A Local Deviation Constraint based Non-rigid Structure from Motion Approach

Xia Chen, Zhan-Li Sun*, Kin-Man Lam, and Zhigang Zeng

Abstract

In many traditional non-rigid structure from motion (NRSFM) approaches, the estimation results of part feature points may deviate from their true values significantly because only the overall estimation error is considered in the models. Aiming at this issue, a local deviation-constrained based column-space-fitting approach is presented in this paper to alleviate the estimation deviation. In our work, an effective model is first constructed by two terms, i.e. the overall estimation error computed by a linear subspace representation, and the constraint term based on the variance of reconstruction error for each frame. Moreover, an Augmented Lagrange Multipliers (ALM) iterative algorithm is presented to solve the optimization of the proposed model. Because both the overall estimation error and the local deviation are utilized, the proposed method can achieve a good estimation performance and a relative uniform estimation error distribution for different feature points. Experimental results on several widely used synthetic sequences and real sequences demonstrate the effectiveness and feasibility of the proposed algorithm.

Index Terms

Non-rigid structure from motion, 3D face reconstruction, column-space-fitting, Augmented Lagrange Multipliers.

The work was supported by a grant from National Natural Science Foundation of China (No. 61370109), a key project of support program for outstanding young talents of Anhui province university (No. gxyqZD2016013), a grant of science and technology program to strengthen police force (No. 1604d0802019), and a grant for academic and technical leaders and candidates of Anhui province (No. 2016H090).

Xia Chen and Zhan-Li Sun are with School of Electrical Engineering and Automation, Anhui University, China.

Kin-Man Lam is with Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong.

*Corresponding author. (e-mail: zhlsun2006@126.com)

I. Introduction

Nowadays, recovering 3D object shapes from 2D images has become a valuable measure to enhance the tasks in computer vision, such as face recognition [1]–[3], etc. As a fundamental method of 3D reconstruction, non-rigid structure from motion (NRSFM) provides an approach to jointly estimate 3D object shapes and the relative camera motions from the corresponding 2D points in a sequence of images. Because of the lack of prior information on 3D shape deformation, NRSFM is still an intractable and underconstrainted problem.

In order to alleviate the uncertainty, more prior information and constraints have been gradually explored in 3D reconstruction model. As a remarkable work, a matrix factorization method was proposed in [4] to represent the unknown 3D shapes as a linear combination of a small number K of 3D shape bases. In the matrix factorization method, the decomposed motion factor and shape basis are constrained to be of a low-rank $3K$. Due to its simplicity, many works have been developed subsequently based on the low rank shape model. In [5], a closed form solution was presented by combining both the rotation constraint and the low rank constraint. In [6], a Gaussian prior is assumed for the shape coefficients, and the optimization is solved using the expectation-maximization (EM) algorithm. Considering the approximate symmetry of facial feature points, an effective depth estimation model was proposed in [7] based on the constraint independent component analysis.

In [8], a non-rigid structure from motion factorization model is proposed by solving a very small semi-definite programming and a nuclear-norm minimization problem. A reconstruction-based metric learning method was presented in [9] to learn a discriminative distance metric for unconstrained face verification. A sequential non-rigid structure from motion model was proposed in [10] by utilizing the physical priors of object's surface. When the non-rigid object has degenerate deformations, the extra degrees-of-freedom will yield spurious shape deformations due to non-negligible noise in real applications. To deal with this problem, a low-rank shape deformation model was proposed to represent 3D structures of degenerate deformations by considering both the rank-deficient nature and the low-rank property [11]. In [12], a dense NRSFM model was given as an energy-based formulation by incorporating the physical, discontinuity-preserving deformation prior.

In order decrease the unknow parameters, the 3D point trajectories are modeled compactly

as the Discrete Cosine Transform basis under the smoothing constraint [13]. Nevertheless, due to the limitation of rank 3*K*, the high-frequency deformation cannot be well modeled for the trajectory representation. In [14], a smoothly deforming 3D shape was modeled as a single point moving along a smooth time-trajectory within a linear shape space. This representation provides a better reconstruction of high-frequency deformation without relaxing the rank-3*K* constraint. A column-space-fitting (CSF) method was developed to obtain the optimized solution [15]. Simulations on multiple sequences have demonstrated that the CSF algorithm can achieve a very good estimation performance for deformed objects.

In most traditional non-rigid structure from motion (NRSFM) models, a good estimation performance was generally achieved by minimizing the overall estimation error of feature points. Because only the overall estimation error is considered, the estimation results of part feature points may deviate from their true values far away. In order to solve this problem, a local deviation-constrained based column-space-fitting approach is presented in this paper to decrease the estimation deviation. In our work, an effective model is first constructed by two terms, i.e. the overall estimation error computed by a linear subspace representation, and the constraint term based on the variance of reconstruction error for each frame. Furthermore, an Augmented Lagrange Multipliers (ALM) iterative algorithm is developed to optimize the proposed model. Because both the overall estimation error and the local deviation are utilized, a good estimation performance can be achieved for the proposed method. Moreover, the estimation errors are uniform for different feature points. In addition, the convergence analysis is carried out for the proposed algorithm.

