The following publication X. Chen, Z. -L. Sun, Z. Zeng and K. -M. Lam, "An Effective Approach for Nonrigid Structure From Motion With Complex Deformation," in IEEE MultiMedia, vol. 27, no. 3, pp. 88-98, 1 July-Sept. 2020 is available at https://doi.org/10.1109/MMUL.2020.3002072.

An Effective Approach for Non-rigid Structure From Motion with Complex Deformation

Xia Chen, Zhan-Li Sun*, *Member, IEEE,* Zhigang Zeng, *Fellow, IEEE,*
and Kin-Man Lam, *Senior Member, IEEE*

***Abstract*—In this paper, an effective approach is proposed for non-rigid structure from motion (NRSFM) to deal with the image sequence with a complex deformation. In the proposed method, a 3D structure is first divided into multiple local trajectory groups by utilizing the distance weights of the pairwise points. For each local trajectory group, a weak estimator is constructed by means of the Kernel Shape Trajectory Approach (KSTA) due to its good performance for the complex deformation. In order to improve the estimation accuracy, a Feature Vector Selection (FVS) algorithm based adaptive rank selection strategy is designed to choose the approximately optimal rank parameter for KSTA. For each 3D point, the final estimation is derived from the outputs of the weak estimators by solving a sparse optimization problem. The experimental results on several commonly used sequences demonstrate the effectiveness and feasibility of the proposed method.**

***Index Terms*—Non-rigid structure from motion, 3D image reconstruction, feature vector selection.**

I. Introduction

During the past decades, recovering 3D shape of an object from 2D image sequence is a fundamental problem in computer vision. The reconstructed 3D structure may be very helpful in many applications, such as face recognition [1], human-computer interaction [2], object detection [3], [4], etc. Non-rigid structure from motion (NRSFM) has been widely used to simultaneously recover the 3D time-varying deformed object and the relative camera motion by utilizing the 2D points of an image sequence. Although many remarkable methods have been developed to the NRSFM problem, it is still a very difficult and complex problem, because of the lack of any prior information on the 3D structure deformation.

Recently, a large majority of prior knowledge and constraints have been exploited to decrease the uncertainty of NRSFM. In [5], an unknown deformed 3D shape was assumed to lie within a linear shape space spanned by a small number (K) of 3D basis shapes. Furthermore, an object independent basis based point trajectory approach (PTA) was proposed in [6] to describe the instantaneous 3D structure in a trajectory space. The number of the unknown variables was reduced by using the Discrete Cosine Transform (DCT) as the predefined trajectory basis. However, due to the limitation of rank 3*K*, the high-frequency deformation cannot be well modeled for the trajectory representation. In [7], a Column Space Fitting (CSF) algorithm was developed to better model the 3D shape deformation with a few low-frequency components in the DCT domain. The smooth time trajectory of each point was considered as a point in a low-dimensional linear shape space. Subsequently, an improved version, named CSF2, was reported in [8] by means of the K complementary spaces of rank-3. In [9], a Kernel Shape Trajectory Approach (KSTA) was proposed to model complex and deformable 3D shapes. Furthermore, a Rotation Invariant Kernels (RIK) based method was developed to deal with the data without temporal ordering [10]. In [11], a Block Matrix Method (BMM) was proposed based on the simple low-rank prior, by solving a semi-definite programming and a trace-norm minimization problem.

In the real world, a fixed rank or smooth deformation is hard to describe the shape space exactly. A Procrustean normal distribution (PND) model based on Procrustes alignment was proposed in [12] to represent the distribution of shape deformations without any additional constraints or prior knowledge. The PND model can strictly separate the motion and the deformation components. In [13], a Procrustean Markov Process (PMP) approach was presented by incorporating a first-order Markov model and the smoothness constraint in PND.

Inspired by the ensemble learning, many weak-estimation integration methods were proposed to effectively improve the performance of NRSFM. In [14], for each point on the surface, the global signature was defined via the eigenvectors on both the bending and stretching shapes. Given the global signatures, a consensus segmentation algorithm was used to divide the points into multiple clusters. Unlike the explicitly-divided local patches, a consensus of non-rigid reconstruction (CNR) algorithm was proposed in [15] by randomly selecting a large number of small trajectory sets from the complete input sequences. All individual reconstructions were obtained via the BMM. Then, for each 3D point, a strong reconstructor was derived from the outputs of weak estimators via a sparse optimization problem.

In this paper, a KSTA based integrated algorithm is reported to deal with the problem of NRSFM with complex deformation. The main works of the proposed method can be summarized as two aspects. For each extracted local trajectory group, a weak estimator is constructed via the KSTA algorithm due to its good performance for the complex deformation. In order to adaptively find the approximately optimal rank parameter, a Feature Vector Selection (FVS) based selection strategy is devised to improve the estimation accuracy.

