

Research Article

New Robust Part-Based Model with Affine Transformations for Facial Landmark Localization and Detection in Big Data

Chentao Zhang,^{1,2} Habte Tadesse Likassa ,³ Peidong Liang ,² and Jielong Guo⁴

¹Fujian (Quanzhou)-HIT Research Institute of Engineering and Technology, Quanzhou 362000, China

²Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen, China

³Department of Statistics, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia

⁴Fujian Institute of Research on the Structure of Matter Fuzhou, Chinese Academy of Sciences, Fuzhou, China

Correspondence should be addressed to Peidong Liang; 420374727@qq.com

Received 31 March 2021; Revised 7 October 2021; Accepted 9 October 2021; Published 3 November 2021

Academic Editor: Angelos Markopoulos

Copyright © 2021 Chentao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we developed a new robust part-based model for facial landmark localization and detection via affine transformation. In contrast to the existing works, the new algorithm incorporates affine transformations with the robust regression to tackle the potential effects of outliers and heavy sparse noises, occlusions and illuminations. As such, the distorted or misaligned objects can be rectified by affine transformations and the patterns of occlusions and outliers can be explicitly separated from the true underlying objects in big data. Moreover, the search of the optimal parameters and affine transformations is cast as a constrained optimization programming. To mitigate the computations, a new set of equations is derived to update the parameters involved and the affine transformations iteratively in a round-robin manner. Our way to update the parameters compared to the state of the art of the works is relatively better, as we employ a fast alternating direction method for multiplier (ADMM) algorithm that solves the parameters separately. Simulations show that the proposed method outperforms the state-of-the-art works on facial landmark localization and detection on the COFW, HELEN, and LFPW datasets.

1. Introduction

Localization of detailed facial features arises in a variety of applications including occlusion coherence for prediction in high-dimensional images [1], landmark localization [2–4], head pose estimation [5, 6], face image alignment [7], low-rank estimation [8–10], and object detection [11–14], etc. However, high-dimensional data are easily impacted by outliers and occlusions. It is thus of importance to develop a deep robust part-based models combining the robust regression algorithm with affine transformation for separating the landmark localization from the occlusions as in [15] and detection [16]. Thus, developing a new algorithm that is resilient with adverse effects while processing big data particularly for face detection and landmark localization is highly indispensable.

A number of algorithms have been suggested for robust facial landmark localization and object detection [17–24]. For instance, [18, 25, 26] addressed these problems using a regression approach. However, these approaches are mostly an approximate and, additionally, they are no longer pose-invariant. To improve the performance of the algorithm and overcome the aforementioned setback, deformable part models (DPMs) based on high-dimensional images [27–30] were taken into account. However, they can not deal with a huge amount of outliers and occlusions. Zhang et al. [31] proposed a tasks-constrained deep convolutional network (TCDCN) through multitask learning which directly conducted the simulations based on high-dimensional images. However, it ignores the potential annoying effects and the correlation between images which jeopardize the robustness of the algorithm. Plus, the issue of the

misalignment problem is also not tackled. Moreover, to tackle the misalignment problem, [8, 9, 32–34] addressed several algorithms via affine transformation and the $L_{2,1}$ norms. To circumvent this dilemma, Martinez et al. [35] considered a robust facial landmark detection (RFLD) based on high-dimensional image data via $L_{2,1}$ norm regularization against poor initializations. This approach, however, may lead to a suboptimal solution by performing a dimension reduction and a regression process. Zhang et al. [36] addressed joint face detection alignment using multitask cascaded convolutional networks and [37] feature selection with multiview data; however, affine transformation is not taken into consideration. While [38] suggested a fast and accurate facial landmark detection network through using an augmented (EMTCNN) instead of [36], yet the performance is not promising. To boost the performance of the algorithms, [38–40] proposed a novel deep network method for detecting facial region and landmarks to learn from low-dimensional image representations; however, the proposed techniques seek affine transformation to deal with occlusions and illuminations. To tackle this dilemma, [8, 32, 33] proposed new methods assisted via affine transformation in high-dimensional images to estimate the optimal parameters corresponding to the low rank recovery.

To deal with outliers and occlusions, Tzimiropoulos and Pantic [41] proposed an active appearance model via a fast simultaneous inverse composition (Fast-SIC) based on image datasets for landmark localization by considering the shape and appearance of images to enhance the face fitting accuracy. However, its localization outputs are distorted due to the presence of outliers and occlusions that is also why the performance of the method is not promising. To solve this problem, [1, 42] proposed hierarchical part models (HPMs) relying on high-dimensional images to explicitly model partial occlusions and detect facial points and occlusions simultaneously based on high-dimensional images using the landmark localization error. In addition to this, [43, 44] proposed a latent multiview self-representations for clustering via the tensor nuclear norm, while [44] addressed the robust sparse low-rank embedding for image dimension reduction. But, in real scenarios, the outliers and occlusion patterns can be very diverse and almost unpredictable. Nada et al. [45] considered a new annotated unconstrained face detector (UF-DD) to enhance the performance; however, it lacks robustness to large-scale variations and is required to perform an exhaustive search of the optimal transformation to obtain satisfactory performance.

2. Characteristics of the State-of-the-Art Methods

In this section, we illustrate the characteristics of the five state-of-the-art works such as TCDCN [31], RFLD [35], Fast-SIC [41], HPM [42], and UF-DD [45] with the proposed approach, which further adds affine transformations and multiple subspaces. The merits of the proposed method are quite different from the baselines in terms of its novelty and updating the parameters. For instance, TCDCN [31] addressed the idea of facial landmark location to deal with

occlusion and pose variation via deep multitask learning, but it yet requires to incorporate affine transformation to explore deep multitask learning in dense landmark detection and high-dimensional image recovery. This method considers a few regularization parameters with about six different labels that fail to penalize the complexity of the weights. To tackle the dilemmas of gross errors and variation in images, RFLD [35] proposed a novel regression method that substitutes the commonly used least squares regression developed by [31] via the use of the $L_{2,1}$ norm, but it lacks performance when there are a large number of landmarks. Moreover, there are seven numbers of parameters taken into account, which increases the computational complexity of the method. To alleviate the problems in facial deformable models for the unconstrained images, Fast-SIC was developed by [41], where six different parameters including the vector appearance, localization, and warping parameters involved in problem formulation, which makes the method difficult to optimize the parameters. To solve this setback, [42] proposed HPM for facial landmark localization and occlusion estimation to boost the performance, with five parameters and a number of network hierarchical structure of layers which is increasing the computational complexity of the algorithm. To tackle the setback of outliers and heavy sparse noises, occlusions and illuminations, [45] proposed a novel UF-DD method for face detection where five numbers of parameters are involved; however, the impacts of large outliers and heavy sparse noises are not alleviated. As a new annotated unconstrained face detector is addressed in [45], the time complexity is relatively better over RFLD [35], Fast-SIC [41], and HPM [42]. In the latest work, there are about seven number of parameters involved including affine transformation. As affine transformation is newly added in the new model, it shares the advantage of pruning out the potential impacts of occlusions and illuminations, outliers, and heavy sparse noises in landmark localization's and facial detection from which we note that the time complexity of the proposed method is highly reduced and the performance is enhanced. Another advantage of the new part-based model over the state-of-the-art works is it considers multiple subspace parameters to constrain each parameters and shares the advantages of positive semidefinite. However, the new technique seeks to establish the spatial dependency between different images that can be captured incorporating the spatial weight matrix, as such the performance of the new part-based model be boosted.

