

Received August 29, 2019, accepted September 30, 2019, date of publication October 3, 2019, date of current version October 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945356

An Accelerated Procrustean Markov Process Model With Coherent Constraint for Non-Rigid Structure From Motion

YING ZHANG^{1,2}, XIA CHEN^{1,2}, ZHAN-LI SUN¹ , (Member, IEEE), KIN-MAN LAM³, (Senior Member, IEEE), AND ZHIGANG ZENG^{4,5} , (Senior Member, IEEE)

¹School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China

²Anhui Province Key Laboratory of Non-Destructive Evaluation, Hefei ZC Optoelectronic Technologies Ltd., Anhui 230031, China

³Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong

⁴School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

⁵Key Laboratory of Image Processing and Intelligent Control of Education Ministry of China, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Zhan-Li Sun (zhlsun2006@126.com)


This work was supported in part by the National Natural Science Foundation of China under Grant 61972002, and in part by the Anhui Province Key Laboratory of Non-Destructive Evaluation under Grant CGHBMWSJC07.

ABSTRACT Non-Rigid Structure from Motion (NRSfM) is the task of reconstructing the 3D point set of a non-rigid object from an ensemble of images with 2D correspondences, which has been a long-lasting challenging research topic. Compared to the state-of-the-art methods for NRSfM, the Procrustean Markov Process (PMP) model has obtained a relatively good performance. However, the estimation error and the convergence time of the PMP model will increase simultaneously when noise is present. To address this problem, in this paper, a coherent constraint is constructed to suppress the noise in the initialization step of the PMP algorithm. Moreover, an Accelerated Expectation Maximization (AEM) algorithm is devised to optimize the PMP estimation model. Experimental results on several widely used sequences demonstrate that our proposed algorithm achieves state-of-the-art performance, as well as its effectiveness and feasibility.

INDEX TERMS Non-rigid structure from motion, accelerated expectation maximization algorithm, coherent constraint.

I. INTRODUCTION

Reconstructing the 3D object shapes from a set of 2D images has become a valuable approach to enhance the tasks in computer vision, such as virtual reality [1], object recognition [2], biometrics [3], human-computer interaction [4], [5], etc. Non-Rigid Structure from Motion (NRSfM) provides a useful approach to simultaneously estimate the 3D time-varying deformed object and the relative camera motion from the corresponding 2D observation points in a sequence of images. Although many effective algorithms have been proposed for NRSfM in the past few decades, it is still a very complex and ill-posed problem due to the lack of prior information about the 3D structure.

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

In order to solve the uncertainty in NRSfM, many different a priori information, assumptions and constraints have been utilized in reconstructing the 3D shapes. Inspired by the factorization technique for Structure from Motion (SfM) [6], a low-rank constraint was proposed in [7] to model the unknown time-varying deformable 3D shapes, represented as a linear combination of a small number of 3D shape bases. In the matrix factorization method, the 2D observed matrix was factored into a 3D pose matrix and a 3D shape basis matrix. Subsequently, many works have been proposed based on the low-rank shape model. In [8], a closed-form solution was reported, which considers both the low-rank constraint and the rotation constraint. An approximate rank- $3K$ solution was derived in [9] by utilizing a Gaussian prior and a probabilistic principal component analysis shape model. In [10], an approximate rank-3 solution was proposed to

solve NRSfM by considering very small semi-definite programming and a nuclear-norm minimization problem. Furthermore, a multilinear factorization algorithm was presented in [11], which incorporates the shape basis assumption and a time-independent latent smoothness characteristic of the unknown 3D non-rigid shapes.

A dual approach to the shape basis representation was proposed in [12] to reduce the number of unknown parameters. The dual approach assumes that the 3D point trajectories are constrained to lie in a linear trajectory space. The linear space is compactly spanned by $3K$ predefined independent basis trajectories, obtained via the Discrete Cosine Transform (DCT). Nevertheless, the rank- $3K$ constraint has limited capability to model high-frequency deformation, represented by trajectories. In [13] and [14], a better reconstruction of high-frequency deformation was achieved without relaxing the rank- $3K$ constraint, by modeling a smoothly deforming 3D shape as a single point moving along a smooth time-trajectory within a linear shape space. The predefined DCT was applied to represent the coefficients of shape basis.

Based on the trajectory representation, a scalable monocular surface reconstruction method was proposed in [15] to solve the NRSfM problem, for both sparse and dense data. The optimized solution was obtained through singular value thresholding, proximal gradient and alternating direction method of multipliers. In [16], the dense NRSfM problem with complex non-rigid deformations was solved based on the Grassmann manifold. The complex non-rigid deformation was assumed lying on a union of local linear subspaces, both spatially and temporally. In addition, a scalable, efficient, and accurate solution was proposed in [17] to solve the NRSfM problem, by combining the existing point-trajectory low-rank models with a probabilistic framework for matrix normal distributions.

For the trajectory-based methods, how to determine the optimal number of shape bases is a difficult problem. In [18], a Procrustean Normal Distribution (PND) was proposed to represent the distribution of shape deformations by strictly separating the motion and deformation components. The 3D structure can be accurately reconstructed via an EM algorithm, without requiring any additional constraints or prior knowledge. Although [18] and the improved version (PND2) [19] have achieved a relatively good reconstruction performance on most commonly used datasets, they do not work well for shapes with some large drastic deformations and noise, due to the lack of smoothing constraints.

In [20], a Procrustean Markov Process (PMP) model was proposed to enforce the smoothness constraint between two adjacent frames. The sequence of 3D shapes is considered as a simple stationary Markov process based on Procrustes alignment. Nevertheless, the convergence of the EM algorithm is relatively slow. Moreover, the PMP model is sensitive to noise. In this paper, an accelerated PMP model with a coherent constraint is proposed to improve the robustness and the convergence speed of the EM algorithm. Experimental results on several commonly used sequences

verify the effectiveness and feasibility of the proposed algorithm.

The key contributions of the proposed approach are two aspects, as follows:

- In order to suppress the noise, a coherent constraint, corresponding to a displacement function, is proposed to preserve the global structure of each shape by constraining adjacent points to move coherently.
- An Accelerated Expectation Maximization (AEM) algorithm is proposed to achieve faster convergence, when optimizing the PMP estimation model.

The remainder of the paper is organized as follows. A detailed description of the proposed method is presented in Section II. Experimental results are given in Section III. Finally, conclusions are made in Section IV.