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

Zhigang Zeng is with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China.

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

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

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

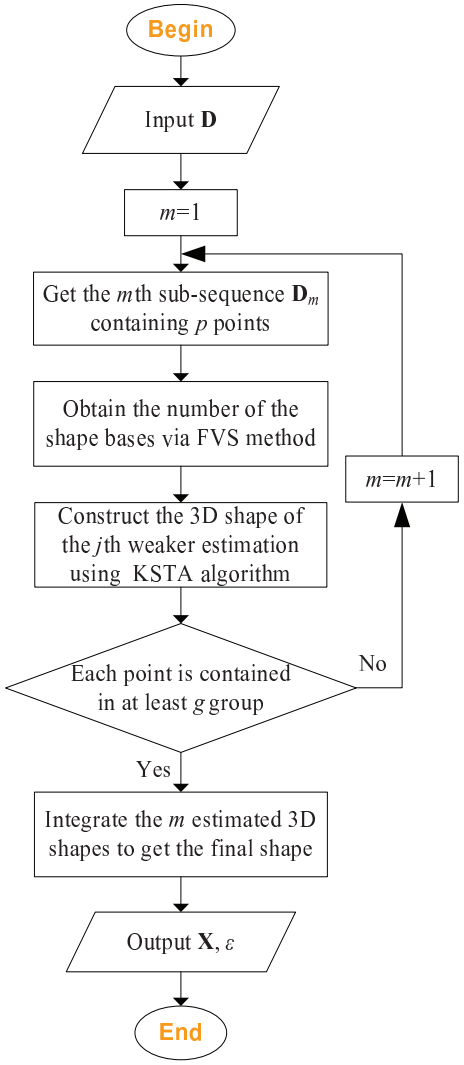

Fig. 1. Flowchart of the proposed KSTA-based integrated CNR approach for NRSFM.

Figure 1 shows the flowchart of the proposed KSTA based integrated CNR approach (CNR-KF) for NRSFM with complex deformation. There are four main steps in the proposed method: extract local trajectory groups, construct the KSTA-based weak estimator, rank parameter selection with FVS, and integration of weak estimators. A detailed description of these four steps is presented in the following subsections.

A. Local Trajectory Group Extraction

Assume that there are n input 2D points tracked over F images in a sequence. Denote $[x_{ij}, y_{ij}]^T$ $(i = 1, 2, \cdots, F; j = 1, 2, \cdots, n)$ as the 2D projection of the j^{th} 3D point in the i^{th} frame of the sequence. Then, the observation matrix $\mathbf{w}_i$ of the i^{th} frame can be represented as a $2 \times n$ matrix,

$$\mathbf{w}_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdots & x_{i,n} \\ y_{i,1} & y_{i,2} & \cdots & y_{i,n} \end{bmatrix}. \tag{1}$$

Correspondingly, the $2F \times n$ observation matrix $\mathbf{D}$ can be obtained by stacking the F frames together,

$$\mathbf{D} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{F,1} & x_{F,2} & \cdots & x_{F,n} \\ y_{F,1} & y_{F,2} & \cdots & y_{F,n} \end{bmatrix}. \tag{2}$$

For convenience, denote $\mathbf{w}_{i,j}$ as the j^{th} 2D point observed on the i^{th} image. Then, the trajectory of the j^{th} point across F frames can be represented as a $2F \times 1$ column vector, i.e.,

$$(\mathbf{w}_{1,j}, \cdots, \mathbf{w}_{i,j}, \cdots, \mathbf{w}_{F,j})^T. \tag{3}$$

In order to reduce the complexity of deformations, multiple local trajectory groups are separated from the observation matrix $\mathbf{D}$. In each trajectory group, the selected points should be close to each other. Moreover, the trajectories of p feature points over F images are similar.

Let $\mathbf{D}_m$ as the m^{th} trajectory group containing p point trajectories, and g_m as the set of points in the m^{th} group. The first trajectory in the m^{th} group is chosen based on a uniform distribution. The other $p-1$ trajectories are chosen by using the following weight w_j [15],

$$w_j = \begin{cases} \exp\left(-\frac{\lambda}{2p} \sum_{j' \in g_m} l_{jj'}\right) & \text{if} j \notin g_m, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

where λ is a predefined parameter. The symbol $l_{jj'}$ represents the pairwise Mahalanobis distance between the j^{th} point and the j'^{th} point,

$$l_{jj'} \triangleq \sum_{i=1}^{F} \|\mathbf{w}_{i,j} - \mathbf{w}_{i,j'}\|, \tag{5}$$

where $1 \le j \le n$, $1 \le j' \le p$. Note that the grouping step will continue until each point is included in at least g groups.