In this paper, we present a new robust regression algorithm for facial landmark localization and face detection. Invoked by the novel idea of affine transformations in [8], and to be more resilient to occlusions and outliers, the new algorithm incorporates affine transformations with the robust regression based on the hierarchical part-based models [1, 42]. As such, the distorted or misaligned images can be rectified by affine transformations and the patterns of occlusions and outliers can be explicitly separated. The proposed method via affine transformation contributes in clearly separating the facial landmark localization and occlusion estimation. The search of the optimal parameters and the affine transformations is cast as a constrained

optimization programming. To mitigate the computational overhead, a new set of equations are derived to update the parameters involved and the affine transformations iteratively using ADMM approach in a round-robin manner. Conducted simulations show that the proposed method outperforms the state-of-the-art works on face detection and landmark localization on some common on the COFW, HELEN, and LFPW datasets. The major contributions of this work include the following:

- (1) The affine transformations are incorporated in the robust regression based on the part-based models to take advantages of both schemes in the learning process
- (2) The affine transformations are aggregated with the low-rank-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
- (3) The ADMM method is employed to solve the new convex optimization problem, and a set of updating equations is derived to iteratively update the optimization variables and the affine transformations
- (4) New set of updating equations is established to iteratively solve the constrained optimization problem
- (5) We conduct experiments on several benchmark datasets, and the experimental results demonstrate the effectiveness of our new method

3. Problem Formulation

Given n images $\{\mathbf{I}_i^0\} \in \mathfrak{R}^{w \times h}$, $i = 1, \dots, n$, all of which contain the same object and are linearly correlated, where w and h denote the weight and height of the images, respectively. Based on the part-based models [1], we approximate $\mathbf{A} = [\text{vec}(\mathbf{I}_1^0) | \dots | \text{vec}(\mathbf{I}_n^0)] \in \mathfrak{R}^{l \times n}$ by a shape parameter, where $\text{vec}(\cdot)$ is the vector stacking operator [46] and l is the number of landmarks. However, in practice, the images $\{\mathbf{I}_i^0\}$ are usually contaminated by occlusions and outliers, so the images $\mathbf{M} \in \mathfrak{R}^{l \times n}$ can be represented as $\mathbf{M} = \mathbf{A}\mathbf{B} + \mathbf{E}$, where $\mathbf{M} \in \mathfrak{R}^{l \times n}$ is a corrupted high-dimensional images, $\mathbf{E} \in \mathfrak{R}^{l \times n}$ is a term accounting for the occlusions and outliers, and $\mathbf{B} \in \mathfrak{R}^{n \times n}$ is an appearance parameter.

To solve the misalignment problem incurred by the outliers and occlusions in \mathbf{M} , the original images need to be linearized, which can be done by affine transformations. Applying a set of affine transformations, denoted by $\boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_n$ to \mathbf{M} [33, 47], then, we then have \mathbf{M}_τ^o , where \mathbf{M}_τ^o denotes the transformed data. Assume that the differences between the consecutive affine transformations are small, so we can approximate \mathbf{M}_τ^o by $\mathbf{M}_\tau^o + \sum_{i=1}^n \mathbf{J}_i \Delta \boldsymbol{\tau}_i \mathbf{v}_i^T$, $\Delta \boldsymbol{\tau} \in \mathfrak{R}^{q \times n}$, in which q denotes the number of parameters, $\mathbf{J}_i =$

$\partial \text{vec}(\mathbf{I}_i \boldsymbol{\tau}_i) / \partial \boldsymbol{\tau}_i \in \mathfrak{R}^{l \times p}$ denotes the Jacobian of the i th training images with respect to $\boldsymbol{\tau}_i$, and \mathbf{v}_i denotes the standard basis for \mathfrak{R}^n . Thus, consequently, the main objective is to minimize the localization error via incorporating a novel idea of affine transformation for better performance, so the overall problem can thus be posted as the following constrained optimization problem

$$\min_{\mathbf{A}, \mathbf{L}, \hat{\boldsymbol{\beta}}, \mathbf{E}, \mathbf{B}, \Delta \boldsymbol{\tau}} \frac{\gamma}{2} \|\mathbf{U}\|_F^2 + \|\mathbf{A}\|_* + \|\mathbf{B}\|_* + \lambda_1 \|\mathbf{G}\|_1 + \lambda_2 \|\mathbf{E}\|_1, \quad (1)$$

$$\text{s.t. } \mathbf{Q} = \mathbf{A}\mathbf{B} + \mathbf{E}, \hat{\boldsymbol{\beta}} = [\mathbf{A}\mathbf{B}; \mathbf{1}^T], \mathbf{B} = \mathbf{G}, \mathbf{G} \pm \mathbf{0}, \quad (2)$$

where $\mathbf{U} = \mathbf{W}(\mathbf{R} - \mathbf{L}\hat{\boldsymbol{\beta}})$, $\mathbf{Q} = \mathbf{M}_\tau^o + \sum_{i=1}^n \mathbf{J}_i \Delta \boldsymbol{\tau}_i \mathbf{v}_i^T$, $\mathbf{W} \in \mathfrak{R}^{d_r \times d_r}$ is a weight matrix, $\mathbf{L} \in \mathfrak{R}^{d_r \times (l+1)}$ is the regression matrix mapping $\mathbf{M}_\tau^o \in \mathfrak{R}^{l \times n}$ to $\mathbf{R} \in \mathfrak{R}^{d_r \times n}$, $\hat{\boldsymbol{\beta}} \in \mathfrak{R}^{(l+1) \times n}$ is a regression coefficient that controls the appearance variation, $\mathbf{1}_n^T \in \mathfrak{R}^n$ denotes as a vector of all ones, $\mathbf{E} \in \mathfrak{R}^{l \times n}$, λ_1 , γ and λ_2 are the regularization parameters, $\|\mathbf{A}\|_* = \sum_{i=1}^{\min(l, n)} \sigma_i(\mathbf{A})$ with $\sigma_i(\mathbf{A})$ being the singular values of \mathbf{A} , $\|\mathbf{A}\|_F^2 = \text{Trace}(\mathbf{A}^T \mathbf{A})$, and $\|\mathbf{E}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^l |\mathbf{E}_{ij}|$.