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

A. Formulation of the LDS-CSF Model

Assume that $[x_{t,j}, y_{t,j}]^T$ $(t = 1, 2, \cdots, T, j = 1, 2, \cdots, n)$ is the 2D projection of the *j*th 3D point observed on the *t*th image, the n input 2D point tracks of T images can be represented as

a $2T \times n$ observation matrix $\mathbf{W}$, i.e.,

$$\mathbf{W} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T,1} & x_{T,2} & \cdots & x_{T,n} \\ y_{T,1} & y_{T,2} & \cdots & y_{T,n} \end{pmatrix}. \tag{1}$$

According to the linear subspace model, $\mathbf{W}$ can be factorized as:

$$\mathbf{W} = \mathbf{MS} = \underbrace{\mathbf{D}(\mathbf{C} \otimes \mathbf{I}_3)}_{\mathbf{M}} \mathbf{S} = \underbrace{\mathbf{D}(\underbrace{\mathbf{\Omega}_d \mathbf{X}}_{\mathbf{C}} \otimes \mathbf{I}_3)}_{\mathbf{M}} \mathbf{S}, \tag{2}$$

where $\mathbf{M} \in R^{2T \times 3K}$ represents the camera motion matrix, $\mathbf{S} \in R^{3K \times n}$ denotes the K 3D shape bases. The matrices $\mathbf{D} \in R^{2T \times 3T}$, $\mathbf{C} \in R^{T \times K}$ and $\mathbf{I}_3$ represent a block-diagonal rotation matrix, a shape coefficient matrix and a 3×3 identity matrix, respectively. The operation $\mathbf{C} \otimes \mathbf{I}_3$ is the Kronecker product of $\mathbf{C}$ and $\mathbf{I}_3$. The matrix $\mathbf{\Omega}_d \in R^{T \times d}$ and the unknown factor $\mathbf{X} \in R^{d \times K}$ denote the DCT basis matrix and the corresponding coefficient matrix, respectively.

For the x and y coordinates of feature points on the tth frame, the reconstruction errors $(e_{x_{t,j}}, e_{y_{t,j}})$ of the jth feature point between the true values $(x_{t,j}, y_{t,j})$ and the estimated results $(x^*_{t,j}, y^*_{t,j})$ can be given by,

$$e_{x_{t,j}} = x_{t,j} - x^*_{t,j}, \tag{3}$$

and

$$e_{y_{t,j}} = y_{t,j} - y^*_{t,j}. \tag{4}$$

respectively. Correspondingly, the mean value of the reconstruction errors for n feature points can be computed as,

$$\overline{e}_{x_t} = \frac{1}{n} \sum_{j=1}^{n} e_{x_{t,j}}, \tag{5}$$

and

$$\overline{e}_{y_t} = \frac{1}{n} \sum_{j=1}^{n} e_{y_{t,j}}. \tag{6}$$

respectively. Furthermore, for the tth frame, the standard deviations σ_{tx} and σ_{ty} of reconstruction errors can be computed as,

$$\sigma_{tx} = \left(\frac{1}{n} \sum_{j=1}^{n} (e_{x_{t,j}} - \overline{e}_{x_t})^2 \right)^{\frac{1}{2}}, \tag{7}$$

and

$$\sigma_{ty} = \left(\frac{1}{n} \sum_{j=1}^{n} (e_{y_{t,j}} - \overline{e}_{y_t})^2 \right)^{\frac{1}{2}}, \tag{8}$$

respectively. For different feature points, we can see from (7) and (8) that the estimation results are closer to the true values as a whole when σ_{tx} and σ_{ty} are smaller. Thus, σ_{tx} and σ_{ty} can be used as the indices to constraint the local deviation extent of the estimation results.

In terms of (2), the local deviation constraint-based column-space-fitting (LDS-CSF) model can be formulated as,

$$\begin{aligned} &\min \ \ \|\mathbf{W} - \mathbf{W}^*\|_F^2, \\ &\text{s.t.} \ \ \frac{1}{2T} \sum_{t=1}^{T} (\sigma_{tx}^2 + \sigma_{ty}^2) = 0 \end{aligned} \tag{9}$$

where $\mathbf{W}^* = \mathbf{MS}$.

B. Optimization Scheme of the LDS-CSF Model

For convenience, we first define some simplified notations before solving the model (9). Let $\mathbf{w}_j \in R^{2T \times 1}$ and $\mathbf{s}_j \in R^{3K \times 1}$ denote the jth column of the 2D observation matrix $\mathbf{W}$ and 3D shape bases $\mathbf{S}$, respectively. The 2D reprojection error $\mathbf{r}_j$ of the jth column can be defined as,

$$\mathbf{r}_j = \mathbf{w}_j - \mathbf{M}\mathbf{s}_j = \mathbf{w}_j - \mathbf{M}\mathbf{M}^{\dagger}\mathbf{w}_j. \tag{10}$$

Furthermore, denote

$$f_1 = \frac{1}{2} \sum_{j=1}^{n} \mathbf{r}_j^T \mathbf{r}_j, \tag{11}$$

and

$$f_2 = \frac{1}{2n} \sum_{j=1}^{n} (\mathbf{r}_j - \frac{1}{n} \sum_{j=1}^{n} \mathbf{r}_j)^2. \tag{12}$$

Then, the LDS-CSF model (9) can be rewritten as,

$$\begin{aligned} &\min f_1 \\ &\text{s.t.} \ f_2 = 0. \end{aligned} \tag{13}$$

where

$$f_1 = \frac{1}{2} \sum_{j=1}^{n} \mathbf{r}_j^T \mathbf{r}_j, \tag{14}$$

$$f_2 = \frac{1}{2n}\sum_{j=1}^{n}(\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{r}_j)^2. \tag{15}$$

Initially, the rotation matrix $\mathbf{D}$ is computed via an Euclidean upgrade method.

It can be seen from (2) that $\mathbf{\Omega}_d$ is a predefined DCT basis matrix. Once the factor $\mathbf{X}$ is given, the factor $\mathbf{M}$ can be determined. Given $\mathbf{M}$, the shape basis $\mathbf{s}_j$ can be estimated by,

$$\mathbf{s}_j = \mathbf{M}^{\dagger}\mathbf{w}_j. \tag{16}$$

It means that $\mathbf{X}$ is the only parameter to be optimized.

In terms of Augmented Lagrange Multipliers (ALM) iterative algorithm [17]–[19], the LDS-CSF model (13) can be reformulated as,

$$L(\mathbf{X}) = f_1 - \lambda f_2 + \frac{\rho}{2}f_2^2, \tag{17}$$

where $\rho > 0$ and λ are the weights of the penalty term and Lagrange multiplier, respectively.