B. The KSTA-Based Weak Estimator

Each extracted local trajectory group $\mathbf{D}_m$ is used as the input of the KSTA algorithm [9] to estimate the 3D deformed shapes of p points. According to the nonlinear shape space model [9], $\mathbf{D}_m$ is factorized as follows,

$$\mathbf{D}_m = \mathbf{M}\hat{\mathbf{S}} = \underbrace{\mathbf{R}(\mathbf{K}_{\text{c,b}} \otimes \mathbf{I}_3)}_{\mathbf{M}} \hat{\mathbf{S}}, \tag{6}$$

where the $\mathbf{M} \in \mathbb{R}^{2F \times 3K}$ is a motion matrix, and $\hat{\mathbf{S}} \in \mathbb{R}^{3K \times p}$ is a shape basis matrix including K shape bases. The $\mathbf{R} \in \mathbb{R}^{2F \times 3F}$ and $\mathbf{I}_3 \in \mathbb{R}^{3 \times 3}$ represent a rotation matrix and a 3×3 identity matrix, respectively. The operator $\otimes$ denotes the Kronecker product. The radial basis function (RBF) kernel matrix $\mathbf{K}_{\text{c,b}} \in \mathbb{R}^{F \times K}$ is derived from [7] by using the Moore-Penrose pseudo-inverse, a kernel trick and a low-rank constraint. The $\mathbf{K}_{\text{c,b}}$ measures the similarity between a shape

matrix $\mathbf{C} \in \mathbb{R}^{F\times h}$ and a basis shape matrix $\mathbf{B} \in \mathbb{R}^{K\times d}$. The $\mathbf{C}$ describes a smooth time-trajectory within a h-dimensional shape space,

$$\mathbf{C} = \mathbf{\Omega}_d \mathbf{X}, \tag{7}$$

where $\mathbf{\Omega}_d \in \mathbb{R}^{F\times d}$ is a small number d of a predefined DCT trajectory basis. As in [7], the t_k^{th} row and the i^{th} column of $\mathbf{\Omega}_d$ ($\omega_{t_k,i}$) denotes the i^{th}-frequency cosine wave observed at the time t_k,

$$\omega_{t_k,i} = \frac{\alpha_i}{\sqrt{F}}\cos(\frac{\pi(2t_k-1)(i-1)}{2F}), \tag{8}$$

where $t_k \in [1, F], i \in [1, d]$. The $\alpha_i = 1$ for $i = 1$, otherwise $\alpha_i = \sqrt{2}$.

The $\mathbf{X} \in \mathbb{R}^{d\times h}$ is the unknown coefficient of the shape trajectory. And $\mathbf{B}$ is defined as,

$$\mathbf{B} = \begin{bmatrix} \omega(t_1)^T & \omega(t_2)^T & \cdots & \omega(t_K)^T \end{bmatrix}^T \mathbf{X}, \tag{9}$$

where $\omega(t_k)$ ($k = \{1, 2, \cdots, K\}, t_k \in [1, F]$) is the frequency cosine wave at an unknown time t_k [9]. Let $\mathbf{c}_t^T$ and $\mathbf{b}_k^T$ as the t^{th} row of $\mathbf{C}$ and the k^{th} row of $\mathbf{B}$, respectively. Then, the $(t,k)^{th}$ entry $\kappa(\mathbf{c}_t^T, \mathbf{b}_k^T)$ of the non-linear kernel function $\mathbf{K}_{\text{c,b}}$ can be represented as,

$$\kappa(\mathbf{c}_t^T, \mathbf{b}_k^T) = e^{-\gamma\left\|\mathbf{c}_t^T - \mathbf{b}_k^T\right\|_2^2}, \tag{10}$$

where $\|\cdot\|_2$ denotes the l_2 norm. The γ is an unknown scale parameter

The 3D shape bases can be obtained by $\hat{\mathbf{S}} = \mathbf{M}^\dagger \mathbf{D}_m$. Given $\hat{\mathbf{S}}$, the matrix $\mathbf{M}$ in (6) can be solved by minimizing the following residual error,

$$f(\mathbf{M}) = \left\|\mathbf{D}_m - \mathbf{M}\mathbf{M}^\dagger \mathbf{D}_m\right\|_F^2, \tag{11}$$

where $\dagger$ and $\|\cdot\|_F$ denote the Moore-Penrose pseudo-inverse and the F norm, respectively. The unknown parameters $\{\mathbf{X}, t_k, \gamma\}$ in (7), (9) and (10) can be optimized by using the Gauss-Newton algorithm [9]. Finally, the unknown 3D shape $\mathbf{S}$ can be given as,

$$\mathbf{S} = (\mathbf{K}_{\text{c,b}} \otimes \mathbf{I}_3)\mathbf{M}^\dagger \mathbf{D}_m. \tag{12}$$

