spectral graph regularization with the robust PCA. Shaker *et al.* [33] proposed an online sequential framework to recover the low-rank component by pruning the sparse corruptions. Hu *et al.* [34] introduced an approximate of the low-rank assumption for the matrix via a low-rank regularization to solve the face image denoising problems. Zhang and Lerman [35] proposed a robust subspace recovery method to reduce the influence of outliers and sparse noise. However, its complexity is high when the outliers and sparse noise are heavy in the data.

Extensive approaches have also been investigated to solve the low-rank subspace decomposition problems. Zhang *et al.* [36] addressed a linear subspace learning approach via the low-rank decomposition. Zhao *et al.* [37] proposed a robust discriminant low-rank representation to explore the multiple subspace structures. Ma *et al.* [38] proposed a generalized Principal Component Analysis (GPCA), which utilized an algebraic approach to model the data drawn from a union of subspaces. Liu *et al.* [26] proposed a low-rank representation to identify the subspace structures from the corrupted data. Recently, Rao *et al.* [39] introduced a compressed sensing technique to subspace segmentation. Elhamifar and Vidal [40] considered a Sparse Subspace Clustering (SSC) approach which used the sparsest representation produced by L_1-minimization [41] to define the affinity matrix of an undirected graph. The subspace segmentation is performed by spectral clustering algorithms such as the normalized cuts [42] and subspace clustering [42], [43]. These algorithms, however, are not robust against occlusions and illuminations. To resolve this drawback, [44] proposed a transformation invariant subspace clustering framework by jointly aligning data samples and learning the subspace representation. Shen *et al.* [45] addressed an online low-rank subspace clustering by basis dictionary pursuit to reduce complexity. Wu *et al.* [46] proposed an orthogonal subspace decomposition algorithm to mitigate the impact of noise in motion segmentation. Li *et al.* [21] arranged the input images as a 3D tensor, and claimed that severe intensity distortions and partial occlusions can be separated in the gradient and frequency domains. The optimally aligned image tensor is achieved by simultaneously specifying a frequency

tensor and a gradient error tensor. It is, however, very time-consuming. Ding and Fu [47] proposed a multi-view subspace learning algorithm through dual low-rank decomposition to seek a low-dimensional view invariant subspace from multi-view data.

To deal with higher-dimensional data, RPCA has also been extended from images to multi-way data [48]. A number of robust tensor PCA methods have been proposed. For example, Liu *et al.* [49] considered an Improved Robust Tensor Principal Component Analysis (IRTPCA) method which further exploited the low-rank structure and used an improved tensor nuclear norm to provide better performance. Compared with our work, apart from different input data type we use different norms and incorporate the affine transformations in low-rank component recovery. Liu *et al.* [48] addressed a low tensor tree rank and total variation minimization method for image completion. However, total variation mainly accounts for the piecewise smoothness, but our focus is mainly on sparsity of the noise and outliers. Also, we use the $L_{2,1}$ norm instead of the L_1 or the L_2 norms in [48], [50].

Recently, non-convex optimization has also drawn lots of research interests, as summarized in [51]–[53]. In some cases, the non-convex regularizers such as L_0, $L_{1/2}$, SCAD, MC, L_q, and $\log -q$ regularizers can indeed provide superior performance over the convex regularizers. However, in some scenarios when the signal is not strictly sparse or when the signal-to-noise ratio is low, the use of the non-convex regularization may not substantially improve the performance [51], [55]. Also, the global solution is not ensured and the analysis of the convergence characteristics algorithms using these non-convex regularizers is not an easy task.

III. PROBLEM FORMULATION

Consider n images, $\{\boldsymbol{I}_i^0\} \in \Re^{w\times h},\ i = 1, \cdots, n$, where w and h denote the weight and height of the images, respectively. All of these images contain the same objects and are highly correlated with each other. In many scenarios, these images are corrupted by occlusions and outliers. We can stack these images into a matrix: $\boldsymbol{M} = [vec(I_1^0) \mid vec(I_2^0) \mid \ldots \mid vec(I_n^0)] \in \Re^{m\times n}$, where $vec(\cdot)$ denotes the vector stacking operator. We can decompose $\boldsymbol{M}$ into a summation of a low-rank component and a sparse error matrix [56], [57]: $\boldsymbol{M} = \boldsymbol{AC} + \boldsymbol{E}$, where we assumed that the subspaces may not be independent from each other or the data are contaminated by large noise, so the low-rank component lies in a union of subspaces, in which $\boldsymbol{A} \in \Re^{m\times n}$ is a clean low-rank dictionary and $\boldsymbol{C} \in \Re^{n\times n}$ is a coefficient matrix as the low-rank representation by $\boldsymbol{A}$, and $\boldsymbol{E} \in \Re^{m\times n}$ denotes a sparse error matrix incurred by outliers or corruptions.

In practice, $\boldsymbol{I}_i^0$ are generally not well aligned, entailing the above low-rank plus sparse decomposition to be imprecise. To take account of this, inspired by [13] we apply affine transformations $\boldsymbol{\tau}_i$ to the potentially misaligned input images $\boldsymbol{I}_i^0$ to get a collection of transformed images $\boldsymbol{I}_i = \boldsymbol{I}_i^0 o \boldsymbol{\tau}_i$, where the operator o indicates the transformation. We can then stack these aligned images into a matrix and obtain $\boldsymbol{M}_{o\boldsymbol{\tau}} = [vec(I_1) \mid vec(I_2) \mid \ldots \mid vec(I_n)] \in \Re^{m\times n}$. The aligned images can be treated as samples taken from a union of low-dimensional subspaces, which, if well aligned, should exhibit a low-rank subspace structure as the rank of the transformed images is as small as possible, up to some outliers and heavy sparse errors. Solving for the variables corresponding to the constraints $\boldsymbol{M}_{o\boldsymbol{\tau}} = \boldsymbol{AC}\,\boldsymbol{\mathcal{C}}\,\boldsymbol{E}$ is intractable due to its nonlinearity. To resolve this dilemma, we assume that the change produced by these affine transformations $\boldsymbol{\tau}_i$ is small and an initial of $\boldsymbol{\tau}_i$ is known. We can then linearize $\boldsymbol{M}_{o\boldsymbol{\tau}}$ by using the first-order Taylor approximation as $\boldsymbol{M}_{o(\boldsymbol{\tau}+\boldsymbol{\Delta\tau})} \approx \boldsymbol{M}_{o\boldsymbol{\tau}} + \sum_{i=1}^{n} \boldsymbol{J}_i \boldsymbol{\Delta\tau} \boldsymbol{v}_i \boldsymbol{v}_i^T$, where $\boldsymbol{M}_{o\boldsymbol{\tau}} \in \Re^{m\times n}$ is the transformed image, $\boldsymbol{\Delta\tau} \in \Re^{p\times n}$ with p being the number of variables, $\boldsymbol{J}_i = \frac{\partial vec(\boldsymbol{I}_i o \boldsymbol{\tau}_i)}{\partial \boldsymbol{\tau}_i} \in \Re^{m\times p}$ denotes the Jacobian of the i^{th} image with respect to $\boldsymbol{\tau}_i$, and $\boldsymbol{v}_i$ is the standard basis for $\Re^n$. In this way, we obtain approximate transformations to recover the low-rank component from the underlying subspaces.