II. METHODOLOGY

The proposed algorithm is composed of three main components: formulation of the PMP model [20], initialization of the PMP model with a coherent constraint, and the optimization of the PMP model using the proposed accelerated EM algorithm.

A. FORMULATION OF THE PMP MODEL

For the i^{th} ($i = 1, \dots, n_s$) frame in an image sequence, the observed 3D structure \mathbf{D}_i can be represented as a collection of 3D coordinates (x, y, z) of n_p feature points, i.e.

$$\mathbf{D}_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,n_p} \\ y_{i,1} & y_{i,2} & \cdots & y_{i,n_p} \\ z_{i,1} & y_{i,2} & \cdots & y_{i,n_p} \end{bmatrix}. \quad (1)$$

Under the orthographic projection model, the z coordinates are unknown for \mathbf{D}_i . Define a $3 \times n_p$ binary weight matrix \mathbf{W}_i , whose elements in the first two rows and the third row are all ones and zeros, respectively. Then, (1) can be represented as,

$$\mathbf{D}_i = \mathbf{W}_i \odot (\mathbf{X}_i - \mathbf{t}_i \mathbf{1}^T) + \mathbf{m}_i, \quad (2)$$

where the $3 \times n_p$ hidden variable \mathbf{X}_i denotes the true 3D shape of the i^{th} frame, $\mathbf{t}_i \in \mathbf{R}^{3 \times 1}$ is the translation, $\mathbf{1} \in \mathbf{R}^{n_p \times 1}$ is a vector with elements of one, and $\mathbf{m}_i \in \mathbf{R}^{3 \times n_p}$ is a zero-mean Gaussian noise with standard deviation σ . The operator \odot denotes the Hadamard product.

As in [20], given the scale s_i and the rotation matrix $\mathbf{R}_i \in \mathbf{R}^{3 \times 3}$, \mathbf{X}_i can be aligned to $\mathbf{Y}_i \in \mathbf{R}^{3 \times n_p}$, as follows:

$$\mathbf{Y}_i = s_i \mathbf{R}_i \mathbf{X}_i. \quad (3)$$

For \mathbf{Y}_i , the first-order linear Markov process can be given as follows:

$$\text{vec}(\mathbf{Y}_i) = \alpha \text{vec}(\mathbf{Y}_{i-1} - \bar{\mathbf{Y}}) + \text{vec}(\bar{\mathbf{Y}}) + \mathbf{n}_i, \quad (4)$$

where $\text{vec}(\mathbf{Y}_i)$ is a vectorization form of \mathbf{Y}_i . \mathbf{Y}_{i-1} and $\bar{\mathbf{Y}} \in \mathbf{R}^{3 \times n_p}$ are the $(i-1)^{\text{th}}$ aligned 3D shape and the mean shape of \mathbf{Y}_i ($i = 1, \dots, n_s$), respectively [20]. The smoothness parameter α is the transition probability of \mathbf{Y}_i moving through the successive time periods. The noise term

$\mathbf{n}_i \in \mathbf{R}^{3n_p \times 1}$ is a Gaussian random vector with independent and identical distribution. In (4), the smoothness assumption can effectively reduce the effect of large deformation.

The aligned 3D shapes \mathbf{Y}_i obeys the procrustean normal distribution [18], i.e.

$$p(\mathbf{Y}_i) \sim \mathcal{N}_{\mathcal{P}}(\bar{\mathbf{Y}}, \boldsymbol{\Sigma}_R), \quad (5)$$

where the symbol $\mathcal{N}_{\mathcal{P}}(\cdot, \cdot)$ denotes the procrustean normal distribution, and $\boldsymbol{\Sigma}_R \in \mathbf{R}^{3n_p \times 3n_p}$ denotes the covariance matrix of $\mathbf{Y}_i (i = 1, \dots, n_s)$. As done in [18], in order to solve the singular problem, $\boldsymbol{\Sigma}_R$ is decomposed as,

$$\boldsymbol{\Sigma}_R = \mathbf{Q}\boldsymbol{\Sigma}\mathbf{Q}^T, \quad (6)$$

where $\mathbf{Q} \in \mathbf{R}^{3n_p \times (3n_p - 7)}$ and $\boldsymbol{\Sigma} \in \mathbf{R}^{(3n_p - 7) \times (3n_p - 7)}$ are an orthogonal matrix and a non-singular positive definite symmetric matrix, respectively.

Combining (4) and (5), the distribution of \mathbf{n}_i is given as follows:

$$\mathbf{n}_i \sim \mathcal{N}(0, \mathbf{Q}\mathbf{H}\mathbf{Q}^T), \quad (7)$$

where the symbol $\mathcal{N}(\cdot, \cdot)$ denotes the normal distribution, and $\mathbf{H} \in \mathbf{R}^{(3n_p - 7) \times (3n_p - 7)}$ is an unknown positive definite symmetric matrix [20]. Furthermore, the mean $\mu_{\mathbf{Y}_i|\mathbf{Y}_{i-1}}$ and the variance $\Sigma_{\mathbf{Y}_i|\mathbf{Y}_{i-1}}$ of the conditional probability $p(\mathbf{Y}_i|\mathbf{Y}_{i-1})$ can be computed as follows:

$$\mu_{\mathbf{Y}_i|\mathbf{Y}_{i-1}} = \alpha \text{vec}(\mathbf{Y}_{i-1}) + \text{vec}(\bar{\mathbf{Y}}), \quad (8)$$

$$\Sigma_{\mathbf{Y}_i|\mathbf{Y}_{i-1}} = \mathbf{Q}\mathbf{H}\mathbf{Q}^T. \quad (9)$$

Considering (5), (8) and (9), the probability $p(\{\mathbf{Y}_i\})$ ($i = \{1, 2, \dots, n_s\}$) can be given as,

$$p(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_{n_s}|\Theta) = p(\mathbf{Y}_1|\Theta) \prod_{j=2}^{n_s} p(\mathbf{Y}_j|\mathbf{Y}_{j-1}, \Theta), \quad (10)$$

where $\Theta = \{\bar{\mathbf{Y}}, \mathbf{R}_i, s_i, \alpha, \mathbf{H}, \boldsymbol{\Sigma}, \mathbf{Q}, \sigma\}$. Referring to [20], the unknown parameters Θ can be estimated by maximizing the following log-likelihood function,

$$\begin{aligned} \log(p(\{\mathbf{D}_i, \mathbf{X}_i\}|\Theta)) &= \log(p(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{n_s}|\Theta)) \\ &+ \sum_{i=1}^{n_s} \log(p(\mathbf{D}_i|\mathbf{X}_i, \Theta)). \end{aligned} \quad (11)$$