Note that the initial value $\mathbf{X}_0$ of $\mathbf{X}$ is estimated by using the following smooth shape trajectories model (STA) within in a linear space [7],

$$\mathbf{D}_m = \mathbf{M}_0\hat{\mathbf{S}}_0 = \underbrace{\mathbf{R}(\mathbf{C}_0 \otimes \mathbf{I}_3)}_{\mathbf{M}_0}\hat{\mathbf{S}}_0 = \mathbf{R}(\underbrace{\mathbf{\Omega}_d\mathbf{X}_0}_{\mathbf{C}_0} \otimes \mathbf{I}_3)\hat{\mathbf{S}}_0, \tag{13}$$

where $\mathbf{M}_0 \in \mathbb{R}^{2F\times 3K_0}$ is an initialized motion matrix, and $\hat{\mathbf{S}} \in \mathbb{R}^{3K_0\times p}$ is an initialized shape basis matrix including K_0 shape bases. The $\mathbf{C}_0 \in \mathbb{R}^{F\times K_0}$ and $\mathbf{X}_0 \in \mathbb{R}^{d\times K_0}$ are an initialized shape coefficient matrix and an unknown coefficient matrix of the shape trajectory, respectively. The matrix $\mathbf{M}_0$ can be solved by minimizing the following residual error,

$$f(\mathbf{M}_0) = \left\|\mathbf{D}_m - \mathbf{M}_0\mathbf{M}_0^\dagger \mathbf{D}_m\right\|_F^2. \tag{14}$$

In (14), the only unknown parameter $\mathbf{X}_0$ can be estimated by using the Gauss-Newton algorithm [7].

C. Rank Parameter Selection with FVS

The rank parameter K, i.e., the number of shape bases, generally has a significant affect on the reconstruction performance of KSTA. Here, a FVS based rank parameter selection approach is presented to automatically select the optimal or an approximately optimal rank parameter for KSTA.

In order to produce an initial value of $\mathbf{B}$, we perform singular value decomposition (SVD) on $\mathbf{C}$,

$$\mathbf{C}\mathbf{C}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{15}$$

where $\mathbf{\Sigma} = \text{diag}\{\lambda_1; \lambda_2; ...; \lambda_F\}$. Denote σ_i ($i = 1, 2, \cdots, F$) as the i^{th} column of $\mathbf{\Sigma}$. For σ_i, a function $\mathbf{f}$ maps $\mathbf{\Sigma}$ into a Hilbert feature space H, i.e.,

$$\begin{aligned} \mathbf{f}: \ \mathbf{\Sigma} &\longrightarrow H, \\ \sigma_i &\longrightarrow \mathbf{f}(\sigma_i). \end{aligned} \tag{16}$$

Let $\mathbf{K} \in \mathbb{R}^{F\times F}$ as a Gauss kernel matrix for $\mathbf{\Sigma}$. The $(i,j)^{th}$ entry k_{ij} of $\mathbf{K}$ can be defined as follows,

$$k_{ij} = \mathbf{f}^T(\sigma_i)\mathbf{f}(\sigma_j) = \mathbf{f}_{\sigma_i}^T \mathbf{f}_{\sigma_j}. \tag{17}$$

A small number K of Features Vectors (FVs) are selected from $\mathbf{\Sigma}$ to form an approximation subspace $\mathbf{\Phi}_K$,

$$\mathbf{\Phi}_K = \{\mathbf{f}_{s_1}, \mathbf{f}_{s_2}, \cdots, \mathbf{f}_{s_K}\}, \tag{18}$$

where $S = \{s_1, s_2, \cdots, s_K\}$. The S is the initial value t_k for $\mathbf{B}$ in KSTA. This subspace $\mathbf{\Phi}_K$ is sufficient to capture the geometrical structure of the whole $\mathbf{\Sigma}$. All column vectors in $\mathbf{\Sigma}$ can be expressed as a linear combination of these selected vectors in (18),

$$\hat{\mathbf{f}}_i = \mathbf{\Phi}_K \cdot \boldsymbol{\alpha}_i, \tag{19}$$

where $\boldsymbol{\alpha}_i = [\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{iK}]^T$ is the weighting factor.