3.1. Proposed Approach. To solve a constrained optimization problem (2), we consider an ADMM given by

$$\begin{aligned} \min_{\mathbf{L}, \hat{\boldsymbol{\beta}}, \mathbf{A}, \mathbf{E}, \mathbf{B}, \Delta \boldsymbol{\tau}} \frac{\gamma}{2} \|\mathbf{U}\|_F^2 + \|\mathbf{A}\|_* + \|\mathbf{B}\|_* + \lambda_1 \|\mathbf{G}\|_1 + \lambda_2 \|\mathbf{E}\|_1 \\ + \langle \mathbf{Y}_1, \mathbf{C} \rangle + \frac{\mu_1}{2} \|\mathbf{C}\|_F^2 + \langle \mathbf{Y}_2, \hat{\boldsymbol{\beta}} - [\mathbf{A}\mathbf{B}; \mathbf{1}^T] \rangle \\ + \frac{\mu_2}{2} \left\| \hat{\boldsymbol{\beta}} - [\mathbf{A}\mathbf{B}; \mathbf{1}^T] \right\|_F^2 + \langle \mathbf{Y}_3, \mathbf{B} - \mathbf{G} \rangle + \frac{\mu_3}{2} \|\mathbf{B} - \mathbf{G}\|_F^2, \end{aligned} \quad (3)$$

where $\mathbf{Y}_1 \in \mathfrak{R}^{l \times n}$, $\mathbf{Y}_2 \in \mathfrak{R}^{(l+1) \times n}$, and $\mathbf{Y}_3 \in \mathfrak{R}^{n \times n}$ are the Lagrangian multipliers, $\mathbf{C} = (\mathbf{Q} - \mathbf{A}\mathbf{B} - \mathbf{E})$, μ_1 , μ_2 , and μ_3 are the penalty parameters, and $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{Trace}(\mathbf{X}^T \mathbf{Y})$. This can be solved by using ADMM. The major necessary condition of ADMM is the convexity issue; we give sufficient conditions under which the algorithm asymptotically reaches the standard first-order necessary conditions for local optimality.

Then, based on a linearized alternating direction method [48], the augmented Lagrangian multiplier in (3) can be reexpressed as

$$\begin{aligned} \min_{\mathbf{L}, \hat{\boldsymbol{\beta}}, \mathbf{A}, \mathbf{E}, \mathbf{B}, \mathbf{G}, \Delta \boldsymbol{\tau}} \frac{\gamma}{2} \|\mathbf{U}\|_F^2 + \|\mathbf{A}\|_* + \|\mathbf{B}\|_* + \lambda_1 \|\mathbf{G}\|_1 + \lambda_2 \|\mathbf{E}\|_1 \\ + \frac{\mu_1}{2} \left\| \mathbf{C} + \frac{\mathbf{Y}_1}{\mu_1} \right\|_F^2 + \frac{\mu_2}{2} \left\| \mathbf{V} + \frac{\mathbf{Y}_2}{\mu_2} \right\|_F^2 + \frac{\mu_3}{2} \left\| \mathbf{B} - \mathbf{G} + \frac{\mathbf{Y}_3}{\mu_3} \right\|_F^2, \end{aligned} \quad (4)$$

where $\mathbf{V} = (\hat{\boldsymbol{\beta}} - [\mathbf{A}\mathbf{B}; \mathbf{1}^T])$. Directly solving (4) is computationally expensive, so to follow, we consider a set of equations to iteratively update the parameters in (4). So typically,

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

Methods	Objective function	Constraints
[51–55]	$\min_{\mathbf{A}, \mathbf{E}} \ \mathbf{A}\ _* + \lambda \ \mathbf{E}\ _1$	$\mathbf{M} = \mathbf{A} + \mathbf{E}$
[56–60]	$\min_{\mathbf{A}, \mathbf{E}} \ \mathbf{A}\ _* + \lambda \ \mathbf{E}\ _1$	$\mathbf{M}_{\text{or}} + \sum_{i=1}^n \mathbf{J}_i \Delta \tau \mathbf{v}_i \mathbf{v}_i^T = \mathbf{A} + \mathbf{E}$
[61–64]	$\min_{\mathbf{B}, \mathbf{E}, \mathbf{G}} \ \mathbf{B}\ _* + \lambda \ \mathbf{E}\ _{2,1}$	$\mathbf{M} = \mathbf{A}\mathbf{B} + \mathbf{E}, \mathbf{B} = \mathbf{G}$
Ours	$\min_{\mathbf{A}, \mathbf{L}, \hat{\boldsymbol{\beta}}, \mathbf{E}, \mathbf{B}, \Delta \tau} (\gamma/2)W \left\ \left(\mathbf{R} - \mathbf{L}\hat{\boldsymbol{\beta}} \right) \right\ _F^2 + \ \mathbf{A}\ _* + \ \mathbf{B}\ _* + \lambda_1 \ \mathbf{G}\ _1 + \lambda_2 \ \mathbf{E}\ _1$	$\mathbf{M}_{\text{r}}^0 + \sum_{i=1}^n \mathbf{J}_i \Delta \tau \mathbf{v}_i \mathbf{v}_i^T = \mathbf{A}\mathbf{B} + \mathbf{E}, \hat{\boldsymbol{\beta}} = [\mathbf{A}\mathbf{B}; \mathbf{1}^T], \mathbf{B} = \mathbf{G}, \mathbf{G} \geq 0$

we can choose to minimize the Lagrangian function to update all involved parameters involved alternatively.