To make the new approach more resilient to outliers and heavy sparse noise, the $L_{2,1}$ norm, which combines the advantages of the L_1 and L_2 norm, is used here. The $L_{2,1}$ regularizer is considered as the rotational invariant of the L_1 norm and can effectively handle outliers [58]. Also, as observed in [59], the $L_{2,1}$ regularizer can also achieve better sparsity promotion than the L_1 norm. The L_1 norm may yield a biased estimation as it ignores the extreme values and cannot handle the collinearity of features. In contrast, the $L_{2,1}$ norm is more stable and has the ability to better preserve the spatial information than the L_1 regularizer, as demonstrated in [58], [59]. Additionally, the $L_{2,1}$ norm is superior to the nonconvex norms when the signals are not sparse or when the matrix is not strictly low-rank [51], [55]. The overall problem can thus be posted as an optimization problem given by

$$\begin{aligned} &\min_{\boldsymbol{A},\boldsymbol{E},\boldsymbol{C},\boldsymbol{Q},\boldsymbol{\Delta\tau}} \|\boldsymbol{A}\|_* + \|\boldsymbol{C}\|_* + \lambda_1 \|\boldsymbol{Q}\|_{2,1} + \lambda_2 \|\boldsymbol{E}\|_{2,1} \\ &\text{s.t}\ \boldsymbol{M}_{o\boldsymbol{\tau}} + \sum_{i=1}^{n} \boldsymbol{J}_i \boldsymbol{\Delta\tau} \boldsymbol{v}_i \boldsymbol{v}_i^T = \boldsymbol{AC} + \boldsymbol{E}, \quad \boldsymbol{C} = \boldsymbol{Q}, \boldsymbol{Q} \succeq \boldsymbol{0} \end{aligned} \tag{1}$$

where $\|\boldsymbol{A}\|_* = \sum_{i=1}^{min(m,n)} \sigma_i(\boldsymbol{A})$ denotes the nuclear norm of $\boldsymbol{A}$, in which $\sigma_i(\boldsymbol{A})$ indicates the singular values of $\boldsymbol{A}$, λ_1 and λ_2 are the regularization parameters, and $\|\boldsymbol{E}\|_{2,1} = \sum_{i=1}^{m}(\sum_{j=1}^{n} \boldsymbol{E}_{ji}^2)^{1/2}$ denotes the $L_{2,1}$ norm of $\boldsymbol{E}$.

The first nuclear norm term in (1) enforces the dictionary $\boldsymbol{A}$ to lie in low-dimensional subspaces. The second and the third terms constrain the representation to be low-rank plus sparse. The fourth term regularizes the error $\boldsymbol{E}$ to be sparse.

For easy reference, a comparison of the proposed approach with other related methods is furnished in Table 1.

IV. PROPOSED METHOD

To solve the optimization problem in (1), we use the augmented Lagrangian multiplier technique and re-write (1) as

$$\begin{aligned} &\mathcal{L}\left(\boldsymbol{A}, \boldsymbol{C}, \boldsymbol{E}, \boldsymbol{Q}, \boldsymbol{\Delta\tau}\right) \\ &\quad = \|\boldsymbol{A}\|_* + \|\boldsymbol{C}\|_* + \lambda_1 \|\boldsymbol{Q}\|_{2,1} \end{aligned}$$

TABLE 1. Comparison of the proposed approach with other related algorithms in terms of the objective function and constraints.

Methods	Objective function	Constraints
[10], [15], [19], [20]	$\min_{A,E} \|A\|_* + \lambda \|E\|_1$	$M = A + E$
[12], [18], [21], [22]	$\min_{A,E} \|A\|_* + \lambda \|E\|_1$	$M_{o\tau} + \sum_{i=1}^{n} J_i \Delta\tau v_i v_i^T = A + E$
[5], [23]–[26]	$\min_{C,E,Q} \|C\|_* + \lambda \|E\|_{2,1}$	$M = AC + E, C = Q$
Ours	$\min_{A,E,C,Q,\Delta\tau} \|A\|_* + \|C\|_* + \lambda_1 \|Q\|_{2,1} + \lambda_2 \|E\|_{2,1}$	$M_{o\tau} + \sum_{i=1}^{n} J_i \Delta\tau v_i v_i^T = AC + E, C = Q, Q \succeq 0$

$$+\lambda_2 \|E\|_{2,1} + \langle Z_1, B - AC - E\rangle + \frac{\mu_1}{2}\|B - AC - E\|_F^2 + \langle Z_2, C - Q\rangle + \frac{\mu_2}{2}\|C - Q\|_F^2 \quad (2)$$

where $Z_1 \in \Re^{m\times n}$ and $Z_2 \in \Re^{n\times n}$ are the Lagrangian multipliers, μ_1 and μ_2 are the penalty parameters, $\langle X, Y\rangle = Trace(X^T Y)$, $||\cdot||_F$ denotes the Frobenius norm, and $B = M_{o\tau} + \sum_{i=1}^{n} J_i \Delta\tau v_i v_i^T$. By using augmented Lagrange multiplier with adaptive penalty [60], Eq. (2) can be rewritten as

$$\begin{aligned}\mathcal{L}(A, C, E, Q, \Delta\tau) &= \|A\|_* + \|C\|_* + \lambda_2 \|E\|_{2,1} + \lambda_1 \|Q\|_{2,1} + \frac{\mu_1}{2}\left\|B - AC - E + \frac{Z_1}{\mu_1}\right\|_F^2 + \frac{\mu_2}{2}\left\|C - Q + \frac{Z_2}{\mu_2}\right\|_F^2 \end{aligned} \quad (3)$$

Solving (3) directly is computationally prohibitive, thereby we consider to iteratively update the variables alternatively via ADMM [60], which decomposes the original problem into a set of subproblems. In each step, we update just one variable while keeping the other variables unchanged. So the above terms like $\|B - AC - E\|_F$ are convex even if it involves the matrix factorization term AC [61].

Firstly, to update A, we fix E, Q, C and $\Delta\tau$, so $A^{(k+1)}$ can be determined by

$$A^{(k+1)} = \underset{A}{\operatorname{argmin}} \mathcal{L}\left\{A, E^{(k)}, C^{(k)}, Q^{(k)}, \Delta\tau^{(k)}\right\} \quad (4)$$

where k is the iteration index. By ignoring all irrelevant terms of A, Eq. (4) can be simplified as

$$A^{(k+1)} = \underset{A}{\operatorname{argmin}}\left\{\|A\|_* + \frac{\mu_1^{(k)}}{2}\left\|B^{(k)} - AC^{(k)} - E^{(k)} + \frac{Z_1^{(k)}}{\mu_1^{(k)}}\right\|_F^2\right\} \quad (5)$$

We can then use the linear augmented direction method with the soft shrinkage operator in [57], [62], [63] and update $A^{(k+1)}$ by

$$A^{(k+1)} = \Omega_{\frac{1}{\kappa_A^{(k)}}}\left\{A^{(k)} - \frac{P_A^{(k)}}{\kappa_A^{(k)}}\right\} \quad (6)$$

where Ω is the singular value threshold, $P_A^{(k)} = (\mu_1^{(k)})(A^{(k)})^T C^{(k)} - \mu_1^{(k)}(B^{(k)} - E^{(k)}) - (Z_1^{(k)} + Z_2^{(k)})(C^{(k)})^T$, $\kappa_A^{(k)} = (\mu_1^{(k)})\frac{\tau_A}{2}$, $\tau_A > \sigma(C^T C)$ is the proximal parameter, in which $\sigma(C^T C)$ denotes the spectral radius of $C^T C$.