B. INITIALIZATION OF PMP MODEL WITH A COHERENT CONSTRAINT

Considering the coherent constraint, a good initial value for Θ can be obtained for PMP via the following optimization model,

$$\begin{aligned} \min_{\Psi} \sum_{i=1}^{n_s} \left\| s_i \mathbf{R}_i \mathbf{X}_i + \mathbf{t}_i \mathbf{1}^T + \mathbf{C}_i \mathbf{G}_i - \bar{\mathbf{X}} \right\|_F^2 + \lambda \text{tr}(\mathbf{C}_i \mathbf{G}_i \mathbf{C}_i^T) \\ \text{s.t. } \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}_3, \quad \left\| \bar{\mathbf{X}} \right\|_F^2 = 1, \end{aligned} \quad (12)$$

where $\mathbf{C}_i \in \mathbf{R}^{3 \times n_p}$ is a coefficient matrix, $\bar{\mathbf{X}}$ is the mean matrix of $\mathbf{X}_i (i = 1, \dots, n_s)$, Ψ is a collection of unknown

parameters $\Psi = \{s_i, \mathbf{R}_i, \mathbf{t}_i, \bar{\mathbf{X}}, \mathbf{X}_i, \mathbf{C}_i\}$, and \mathbf{I}_3 is a 3×3 identity matrix. The matrix $\mathbf{G}_i \in \mathbf{R}^{n_p \times n_p}$ is a kernel matrix, whose element g_{mn} is computed as follows:

$$g_{mn} = \mathbf{G}(\mathbf{X}_{i,m}, \mathbf{X}_{i,n}) = \exp\left(-\frac{\|\mathbf{X}_{i,m} - \mathbf{X}_{i,n}\|_F^2}{2\beta}\right), \quad (13)$$

where $\mathbf{X}_{i,m}$ and $\mathbf{X}_{i,n}$ are the m^{th} and the n^{th} point in \mathbf{X}_i , respectively; and β defines the width of the Gaussian kernel function [25]. The transformation $\mathbf{C}_i \mathbf{G}_i$ is assumed to be a displacement function. The regularization term $\text{tr}(\mathbf{C}_i \mathbf{G}_i \mathbf{C}_i^T)$ is a global structure constraint, following the motion coherence theory [23], [24], which can constrain the smoothness of $\mathbf{C}_i \mathbf{G}_i$ [25]. The parameter λ makes a trade-off between Procrustes alignment and regularization.

First, $\bar{\mathbf{X}}$ and \mathbf{t}_i can be obtained by combining (12) and the constraint term $\bar{\mathbf{X}}\mathbf{1} = 0$,

$$\bar{\mathbf{X}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (s_i \mathbf{R}_i \mathbf{X}_i + \mathbf{t}_i \mathbf{1}^T + \mathbf{C}_i \mathbf{G}_i), \quad (14)$$

$$\mathbf{t}_i = -\frac{1}{n_p} (s_i \mathbf{R}_i \mathbf{X}_i + \mathbf{C}_i \mathbf{G}_i) \mathbf{1}. \quad (15)$$

Substitute (15) into (12) and let $\mathbf{B} = \mathbf{I}_{n_p} - \frac{1}{n_p} \mathbf{1}\mathbf{1}^T$, where $\mathbf{B} \in \mathbf{R}^{n_p \times n_p}$, we can get

$$\min_{\mathbf{R}_i} \left\| s_i \mathbf{R}_i \mathbf{X}_i \mathbf{B} - (\bar{\mathbf{X}} - \mathbf{C}_i \mathbf{G}_i) \mathbf{B} \right\|_F^2. \quad (16)$$

For (16), considering the orthogonal Procrustes problem, we have

$$\mathbf{R}_i = \mathbf{V}_i \mathbf{U}_i^T, \quad (17)$$

where $\mathbf{U}_i \boldsymbol{\Lambda}_i \mathbf{V}_i^T = \text{svd}(\mathbf{X}_i \mathbf{B} (\bar{\mathbf{X}} - \mathbf{C}_i \mathbf{G}_i) \mathbf{B}^T)$, and $\text{svd}(\cdot)$ represents the singular value decomposition [26].

As each shape variation is assumed to be orthogonal to the mean shape [18], we have

$$\text{vec}(s_i \mathbf{R}_i \mathbf{X}_i + \mathbf{C}_i \mathbf{G}_i - \bar{\mathbf{X}})^T \text{vec} \bar{\mathbf{X}} = 0. \quad (18)$$

Considering $\left\| \bar{\mathbf{X}} \right\|_F^2 = 1$, we have

$$\text{vec}(s_i \mathbf{R}_i \mathbf{X}_i + \mathbf{C}_i \mathbf{G}_i)^T \text{vec}(\bar{\mathbf{X}}) = 1. \quad (19)$$

According to (19), s_i can be computed as follows

$$s_i = \frac{1 - \text{vec}(\mathbf{C}_i \mathbf{G}_i)^T \text{vec} \bar{\mathbf{X}}}{\text{vec}(\mathbf{R}_i \mathbf{X}_i \mathbf{B}_i)^T \text{vec} \bar{\mathbf{X}}}. \quad (20)$$

As in [20], the true 3D shape, \mathbf{X}_i , can be decomposed as follows:

$$\mathbf{X}_i = \mathbf{D}_i + \mathbf{L}(\mathbf{z}_i), \quad (21)$$

where $\mathbf{L}(\mathbf{z}_i)$ transforms \mathbf{z}_i into a $3 \times n_p$ matrix, in which the elements of the first two rows are zeros and the elements of the third row are \mathbf{z}_i . Furthermore, $\text{vec}(\mathbf{L}(\mathbf{z}_i)) = \tilde{\mathbf{W}} \mathbf{z}_i$, where $\tilde{\mathbf{W}}$ is a truncated version of $(\mathbf{I} - \text{diag}(\text{vec}(\mathbf{W})))$ removing all-zero columns.