Firstly, the updates of \mathbf{L} and $\hat{\boldsymbol{\beta}}$ are determined, respectively, by

$$\left. \begin{aligned} \mathbf{L}^{k+1} &= \underset{\mathbf{L}}{\operatorname{argmin}} L\left(\mathbf{L}^k, \boldsymbol{\beta}^{\wedge k}, \mathbf{A}^k, \mathbf{E}^k, \mathbf{B}^k, \mathbf{G}^k, \Delta \tau^k\right) \\ \boldsymbol{\beta}^{\wedge k+1} &= \underset{\hat{\boldsymbol{\beta}}}{\operatorname{argmin}} L\left(\mathbf{L}^{k+1}, \boldsymbol{\beta}^{\wedge k}, \mathbf{A}^k, \mathbf{E}^k, \mathbf{B}^k, \mathbf{G}^k, \Delta \tau^k\right) \end{aligned} \right\}, \quad (5)$$

where k is the index of iterations, and by ignoring all irrelevant terms and applying the ordinary least square regression procedures on \mathbf{L} and $\hat{\boldsymbol{\beta}}$, finally, the updates of \mathbf{L} and $\hat{\boldsymbol{\beta}}$ are given by

$$\left. \begin{aligned} \mathbf{L}^{k+1} &= \left(\boldsymbol{\beta}^{\wedge} \boldsymbol{\beta}^{\wedge T} + \gamma \left(\mathbf{I}_{(d_m+1)} \right) \right)^{-1} \mathbf{R} \left(\boldsymbol{\beta}^{\wedge} \right)^T \\ \boldsymbol{\beta}^{\wedge k+1} &= [\mathbf{N}\mathbf{L} + \mu_2 \mathbf{I}_d]^{-1} [\mathbf{N}\mathbf{R} - \mathbf{Z}] \end{aligned} \right\}, \quad (6)$$

where $\mathbf{N} = (\gamma \mathbf{L}^T \mathbf{W}^T \mathbf{W})$ and $\mathbf{Z} = (\mathbf{Y}_2 + \mu_2 [\mathbf{A}\mathbf{B}; \mathbf{1}^T])$ and \mathbf{I}_{d_m+1} is an identity matrix.

Secondly, to update the shape parameter \mathbf{A} , we fix \mathbf{E} , \mathbf{L} , $\hat{\boldsymbol{\beta}}$, \mathbf{G} , \mathbf{B} , and $\Delta \tau$ remain as constants, and \mathbf{A}^{k+1} can be determined by

$$\mathbf{A}^{k+1} = \underset{\mathbf{A}}{\operatorname{argmin}} L\left(\mathbf{L}^{k+1}, \boldsymbol{\beta}^{\wedge k+1}, \mathbf{A}^k, \mathbf{E}^k, \mathbf{B}^k, \mathbf{G}^k, \Delta \tau^k\right). \quad (7)$$

Using the soft shrinkage operator, we can get the update of \mathbf{A} as

$$\mathbf{A}^{k+1} = \Omega_{1/\kappa_A} \left(\mathbf{A}^k - \frac{P_{\mathbf{A}}^k}{\kappa_A} \right), \quad (8)$$

where Ω is the singular value threshold, $P_{\mathbf{A}}^k = (\mu_1 + \mu_2) \mathbf{A}^T \mathbf{B} - \mu_1 (\mathbf{Q} - \mathbf{E}) - (\mathbf{Y}_1 - \mu_2 (\hat{\boldsymbol{\beta}}) + \mathbf{Y}_2) \mathbf{B}^T$, $\kappa_A = ((\mu_1 + \mu_2) \boldsymbol{\eta}_A) / 2$, $\boldsymbol{\eta}_A > \sigma(\mathbf{B}^T \mathbf{B})$ is the proximal parameter, in which $\sigma(\mathbf{B}^T \mathbf{B})$ denotes the spectral radius of $\mathbf{B}^T \mathbf{B}$.

Thirdly, to update \mathbf{E} , we fix $\hat{\boldsymbol{\beta}}$, \mathbf{A} , \mathbf{L} , \mathbf{G} , \mathbf{B} , and $\Delta \tau$, and \mathbf{E}^{k+1} can be determined by

$$\mathbf{E}^{k+1} = \underset{\mathbf{E}}{\operatorname{argmin}} L\left(\mathbf{L}^{k+1}, \boldsymbol{\beta}^{\wedge k}, \mathbf{A}^k, \mathbf{E}^k, \mathbf{B}^k, \mathbf{G}^k, \Delta \tau^k\right). \quad (9)$$

By invoking a linearized alternating direction, \mathbf{E} is updated by

$$\mathbf{E}^{k+1} = \Gamma_{\lambda_1/\mu_1} \left(\mathbf{Q} - \mathbf{A}^k \mathbf{B} + \frac{\mathbf{Y}_1}{\mu_1} \right), \quad (10)$$

where Γ is the soft shrinkage operator [49, 50].

Next, to get the optimal update of \mathbf{B} , we again keep all \mathbf{A} , \mathbf{E} , \mathbf{L} , $\hat{\boldsymbol{\beta}}$, \mathbf{G} , and $\Delta \tau$ as constants, then the updating parameter \mathbf{B} can be determined by

$$\mathbf{B}^{k+1} = \underset{\mathbf{B}}{\operatorname{argmin}} L\left(\mathbf{L}^{k+1}, \boldsymbol{\beta}^{\wedge k+1}, \mathbf{A}^{k+1}, \mathbf{E}^{k+1}, \mathbf{B}^k, \mathbf{G}^k, \Delta \tau^k\right). \quad (11)$$

Then, by employing a singular value threshold operators, the update of \mathbf{B} is given by

$$\mathbf{B}^{k+1} = \Omega_{\frac{1}{\kappa_B}} \left(\mathbf{B}^k - \frac{P_{\mathbf{B}}^k}{\kappa_B} \right), \quad (12)$$

where $\kappa_B = (\mu_1 + \mu_2 + \mu_3) (\boldsymbol{\eta}_B / 2)$, $\boldsymbol{\eta}_B > \sigma(\mathbf{A}\mathbf{A}^T)$ and $P_{\mathbf{B}}^k = \mathbf{A}^k (\mu_1 + \mu_2) \mathbf{A}^{k+1} - \mu_1 (\mathbf{Q} - \mathbf{E}) - \mathbf{Y}_1 - (\mu_2 \hat{\boldsymbol{\beta}} + \mathbf{Y}_2) + \mu_3 (\mathbf{B}^k - \mathbf{G}^k) + \mathbf{Y}_3$.