Once S is available, the weighting factor $\boldsymbol{\alpha}_i$ for σ_i can be obtained by letting the estimated mapping $\hat{\mathbf{f}}_i$ as close as possible to the real mapping $\mathbf{f}_i$,

$$\min_{\boldsymbol{\alpha}_i} \frac{\left\|\mathbf{f}_i - \hat{\mathbf{f}}_i\right\|^2}{\left\|\mathbf{f}_i\right\|^2}. \tag{20}$$

Substitute (19) into (20), and then compute the one-order partial derivative of (20) with respect to $\boldsymbol{\alpha}_i$. By setting the partial derivative as zero, the coefficient $\boldsymbol{\alpha}_i$ can be derived as follows,

$$\boldsymbol{\alpha}_i = (\mathbf{\Phi}_K^T \mathbf{\Phi}_K)^{-1}\mathbf{\Phi}_K \mathbf{f}_i. \tag{21}$$

Then, substitute (21) into (20), we can get

$$\min_{\boldsymbol{\alpha}_i} 1 - \frac{\mathbf{f}_i^T \mathbf{\Phi}_K (\mathbf{\Phi}_K^T \mathbf{\Phi}_K)^{-1}\mathbf{\Phi}_K^T \mathbf{f}_i}{\mathbf{f}_i^T \mathbf{f}_i}. \tag{22}$$

Consider all samples σ_i in $\mathbf{\Sigma}$, the data set S can be obtained by minimizing the following equation [16],

$$\min_S \left(\sum_{\sigma_i \in \mathbf{\Sigma}} \left(1 - \frac{\mathbf{f}_i^T \mathbf{\Phi}_K (\mathbf{\Phi}_K^T \mathbf{\Phi}_K)^{-1}\mathbf{\Phi}_K^T \mathbf{f}_i}{\mathbf{f}_i^T \mathbf{f}_i}\right)\right). \tag{23}$$

Let J_S and J_{S_i} define the global and local fitness of a given set S,

$$J_S = \frac{1}{F}\sum_{\sigma_i \in \mathbf{\Sigma}} J_{S_i}, \tag{24}$$

$$J_{S_i} = \frac{\mathbf{f}_i^T \mathbf{\Phi}_K (\mathbf{\Phi}_K^T \mathbf{\Phi}_K)^{-1} \mathbf{\Phi}_K^T \mathbf{f}_i}{\mathbf{f}_i^T \mathbf{f}_i}. \tag{25}$$

Then, (23) is equivalent to,

$$\max_S (J_S). \tag{26}$$

Note that the maximum value of (26) is 1, and the minimum value of (22) is 0.

The selection of K FVs is an iterative process. A kernel function is first selected to obtain the kernel matrix $\mathbf{K}$ according to (17). Then, all local fitness are computed by using (25), and set K as 1. The sample that maximizes the global fitness J_S is chosen as the first feature vector (FV) in S. Moreover, the lowest local fitness is used to search for the next FV. The new FV can reconstruct all the samples while leading to an even larger global fitness. The selected FVs should be as orthogonal as possible. The iterations are repeated until,

$$J_S < \rho \text{ or } K > \tau. \tag{27}$$

where ρ and τ are a given threshold and the maximum number of iterations, respectively. Finally, we can obtain the number of shape basis K and the initial value t_k for $\mathbf{B}$.

Similarly, the K_0 in the initialization step of KSTA is estimated based on the singular values λ_{0i} $(i = 1, ..., F)$ of $\mathbf{D}_m$. The singular values λ_{0i} are first obtained via SVD,

$$\mathbf{D}_m \mathbf{D}_m^T = \mathbf{U}_0 \mathbf{\Sigma}_0 \mathbf{V}_0^T, \tag{28}$$

where $\mathbf{\Sigma}_0 = \text{diag}\{\lambda_{01}; \lambda_{02}; ...; \lambda_{0F}\}$. Then, K_0 singular values can be selected from $\mathbf{\Sigma}_0$ via the abovementioned strategy. All feature vectors can be approximated as a linear combination of these K_0 feature vectors.

The pseudocode of the adaptive rank parameter selection algorithm for KSTA is given in Algorithm 1.

Algorithm 1 The pseudocode of the adaptive rank parameter selection algorithm for KSTA.