Secondly, to update E, we keep A, Q, C and $\Delta\tau$ as constants, so $E^{(k+1)}$ can be determined by

$$E^{(k+1)} = \underset{E}{\operatorname{argmin}} \mathcal{L}\left\{A^{(k+1)}, E, C^{(k)}, Q^{(k)}, \Delta\tau^{(k)}\right\} \quad (7)$$

Again, by ignoring all irrelevant terms of E, Eq. (7) can be simplified as

$$E^{(k+1)} = \underset{E}{\operatorname{argmin}}\left\{\lambda_2 \|E\|_{2,1} + \frac{\mu_1^{(k)}}{2}\left\|B^{(k)} - A^{(k)}C^{(k)} - E + \frac{Z_1^{(k)}}{\mu_1^{(k)}}\right\|_F^2\right\} \quad (8)$$

By using lemma [64], the update of the i^{th} column of $E^{(k+1)}$, $E_i^{(k+1)}$, is given by

$$E_i^{(k+1)} = \begin{cases} \dfrac{\left\|V_i^{(k)}\right\|_2 - \frac{\lambda_2}{\mu_1^{(k)}}}{\left\|V_i^{(k)}\right\|_2} V_i^{(k)}, & \text{if } \left\|V_i^{(k)}\right\|_2 \geq \dfrac{\lambda_2}{\mu_1^{(k)}} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $\|.\|_2$ denotes the Euclidean norm and $V^{(k)} = (B^{(k)} - A^{(k)}C^{(k)} + \frac{Z_1^{(k)}}{\mu_1^{(k)}})$.

Thirdly, to find an update of C, we keep A, E, Q and $\Delta\tau$ as constants, and thus $C^{(k+1)}$ can be determined by

$$C^{(k+1)} = \underset{C}{\operatorname{argmin}} \mathcal{L}\left\{A^{(k+1)}, E^{(k+1)}, C, Q^{(k)}, \Delta\tau^{(k)}\right\} \quad (10)$$

By ignoring all irrelevant terms of C, Eq. (10) can be simplified as

$$C^{(k+1)} = \underset{C}{\operatorname{argmin}}\left\{\|C\|_* + \frac{\mu_1^{(k)}}{2}\left\|B^{(k)} - A^{(k)}C - E^{(k)} + \frac{Z_1^{(k)}}{\mu_1^{(k)}}\right\|_F^2 + \frac{\mu_2}{2}\left\|C - Q^{(k)} + \frac{Z_2^{(k)}}{\mu_2^{(k)}}\right\|_F^2\right\} \quad (11)$$

To solve this sub-problem, we can employ a linearized augmented Lagrangian multiplier with the singular value threshold [57], [62], [63] Thereby, $C^{(k+1)}$ can be updated by

$$C^{(k+1)} = \Omega_{\frac{1}{\kappa_C^{(k)}}}\left\{C^{(k)} - \frac{P_C^{(k)}}{\kappa_C^{(k)}}\right\} \quad (12)$$

where $\kappa_{\boldsymbol{C}}^{(k)} = (\mu_1^{(k)} + \mu_2^{(k)})\frac{\tau_{\boldsymbol{C}}}{2}$, $\tau_{\boldsymbol{C}} > \sigma(\boldsymbol{A}\boldsymbol{A}^T)$ and $P_{\boldsymbol{C}}^{(k)} = \boldsymbol{A}^{(k)}(\mu_1^{(k)} + \mu_2^{(k)})\boldsymbol{A}^{(k)} - \mu_1^{(k)}(\boldsymbol{B}^{(k)} - \boldsymbol{E}^{(k)}) - \boldsymbol{Z}_1^{(k)} + \mu_2^{(k)}(\boldsymbol{C}^{(k)} - \boldsymbol{Q}^{(k)}) + \boldsymbol{Z}_2^{(k)}$.

Again, to update $\boldsymbol{Q}$, we keep $\boldsymbol{A}$, $\boldsymbol{E}$, $\boldsymbol{C}$ and $\boldsymbol{\Delta\tau}$ as constants and $\boldsymbol{Q}^{(k+1)}$ can then be determined by

$$\boldsymbol{Q}^{(k+1)} = \underset{\boldsymbol{Q}}{\operatorname{argmin}} \mathcal{L}\left\{\boldsymbol{A}^{(k+1)}, \boldsymbol{E}^{(k+1)}, \boldsymbol{C}^{(k+1)}, \boldsymbol{Q}, \boldsymbol{\Delta\tau}^{(k)}\right\} \quad (13)$$

By ignoring all of the irrelevant terms, Eq. (13) can be simplified as

$$\boldsymbol{Q}^{(k+1)} = \underset{\boldsymbol{Q}}{\operatorname{argmin}} \left\{ \begin{array}{l} \lambda_1 \|\boldsymbol{Q}\|_{2,1} \\ +\frac{\mu_2^{(k)}}{2} \left\| \boldsymbol{C}^{(k)} - \boldsymbol{Q} + \frac{\boldsymbol{Z}_2^{(k)}}{\mu_2^{(k)}} \right\|_F^2 \end{array} \right\} \quad (14)$$

Again, by using lemma in [64], the update of the i^{th} column of $\boldsymbol{Q}^{(k+1)}$, $\boldsymbol{Q}_i^{(k+1)}$, is given by

$$\boldsymbol{Q}_i^{(k+1)} = \begin{cases} \frac{\left\|\boldsymbol{P}_i^{(k)}\right\|_2 - \frac{\lambda_1}{\mu_2^{(k)}}}{\left\|\boldsymbol{P}_i^{(k)}\right\|_2} \boldsymbol{P}_i^{(k)}, & \text{if } \left\|\boldsymbol{P}_i^{(k)}\right\|_2 \geq \frac{\lambda_1}{\mu_2^{(k)}} \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

where $\boldsymbol{P}^{(k)} = (\boldsymbol{C}^{(k+1)} + \frac{\boldsymbol{Z}_2^{(k)}}{\mu_2^{(k)}})$.

Lastly, to get an update of $\boldsymbol{\Delta\tau}$, we keep $\boldsymbol{A}$, $\boldsymbol{E}$, $\boldsymbol{C}$ and $\boldsymbol{Q}$ as constants and $\boldsymbol{\Delta\tau}^{(k+1)}$ can be determined by

$$\boldsymbol{\Delta\tau}^{(k+1)} = \underset{\boldsymbol{\Delta\tau}}{\operatorname{argmin}} \mathcal{L}\left\{\boldsymbol{A}^{(k+1)}, \boldsymbol{E}^{(k+1)}, \boldsymbol{C}^{(k+1)}, \boldsymbol{\Delta\tau}, \boldsymbol{Q}^{(k+1)}\right\} \quad (16)$$

By ignoring all irrelevant terms of $\boldsymbol{\Delta\tau}$, we can obtain

$$\boldsymbol{\Delta\tau}^{(k\mathcal{C}1)} = \underset{\boldsymbol{\Delta\tau}}{\operatorname{argmin}} \left\{ \frac{\mu_1^{(k)}}{2} \left\| \boldsymbol{B}^{(k)} - \boldsymbol{A}^{(k+1)}\boldsymbol{C}^{(k+1)} - \boldsymbol{E}^{(k+1)} + \frac{\boldsymbol{Z}_1^{(k)}}{\mu_1^{(k)}} \right\|_F^2 \right\} \quad (17)$$

Solving (17) with the threshold operators [13] and [12], we can get an update of $\boldsymbol{\Delta\tau}^{(k+1)}$ as

$$\boldsymbol{\Delta\tau}^{(k+1)} = \sum_{i=1}^{n} \boldsymbol{J}_i^{+} \left(\boldsymbol{A}^{(k+1)}\boldsymbol{C}^{(k+1)} + \boldsymbol{E}^{(k+1)} - \boldsymbol{M}_{o\tau} - \frac{\boldsymbol{Z}_1^{(k)}}{\mu_1^{(k)}} \right) \boldsymbol{v}_i \boldsymbol{v}_i^T \quad (18)$$

where $\boldsymbol{J}_i^{+}$ denotes the Moores-Penroses Pseudoinverse of $\boldsymbol{J}_i$ [65].