To find the optimal update of \mathbf{G} , we fix \mathbf{A} , \mathbf{E} , \mathbf{L} , $\hat{\boldsymbol{\beta}}$, \mathbf{B} , and $\Delta \tau$ as constants, then \mathbf{G}^{k+1} can be obtained by

$$\mathbf{G}^{k+1} = \underset{\mathbf{G}}{\operatorname{argmin}} L\left(\mathbf{L}^{k+1}, \boldsymbol{\beta}^{\wedge k+1}, \mathbf{A}^{k+1}, \mathbf{E}^{k+1}, \mathbf{B}^{k+1}, \mathbf{G}^k, \Delta \tau^k\right). \quad (13)$$

Similarly, employing the soft threshold operator and an augmented Lagrangian multiplier to subproblem (14), then finally, \mathbf{G} is updated by

$$\mathbf{G}^{k+1} = \Gamma_{\lambda_2/\mu_3} \left(\mathbf{B}^{k+1} + \frac{\mathbf{Y}_3}{\mu_3} \right). \quad (14)$$

Invoked with an affine transformations, we obtained an

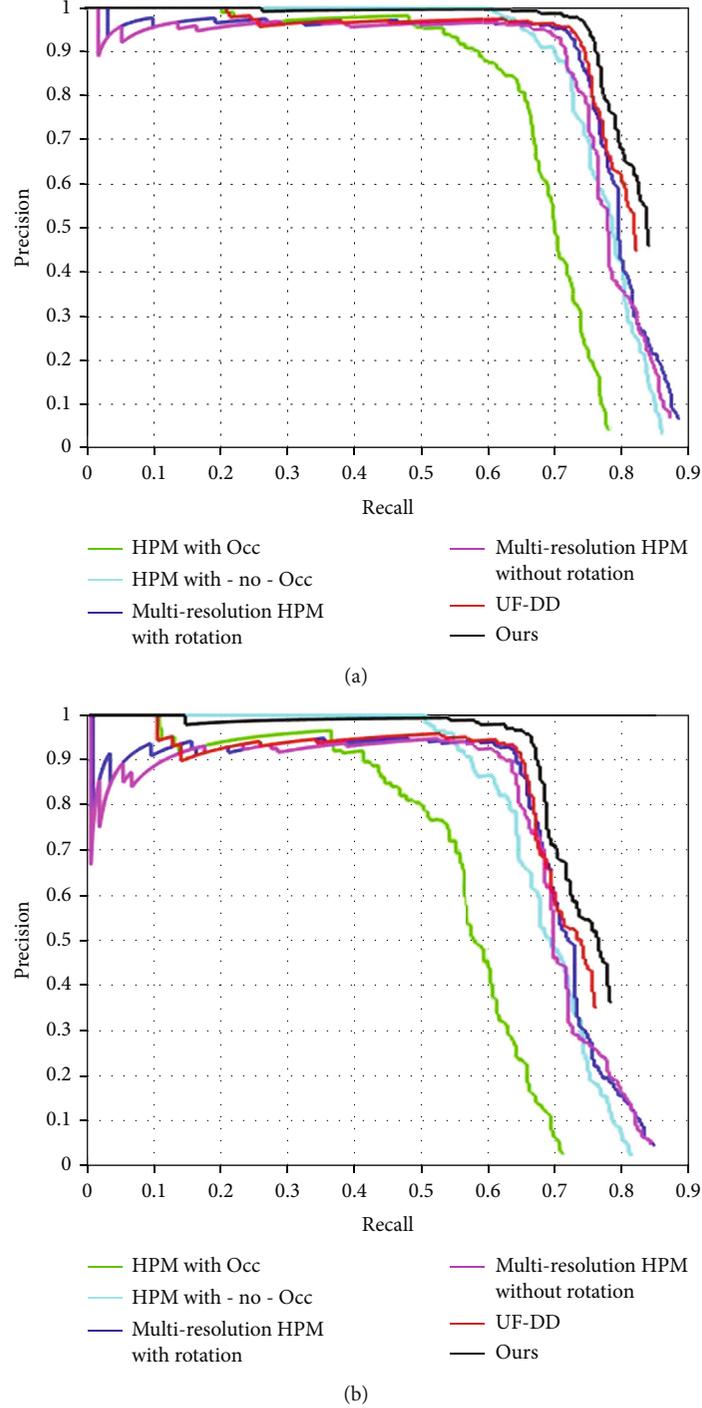


FIGURE 1: Precision-recall curves of the proposed method and baselines methods on the (a) COFW dataset and (b) LFPW dataset.

additional optimal parameter. Therefore, we need to derive an additional update for $\Delta\tau$.

To do so, we update $\Delta\tau$, we fix \mathbf{A} , \mathbf{E} , \mathbf{L} , $\hat{\boldsymbol{\beta}}$, \mathbf{B} , and \mathbf{G} , and $\Delta\tau^{k+1}$ can be obtained by

$$\Delta\tau^{k+1} = \underset{\Delta\tau}{\operatorname{argmin}} L(\mathbf{L}^{k+1}, \hat{\boldsymbol{\beta}}^{k+1}, \mathbf{A}^{k+1}, \mathbf{E}^{k+1}, \mathbf{B}^{k+1}, \mathbf{G}^{k+1}, \Delta\tau^k). \quad (15)$$

Along the same line, the affine transformations $\Delta\tau$ can be updated by

$$\Delta\tau^{k+1} = \sum_{i=1}^n \chi_i^+ \left(\mathbf{A}^{k+1} \mathbf{B}^{k+1} + \mathbf{E}^{k+1} - \mathbf{M}_\tau^o - \frac{\mathbf{Y}_1}{\mu_1} \right) \mathbf{v}_i \mathbf{v}_i^T, \quad (16)$$

where χ_i^+ is a More-Penrose-pseudoinverse of χ_i .

Similarly, the updates of \mathbf{Y}_1 , \mathbf{Y}_2 , and \mathbf{Y}_3 are given, respectively, by

$$\left. \begin{aligned} \mathbf{Y}_1^{k+1} &= \mathbf{Y}_1^k + \mu_1^{k+1} \left(\mathbf{Q} - \mathbf{A}^{k+1} \mathbf{B}^{k+1} - \mathbf{E}^{k+1} \right) \\ \mathbf{Y}_2^{k+1} &= \mathbf{Y}_2^k + \mu_2^{k+1} \left(\hat{\boldsymbol{\beta}} - \left[\mathbf{A}^{k+1} \mathbf{B}^{k+1}; \mathbf{I}^{k+1} \right] \right) \\ \mathbf{Y}_3^{k+1} &= \mathbf{Y}_3^k + \mu_3^{k+1} \left(\mathbf{B}^{k+1} - \mathbf{G}^{k+1} \right) \end{aligned} \right\}. \quad (17)$$