Input: Σ
Output: K, S

```
Set K = 0, S = { }, g = 0, ρ, τ.
Compute the Gauss kernel matrix K.
Set K = 1, and choose the sample that maximizes (26) as the first feature vector (FV). Let s_K as the index of the chosen sample, and S = {S s_K}.
repeat
    Compute (Φ_K^T Φ_K)^{-1} by the selected FV and the partition inversion method.
    Update J_{S_i} by the selected FV,
    Compute J_S by the unselected FV,
    Choose the smaple that minimize (25) as the current FV. Let s_K as the index of the chosen sample, and S = {S s_K}.
    if J_S < ρ, then
        K ← K + 1,
    end if
until J_S < ρ or K > τ.
```

D. Integration of Weak Estimators

Denote $\bar{\mathbf{z}}_i \in \mathbb{R}^{1\times n}$ as the *z*-coordinates of the strong estimator for the i^{th} sample. The translation ambiguity in $\bar{\mathbf{z}}_i$ is eliminated by enforcing the centroid of $\bar{\mathbf{z}}_i$ at origin. Then, $\bar{\mathbf{z}}_i$ can be solved by deriving the median of all local trajectory groups, which is equal to minimize the following l_1 norm [15],

$$\min_{\overline{\mathbf{z}}_i, \mathbf{t}_i} \sum_i^F \left\| \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{E} & -\mathbf{1} \otimes \mathbf{I} \\ \mathbf{1}^T & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \overline{\mathbf{z}}_i^T \\ \mathbf{t}_i \end{bmatrix} \right\|_1 \tag{29}$$

where

$$\begin{aligned} \mathbf{y}_i &= \begin{bmatrix} \mathbf{z}_{i|1} & \mathbf{z}_{i|2} & \cdots & \mathbf{z}_{i|m} & \cdots \end{bmatrix}^T, \\ \mathbf{t}_i &= \begin{bmatrix} t_{i|1} & t_{i|2} & \cdots & t_{i|m} & \cdots \end{bmatrix}^T, \\ \mathbf{E} &= \begin{bmatrix} \mathbf{E}_1^T & \mathbf{E}_2^T & \cdots & \mathbf{E}_m^T & \cdots \end{bmatrix}^T, \end{aligned} \tag{30}$$

where the elements in $\mathbf{z}_{i|m} \in \mathbb{R}^{1\times n}$ is the *z*-coordinates of the i^{th} sample if the point is belong to the m^{th} weak estimator and zero otherwise. The $t_{i|m} \in \mathbb{R}$ is the translation variable, which can eliminate the translation ambiguity in each weak estimator. The $\mathbf{1} \in \mathbb{R}^{n\times 1}$ is a vector with elements of one. The $\mathbf{E}_m \in \mathbb{R}^{n\times n}$ is a diagonal matrix of which the elements is one if its index of row or column is belong to m^{th} local trajectory group and zero otherwise. In order to solve (29) efficiently, an auxiliary variable u is introduced [15]. The unknown parameters can be obtained via the alternating direction method of multipliers (ADMM), a linear least squares and a soft thresholding.

Note that the reflection ambiguity in each groups has been solved by using an align step between the original and recovered 3D shapes in each KSTA-based weak estimator. Therefore, there is no need to use a reflection ambiguity step before integration.

III. Experiments

A. Experimental Data and Set-Up

TABLE I
THE NUMBERS OF FRAMES (F) AND THE NUMBERS OF POINT TRACKS (n) FOR FIVE COMMONLY USED SEQUENCES AND ONE DENSE SET.

Number	Sequence	F	n
1	walking	260	55
2	jaws	240	91
3	dance	264	75
4	FRGC	400	62
5	capoeira	250	41
6	pace	201	1453

The performance of the proposed method is evaluated on five benchmark data with complex deformation: *walking*, *jaws*, *dance*, *Face Recognition Grand Challenge (FRGC)*, *capoeira*, and one dense set *pace*. These sequences are publicly available from [6], [9], [17], [18]. For these sequences, the corresponding number of frames (F) and the number of points tracked (n) are listed in Table I. Figure 2 shows one frame of these six image sequences.

Moreover, five sequences (*dance*, *drink*, *pickup*, *stretch*, and *yoga*) with three rotation angles $(60°, 90°, 120°)$ are also used

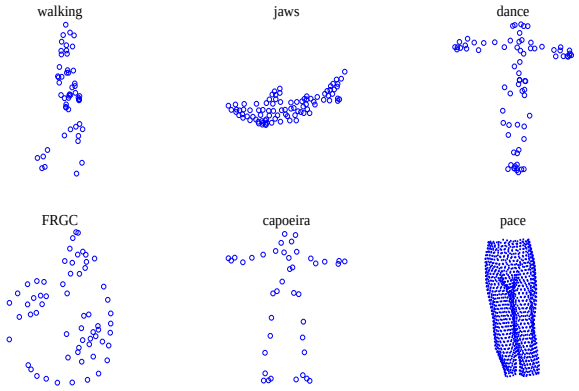

Fig. 2. One frame of six widely used motion sequences.