Following the same steps as the above, the Lagrangian multipliers $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are updated by

$$\boldsymbol{Z}_1^{(k+1)} = \boldsymbol{Z}_1^{(k)} + \mu_1^{(k+1)} \left\{ \boldsymbol{B}^{(k)} - \boldsymbol{A}^{(k+1)}\boldsymbol{C}^{(k+1)} - \boldsymbol{E}^{(k+1)} \right\} \quad (19)$$

and

$$\boldsymbol{Z}_2^{(k+1)} = \boldsymbol{Z}_2^{(k)} + \mu_2^{(k+1)} \left\{ \boldsymbol{C}^{(k+1)} - \boldsymbol{Q}^{(k+1)} \right\} \quad (20)$$

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

$$\mu_1^{(k+1)} = min\left\{\mu_{max}, \rho\mu_1^{(k)}\right\} \quad (21)$$

and

$$\mu_2^{(k+1)} = min\left\{\mu_{max}, \rho\mu_2^{(k)}\right\} \quad (22)$$

where ρ is a properly chosen constant and μ_{max} is a tunable parameter adjusting the convergence of the proposed method.

These updating equations proceed in a round-robin manner until convergence. For easy reference, the updating equations of the proposed algorithm is summarized in Algorithm 1.

Algorithm 1 ADMM for the Proposed Algorithm

Input: Data Matrix $\boldsymbol{M} \in \Re^{\boldsymbol{m}\times\boldsymbol{n}}$, $\boldsymbol{A^0} \in \Re^{m\times n}$, $\boldsymbol{E^0} \in \Re^{m\times n}$, $\boldsymbol{\Delta\tau^0} \in \Re^{p\times n}$, $\boldsymbol{C^0} \in \Re^{n\times n}$, $\boldsymbol{Q^0} \in \Re^{n\times n}$, λ_1, λ_2, ρ

While not converged **Do**

1: Update: $\boldsymbol{A^{(k\mathcal{C}1)}}$ by (6)
2: Update: $\boldsymbol{E^{(k\mathcal{C}1)}}$ by (9)
3: Update: $\boldsymbol{C^{(k\mathcal{C}1)}}$ by (12)
4: Update: $\boldsymbol{Q^{k\mathcal{C}1}}$ by (15)
5: Update: $\boldsymbol{\Delta\tau^{(k\mathcal{C}1)}}$ by (18)
6: Update: $\boldsymbol{Z}_1^{(k+1)}$ by (19)
7: Update: $\boldsymbol{Z}_2^{(k+1)}$ by (20)
8: Update: $\mu_1^{(k+1)}$ by (21)
9: Update: $\mu_2^{(k+1)}$ by (22)

End while

Outputs:$\boldsymbol{A}$,$\boldsymbol{E}$, $\boldsymbol{C}$, $\boldsymbol{Q}$, $\boldsymbol{\Delta\tau}$

V. CONVERGENCE ANALYSIS

This section considers the convergence characteristics of the updating equations addressed in the previous section. Two theorems are established to address this issue.

Theorem 1: If $\left\{\mu_1^{(k)}\right\}$ is non-decreasing and upper-bounded by $\tau_A > \sigma(\boldsymbol{C}\boldsymbol{C}^T)$, then the sequences $\left\{\boldsymbol{A}^{(k)}\right\}$, $\left\{\boldsymbol{E}^{(k)}\right\}$ and $\left\{\boldsymbol{Z}_1^{(k)}\right\}$ by ADMM converge to a Karush-Kuhn-Tucker (KKT) point of (6) and (9).

Theorem 2: If $\left\{\mu_1^{(k)}\right\}$ and $\left\{\mu_2^{(k)}\right\}$ are non-decreasing and upper-bounded by $\tau_{\boldsymbol{C}} > \sigma(\boldsymbol{A}\boldsymbol{A}^T)$, then the sequences $\left\{\boldsymbol{C}^{(k)}\right\}$,$\left\{\boldsymbol{Q}^{(k)}\right\}$, $\left\{\boldsymbol{\Delta\tau}^{(k)}\right\}$, $\left\{\boldsymbol{Z}_1{}^{(k)}\right\}$, $\left\{\boldsymbol{Z}_2{}^{(k)}\right\}$ generated by ADMM converge to a KKT point of (12), (15) and (18).

The proofs of both theorems can be readily extended from [54], [57], [62], [63].

Theorem 1 implies that the variables $\boldsymbol{A}^{(k)}$ and $\boldsymbol{E}^{(k)}$ are guaranteed to converge to the global optimum with an appropriate choice of the Lagrange multipliers $\left\{\boldsymbol{Z}_1^{(k)}\right\}$ and $\left\{\boldsymbol{Z}_2^{(k)}\right\}$, and a sufficiently large penalty parameters $\left\{\mu_1^{(k)}\right\}$. Meanwhile, Theorem 2 shows that the variables $\boldsymbol{C}^{(k)}$, $\boldsymbol{Q}^{(k)}$ and $\boldsymbol{\Delta\tau}^{(k)}$ are guaranteed to converge to the global optimum with an appropriate choice of the Lagrange multipliers $\left\{\boldsymbol{Z}_1^{(k)}\right\}$ and $\left\{\boldsymbol{Z}_2^{(k)}\right\}$, and sufficiently large penalty parameters $\left\{\mu_1^{(k)}\right\}$ and $\left\{\mu_2^{(k)}\right\}$.

TABLE 2. Comparison of the proposed approach with the L_1 or $L_{2,1}$ norm of E and Q, and with and without the affine transformations on three databases.