The updates of the regularization parameters μ_1 , μ_2 , and μ_3 are given by

$$\left. \begin{aligned} \mu_1^{k+1} &= \min \left(\mu_{\max}, \rho \mu_1^k \right) \\ \mu_2^{k+1} &= \min \left(\mu_{\max}, \rho \mu_2^k \right) \\ \text{and } \mu_3^{k+1} &= \min \left(\mu_{\max}, \rho \mu_3^k \right) \end{aligned} \right\}, \quad (18)$$

where ρ is a properly chosen constant and μ_{\max} are tunable parameters adjusting the convergence of the proposed algorithm. We sequentially update \mathbf{L} , $\hat{\boldsymbol{\beta}}$, \mathbf{A} , \mathbf{E} , \mathbf{B} , \mathbf{G} , and $\Delta\boldsymbol{\tau}$ independently by keeping all of the other parameters unchanged. First, the regression matrix \mathbf{L} and $\hat{\boldsymbol{\beta}}$ are updated by (6). Next, we can update \mathbf{A} , \mathbf{E} , \mathbf{B} , \mathbf{G} , and the optimal affine transformation parameters, $\Delta\boldsymbol{\tau}$, by (8), (10), (12), (14), and (16), respectively. Finally, the Lagrangian multipliers \mathbf{Y}_1 , \mathbf{Y}_2 , and \mathbf{Y}_3 are updated by (17) and the regularization parameters μ_1 , μ_2 , and μ_3 are updated by (18). These new updating equations proceed in a round-robin manner until convergence. Since a monotonically decreasing and bounded below sequence will converge, the above algorithm is ensured to converge. It is noteworthy that based on the augmented Lagrangian multiplier, ordinary least squares procedures, and soft-threshold operators, all the updating parameters are achieved. The overall summary of the proposed method compared with other related works is summarized in Table 1.

4. Experimental Results

In this section, some simulations are conducted to assess the effectiveness of the proposed algorithm. Three datasets are considered in the simulations, including the Labeled Face Parts in the Wild (LFPW) [65], the HELEN68 [21], and the more challenging Caltech Occluded Faces in the Wild (COFW) [66] datasets. First, we consider the novel ideas of affine transformation that tackle the adverse effects of outliers, heavy sparse noises, occlusions, and illuminations. After an intensive mathematical derivation, we experimentally supported the effectiveness of the proposed method through some numerical simulations.

4.1. Comparison with the State-of-the-Art Methods. In this subsection, we compare the proposed approach, which adds affine transformations with some recently reported works in terms of precision-recall curves, landmark localization error, and time complexity on the aforementioned three databases.

TABLE 2: Comparison of localization error for facial landmark localization.

Methods	HELEN68	COFW
TCDCN [31]	0.0910	0.0892
RFLD [35]	0.0799	0.0687
Fast-SIC [41]	0.0422	0.0382
HPM [42]	0.0138	0.0178
UF-DD [45]	0.0122	0.0142
Ours	0.0067	0.0063

In this work, five different state-of-the-art methods, including TCDCN [31], RFLD [35], Fast-SIC [41], HPM [42], UF-DD [45], and the proposed algorithm are conducted based on the performance evaluation and time complexity. The results from these baselines are to reimplement of the publicly available codes.

4.2. Face Detection. First, we assess the proposed approach for face detection based on the LFPW and COFW databases in terms of the precision-recall curve. The comparisons of the proposed approach with UF-DD [45] and HPM [42] are shown in Figure 1, from which we can see that UF-DD outperforms HPM, as it downsampled the images with high resolution using an unconstrained face detector to tackle the impact of large variations. As noted in Figure 1, the HPM without occlusion is relatively better compared to the HPM with occlusion, Multiresolution HPM with rotation outperforming to the HPM with occlusion and multiresolution HPM without rotation is relatively outperforming to all other versions of the HPM in face detecting. We can also notice from Figure 1 that the proposed algorithm is superior than all HPM versions and UF-DD by achieving better recall rate on both datasets. The superiority of the new approach is due to an incorporation of an affine transformation and the multiple subspaces and constraining the parameters taking the advantages of positive semidefinite, which makes it to be more robust with a various adverse effects of outliers, illuminations, occlusions, and heavy sparse noises. This is because the new algorithm combines affine transformations with the part-based model so that it is more robust against the impact of occlusions and outliers. To justify the effectiveness of the proposed method, comparison of the localization error is taken into account, from which we note that the proposed method is outperforming to five different baselines. The performance of the proposed method is superior to all the baselines. This is due to affine transformation; first, it aligns the misaligned images and it prunes out the potential impact of annoying effects which boosts the performance of the proposed method.

4.3. Facial Landmark Localization. In this subsection, we compare the proposed algorithm with some state-of-the-art for the facial landmark localization on the HELEN68 and COFW datasets. The comparison with five baselines, TCDCN [31], RFLD [35], Fast-SIC [41], HPM [42], and UF-DD [45] in terms of the root mean square error is as shown in Table 2, from which we can see that TCDCN has



FIGURE 2: Performance evaluation of the proposed method via facial landmark localization on the HELEN68 dataset: (a) TDCDN, (b) RFLD, (c) Fast-SIC, (d) HPM, (e) UF-DD, and (f) ours. The red points indicate the key points which are predicted as being occluded.

the largest error compared with the other baselines, as it ignores the multicollinearity during multitasks to localize the objects. Also, RFLD [35] and Fast-SIC [41] can not provide satisfactory landmark localization accuracy, as their localization outputs are distorted and the patterns of occlusions can not be explicitly separated. We can also notice that HPM outperforms [31, 35, 41], as it utilizes a hierarchical-based structures to explicitly model the occlusions to attain accurate landmark localization. UF-DD outperforms [31, 35, 41], since UF-DD further downsampled the images with high resolution using a face detector. We can also find that the proposed method outperforms all baselines on both datasets. Thus, the results of Table 2 justify the effectiveness of the proposed approach based on the localization error for the facial landmark localization as compared with the main-state-of-the art works. This is because it aggregates robust regression using affine transformations with part-based models to accurately localize facial landmarks even under outliers and occlusions.

As an illustration, we also show three images for facial landmark localization based on the aforementioned algorithms as shown in Figure 2, from which we can see that TCDCN, RFLD, and Fast-SIC provide poor localization. This is because the approaches fail to clearly separate the landmark localization from the occluded ones as TCDCN [31] considers combined multitask learning and ignores multicollinearity while RFLD [35] considers a large number of concatenating histogram of oriented gradient (HOG) descriptors with a multiple initializations which influences the separation, which still makes the separation challenging, when the images are distorted by the presence of occlusions as shown in Figures 2(a)–2(c). Moreover, it is also observed that the Fast-SIC [41] is relatively better compared to TCDCN [31] and RFLD [35] in separating landmark localization from the annoying effects such as occlusion, illuminations, outliers, and heavy sparse noises, at it considers the project out optimization framework. UF-DD [45] which incorporates the new annotated unconstrained face detection and downsampled strategy the images with high resolution provides better localization than HPM [31, 35, 41, 42], as it better separates the landmark localization (blue) and occluded landmarks (red) explicitly as depicted in Figure 2(e). The proposed approach works better under the impact of outliers and occlusion and explicitly separates the landmark localization’s from the occlusions as shown in Figure 2(f). This once again justifies that effectiveness of the combination of the affine transformations and part-based models, which makes the proposed method to be more robust with outliers and images which are highly influenced by occlusions.