According to the Gauss-Newton method, the first order partial derivative of L with respective to $\mathbf{X}$ can be given by,

$$\frac{\partial L}{\partial \mathbf{X}} = \frac{\partial f_1}{\partial \mathbf{X}} - \lambda\frac{\partial f_2}{\partial \mathbf{X}} + \rho f_2\frac{\partial f_2}{\partial \mathbf{X}} \tag{18}$$

Furthermore, the second order partial derivative of L with respective to $\mathbf{X}$ can be computed as,

$$\frac{\partial^2 L}{\partial \mathbf{X}^2} = \frac{\partial^2 f_1}{\partial \mathbf{X}^2} - \lambda\frac{\partial^2 f_2}{\partial \mathbf{X}^2} + \rho f_2(\frac{\partial f_2}{\partial \mathbf{X}})^2 + \rho f_2\frac{\partial^2 f_2}{\partial \mathbf{X}^2} \tag{19}$$

According to (14), we can obtain the first order partial derivative and the second order partial derivative of f_1, i.e.,

$$\partial f_1(\mathbf{X}) = \sum_{j=1}^{n}\mathbf{r}_j^T\partial\mathbf{r}_j, \tag{20}$$

$$\partial^2 f_1(\mathbf{X}) \approx \sum_{j=1}^{n}\partial\mathbf{r}_j^T\partial\mathbf{r}_j. \tag{21}$$

In terms of (15), we can obtain the first order partial derivative and the second order partial derivative of f_2, i.e.,

$$\partial f_2(\mathbf{X}) = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{r}_j)^T(\partial\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\partial\mathbf{r}_j), \tag{22}$$

$$\partial^2 f_2(\mathbf{X}) \approx \frac{1}{n}\sum_{j=1}^{n}(\partial\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\partial\mathbf{r}_j)^T(\partial\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\partial\mathbf{r}_j), \tag{23}$$

Define

$$\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{M}\mathbf{M}^{\dagger}, \tag{24}$$

(10) can be rewritten as,

$$\mathbf{r}_j = \mathbf{P}^{\perp}\mathbf{w}_j. \tag{25}$$

Then, we can obtain

$$\partial \mathbf{r}_j = -\mathbf{J}_j vec(\partial \mathbf{x}), \tag{26}$$

where Jacobian matrices is defined as [15],

$$\mathbf{J}_j = [\mathbf{s}_j^T \otimes \mathbf{P}^{\perp}(\mathbf{B} \otimes \mathbf{I}_3)]\mathbf{V}. \tag{27}$$

Furthermore, (20)-(23) can be rewritten as,

$$\partial f_1(\mathbf{X}) = -\sum_{j=1}^{n}(\mathbf{J}_j^T\mathbf{r}_j)^T vec(\partial \mathbf{X}), \tag{28}$$

$$\partial^2 f_1(\mathbf{X}) \approx \sum_{j=1}^{n}(vec(\partial \mathbf{X}))^T(\mathbf{J}_j^T\mathbf{J}_j)vec(\partial \mathbf{X}), \tag{29}$$

$$\partial f_2(\mathbf{X}) = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{r}_j)^T(-\mathbf{J}_j + \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}_j)vec(\partial \mathbf{X}), \tag{30}$$

$$\partial^2 f_2(\mathbf{X}) \approx \frac{1}{n}\sum_{j=1}^{n}(vec(\partial \mathbf{X}))^T(\mathbf{J}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}_j)^T(\mathbf{J}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}_j)vec(\partial \mathbf{X}). \tag{31}$$

Correspondingly, the gradient g and Hessian matrix $\mathbf{H}$ of L are given as follows,

$$g = g_1 - \lambda g_2 + \rho f_2 g_2, \tag{32}$$

$$\mathbf{H} = \mathbf{H}_1 - \lambda \mathbf{H}_2 + \rho g_2^T g_2 + \rho f_2 \mathbf{H}_2, \tag{33}$$

where

$$g_1 = -\sum_{j=1}^{n}\mathbf{J}_j^T\mathbf{r}_j \tag{34}$$

$$g_2 = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{r}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{r}_j)^T(-\mathbf{J}_j + \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}_j) \tag{35}$$

$$\mathbf{H}_1 = \sum_{j=1}^{n}\mathbf{J}_j^T\mathbf{J}_j \tag{36}$$

$$\mathbf{H}_2 = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{J}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}_j)^T(\mathbf{J}_j - \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}_j) \tag{37}$$

After obtaining g and $\mathbf{H}$, the vectorized form of the adjustment matrix $\triangle\mathbf{X}$ can be given by,

$$vec(\triangle\mathbf{X}) \leftarrow (\mathbf{H} + \delta\mathbf{I})^{-1}g, \tag{38}$$

where δ is the damping coefficient. After obtaining $\triangle\mathbf{X}$, $\mathbf{X}$ can be updated as follows,

$$\mathbf{X} \leftarrow \mathbf{X} - \triangle\mathbf{X}. \tag{39}$$

In the ALM algorithm, the unknown parameters λ, ρ and $\mathbf{X}$ are optimized alternately. After initiation, λ and ρ are first fixed. Then, the matrix $\mathbf{X}$ is optimized via the Gauss-Newton iteration algorithm, i.e. (38) and (39), untill

$$L(\mathbf{X}) - L(\mathbf{X} - \triangle\mathbf{X}) < \epsilon_0, \tag{40}$$

where ϵ_0 is a small enough value. Afterwards, the weight ρ of the penalty term and the Lagrange multiplier λ can be updated with the optimized $\mathbf{X}$. The pseudocode of the LDS-CSF algorithm is given in Algorithm 1.

C. Convergence analysis

The convergence of Algorithm 1 is analyzed in this subsection. From (9) and (14), we can see that f_1 in (13) is exactly the cost function of CSF [14], [15]. Therefore, we only need address the convergence of f_{2_i} and λ_i in (13) here. Referring to Theorem 2 of [17] and Theorem 3.4 of [20], the **Theorem 1** about convergence can be expressed as follows.