Method	Handwritten digits	Natural face images	Video face images
L_1 norm	154.0252	199.1704	188.0995
L_1 norm with affine transformations	167.6651	212.9273	199.8336
$L_{2,1}$ norm with affine transformations	204.1662	235.1174	203.5035

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we first assess the convergence behavior of the proposed algorithm. Afterward, we justify the effectiveness of the new algorithm for image recovery on some public databases. Our comparison is in terms of two widespread metrics: Peak Signal to Noise Ratio (PSNR) and Structural Similarly Index (SSIM). The PSNR is defined as

$$\text{PSNR}(f,\hat{f}) = 10log_{10}\frac{255^2}{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{i=1}^{n}(f_{ij}-\hat{f}_{ij})^2} \tag{23}$$

where f and $\hat{f}$ denote the original image and the recovered image, the size of both of which is $m \times n$. On the other hand, the SSIM is a perceptual image quality assessment measure and is defined as

$$\text{SSIM}(f,\hat{f}) = \frac{(2\mu_f\mu_{\hat{f}}+C_1)(2\sigma_{f,\hat{f}}+C_2)}{(\mu_f^2+\mu_{\hat{f}}^2+C_1)(\sigma_f^2+\sigma_{\hat{f}}^2+C_2)} \tag{24}$$

where $\sigma_{f,\hat{f}}$ is the correlation coefficient between f and $\hat{f}$, μ_f and σ_f^2 denote the mean and variance corresponding to f, respectively, and C_1 and C_2 are constants mostly taken as one to stabilize the algorithm. The three databases considered include the MINST database [66], the Wild database [67], and the Al Gore talking video [13].

A. EXPERIMENTAL CONVERGENCE PERFORMANCE

In this section, we provide some experimental analysis of the proposed algorithm based on three sets of images. The first set of image is the handwritten digits from the MINST database, the second one is the Natural Face images from the Wild database, and the third one is the video face images from the Al Gore talking video. As [49], the convergence characteristic is assessed by comparing the Relative Square Error (RSE) vs. the iteration number, where RSE is defined as

$$\frac{\left\|\boldsymbol{f}\hat{-}f\right\|_F}{\|f\|_F} \tag{25}$$

in which $\hat{\boldsymbol{f}}$ is the recovered image and $\boldsymbol{f}$ is the original one. The resulting convergence curves are as shown in Fig. 1 based on three sets of images, from which we can see that the RSEs of the proposed algorithm for all images decrease with the iteration number and then reach at a constant after a few iterations. This fact justifies the convergence of the proposed algorithm.

B. ABLATION STUDIES

First, we investigate the performance of the proposed approach using either the L_1 or the $L_{2,1}$ norms of $\boldsymbol{E}$ and $\boldsymbol{Q}$, and with and without the affine transformations on three databases, as shown in Table 2, from which we can see that the PSNR performance can be boosted with the incorporation of the affine transformations in the low-rank plus sparse decomposition. This is because the $L_{2,1}$ norm is considered, which is better than the L_1 norm in dealing with outliers. Also, the performance can be further improved if we consider the $L_{2,1}$ norm instead of the L_1 norm of $\boldsymbol{Q}$ and $\boldsymbol{E}$ in the objective function in (1). The performance improvement on the handwritten digits is more substantial. This is because the $L_{2,1}$ norm can remove the correlated samples across the images, so it is more resilient to outliers and large variations in the images, which are more pronounced in the handwritten digits.

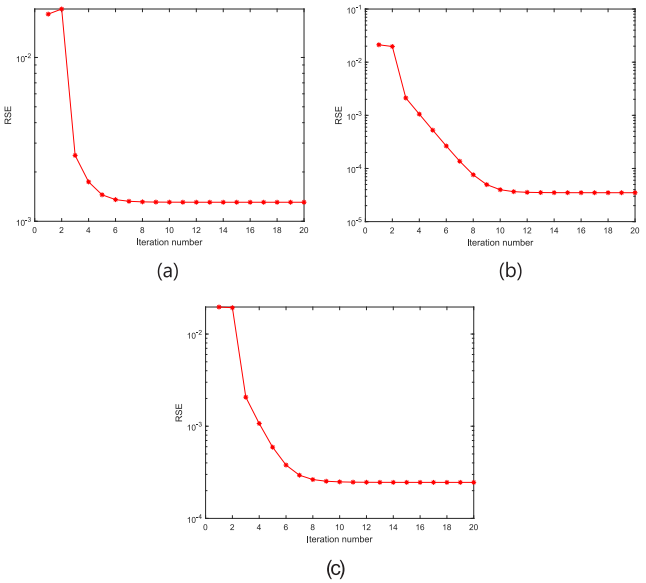

FIGURE 1. Convergence curves of the proposed algorithm on (a) Handwritten images; (b) Natural Face images;(c) Video Face images.

C. COMPARISON WITH THE STATE-OF-THE-ART METHODS

In this subsection, we compare the proposed approach, which adds in the affine transformations and uses the $L_{2,1}$ norm, with some recently reported works in terms of PSNR on the aforementioned three databases. Five different state-of-the-art methods, including RASL [13], IA-RPCA [17], PSVT [14], NQLSD [12], IGO-RIA [18] and the proposed algorithm are conducted for comparison. The results from these baselines are to re-implement of the publicly available codes.

1) HANDWRITTEN DIGITS

In this experiment, 30 handwritten digits of the size 29×29 are taken from the MINST database [66]. We compare the PSNR performance of the proposed method with the aforementioned five baselines, as shown in Table 3 and Table 4, from which we can see that NQLSD has better performance than IA-RPCA, as NQLSD employs the local linear approximation with a quadratic penalty approach to tackle the potential setback of outliers and sparse noise in the images. We can notice that PSVT yields better performance than RASL, NQLSD and IA-RPCA. This is because PSVT employs the truncated nuclear norm rather than the nuclear norm to better deal with outliers and sparse noise. IGO-RIA is superior to the other four baselines, as it considers an iterative linearization to minimize the L_1 norm of the sparse error and the optimization variables are adaptively updated based on the incremental thin singular value decomposition. We can also observe from Tables 3 and 4 that the proposed approach provides the best performance. This is because it incorporates a set of affine transformations and employs the $L_{2,1}$ norm to simultaneously align and recover the hand written digits, both of which enable the new algorithm to be more robust against outliers, heavy sparse noise, occlusions, and illuminations.

As an illustration, some visual images of the recovered handwritten digits based on the aforementioned methods are shown in Fig. 2, from which we can see that the proposed method is better aligned and recovers the handwritten images better as compared with the other five baselines. As shown in Fig. 2 (g), the recovered handwritten images provide clearer visual quality by properly removing the adverse effects such as outliers and heavy sparse noise. This is in agreement with the results in Tables 3 and 4, and further justifies that the proposed approach is more resilient to outliers and heavy sparse noise.

2) NATURAL FACE IMAGES

Next, we conduct simulations on more challenging images taken from the Labeled Natural Faces in the Wild database [67]. In this experiment, 7 Natural Face images with the size 80×60 are considered. We compare the proposed method with the aforementioned five baselines in terms of PSNR for image recovery. The comparison results are given in Tables 3 and 4, from which we can see that PSVT outperforms RASL, NQLSD and IA-RPCA, as it considers the partial sum of the singular values instead of the nuclear norm of the low-rank component to reduce the errors between the original distorted Natural Face images and the recovered one. IGO-RIA provides the second best performance, as it can cope with outliers and heavy sparse noise via the L_1 norm and the subspace learning into a unified online framework. We can also find that the proposed method provides the largest PSNR compared with the other baselines. This is because the new approach uses the affine transformations to fix the images in the low-rank plus sparse decomposition and employs the $L_{2,1}$ norm, so it is more robust to various annoying effects.