In summary, the proposed method is relatively better in landmark localization and facial detection; this is due to affine transformation considered detects the potential impact of outliers, heavy sparse noises, occlusions and illuminations. However, the major drawback of the proposed method is it needs to account the spatial dependency between images through incorporating a spatial weight matrix in the mathematical formulation.

TABLE 3: Comparison of time complexity in minutes.

Methods	Time complexity
TCDCN [31]	4.34
RFLD [35]	4.56
Fast-SIC [41]	4.78
HPM [42]	5.67
UF-DD [45]	4.20
Ours	3.56

5. Computational Complexity

The time complexity of the proposed method as compared to the state-of-the-art works is described in this section. On a very standard desktop computer, the computational load of the baselines mainly TCDCN [31], RFLD [35], Fast-SIC [41], HPM [42], and UF-DD [45] along with the proposed method is given in Table 3, from which we note that the proposed method has small number of running time due to in the proposed method there are less number of parameters involved in updating the parameters. Additionally, our algorithm can handle batches of over one hundred images in a few minutes on a standard PC as the number of parameters involved is small as compared to the state-of-the-art works. The new algorithm is guaranteed faster time complexity compared to the state-of-the-art algorithms as shown in Table 3. This is due to affine transformation removing the extreme values, and various regularization parameters are also involved to make the situation of the proposed model stable and gross errors including outliers, occlusions, and illuminations.

6. Conclusions

In this paper, we developed a new model for facial landmark localization and detection, comprising an affine transformation and ADMM methods. The new algorithm combines the efficacious affine transformation-assisted robust regression with the part-based models to enjoy the advantages of both schemes. The problem is formulated as a constrained optimization programming, and a set of equations are established to iteratively update the parameters and the affine transformations. 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. Simulations justify the effectiveness of the new algorithm on some common on the COFW, HELEN, and LFPW datasets compared to the state-of-the-art works. To minimize the time complexity, the ADMM approach is further considered in a round-robin manner. In this work, the affine transformation taken into account is to correct the alignments of the individual images, but not consider the issue of spatial dependency between images. This can be captured as one of our future work through incorporating the spatial weight matrix between images. Another future work is to excel this new part-based model in high-dimensional tensor datasets.

Data Availability

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

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 Scientific and Technological Program of Quanzhou City under Grant no. 2019CT009. We thank Addis Ababa University, Ethiopia.

References

- [1] G. Ghiasi and C. C. Fowlkes, "Occlusion coherence: localizing occluded faces with a hierarchical deformable part model," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2385–2392, Columbus, OH, USA, June 2014.
- [2] Y. Wu and Q. Ji, "Facial landmark detection: a literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019.
- [3] H. Zhang, Q. Li, Z. Sun, and Y. Liu, "Combining data-driven and model-driven methods for robust facial landmark detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2409–2422, 2018.
- [4] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121–135, 2019.
- [5] X. Xu and I. A. Kakadiaris, "Joint head pose estimation and face alignment framework using global and local CNN features," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 642–649, Washington, DC, USA, May 2017.
- [6] Y. Cai, L. Ge, J. Cai, N. M. Thalmann, and J. Yuan, "3d hand pose estimation using synthetic data and weakly labeled RGB images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 3739–3753, 2020.
- [7] J. Deng, G. Trigeorgis, Y. Zhou, and S. Zafeiriou, "Joint multi-view face alignment in the wild," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3636–3648, 2019.
- [8] 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, Article ID 8136384, 9 pages, 2020.
- [9] H. T. Likassa and W.-H. Fang, "Robust regression for image alignment via subspace recovery techniques," in *Proceedings of the 2018 VII International Conference on Network, Communication and Computing - ICNCC 2018*, pp. 288–293, Taipei City Taiwan, 2018.
- [10] W. Yu, J. Lin, Q. Cai et al., "A low rank promoting prior for unsupervised contrastive learning," 2021, <https://arxiv.org/abs/2108.02696>.
- [11] P. Liang, H. T. Likassa, C. Zhang, and J. Guo, "New robust PCA for outliers and heavy sparse noises' detection via affine transformation, the and norms, and spatial weight matrix in high-dimensional images: from the perspective of signal processing," *International Journal of Mathematics and Mathematical Sciences*, vol. 2021, Article ID 3047712, 9 pages, 2021.
- [12] Q. Yin, T. Liu, Z. Lin, W. An, and Y. Guo, "Moving object detection in satellite videos via spatial-temporal tensor model and weighted Schatten p-norm minimization," *IEEE Geoscience and Remote Sensing Letters*, p. 1, 2021.
- [13] Y. Gao, Z. Kuang, G. Li, W. Zhang, and L. Lin, "Hierarchical reasoning network for human-object interaction detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 8306–8317, 2021.
- [14] Y. Shi, Y. Li, X. Wei, and Y. Zhou, "A faster-RCNN based chemical fiber paper tube defect detection method," in *2017 5th International Conference on Enterprise Systems (ES)*, pp. 173–177, Beijing, China, September 2017.
- [15] H. Hu, C. Wang, T. Jiang, Z. Guo, and Y. Han, "Robust and efficient facial landmark localization," in *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–7, Shenzhen, China, July 2021.
- [16] J. Wan, Z. Lai, J. Li, J. Zhou, and C. Gao, "Robust facial landmark detection by multiorder multiconstraint deep networks," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2021.
- [17] H. K. Galoogahi and T. Sim, "Correlation filter cascade for facial landmark localization," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–8, Lake Placid, NY, USA, March 2016.
- [18] X. Yu, J. Huang, S. Zhang, W. Yan, and D. N. Metaxas, "Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model," in *2013 IEEE International Conference on Computer Vision*, pp. 1944–1951, Sydney, NSW, Australia, December 2013.
- [19] Z. Shao, S. Ding, H. Zhu, C. Wang, and L. Ma, "Face alignment by deep convolutional network with adaptive learning rate," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1283–1287, Shanghai, China, March 2016.
- [20] W. Chen, Q. Zhou, and R. Hu, "Face alignment by combining residual features in cascaded hourglass network," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 196–200, Athens, Greece, October 2018.
- [21] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," *Proceedings of the European Conference on Computer Vision*, , pp. 679–692, Springer, 2012.
- [22] B. V. Frade and E. R. Nascimento, "A two-step learning method for detecting landmarks on faces from different domains," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2655–2659, Athens, Greece, October 2018.
- [23] S. Honari, P. Molchanov, S. Tyree, P. Vincent, C. Pal, and J. Kautz, "Improving landmark localization with semi-supervised learning," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, July 2018.
- [24] G.-S. Hsu, K.-H. Chang, S.-C. Huang, and S.-L. Chung, "Face detection and landmark localization using bilayer tree structured model," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4303–4307, Quebec City, QC, Canada, September 2015.
- [25] J. Lv, X. Shao, J. Xing, C. Cheng, and X. Zhou, "A deep regression architecture with two-stage reinitialization for high