Considering (15) and (21), (12) can be rewritten as follows:

$$\sum_{i=1}^{n_s} \left\| (s_i \mathbf{R}_i (\mathbf{D}_i + \mathbf{L}(\mathbf{z}_i)) + \mathbf{C}_i \mathbf{G}_i) \mathbf{B} - \bar{\mathbf{X}} \right\|_F^2 + \lambda \text{tr}(\mathbf{C}_i \mathbf{G}_i \mathbf{C}_i^T). \quad (22)$$

Then, we compute the one-order partial derivative of (22) with respect to \mathbf{C}_i and \mathbf{z}_i , respectively. As a result, \mathbf{C}_i and \mathbf{z}_i can be derived by setting these two partial derivatives to zeros, as follows:

$$\mathbf{z}_i = \left[\tilde{\mathbf{W}}^T (\mathbf{B} \otimes \mathbf{I}_3) \tilde{\mathbf{W}} \right]^\dagger \times \left[\tilde{\mathbf{W}}^T (\mathbf{B} \otimes \mathbf{I}_3) \text{vec} \left(\frac{1}{s_i} \mathbf{R}_i^T (\bar{\mathbf{X}} + \mathbf{C}_i \mathbf{G}_i) - \mathbf{D}_i \right) \right], \quad (23)$$

$$\mathbf{C}_i = \left[\left(\frac{1}{s_i} \mathbf{R}_i^T \bar{\mathbf{X}} - \mathbf{X}_i \mathbf{B} \right) \right] \left[\mathbf{G} \mathbf{B} + \frac{\lambda}{s_i^2} \mathbf{I}_{3n_p} \right]^\dagger, \quad (24)$$

where the operators \otimes and \dagger denote the Kronecker product and the pseudo inverse, respectively.

C. THE PMP MODEL OPTIMIZATION USING AN ACCELERATED EM ALGORITHM

For the model (11), an accelerated EM algorithm is proposed to derive the solutions. Let $\mu_{i|n_s}$ and $\mathbf{C}_{i|n_s}$ be the mean and covariance of $p(\mathbf{X}_i | \mathbf{D}_i, \dots, \mathbf{D}_{n_s})$, respectively. The cross-covariance of $\text{vec}(\mathbf{X}_i)$ and $\text{vec}(\mathbf{X}_{i+1})$ is denoted as $\mathbf{C}_{i,i+1|n_s}$. The variables $\mu_{i|n_s}$, $\mathbf{C}_{i|n_s}$ and $\mathbf{C}_{i,i+1|n_s}$ can be computed by the Kalman smoothing in the E-step [20].

In the M-step, all the unknown parameters Θ in (11) are updated by maximizing the expectation of (11), i.e.

$$\mathbf{J}(\Theta | \Theta^t) = E \left[\log (p(\{\mathbf{D}_i, \mathbf{X}_i\} | \Theta)) | \Theta^t \right]. \quad (25)$$

Then, each element of Θ^t at the $(t + 1)^{\text{th}}$ iteration can be obtained as follows:

$$\Theta^{t+1} = \arg \max_{\theta \in \Theta} \mathbf{J}(\Theta | \Theta^t). \quad (26)$$

The E-step and M-step of the original EM algorithm are repeated to produce a series of estimates $(\Theta^{t+1}, \Theta^t, \Theta^{t-1})$.

Denote ϕ as a vectorized variable of Θ . Referring to [27], for the accelerated EM algorithm, ϕ can be updated as follows:

$$\phi_{new}^t = \phi^t + \left[\left[\phi^{t-1} - \phi^t \right]^{-1} \right] + \left[\phi^{t+1} - \phi^t \right]^{-1} \right]^{-1}, \quad (27)$$

where the operation $[\cdot]^{-1}$ is defined as follows:

$$[\cdot]^{-1} = \frac{1}{\|\cdot\|^2}. \quad (28)$$

In (27), a problem is addressed here. In (11), \mathbf{H} and Σ are both required to be positive definite symmetric matrices. In order to satisfy this condition, the upper or lower triangular part of \mathbf{H}^t and Σ^t are first extracted and vectorized. After updated by (27), they are transformed into the corresponding positive definite symmetric matrices.

TABLE 1. The numbers of frames (n_s) and the numbers of point tracks (n_p) for eleven motion capture sequences.

Number	Sequence	n_s	n_p
1	walking	260	55
2	jaws	240	91
3	dance	264	75
4	face1	74	37
5	face2	316	40
6	pickup	357	41
7	stretch	370	41
8	yoga	307	41
9	drink	1102	41
10	FRGC	400	62
11	capoeira	250	41

As a result, we can obtain a set of new variables Θ_{new}^t according to (26) and (27). Denote \mathbf{b}_{new}^t as,

$$\mathbf{b}_{new}^t = \left[\text{vec}(\bar{\mathbf{Y}}_{new}^t); \text{vec}(\mathbf{R}_{new}^t); \text{vec}(s_{new}^t); \alpha_{new}^t; \text{vec}(\mathbf{H}_{new}^t); \text{vec}(\Sigma_{new}^t); \text{vec}(\mathbf{Q}_{new}^t); \sigma_{new}^t \right], \quad (29)$$

where $\mathbf{b}_{new}^t \in \mathbf{R}^{c \times 1}$, and $c = (3n_p - 7)(6n_p - 7) + 3n_p + 10n_s + 2$.

The iterations are repeated until,

$$e_\Theta = \|\mathbf{b}_{new}^t - \mathbf{b}_{old}^t\|_F^2 < \rho, \text{ or } t > \tau, \quad (30)$$

where τ is the maximum number of iterations. The pseudocode of the PMP-CAEM algorithm is given in Algorithm 1.

Algorithm 1 The Pseudocode of the PMP-CAEM Algorithm

- 1: Initialize $\Theta^0 = \{\bar{\mathbf{Y}}^0, \mathbf{R}_i^0, s_i^0, \alpha^0, \mathbf{H}^0, \Sigma^0, \mathbf{Q}^0, \sigma^0\}$, \mathbf{b}^0 .
- 2: Set $\rho = 1e - 05$, $\tau = 1e + 03$.
- 3: $\Theta^1 \leftarrow \Theta^0$, $\mathbf{b}^1 \leftarrow \mathbf{b}^0$,
- 4: $\Theta_{old}^1 \leftarrow \Theta^1$, $\mathbf{b}_{old}^1 \leftarrow \mathbf{b}^1$
- 5: $t \leftarrow 0$,
- 6: **repeat**
- 7: Compute Θ_{new}^t by (26) and (27),
- 8: Compute e_Θ by (30),
- 9: $\Theta_{old}^t \leftarrow \Theta_{new}^t$, $\mathbf{b}_{old}^t \leftarrow \mathbf{b}_{new}^t$,
- 10: $\Theta^{t-1} \leftarrow \Theta_{new}^t$,
- 11: $\Theta^t \leftarrow \Theta^{t+1}$,
- 12: Update $t \leftarrow t + 1$,
- 13: **until** $e_\Theta < \rho$ or $t > \tau$.