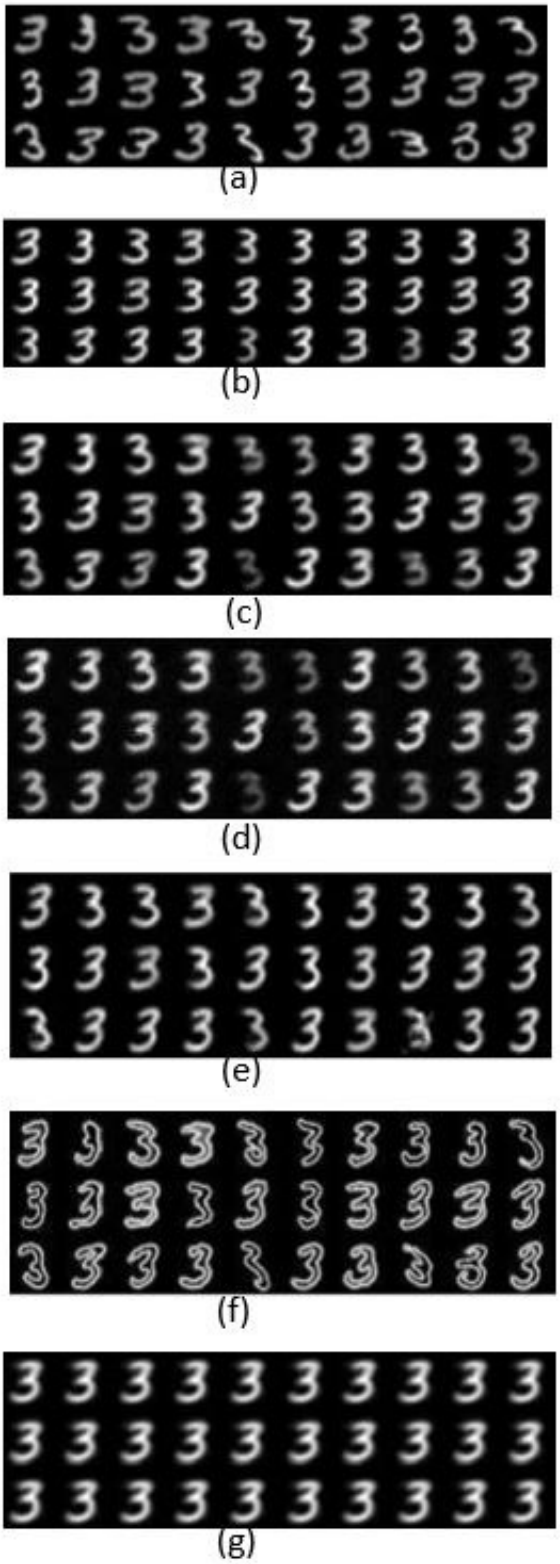

FIGURE 2. Some recovered handwritten digits: (a) Original; (b) RASL; (c) PSVT;(d) NLQSD; (e) IA-RPCA; (f) IGO-RIA; (g) Ours.

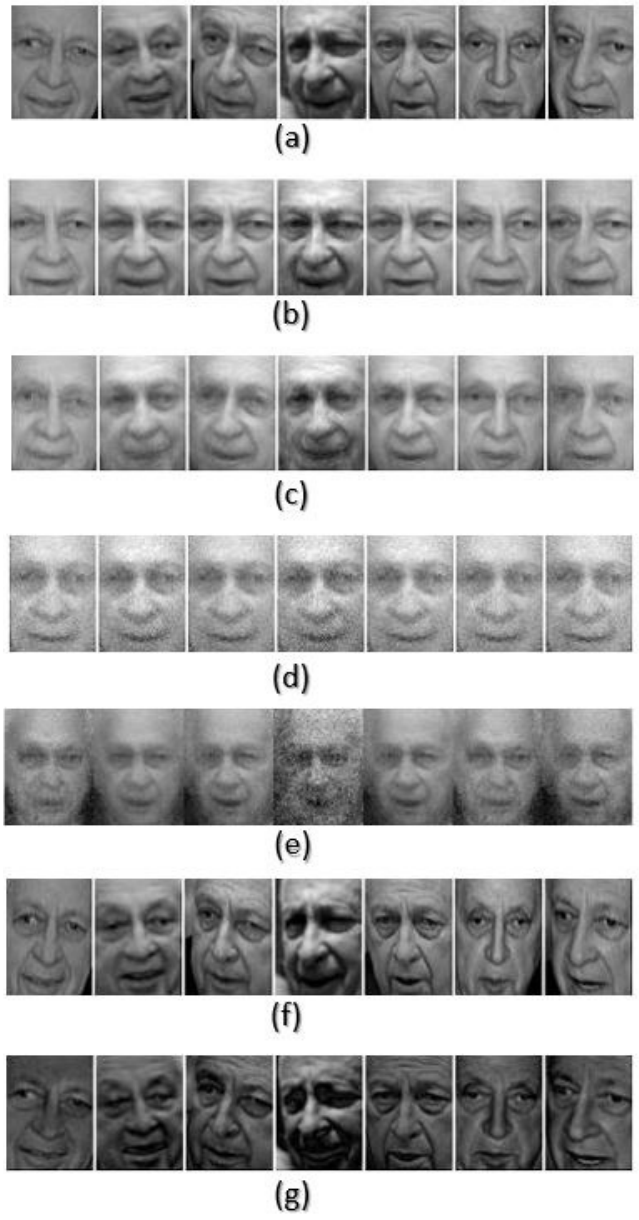

FIGURE 3. Some recovered Natural Face images:(a) Original; (b) RASL; (c) PSVT;(d) NLQSD; (e) IA-RPCA; (f) IGO-RIA (g) Ours.

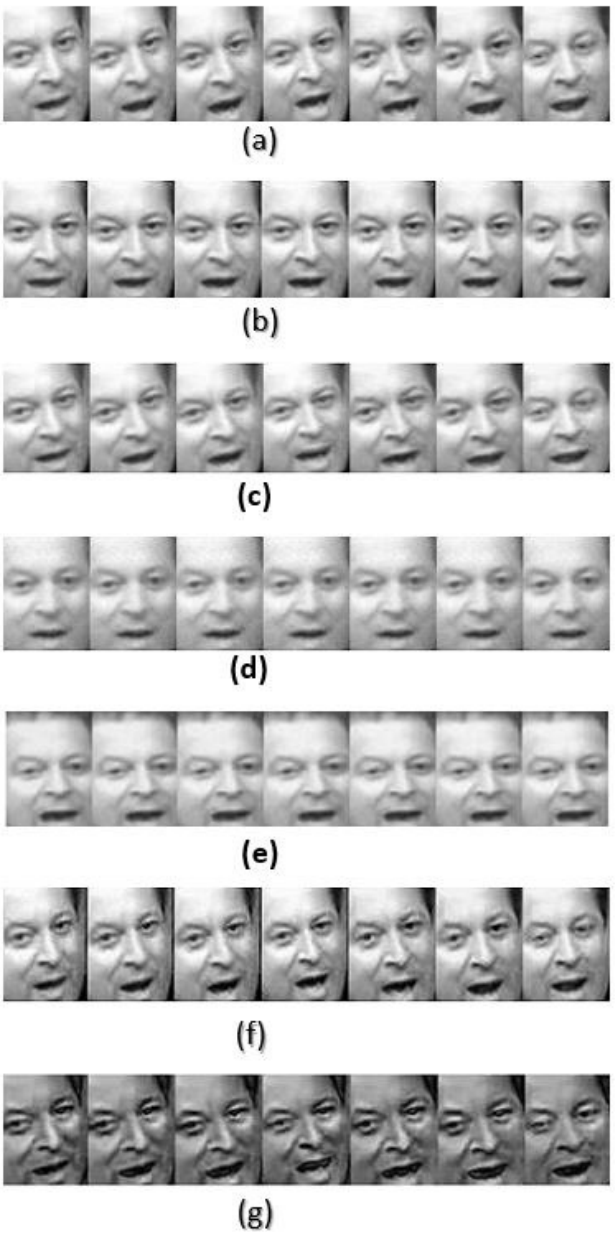

FIGURE 4. Some recovered face images from Al Gore talking:(a) Original; (b) RASL; (c) PSVT;(d) NLQSD; (e) IA-RPCA; (f) IGO-RIA (g) Ours.