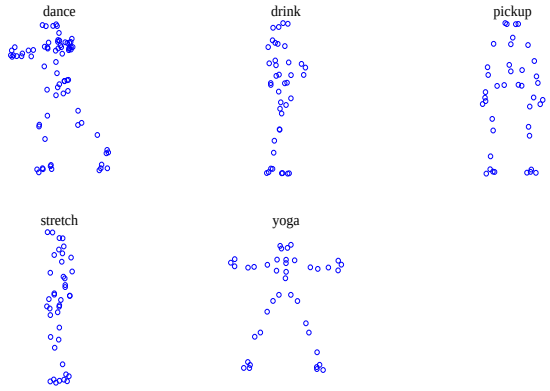

Fig. 3. One frame of the five widely used motion sequences with rotation when the degree of rotation is setting as 60.

TABLE II
THE NUMBERS OF FRAMES (F) AND THE NUMBERS OF POINT TRACKS (n) FOR FIVE MOTION CAPTURE SEQUENCES WITH VARIOUS DEGREES OF ROTATIONS.

Number	Sequence	F	n
1	dance	264	75
2	drink	1102	41
3	pickup	357	41
4	stretch	370	41
5	yoga	307	41

to test the performance to the data with some drastic rigid motion. These sequences are publicly available from [15]. Table II lists the corresponding number of frames (F) and the number of points tracked (n) for five data. Figure 3 shows one frame of the five widely used motion sequences when the rotation angle is setting as 60. All simulations were conducted using MATLAB, running on an ordinary personal computer.

In order to measure the reconstruction performance, the normalized error of the 3D coordinates is used as the performance index, i.e.,

$$\varepsilon = \frac{1}{F}\sum_{i=1}^{F}\frac{\|\mathbf{S}_i - \tilde{\mathbf{S}}_i\|_F^2}{\|\mathbf{S}_i\|_F^2}, \tag{31}$$

where $\tilde{\mathbf{S}}_i$ and $\mathbf{S}_i$ denote the estimated 3D shape and the true 3D shape, respectively. The smaller ε, the higher accuracy of the estimations.

B. Comparisons to Recently Reported Results

In order to measure the effectiveness of the proposed method (CNR-KF), we compare it with several existing NRSFM algorithms, including the well-known KSTA [9], CNR [15], the column space fitting method (denoted as CSF) [7], an improved version of CSF (denoted as CSF2) [8], a kernel-based method (denoted as RIK) [10], block matrix method (denoted as BMM) [11], the procrustean normal distribution method (denoted as PND2) [19] and the procrustean markov process method (denoted as PMP) [13].

Except for CNR, PMP, and PND2, the low rank parameter K has a significant influence on the final estimation performance. For a fair comparison, the parameter K is successively set as $\{1, 2, \cdots, 13\}$ for six methods. The parameter value corresponding to the smallest estimation error is selected as the approximate optimum parameter value of K.

Tables III shows 3D reconstruction errors ε of the nine methods for six sequences, respectively. 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. Compared to other methods, we can see from Table III that, the reconstruction errors of CNR-KF are the smallest for *walking*, *dance*, *capoeira*, and *pace* sequences. All these four data contain large deformations or complex deformations. Compared to CNR, a non-linear model is designed as a weak reconstructor. From Table III, we can see that the accuracy of 3D reconstruction using the nonlinear model is obviously higher than that of using the linear model for complex deformation. For *jaws* and *FRGC*, the performances of other eight algorithms are better than that of CNR-KF because the deformations are relative small.

From Table IV, we can see that the 3D reconstruction error ε of CNR-KF is generally lower than that of other approaches for most sequences. Note that the deformations in some sequences are not complicated. However, the most existing algorithms do not work well for some data with large rigid motion. Thus, the proposed algorithm is more robust than other algorithms for sequences with drastic rigid motion.

C. Discussion

In order to test the effectiveness of the FVS algorithm based rank selection strategy, we performed the experiments with a simplified approach, i.e. the rank selection strategy is not adopted in the proposed method, denoted as CNR-K. The parameter K is successively set as $\{2, 3, 4, 5, 6\}$ for CNR-K. Let μ denote the corresponding mean for CNR-K on each sequence over all K.

Table V and VI tabulate the 3D reconstruction errors ε of six original sequences and five sequences with different rotation angles for CNR-KF and CNR-K when K is set as different values, respectively. The reconstruction errors of CNR-KF are lower than the mean errors of CNR-K for most sequences. Thus, the FVS can effectively select an approximately optimal value of K for KSTA.