Theorem 1. If $\rho_i \to +\infty$ and $\sum_{i=1}^{\infty}\rho_i^{-1} = +\infty$, the local deviation constraint f_{2_i} obtained in Algorithm 1 converges to the optimal value of (9), and λ_i converges to the optimal Lagrange multiplier.

Proof. Assume that f_2^* is the optimal value of (9), and λ^* is the optimal Lagrange multiplier. According to (9), Algorithm 1 and Lemma 3.2 in [20], we can obtain that

$$\begin{cases} f_2^* = 0, \\ f_{2_i} \geq 0, \\ \lambda_{i+1} - \lambda_i = -\rho_i f_{2_{i+1}} \leq 0, \\ \lambda^* \in \rho_i f_2^*, \\ \lambda_{i+1} \in \rho_i f_{2_{i+1}}. \end{cases} \tag{42}$$

Algorithm 1 The pseudocode of the LDS-CSF algorithm.

Set $i = 0$, $\epsilon_1 = 5 \times 10^{-6}, \epsilon_2 = 10^{-6}$.
Initialize λ_0, $\rho_0 > 0$, $0 < \gamma < 1$, $\beta > 1$.
Initialize $\mathbf{X}_0$, $\mathbf{S}_0$, f_0, f_{1_0}, f_{2_0}.
repeat
    Using the Gauss-Newton algorithm to obtain the optimized $\mathbf{X}$ by fixing λ_i and ρ_i,

$$(\mathbf{X}_{i+1}, L(\mathbf{X})_{i+1}, f_{1_{i+1}}, f_{2_{i+1}}) = argmin\ L(\mathbf{X}_i, \lambda_i, \rho_i) \tag{41}$$

    Updating λ_i and ρ_i by fixing $\mathbf{X}$,
    if $f_{2_{i+1}} < \gamma f_{2_i}$ **then**
        $\lambda_{i+1} = \lambda_i - \rho_i f_{2_{i+1}}$
        $\rho_{i+1} = \rho_i$
    else
        $\lambda_{i+1} = \lambda_i$
        $\rho_{i+1} = \beta \rho_i$
    end if
    Update $i \leftarrow i + 1$
until $f_1 < \epsilon_1 \ \ \& \ \ f_2 < \epsilon_2$

If $\rho_i \to +\infty$ and $\rho_{i+1} > \rho_i$,

$$\begin{aligned}
\rho_i^{-2}\|\lambda_{i+1} - \lambda_i\|_F^2 &= \rho_i^{-2}\|\lambda^* - \lambda_i\|_F^2 - \rho_i^{-2}\|\lambda_{i+1} - \lambda^*\|_F^2 + 2\rho_i^{-2}\left\langle \lambda_{i+1} - \lambda_i, \lambda_{i+1} - \lambda^* \right\rangle \\
&= \rho_i^{-2}\|\lambda^* - \lambda_i\|_F^2 - \rho_i^{-2}\|\lambda_{i+1} - \lambda^*\|_F^2 - 2\left\langle f_{2_{i+1}}, f_{2_{i+1}} - f_2^* \right\rangle \\
&= \rho_i^{-2}\|\lambda^* - \lambda_i\|_F^2 - \rho_i^{-2}\|\lambda_{i+1} - \lambda^*\|_F^2 - 2\left\langle f_{2_{i+1}}, f_{2_{i+1}} \right\rangle \\
&\leq \rho_i^{-2}\|\lambda^* - \lambda_i\|_F^2 - \rho_i^{-2}\|\lambda_{i+1} - \lambda^*\|_F^2 \\
&\leq \rho_i^{-2}\|\lambda^* - \lambda_i\|_F^2 - \rho_{i+1}^{-2}\|\lambda^* - \lambda_{i+1}\|_F^2,
\end{aligned} \tag{43}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. By combining the third condition of (42) and $\rho_{i+1} > \rho_i$, we know that $\rho_i^{-2}\|\lambda^* - \lambda_i\|_F^2$ is non-increasing. Then, we have

$$\sum_{i=1}^{+\infty} \rho_i^{-2}\|\lambda_{i+1} - \lambda_i\|_F^2 < +\infty. \tag{44}$$

Therefore, any value of f_{2_i} is a feasible solution of (9) on account of

$$\|f_{2_i}\|_F^2 = \rho_i^{-2}\|\lambda_i - \lambda_{i-1}\|_F^2 \to 0. \tag{45}$$

Furthermore, using the same proof of Theorem 3.4 in [20], we have

$$\begin{aligned}
&\|f_{2_{i+1}} - f_2^*\|_F^2 + \rho_i^{-2}\|\lambda_{i+1} - \lambda^*\|_F^2 \\
&= \|f_{2_i} - f_2^*\|_F^2 - \|f_{2_{i+1}} - f_{2_i}\|_F^2 + 2\left\langle f_{2_{i+1}} - f_2^*, f_{2_{i+1}} - f_{2_i}\right\rangle \\
&\quad + \rho_i^{-2}(\|\lambda_i - \lambda^*\|_F^2 - \|\lambda_{i+1} - \lambda_i\|_F^2 + 2\left\langle\lambda_{i+1} - \lambda^*, \lambda_{i+1} - \lambda_i\right\rangle) \\
&= \|f_{2_i} - f_2^*\|_F^2 + \rho_i^{-2}\|\lambda_i - \lambda^*\|_F^2 - \|f_{2_{i+1}} - f_{2_i}\|_F^2 - \rho_i^{-2}\|\lambda_{i+1} - \lambda_i\|_F^2 \\
&\quad + 2\left\langle f_{2_{i+1}} - f_2^*, f_{2_{i+1}} - f_{2_i}\right\rangle + 2\rho_i^{-2}\left\langle\lambda_{i+1} - \lambda^*, \lambda_{i+1} - \lambda_i\right\rangle,
\end{aligned} \tag{46}$$