Again, as an illustration, some recovered Natural Face images based on the proposed method and the aforementioned baselines are given in Fig. 3, where the Natural Face images with different corruptions are depicted in Fig. 3 (a). The recovered images by the aforementioned algorithms are shown in Fig. 3 (b)-(g), from which we can see that the visual quality of the proposed method is better than all of the baselines. This is consistent with the numerical results in Tables 3 and 4.

3) VIDEO FACE IMAGES

Finally, we conduct an experiment on more complicated face images in videos taken from the Al Gore talking [13]. In this video, 7 different video face images with the size 232×312 are considered. The comparison of PSNR using the proposed method and the other five baselines is given in Tables 3 and 4, from which we can see that IGO-RIA yields the second best performance. This is because IGO-RIA combines the subspace learning with iterative linearization to minimize the L_1-norm of the sparse error. We can also notice from Tables 3 and 4 that the proposed method still outperforms all baselines, as it includes affine transformations and utilizes the $L_{2,1}$ norm to render more robust subspace recovery.

As an illustration, some recovered face images based on the above algorithms are provided in Fig. 4, from which we can find that the proposed method again produces recovered

TABLE 3. Performance comparisons on different databases in terms of PSNR.

Method	Handwritten digits	Natural face images	Video face images
RASL [13]	195.7547	196.8332	196.1255
PSVT [14]	196.3255	196.9299	196.1443
NQLSD [12]	186.7870	196.4645	195.0887
IA-RPCA [17]	178.1612	192.2816	196.3044
IGO-RIA [18]	200.6199	202.3715	202..2241
Ours	204.1662	235.1174	203.5035

TABLE 4. Performance comparisons on different databases in terms of SSIM.

Method	Handwritten digits	Natural face images	Video face images
RASL [13]	0.7547	0.7529	0.7564
PSVT [14]	0.7586	0.7531	0.7533
NQLSD [12]	0.5980	0.7318	0.7490
IA-RPCA [17]	0.5605	0.7306	0.7820
IGO-RIA [18]	0.9626	0.9683	0.9617
Ours	0.9806	0.9862	0.9804

images with better visual quality as compared with the five baselines. These results are once again in agreement with the numerical results in Tables 3and 4.

VII. CONCLUSIONS

In this paper, we have developed a new algorithm for robust image recovery by combining the affine transformations with low-rank plus sparse image representation and utilizing the $L_{2,1}$ norm. The search of the affine transformations and the optimization variables is formulated as a constrained convex optimization problem. The ADMM approach is then employed and a new set of equations is established to iteratively update the optimization variables and the affine transformations. Moreover, the convergence of these new updating equations is experimentally verified as well. Conducted simulations show that the new algorithm outperforms the state-of-the-art methods in terms of accuracy on three public databases.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the associate editor and to the anonymous reviewers for carefully reviewing the manuscript, for many thoughtful comments, and for bringing more related references to their attention, which have enhanced the readability and quality of this manuscript.

REFERENCES

[1] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 156–171, Jan. 2017.

[2] H. T. Likassa and W.-H. Fang, "Robust regression for image alignment via subspace recovery techniques," in *Proc. ACM Int. Conf. Netw., Commun. Comput.*, 2018, pp. 288–293.

[3] G. Chen, X.-Y. Liu, L. Kong, J.-L. Lu, W. Shu, and M.-Y. Wu, "JSSDR: Joint-sparse sensory data recovery in wireless sensor networks," in *Proc. IEEE Int. Conf. Wireless Mobile Comput., Netw. Commun.*, Oct. 2013, pp. 367–374.

[4] X. Xiang and T. D. Tran, "Linear disentangled representation learning for facial actions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 12, pp. 3539–3544, Dec. 2018.

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

[6] M. K. Chung, H. Lee, P. T. Kim, and J. C. Ye, "Sparse topological data recovery in medical images," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Jun. 2011, pp. 1125–1129.

[7] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust PCA in image and video processing," *Proc. IEEE*, vol. 106, no. 8, pp. 1427–1457, Aug. 2018.

[8] S. Wang, Y. Wang, Y. Chen, P. Pan, Z. Sun, and G. He, "Robust PCA using matrix factorization for background/foreground separation," *IEEE Access*, vol. 6, pp. 18945–18953, 2018.

[9] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.

[10] F. De la Torre and M. J. Black, "Robust parameterized component analysis: Theory and applications to 2D facial appearance models," *Comput. Vis. Image Understand.*, vol. 91, nos. 1–2, pp. 53–71, 2003.

[11] S. E. Ebadi and E. Izquierdo, "Approximated RPCA for fast and efficient recovery of corrupted and linearly correlated images and video frames," in *Proc. IEEE Int. Conf. Syst., Signals Image Process.*, Sep. 2015, pp. 49–52.

[12] X. Chen, Z. Han, Y. Wang, Y. Tang, and H. Yu, "Nonconvex plus quadratic penalized low-rank and sparse decomposition for noisy image alignment," *Sci. China Inf. Sci.*, vol. 59, no. 5, May 2016, Art. no. 052107.

[13] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.

[14] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 4, pp. 744–758, Apr. 2016.

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

[16] J. He, D. Zhang, L. Balzano, and T. Tao, "Iterative Grassmannian optimization for robust image alignment," *Image Vis. Comput.*, vol. 32, no. 10, pp. 800–813, 2014.

[17] W. Song, J. Zhu, Y. Li, and C. Chen, "Image alignment by online robust PCA via stochastic gradient descent," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1241–1250, Jul. 2016.

[18] Q. Zheng, Y. Wang, and P. A. Heng, "Online subspace learning from gradient orientations for robust image alignment," *IEEE Trans. Image Process.*, vol. 28, no. 7, pp. 3383–3394, Jul. 2019.

[19] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM J. Optim.*, vol. 21, no. 2, pp. 57–81, 2011.