III. EXPERIMENTS

A. EXPERIMENT DATASETS AND SET-UP

The performance of the proposed method is evaluated on eleven widely used motion sequences: walking, jaws, dance, face1, face2, pickup, stretch, yoga, drink, Face Recognition Grand Challenge (FRGC), and capoeira. These sequences are publicly available from [9], [12], [18], [21]. Note that the FRGC is a 3D facial-landmark dataset from [18], by adding random rotation and scaling to the original FRGC 2.0 database without temporal dependence [22]. For these sequences, the corresponding number of frames (n_s) and the number of points tracked (n_p) are listed in Table 1.

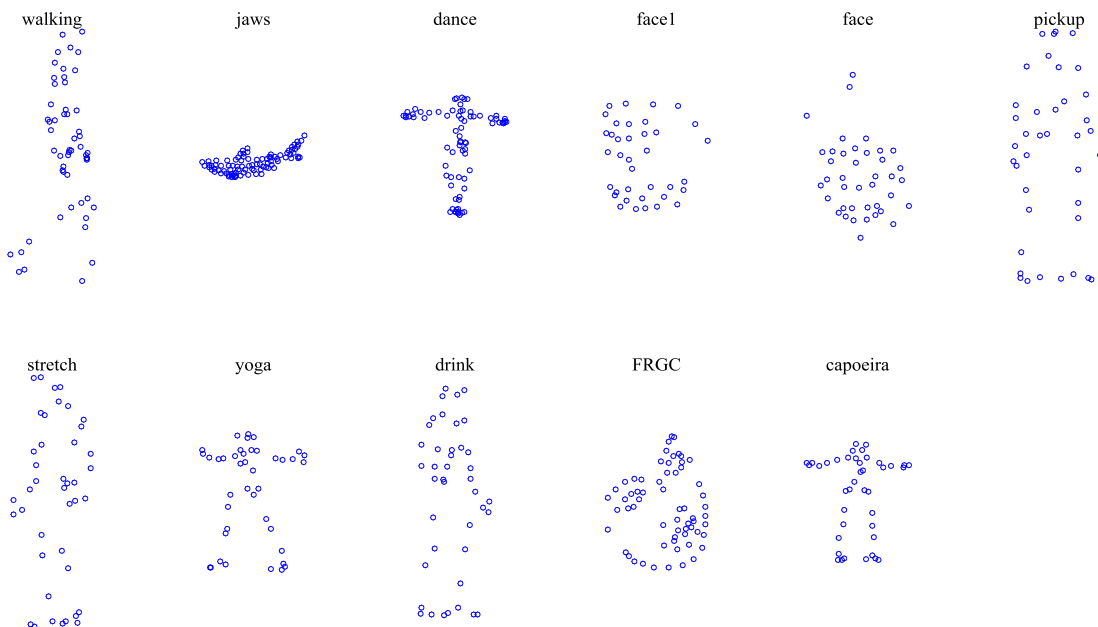


FIGURE 1. One frame of the eleven widely used motion sequences.

TABLE 2. The 3D reconstruction error ε of eleven sequences without noise for six different methods, and the corresponding mean and standard deviation ($\mu \pm \sigma$) for each method on all the sequences.

Sequence	CSF	CSF2	BMM	PND2	PMP	PMP-CAEM
walking	0.1050	0.0695	0.0805	0.0407	0.0424	0.0428
jaws	0.0048	0.0259	0.1448	0.0272	0.0096	0.0096
dance	0.1808	0.1397	0.1360	0.1247	0.1278	0.1280
face1	0.0433	0.0343	0.0362	0.0215	0.0198	0.0204
face2	0.0677	0.0206	0.0197	0.0150	0.0166	0.0162
pickup	0.0170	0.0213	0.0302	0.0133	0.0127	0.0126
stretch	0.0246	0.0219	0.0235	0.0150	0.0124	0.0123
yoga	0.0211	0.0207	0.0223	0.0128	0.0128	0.0129
drink	0.0059	0.0071	0.0149	0.0031	0.0018	0.0019
FRGC	0.1909	0.1909	0.1147	0.0731	0.0727	0.0726
capoeira	0.2258	0.3309	0.2544	0.3116	0.3132	0.3126
$\mu \pm \sigma$	0.0803 \pm 0.0103	0.0806 \pm 0.0067	0.0798 \pm 0.0058	0.0598 \pm 0.0082	0.0584 \pm 0.0085	0.0596 \pm 0.0080

Figure 1 shows one frame of these eleven image sequences. All simulations were conducted using MATLAB, running on an ordinary personal computer.

In order to evaluate the reconstruction performance, the normalized error ε of the 3D coordinates between the estimated 3D shape ($\tilde{\mathbf{X}}_i$) and the ground-true 3D shape (\mathbf{X}_i) is used as the performance index, i.e.,

$$\varepsilon = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{\|\mathbf{X}_i - \tilde{\mathbf{X}}_i\|_F^2}{\|\mathbf{X}_i\|_F^2}. \quad (31)$$

Smaller ε means that the estimation is more accurate.

B. COMPARISON TO RECENTLY REPORTED RESULTS

In order to evaluate the effectiveness of the proposed method, denoted as PMP-CAEM, we compare it with several state-of-the-art NRSfM algorithms, including the well-known block matrix method (denoted as BMM) [10], the

TABLE 3. The computation runtimes (seconds) of eleven sequences without noise for the six methods.