Considering the fifth condition in (42), we can get,

$$\begin{aligned}
&\left\langle f_{2_{i+1}} - f_2^*, f_{2_{i+1}} - f_{2_i}\right\rangle + \rho_i^{-2}\left\langle\lambda_{i+1} - \lambda^*, \lambda_{i+1} - \lambda_i\right\rangle \\
&= \left\langle f_{2_{i+1}} - f_2^*, f_{2_{i+1}} - f_{2_i}\right\rangle + \rho_i^{-2}\left\langle\rho_i(f_{2_{i+1}} - f_2^*), -\rho_i f_{2_{i+1}}\right\rangle \\
&= \left\langle f_{2_{i+1}} - f_2^*, f_{2_{i+1}} - f_{2_i}\right\rangle - \left\langle f_{2_{i+1}} - f_2^*, f_{2_{i+1}}\right\rangle \\
&= -\left\langle f_{2_{i+1}} - f_2^*, f_{2_i}\right\rangle \\
&= -\left\langle f_{2_{i+1}}, f_{2_i}\right\rangle \le 0.
\end{aligned} \tag{47}$$

Thus, $\|f_{2_{i+1}} - f_2^*\|_F^2 + \rho_i^{-2}\|\lambda_{i+1} - \lambda^*\|_F^2$ is non-increasing, which means that

$$\|f_{2_i} - f_2^*\|_F^2 + \rho_i^{-2}\|\lambda_i - \lambda^*\|_F^2 \to 0. \tag{48}$$

Then, we have that

$$\lim_{i\to+\infty} f_{2_i} = f_2^*, \qquad \lim_{i\to+\infty}\lambda_i = \lambda^*, \tag{49}$$

Therefore, f_{2_i} and λ_i converge to the optimal values in Algorithm 1.

III. Experiments

A. Experimental Data and Set-Up

The performance of the proposed method is evaluated on twelve widely used motion sequences. Among these data, there are eight synthetic image sequences (jaws, walking, face2, face1, stretch, pickup, yoga, drink) and four real-image sequences (dance, cubes, matrix, dinosaur). For these sequences, the corresponding number of frames (T) and the number of points tracked (n) are

TABLE I

THE NUMBERS OF FRAMES (T) AND THE NUMBERS OF POINT TRACKS (n) FOR SEVEN MOTION CAPTURE SEQUENCES.

Number	Sequence	T	n
1	jaws	240	91
2	walking	260	55
3	face2	316	40
4	face1	74	37
5	stretch	370	41
6	pickup	357	41
7	yoga	307	41
8	drink	1102	41
9	dance	264	75
10	matrix	105	30
11	cubes	200	14
12	dinosaur	231	49

listed in Table I. Note that these sequences are publicly available from [8], [14]–[16]. Figures 1 and 2 show one frame of the eight synthetic image sequences and the four real-image sequences, respectively.

In order to measure the estimation performance, the average mean-square error $err3D$ of the 3D coordinates between the estimated 3D shape ($\tilde{\mathbf{S}}$) and the true 3D shape ($\mathbf{S}$) is used as the performance index, i.e.,

$$err3D = \frac{1}{T}\sum_{t=1}^{T}\frac{\|\tilde{\mathbf{S}}_t - \mathbf{S}_t\|_F^2}{\|\mathbf{S}_t\|_F^2}. \tag{50}$$

It should be pointed out that, due to the ambiguity of the camera motion, a Procrustes alignment method is used to align the estimated 3D shape. Smaller $err3D$ means that the estimations are more accurate.

B. Comparisons to Recently Reported Results

In order to evaluate the effectiveness of the proposed method (LDS-CSF), we compare it with several existing NRSFM algorithms, including the well-known the block matrix method (denoted as BMM) [8], the rotation invariant kernel (denoted as RIK) [16], the column space

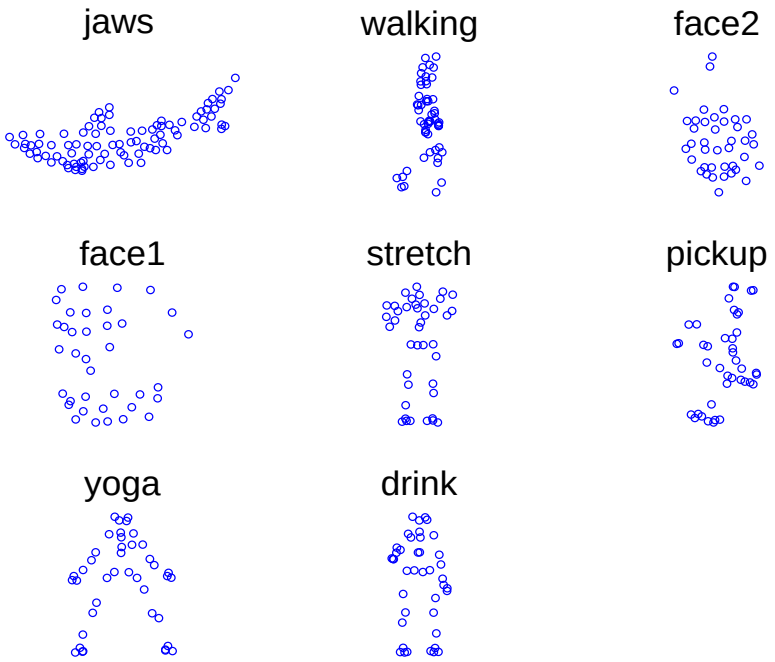

Fig. 1. One frame of the eight synthetic image sequences.

fitting method (denoted as CSF) [14] and the CSF2 method [15]. In experiments, the parameters of the various methods are set as the default values in the program package provided by the corresponding authors. Table II shows the mean and standard deviation of 3D reconstruction error of five methods for 12 sequences. In order to easily compare the performances of different algorithms, the best result and the second-best result for each sequence are highlighted in red and blue, respectively. It can be seen from Table II that the estimation errors of LDS-CSF are smaller than that of other method in general. Further, Tables III shows the corresponding decreasing percentages of LDS-CSF compared to BMM, RIK, and CSF2, respectively. From Tables II and III, we can see that LDS-CSF has a better performance than BMM, RIK, CSF and CSF2 for 12 sequence.