[20] R. Vidal, Y. Ma, and S. S. Sastry, "Robust principal component analysis?" in *Generalized Principal Component Analysis* (Interdisciplinary Applied Mathematics 40). Springer, 2016, pp. 63–122.
[21] Y. Li, C. Chen, F. Yang, and J. Huang, "Deep sparse representation for robust image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 4894–4901.
[22] A. Baghaie, R. M. D'Souza, and Z. Yu, "Sparse and low rank decomposition based batch image alignment for speckle reduction of retinal OCT images," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 226–230.
[23] Q. Zheng, Y. Wang, and P. A. Heng, "Online robust image alignment via subspace learning from gradient orientations," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1771–1780.
[24] S. Xiao, M. Tan, D. Xu, and Z. Y. Dong, "Robust kernel low-rank representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2268–2281, Nov. 2016.
[25] C. Lu, Z. Lin, and S. Yan, "Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 646–654, Feb. 2015.
[26] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[27] A. Vedaldi, G. Guidi, and S. Soatto, "Joint data alignment up to (lossy) transformations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[28] S.-H. Lai and M. Fang, "Robust and efficient image alignment with spatially varying illumination models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1999, pp. 167–172.
[29] D. Hsu, S. M. Kakade, and T. Zhang, "Robust matrix decomposition with sparse corruptions," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7221–7234, Nov. 2011.
[30] M. Rahmani and G. K. Atia, "Randomized subspace learning approach for high dimensional low rank plus sparse matrix decomposition," in *Proc. 49th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2015, pp. 1796–1800.
[31] A. Podosinnikova, S. Setzer, and M. Hein, "Robust PCA: Optimization of the robust reconstruction error over the stiefel manifold," in *Proc. German Conf. Pattern Recognit.*, 2014, pp. 121–131.
[32] N. Shahid, N. Perraudin, V. Kalofolias, G. Puy, and P. Vandergheynst, "Fast robust PCA on graphs," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 740–756, Jun. 2016.
[33] M. Shakeri and H. Zhang, "COROLA: A sequential solution to moving object detection using low-rank approximation," *Comput. Vis. Image Understand.*, vol. 146, pp. 27–39, May 2016.
[34] Z. Hu, F. Nie, L. Tian, R. Wang, and X. Li, "A comprehensive survey for low rank regularization," 2018, *arXiv:1808.04521*. [Online]. Available: https://arxiv.org/abs/1808.04521
[35] T. Zhang and G. Lerman, "A novel M-estimator for robust PCA," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 749–808, 2014.
[36] F. Zhang and J. Yang, "A linear subspace learning approach via low rank decomposition," in *Proc. IEEE Int. Conf. Innov. Bio-Inspired Comput. Appl.*, Dec. 2011, pp. 81–84.
[37] X. Zhao, G. An, Y. Cen, H. Wang, and R. Zhao, "Robust discriminant low-rank representation for subspace clustering," *Soft Comput.*, vol. 23, no. 16, pp. 7005–7013, 2019.
[38] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, pp. 413–458, 2008.
[39] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1832–1845, Oct. 2010.
[40] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2790–2797.
[41] 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 *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 2080–2088.
[42] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Departmental Papers (CIS)*, p. 107, 2000.
[43] Y.-X. Wang and H. Xu, "Noisy sparse subspace clustering," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 320–360, 2016.
[44] Q. Li, Z. Sun, Z. Lin, R. He, and T. Tan, "Transformation invariant subspace clustering," *Pattern Recognit.*, vol. 59, pp. 142–155, Nov. 2016.
[45] J. Shen, P. Li, and H. Xu, "Online low-rank subspace clustering by basis dictionary pursuit," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 622–631.
[46] Y. Wu, Z. Zhang, T. S. Huang, and J. Y. Lin, "Multibody grouping via orthogonal subspace decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2001, p. II.
[47] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proc. Conf. Artif. Intell.*, 2016, pp. 1181–1187.
[48] Y. Liu, Z. Long, and C. Zhu, "Image completion using low tensor tree rank and total variation minimization," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 338–350, Feb. 2019.
[49] Y. Liu, L. Chen, and C. Zhu, "Improved robust tensor principal component analysis via low-rank core matrix," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1378–1389, Dec. 2018.
[50] 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 Trans. Pattern Anal. Mach. Intell.*, to be published.
[51] F. Wen, L. Chu, P. Liu, and R. C. Qiu, "A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning," *IEEE Access*, vol. 6, pp. 69883–69906, 2018.
[52] F. Wen, L. Adhikari, L. Pei, R. F. Marcia, P. Liu, and R. C. Qiu, "Nonconvex regularization-based sparse recovery and demixing with application to color image inpainting," *IEEE Access*, vol. 5, pp. 11513–11527, 2017.
[53] P. Jain and P. Kar, "Non-convex optimization for machine learning," *Found. Trends Mach. Learn.*, vol. 10, nos. 3–4, pp. 143–336, 2017.
[54] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
[55] Y. Li, Y. Lin, X. Cheng, Z. Xiao, F. Shu, and G. Gui, "Nonconvex penalized regularization for robust sparse recovery in the presence of $S\alpha S$ noise," *IEEE Access*, vol. 6, pp. 25474–25485, 2018.
[56] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: https://arxiv.org/abs/1009.5055
[57] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
[58] C. Ding, D. Zhou, X. He, and H. Zha, "R_1-PCA: Rotational invariant L_1-norm principal component analysis for robust subspace factorization," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 281–288.
[59] R. Li, X. Wang, L. Lei, and Y. Song, "$L_2$1-norm based loss function and regularization extreme learning machine," *IEEE Access*, vol. 7, pp. 6575–6586, 2018.
[60] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[61] Y. Zhang, D. Shi, J. Gao, and D. Cheng, "Low-rank-sparse subspace representation for robust regression," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 6, Jul. 2017, pp. 2972–2981.
[62] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2012, pp. 2328–2335.
[63] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[64] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 569–592, 2009.
[65] P. Courrieu, "Fast computation of moore-penrose inverse matrices," 2008, *arXiv:0804.4809*. [Online]. Available: https://arxiv.org/abs/0804.4809
[66] Y. LeCun, C. Cortes, and C. J. C. Burges. (1998). *The MNIST Database of Handwritten Digits*. [Online]. Available: http://yann.lecun.com/exdb/mnist/
[67] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, 2007.

HABTE TADESSE LIKASSA received the B.Sc. degree from the University of Gondar, Ethiopia, in 2008, and the M.Sc. degree from Addis Ababa University, Ethiopia, in 2011. He is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan. He has been serving as a Graduate Assistant (2009), and also a Lecturer with Debre Birhan University, Ethiopia (2012–2013). He has also worked as a Lecturer and the Head of the Department of Statistics with Ambo University, Ethiopia, from 2014 to 2016. His research areas of interests include robust methods, image processing, statistical signal processing, and machine learning.

WEN-HSIEN FANG received the B.S. degree in electrical engineering from National Taiwan University, in 1983, and the M.S.E. and Ph.D. degrees from the University of Michigan, Ann Arbor, MI, USA, in 1988 and 1991, respectively, in electrical engineering and computer science. In fall 1991, he joined the faculty of the National Taiwan University of Science and Technology, where he currently holds a position as a Professor with the Department of Electronic and Computer Engineering. His research interests include various facets of signal processing applications, including statistical signal processing, multimedia signal processing, and machine learning.

JENQ-SHIOU LEU received the B.S. degree in mathematics and the M.S. degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, and the Ph.D. degree on a part-time basis in computer science from National Tsing Hua University, HsingChu, Taiwan. He was with Rising Star Technology, Taiwan, as an R&D Engineer, from 1995 to 1997, and worked in the telecommunication industry (Mobitai Communications and Taiwan Mobile), from 1997 to 2007, as the Assistant Manager. In February 2007, he joined the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology (NTUST), as an Assistant Professor. From February 2011 to January 2014, he was an Associate Professor. Since February 2014, he has been a Professor. His research interests include heterogeneous network integration, mobile service and platform design, cybersecurity, distributed computing, green and orange technology integration.