- performance facial landmark detection,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3691–3700, Honolulu, HI, USA, July 2017.
- [26] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, and X. Wu, “Face detection, bounding box aggregation and pose estimation for robust facial landmark localisation in the wild,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2106–2115, Honolulu, HI, USA, July 2017.
- [27] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, “Robust discriminative response map fitting with constrained local models,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3444–3451, Portland, OR, USA, June 2013.
- [28] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.
- [29] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886, Providence, RI, USA, June 2012.
- [30] G. Tzimiropoulos and M. Pantic, “Gauss-newton deformable part models for face alignment in-the-wild,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1851–1858, Columbus, OH, USA, June 2014.
- [31] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, “Facial landmark detection by deep multi-task learning,” *Proceedings of the European Conference on Computer Vision*, pp. 94–108, Springer, 2014.
- [32] 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 *2018 3rd International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, pp. 1–4, Yilan, Taiwan, April 2018.
- [33] 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.
- [34] 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, Article ID 1286909, 10 pages, 2020.
- [35] B. Martinez and M. F. Valstar, “ $L_{2,1}$ -based regression and prediction accumulation across views for robust facial landmark detection,” *Image and Vision Computing*, vol. 47, pp. 36–44, 2016.
- [36] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [37] R. Zhang, F. Nie, X. Li, and X. Wei, “Feature selection with multi-view data: a survey,” *Information Fusion*, vol. 50, pp. 158–167, 2019.
- [38] H.-W. Kim, H.-J. Kim, S. Rho, and E. Hwang, “Augmented EMTCNN: a fast and accurate facial landmark detection network,” *Applied Sciences*, vol. 10, no. 7, p. 2253, 2020.
- [39] X. Wei, H. Shen, and M. Kleinsteuber, “Trace quotient meets sparsity: a method for learning low dimensional image representations,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5268–5277, Las Vegas, NV, USA, June 2016.
- [40] T. Kim, J. Mok, and E. Lee, “Detecting facial region and landmarks at once via deep network,” *Sensors*, vol. 21, no. 16, p. 5360, 2021.
- [41] G. Tzimiropoulos and M. Pantic, “Fast algorithms for fitting active appearance models to unconstrained images,” *International Journal of Computer Vision*, vol. 122, no. 1, pp. 17–33, 2017.
- [42] K. Yuen and M. M. Trivedi, “An occluded stacked hourglass approach to facial landmark localization and occlusion estimation,” *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 4, pp. 321–331, 2017.
- [43] G.-F. Lu and J. Zhao, “Latent multi-view selfrepresentations for clustering via the tensor nuclear norm,” *Applied Intelligence*, pp. 1–13, 2021.
- [44] Z. Liu, Y. Lu, Z. Lai, W. Ou, and K. Zhang, “Robust sparse low-rank embedding for image dimension reduction,” *Applied Soft Computing*, vol. 113, p. 107907, 2021.
- [45] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, “Pushing the limits of unconstrained face detection: a challenge dataset and baseline results,” 2018, <https://arxiv.org/abs/1804.10275>.
- [46] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3 edition, 2012.
- [47] Yigang Peng, A. Ganesh, J. Wright, Wenli Xu, and 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.
- [48] Z. Lin, R. Liu, and S. Zhixun, “Linearized alternating direction method with adaptive penalty for low-rank representation,” *Advances in Neural Information Processing Systems*, pp. 612–620, 2011.
- [49] 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.
- [50] 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 *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2328–2335, Providence, RI, USA, June 2012.
- [51] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, vol. 5, Springer, 2016.
- [52] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, “On the applications of robust PCA in image and video processing,” *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1427–1457, 2018.
- [53] M. Iliadis, H. Wang, R. Molina, and A. K. Katsaggelos, “Robust and low-rank representation for fast face identification with occlusions,” *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2203–2218, 2017.
- [54] Y. Lu, C. Yuan, X. Li, Z. Lai, D. Zhang, and L. Shen, “Structurally incoherent low-rank 2DLPP for image classification,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 6, pp. 1701–1714, 2018.
- [55] Z. Guo and C.-M. Pun, “RPCA-induced self-representation for subspace clustering,” *Neurocomputing*, vol. 437, pp. 249–260, 2021.
- [56] Q. Zheng, Y. Wang, and P. A. Heng, “Online subspace learning from gradient orientations for robust image alignment,” *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3383–3394, 2019.
- [57] H. Wang, R. Cheng, J. Zhou, L. Tao, and H. K. Kwan, “Multi-stage model for robust face alignment using deep neural networks,” *Cognitive Computation*, pp. 1–17, 2021.

- [58] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Subspace learning from image gradient orientations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2454–2466, 2012.
- [59] Q. Guo, Y. Zhang, S. Qiu, and C. Zhang, "Accelerating patch-based low-rank image restoration using KD-forest and Lanczos approximation," *Information Sciences*, vol. 556, pp. 177–193, 2021.
- [60] X. Lin, Y. Liang, J. Wan, C. Lin, and S. Z. Li, "Region-based context enhanced network for robust multiple face alignment," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3053–3067, 2019.
- [61] S. Xiao, M. Tan, D. Xu, and Z. Y. Dong, "Robust kernel low-rank representation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2268–2281, 2015.
- [62] M. Sun, S. Wang, P. Zhang et al., "Projective multiple kernel subspace clustering," *IEEE Transactions on Multimedia*, p. 1, 2021.
- [63] J. Lou, X. Cai, Y. Wang, H. Yu, and S. Canavan, "Multi-subspace supervised descent method for robust face alignment," *Multimedia Tools and Applications*, vol. 78, no. 24, pp. 35455–35469, 2019.
- [64] Y. Cai, Z. Zhang, Z. Cai, X. Liu, X. Jiang, and Q. Yan, "Graph convolutional subspace clustering: a robust subspace clustering framework for hyperspectral image," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4191–4202, 2020.
- [65] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2144–2151, Barcelona, Spain, November 2011.
- [66] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *2013 IEEE International Conference on Computer Vision*, pp. 1513–1520, Sydney, NSW, Australia, December 2013.