Sequence	CSF	CSF2	BMM	PND2	PMP	PMP-CAEM
walking	9.7	18.2	1381.8	281.3	237.4	146.5
jaws	0.1	18.5	192.4	629.9	353.2	437.2
dance	35.5	34.2	1083.0	531.7	510.2	213.7
face1	4.5	4.0	8.8	34.3	29.5	13.6
face2	6.4	37.3	45.9	126.0	117.0	73.5
pickup	10.2	107.2	134.9	211.2	158.6	102.2
stretch	12.1	81.0	1059.4	217.4	168.5	139.0
yoga	21.2	47.3	843.2	236.8	123.6	49.6
drink	47.2	563.6	1439.9	577.6	350.4	138.0
FRGC	31.1	32.2	96.1	318.4	258.0	238.1
capoeira	7.4	12.6	111.9	149.7	117.3	68.7

column-space-fitting method (denoted as CSF) [13], the CSF2 method [14], the procrustean normal distribution method (denoted as PND2) [19] and the procrustean Markov process method (denoted as PMP) [20].

TABLE 4. The 3D reconstruction errors ε of the six methods on eleven sequences when a_{noise} is set at 0.26.

Sequence	CSF	CSF2	BMM	PND2	PMP	PMP-CAEM
walking	0.5112	0.5137	0.4759	1.0359	0.4503	0.4130
jaws	0.7233	1.6459	0.3616	0.7870	0.4527	0.4381
dance	0.5359	1.9010	0.5149	1.0398	0.4570	0.4571
face1	0.5380	0.4944	0.3627	0.7121	0.3402	0.3364
face2	0.6021	0.6467	0.6493	0.9700	0.5474	0.5473
pickup	0.4390	0.4136	0.4662	0.6883	0.3502	0.3523
stretch	0.4696	0.4572	0.4561	0.7111	0.4365	0.4076
yoga	0.4571	0.4393	0.4709	0.6681	0.2987	0.2940
drink	0.3685	0.3548	0.3917	0.6339	0.3221	0.3218
FRGC	0.5915	0.5917	0.4451	0.8351	0.3720	0.3717
capoeira	0.5094	0.4977	0.4930	0.7329	0.4619	0.4604
$\mu \pm \sigma$	0.5223 \pm 0.0090	0.7233 \pm 0.2792	0.4625 \pm 1.0064	0.8013 \pm 0.0221	0.4081 \pm 0.0058	0.4000 \pm 0.0055

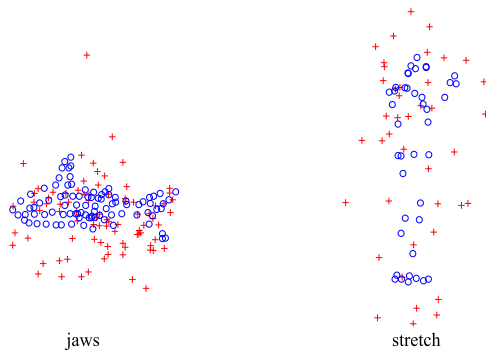


FIGURE 2. One frame of the sequences *jaws* and *stretch* with and without noise. The symbols ‘o’ and ‘+’ represent the observed ground truth and the points with noise, respectively.

Among these methods, the low-rank parameter K has a significant influence on the final estimation performance for CSF, CSF2 and BMM. For a fair comparison, the parameter K is successively set at $\{1, 2, \dots, 13\}$ for these three methods. The parameter value corresponding to the smallest estimation error is selected as the approximate optimum parameter value of K .

Table 2 shows the 3D reconstruction errors ε on eleven sequences without noise for the six methods. In order to easily compare the performances of different algorithms, the best result and the second-best result in Table 2 are highlighted in red and blue, respectively. The reconstruction errors of PND2, PMP and PMP-CAEM are generally lower than that of other methods for most sequences. Moreover, the reconstruction errors of PMP-CAEM are close to that of PMP. Table 3 shows the computation times (seconds) of the different methods on the eleven sequences without noise. We can see that the computation runtimes of PND2, PMP and PMP-CAEM are obviously longer than that of other methods. However, the computation times of PMP-CAEM are significantly lower than that of PMP. This shows that the computation runtimes required by PMP can be greatly reduced by the use of the accelerated expectation maximization algorithm.

TABLE 5. The number of iterations for the EM algorithm in PMP and PMP-CAEM when a_{noise} is set at 0.26.

Sequence	PMP	PMP-CAEM
walking	1000	273
jaws	457	164
dance	698	438
face1	93	77
face2	267	258
pickup	816	678
stretch	693	272
yoga	469	358
drink	1000	1000
FRGC	243	183
capoeira	507	432

In order to investigate the robustness to noise, we conducted the experiments with the addition of the Gaussian noise on the original sequences. The standard deviation or level of the Gaussian noise is set as $\sigma_{noise} = a_{noise} \max_{i,j,k} \{|d_{ijk}|\}$, where the noise rate a_{noise} is set at $\{0.2, 0.22, 0.24, 0.26, 0.28\}$, respectively, and d_{ijk} is the $(j, k)^{th}$ elements of \mathbf{D}_i , where $i = 1, \dots, n_s$ and $j = 1, 2, 3; k = 1, \dots, n_p$. Figure 2 shows one frame of the sequences *jaws* and *stretch* with and without noise. The symbols ‘o’ and ‘+’ represent the points of the ground truth and points with noise, respectively. We can see that the positions of the points are randomly changed when noise is added to the original data.

As an example, Table 4 shows the 3D reconstruction errors ε of the six methods on the eleven sequences for the six methods when a_{noise} is set at 0.26. We can see from Tables 2 and 4 that the reconstruction errors of the various algorithms are significantly increased when noises are added. The reconstruction errors of PMP and PMP-CAEM are obviously lower than that of other methods for most of the sequences. For CSF, CSF2, and BMM, a 3D shape is assumed to be composed by a linear combination of K shape bases. Such a model cannot achieve a satisfactory result because the deformation and translation caused by the noise are random and irregular for the different points.