Taking the five frames of *face1* as an example, Fig. 3 shows the comparisons of the true values and the estimated values of the reconstructed feature points for the various methods. We can see that the feature points estimated by LDS-CSF are closer to the true values than those

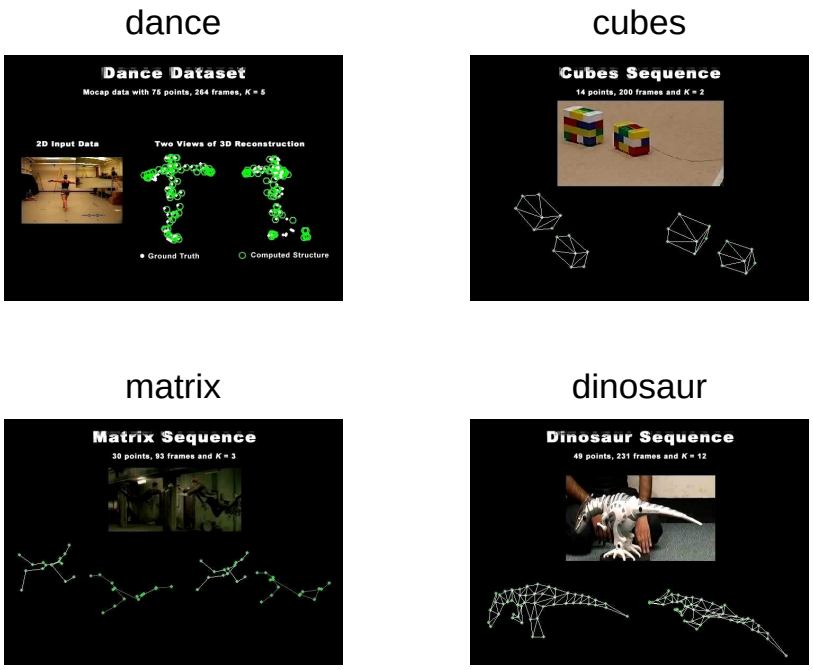

Fig. 2. One frame of the four real-image sequences.

estimated by the other four methods, which coincides with the performance indices of the 3D reconstruction errors.

Moreover, Fig. 4 shows the reconstruction error comparisons of the feature points for one frame of the sequence *face1* of the various methods. For the methods BMM, RIK, CSF, CSF2, we can see from Fig. 4 that the reconstruction errors of one section of feature points are smaller, but the reconstruction errors of the other section of feature points are larger. This indicates that the estimation results of a part of feature points deviate from their true values far away. Nevertheless, from Fig. 4, it can be seen that reconstruction errors of the LDS-CSF model are more evenly distributed than that of other methods for different feature points. This means that the proposed method can effectively decrease the local deviations.

IV. Conclusions

A local deviation-constrained based column-space-fitting approach is presented in this paper to alleviate the estimation deviation. The proposed method is demonstrated to be able to achieve

TABLE II

THE MEAN AND STANDARD DEVIATION ($\mu \pm \delta$) OF 3D RECONSTRUCTION ERROR OF 12 SEQUENCES FOR FIVE METHODS.

Sequence	BMM	RIK	CSF	CSF2	LDS-CSF
jaws	0.1456±0.0808	0.0306±0.0307	0.0048±0.0109	0.0259±0.0207	0.0096±0.0079
walking	0.0891±0.0510	0.1123±0.1015	0.1229±0.0553	0.0695±0.0423	0.0638±0.0294
face2	0.0213±0.0070	0.0262±0.0088	0.0235±0.0082	0.0209±0.0074	0.0203±0.0071
face1	0.0398±0.0155	0.0394±0.0136	0.0434±0.0208	0.0339±0.0110	0.0310±0.0087
stretch	0.0562±0.0143	0.0505±0.0160	0.0399±0.0160	0.0379±0.0135	0.0385±0.0134
pickup	0.1286±0.0929	0.1419±0.0848	0.1396±0.0868	0.1385±0.0969	0.1227±0.0772
yoga	0.0809±0.0554	0.0883±0.0551	0.0866±0.0532	0.0868±0.0540	0.0832±0.0531
drink	0.0169±0.0132	0.0163±0.0070	0.0132±0.0066	0.0131±0.0066	0.0131±0.0063
dance	0.1445±0.0636	0.1819±0.0653	0.1806±0.0651	0.1374±0.0640	0.1159±0.0412
cubes	0.4860±0.0774	0.0738±0.0344	0.1185±0.0464	0.0726±0.0319	0.0697±0.0294
matrix	0.3142±0.0683	0.3254±0.1062	0.3483±0.0863	0.3242±0.0993	0.3097±0.0926
dinosaur	0.1818±0.0534	0.2715±0.0774	0.4576±0.3115	0.4422±0.2908	0.1629±0.0756

a better estimation performance as a whole. Moreover, the local deviation constraint is verified to be effective to enhance the estimation stability of different feature points. The experimental results of the widely used synthetic image sequences and real image sequences have demonstrated the effectiveness and feasibility of the proposed algorithm.

REFERENCES

[1] L. Zou, S. Cheng, Z. Xiong, and K. R. Castleman, "3-D face recognition based on warped example faces," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 513-528, 2007.

[2] I. Mpiperis, S. Malassiotis, and M. G. Strintzis, "Bilinear models for 3-D face and facial expression recognition," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 498-511, 2008.

[3] S. Berretti, A. D. Bimbo, and P. Pala, "Sparse matching of salient facial curves for recognition of 3-D faces with missing parts," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 374-389, 2013.

[4] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3D shape from image streams," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 690-696, 2000.

TABLE III

THE 3D RECONSTRUCTION ERROR DECREASING PERCENTAGES (%) OF LDS-CSF COMPARED TO BMM, RIK AND CSF2 FOR 12 SEQUENCES.

sequence	(1-$\frac{LDS-CSF}{BMM}$)*100	(1-$\frac{LDS-CSF}{RIK}$)*100	(1-$\frac{LDS-CSF}{CSF2}$)*100
jaws	93.45	68.63	62.93
walking	28.40	43.19	8.20
face2	4.69	22.52	2.87
face1	22.11	21.32	8.55
stretch	31.49	23.76	-1.58
pickup	4.59	13.53	11.41
yoga	-2.84	5.78	4.15
drink	22.49	19.63	0
dance	19.79	36.28	15.65
cubes	85.66	7.68	4.00
matrix	1.43	4.82	4.47
dinosaur	10.40	40.00	63.16

[5] J. Xiao, J. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 233-246, 2006.

[6] L. Torresani, A. Hertzmann, and C. Bregler,"Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 878-892, 2008.

[7] Z. L. Sun, and M. K. Lam, "Depth estimation of face images based on the constrained ICA model," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 360-370, 2011.

[8] Y. Dai, H. Li, and M. He, "A simple prior-free method for non-rigid structure-from-motion factorization,"*International Journal of Computer Vision*, vol. 107, no. 2, pp. 101-122, 2014.

[9] J. W. Lu, G. Wang, W. H. Deng, and K. Jia, "Reconstruction-based metric learning for unconstrained face verification," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 79-89, 2015.

[10] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel, "Sequential non-rigid structure from motion using physical priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 979-994, 2016.

[11] Z. Zhou, F. Shi, J. Xiao, and W. Wu, "Non-rigid structure-from-motion on degenerate deformations with low-rank shape deformation model," *IEEE Transactions on Multimedia*, vol. 17, no. 2, pp. 171-185, 2015.

[12] A. M. Gallardo, T. Collins, A. Bartoli, and F. Mathias, "Dense non-rigid structure-from-motion and shading with unknown," *Proceedings of IEEE International Conference on Computer Vision*, pp. 3884-3892, 2017.

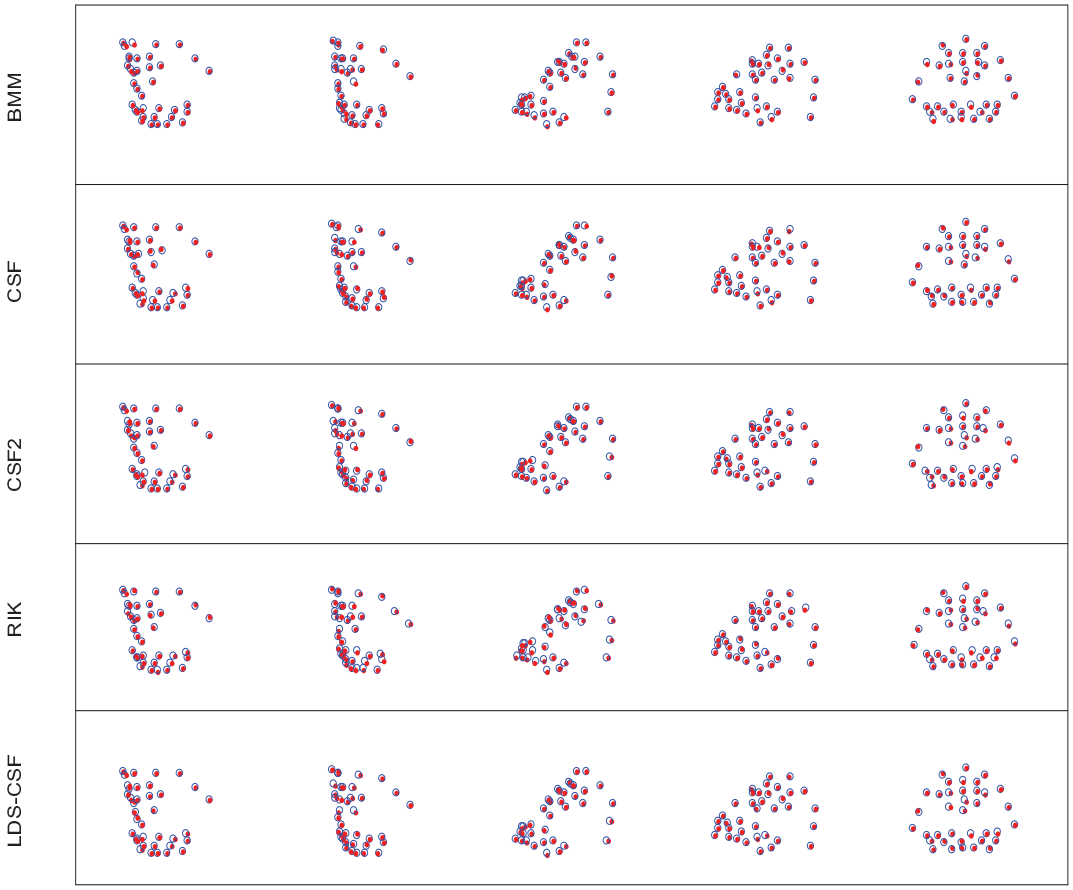

Fig. 3. The comparisons of the true values and the estimated values of the reconstructed feature points for the various methods.

[13] I. Akhter, Y. A. Sheikh, S. Khan, and T. Kanade, "Trajectory space: a dual representation for nonrigid structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1442-1456, Jul. 2011.

[14] P. F. U. Gotardo and A. M. Martinez, "Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 2051-2065, 2011.