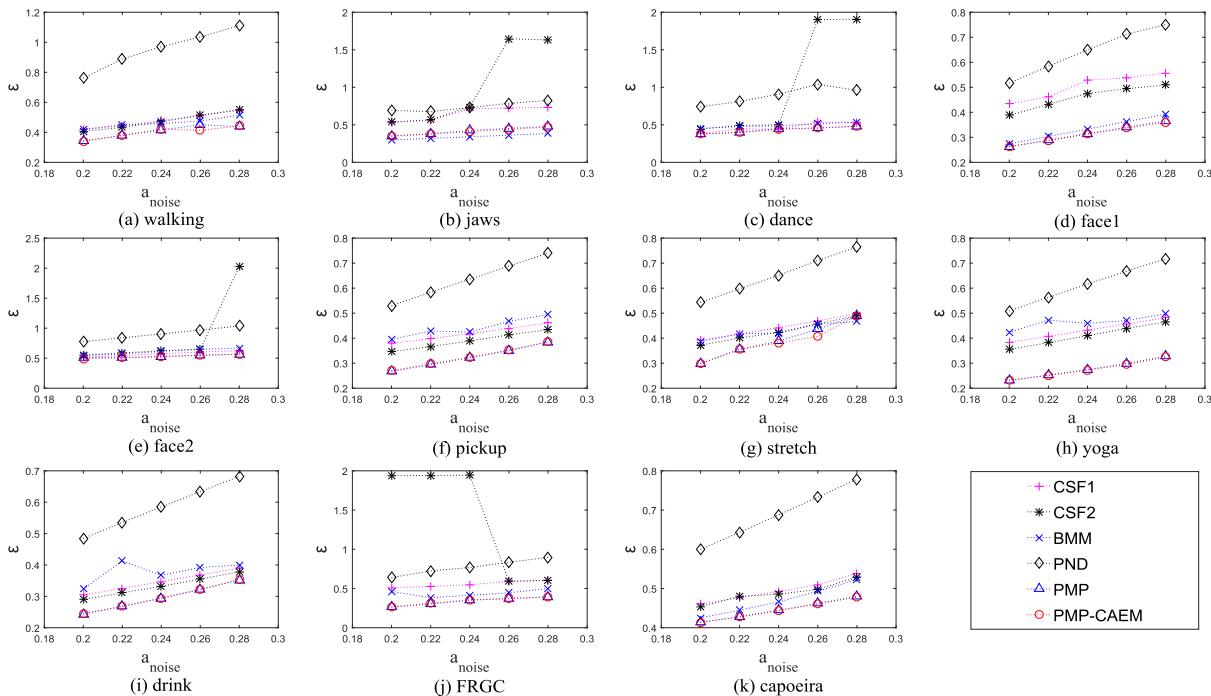


FIGURE 3. The 3D reconstruction errors ϵ of the six methods on the eleven sequences, when a_{noise} is set at different values.

TABLE 6. The mean and standard deviation ($\mu \pm \sigma$) of 3D reconstruction errors ϵ of the six methods on the eleven sequences, when a_{noise} is set at different values.

Sequence	CSF	CSF2	BMM	PND2	PMP	PMP-CAEM
walking	0.4794±0.0518	0.4751±0.0586	0.4636±0.0346	0.9535±0.1349	0.4051±0.0443	0.3992±0.0401
jaws	0.6522±0.0992	1.0239±0.5666	0.3407±0.0340	0.7419±0.0628	0.4179±0.0511	0.4083±0.0490
dance	0.4680±0.0653	1.0492±0.7814	0.4935±0.0318	0.8921±0.1186	0.4335±0.0434	0.4303±0.0403
face1	0.5039±0.0530	0.4598±0.0497	0.3330±0.0468	0.6427±0.0947	0.3144±0.0407	0.3119±0.0388
face2	0.5727±0.0451	0.8874±0.6379	0.6095±0.0522	0.9054±0.1026	0.5332±0.0239	0.5278±0.0303
pickup	0.4190±0.0325	0.3899±0.0351	0.4419±0.0390	0.6355±0.0833	0.3236±0.0462	0.3259±0.0452
stretch	0.4436±0.0420	0.4286±0.0464	0.4294±0.0323	0.6534±0.0878	0.3939±0.0733	0.3861±0.0707
yoga	0.4324±0.0390	0.4103±0.0437	0.4646±0.0268	0.6148±0.0825	0.2769±0.0379	0.2749±0.0364
drink	0.3470±0.0339	0.3332±0.0349	0.3791±0.0352	0.5842±0.0782	0.2959±0.0430	0.2965±0.0429
FRGC	0.5558±0.0407	1.4037±0.7356	0.4408±0.0446	0.7724±0.0994	0.3361±0.0523	0.3399±0.0511
capoeira	0.4955±0.0301	0.4891±0.0274	0.4702±0.0388	0.6884±0.0707	0.4458±0.0263	0.4445±0.0258

For PMP and PMP-CAEM, the smoothing constraint can suppress the partial deformation and deviation caused by noise. Different from PMP and PMP-CAEM, the smoothing constraint is not considered in the PND2 model. Therefore, the noise has a more serious effect on its final estimation results. From Table 4, it can be seen that the performance of PND2 is not yet as good as PMP and PMP-CAEM. Moreover, the reconstruction errors of PMP-CAEM are lower than that of PMP.

Table 5 shows the number of iterations for the EM algorithm used in PMP and the accelerated EM used in PMP-CAEM, when a_{noise} is set at 0.26. The number of iterations of PMP-CAEM is obviously lower than that of PMP. Therefore, the accelerated expectation maximization algorithm can significantly decrease the convergence time of PMP.

Figure 3 shows the 3D reconstruction errors ϵ of the six methods on the eleven sequences, when a_{noise} is set at different values. Table 6 tabulates the corresponding mean and standard deviation ($\mu \pm \sigma$) of the 3D reconstruction errors for different noise rates. The reconstruction errors of PMP and PMP-CAEM are obviously lower than that of other methods for most sequences. Moreover, the reconstruction errors of PMP-CAEM are lower than that of PMP, due to the use of the coherence constraint.

IV. CONCLUSION

In this paper, an accelerated PMP model, with a coherent constraint, is proposed for non-rigid structure from motion. The experimental results demonstrated that the proposed method can simultaneously decrease the estimation error and the convergence time of EM algorithm for PMP.

ACKNOWLEDGMENT

(Ying Zhang and Xia Chen contributed equally to this work, joint first authors.)