[15] P. F. U. Gotardo and A. M. Martinez, "Non-rigid structure from motion with complementary rank-3 spaces," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3065-3072, 2011.

[16] O. C. Hamsici, P. F. U. Gotardo, and A. M. Martinez, "Learning spatially-smooth mappings in non-rigid structure from motion," *European Conference on Computer Vision*, pp. 260-273, 2012.

[17] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *UIUC Technicial Report UIUL-ENG-09-2214*, 2010.

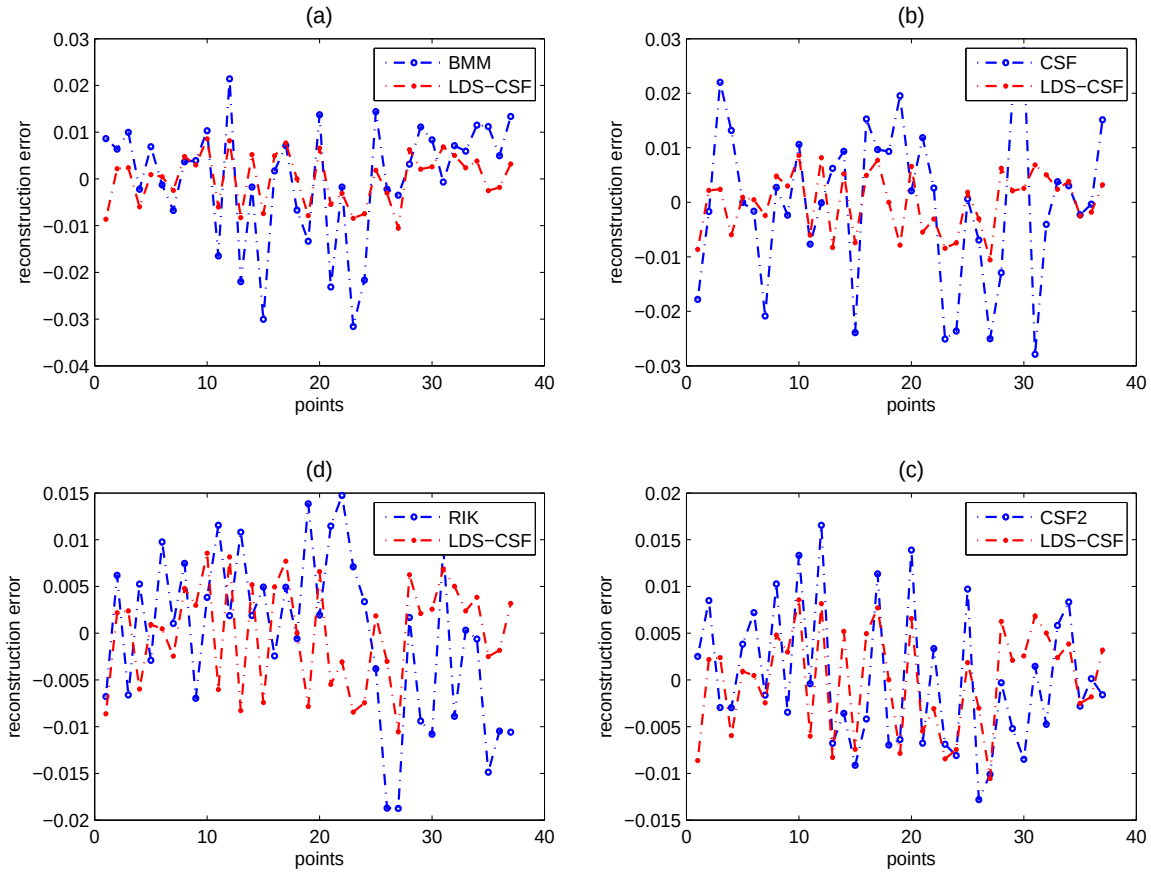

Fig. 4. The reconstruction error comparisons of the feature points for one frame of the sequence *face1* of the various methods.

[18] M. R. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303-320, 1969.

[19] D. P. Bertsekas, "Constrained optimization and Lagrange multiplier methods," *Academic press*, 2014.

[20] C. Li, C. L. Wang, and J. Wang, "Convergence analysis of the augmented Lagrange multiplier algorithm for a class of matrix compressive recovery," *Applied Mathematics Letters*, vol. 59, pp. 12-17, 2016.

PLACE PHOTO HERE

Xia Chen is now pursuing her Ph.D. degree in the School of Electrical Engineering and Automation at Anhui University. Her research interests include machine learning, and image and signal processing.

PLACE PHOTO HERE

Zhan-Li Sun received the Ph.D. degree from the University of Science and Technology of China, in 2005.

Since 2006, he has worked with The Hong Kong Polytechnic University, Nanyang Technological University, and National University of Singapore. He is currently a Professor with School of Electrical Engineering and Automation, Anhui University, China. His research interests include machine learning, and image and signal processing.

PLACE PHOTO HERE

Kin-Man Lam received the Associateship in Electronic Engineering with distinction 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, University of Sydney, Sydney, Australia, in August 1996.

From 1990 to 1993, he was a Lecturer at the Department of Electronic Engineering, The Hong Kong Polytechnic University. He joined the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, as an Assistant Professor in October 1996, became an Associate Professor in 1999, and has been a Professor since 2010. He has been a member of the organizing committee and program committee of many international conferences. Currently, Dr. Lam is a General Co-chair of the IEEE International Conference on Signal Processing, Communications, & Computing (ICSPCC2012). He is a BoG member of the Asia-Pacific Signal and Information Processing Association (APSIPA) and the Director-Student Services of the IEEE Signal Processing Society. Dr. Lam also serves as an Associate Editor of *IEEE Transactions on Image Processing*, *APSIPA Transactions on Signal and Information Processing*, and *EURASIP International Journal on Image and Video Processing*. His current research interests include human face recognition, image and video processing, and computer vision.