REFERENCES

- [1] J. Yang, T. Liu, B. Jiang, H. Song, and W. Lu, "3D panoramic virtual reality video quality assessment based on 3D convolutional neural networks," *IEEE Access*, vol. 6, pp. 38669–38682, 2018.
- [2] C.-Y. Tsai and S.-H. Tsai, "Simultaneous 3D object recognition and pose estimation based on RGB-D images," *IEEE Access*, vol. 6, pp. 28859–28869, 2018.
- [3] T. N. Tan and H. Lee, "High-secure fingerprint authentication system using ring-lwe cryptography," *IEEE Access*, vol. 7, pp. 23379–23387, 2019.
- [4] J. Deng, G. Pang, Z. Zhang, Z. Pang, H. Yang, and G. Yang, "cGAN based facial expression recognition for human-robot interaction," *IEEE Access*, vol. 7, pp. 9848–9859, 2019.
- [5] H. Son, S. Shin, S. Choi, S.-Y. Kim, and J. R. Kim, "Interacting auto-multiscopic 3D with haptic paint brush in immersive room," *IEEE Access*, vol. 6, pp. 76464–76474, 2018.
- [6] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, Nov. 1992.
- [7] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3D shape from image streams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2000, pp. 690–696.
- [8] J. Xiao, J. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," *Int. J. Comput. Vis.*, vol. 67, no. 2, pp. 233–246, Apr. 2006.
- [9] L. Torresani, A. Hertzmann, and C. Bregler, "Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 878–892, May 2008.
- [10] Y. Dai, H. Li, and M. He, "A simple prior-free method for non-rigid structure-from-motion factorization," *Int. J. Comput. Vis.*, vol. 107, no. 2, pp. 101–122, 2014.
- [11] Q. Dong and H. Wang, "Latent-smoothness nonrigid structure from motion by revisiting multilinear factorization," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3557–3570, Sep. 2018.
- [12] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Trajectory space: A dual representation for nonrigid structure from motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1442–1456, Jul. 2011.
- [13] P. F. U. Gotardo and A. M. Martinez, "Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2051–2065, Oct. 2011.
- [14] P. F. U. Gotardo and A. M. Martinez, "Non-rigid structure from motion with complementary rank-3 spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 3065–3072.
- [15] M. D. Ansari, V. Golyanik, and D. Stricker, "Scalable dense monocular surface reconstruction," in *Proc. Int. Conf. 3D Vis. (3DV)*, Oct. 2017, pp. 78–87.
- [16] S. Kumar, A. Cherian, Y. Y. Dai, and H. Li, "Scalable dense non-rigid structure-from-motion: A grassmannian perspective," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 254–263.
- [17] A. Agudo and F. Moreno-Noguer, "A scalable, efficient, and accurate solution to non-rigid structure from motion," *Comput. Vis. Image Understand.*, vol. 167, pp. 121–133, Feb. 2018.
- [18] M. Lee, J. Cho, C.-H. Choi, and S. Oh, "Procrustean normal distribution for non-rigid structure from motion" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1280–1287.
- [19] M. Lee, J. Cho, and S. Oh, "Procrustean normal distribution for non-rigid structure from motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1388–1400, Jul. 2017.
- [20] M. Lee, C.-H. Choi, and S. Oh, "A procrustean Markov process for non-rigid structure recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1550–1557.
- [21] P. F. U. Gotardo and A. M. Martinez, "Kernel non-rigid structure from motion," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 802–809.
- [22] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 947–954.
- [23] A. L. Yuille and N. M. Grzywacz, "The motion coherence theory," in *Proc. Int. Conf. Comput. Vis.*, Dec. 1988, pp. 344–353.
- [24] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.
- [25] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
- [26] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
- [27] M. Kuroda, Z. Geng, and M. Sakahihara, "Improving the vector ϵ acceleration for the EM algorithm using a re-starting procedure," *Comput. Statist.*, vol. 30, no. 4, pp. 1051–1077, Dec. 2015.



YING ZHANG is currently pursuing the master's degree with the School of Electrical Engineering and Automation, Anhui University. Her research interests include machine learning, image processing, and signal processing.



XIA CHEN is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation, Anhui University. Her research interests include machine learning, image processing, and signal processing.



ZHAN-LI SUN (M'19) received the Ph.D. degree from the University of Science and Technology of China, in 2005. Since 2006, he has been with The Hong Kong Polytechnic University, Nanyang Technological University, and the National University of Singapore. He is currently a Professor with the School of Electrical Engineering and Automation, Anhui University, China. His research interests include machine learning, image processing, and signal processing. He serves as an Associate Editor of IEEE ACCESS.



KIN-MAN LAM (SM'14) received the Associateship (Hons.) in electronic engineering from The Hong Kong Polytechnic University (formerly called Hong Kong Polytechnic), in 1986, the M.Sc. degree in communication engineering from the Department of Electrical Engineering, Imperial College of Science, Technology and Medicine, London, U.K., in 1987, and the Ph.D. degree from the Department of Electrical Engineering, The University of Sydney, Sydney, Australia, in August 1996. From 1990 to 1993, he was a Lecturer with the Department of Electronic Engineering, The Hong Kong Polytechnic University. In October 1996, he joined the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, as an Assistant Professor, became an Associate Professor, in 1999, and has been a Professor since 2010. His current research interests include human face recognition, image and video processing, and computer vision. He was a Member of the organizing committee and program committee of many international conferences. He was also a BoG Member of the Asia-Pacific Signal and Information Processing Association (APSIPA) and the Director-Student Services of the IEEE Signal Processing Society. He is currently a General Co-Chair of the IEEE International Conference on Signal Processing, Communications, and Computing (ICSPCC2012). He also serves as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, APSIPA *Transactions on Signal and Information Processing*, and the EURASIP *International Journal on Image and Video Processing*.



ZHIGANG ZENG (SM'07) received the Ph.D. degree in systems analysis and integration from the Huazhong University of Science and Technology, Wuhan, China, in 2003. He is currently a Professor with the School of Automation, Huazhong University of Science and Technology, and also with the Key Laboratory of Image Processing and Intelligent Control of the Education Ministry of China, Wuhan, China. He has authored or coauthored more than 100 international journal articles. His current research interests include the theory of functional differential equations and differential equations with discontinuous right-hand sides, and their applications to the dynamics of neural networks, memristive systems, and control systems. Dr. Zeng has been a Member of the Editorial Board of *Neural Networks*, since 2012, *Cognitive Computation*, since 2010, and *Applied Soft Computing*, since 2013. He was an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, from 2010 to 2011, has been an Associate Editor for the IEEE TRANSACTIONS ON CYBERNETICS, since 2014, and the IEEE TRANSACTIONS ON FUZZY SYSTEMS, since 2016.

...