TABLE III
THE 3D RECONSTRUCTION ERROR ε OF SIX SEQUENCES FOR THE NINE METHODS.

Sequence	KSTA	RIK	CSF2	CSF	BMM	CNR	PND2	PMP	CNR-KF
walking	0.0795	0.0754	0.0695	0.1050	0.0805	0.0397	0.0407	0.0424	0.0298
jaws	0.0069	0.0306	0.0259	0.0048	0.1448	0.0832	0.0272	0.0096	0.0350
dance	0.1730	0.1211	0.1397	0.1808	0.1360	0.0720	0.1247	0.1278	0.0549
FRGC	0.1891	0.1759	0.1909	0.1909	0.1147	0.1462	0.0731	0.0727	0.1277
capoeira	0.2671	0.2745	0.3309	0.2258	0.2544	0.1708	0.3116	0.3132	0.1163
pace	0.1080	0.0910	0.1045	0.1165	0.0996	0.0728	0.0748	0.0752	0.0663

TABLE IV
THE 3D RECONSTRUCTION ERROR ε OF FIVE SEQUENCES WITH DIFFERENT ROTATION ANGLES FOR THE NINE METHODS.

Sequence	Degree	KSTA	RIK	CSF2	CSF	BMM	CNR	PND2	PMP	CNR-KF
dance	60°	0.1436	0.1045	0.1110	0.1646	0.1036	0.0797	0.1028	0.2952	0.0582
	90°	0.1641	0.1055	0.1113	0.1749	0.1027	0.0682	0.0987	0.2711	0.0555
	120°	0.1722	0.1189	0.1251	0.1731	0.0997	0.0669	0.1071	0.2401	0.0554
drink	60°	0.0532	0.0597	0.0623	0.0619	0.0377	0.0431	0.0575	0.2302	0.0357
	90°	0.0398	0.0604	0.0476	0.0518	0.0372	0.0351	0.0420	0.0821	0.0280
	120°	0.0497	0.0603	0.0600	0.0539	0.0360	0.0304	0.0411	0.1067	0.0295
pickup	60°	0.1745	0.2225	0.1945	0.2825	0.1275	0.1279	0.2084	0.5319	0.1260
	90°	0.1376	0.1920	0.1826	0.2453	0.0794	0.0901	0.1366	0.2515	0.0917
	120°	0.1513	0.1955	0.1551	0.1793	0.1068	0.0963	0.1445	0.3316	0.1055
stretch	60°	0.0997	0.1109	0.1198	0.1415	0.0925	0.0944	0.1370	0.1828	0.0761
	90°	0.0751	0.0955	0.0951	0.0734	0.0752	0.0769	0.6442	0.1299	0.0537
	120°	0.0889	0.0993	0.1096	0.0902	0.0883	0.0849	0.1002	0.1916	0.0789
yoga	60°	0.3018	0.2511	0.2578	0.3165	0.1513	0.1872	0.2205	0.3864	0.1205
	90°	0.3469	0.1725	0.1658	0.3165	0.1445	0.1187	0.2068	0.2160	0.1503
	120°	0.2022	0.2035	0.1743	0.2361	0.1158	0.1110	0.2309	0.3341	0.0963

TABLE V
THE 3D RECONSTRUCTION ERROR ε OF SIX SEQUENCES WHEN K IS SETTING SOME DIFFERENT VALUES, AND THE CORRESPONDING MEAN (μ) FOR CNR-K ON EACH SEQUENCE OVER ALL K.

Sequence	CNR-K						CNR-KF
	K=2	K=3	K=4	K=5	K=6	μ	
walking	0.0371	0.0358	0.0325	0.0411	0.0469	0.0384	0.0298
jaws	0.0916	0.0965	0.0838	0.0933	0.0758	0.0879	0.0350
dance	0.0707	0.0541	0.0504	0.0528	0.0551	0.0562	0.0549
FRGC	0.1276	0.1280	0.1284	0.1284	0.1283	0.1281	0.1277
capoeira	0.1792	0.1249	0.1204	0.1108	0.1033	0.1253	0.1153
pace	0.0715	0.0660	0.0660	0.0662	0.0661	0.0671	0.0663

TABLE VI
THE 3D RECONSTRUCTION ERROR ε OF FIVE SEQUENCES WITH DIFFERENT ROTATION ANGLES WHEN K IS SETTING AS DIFFERENT VALUES, AND THE CORRESPONDING MEAN (μ) FOR CNR-K ON EACH SEQUENCE OVER ALL K.

Sequence	Degree	CNR-K						CNR-KF
		K=2	K=3	K=4	K=5	K=6	μ	
dance	60°	0.0697	0.0571	0.0549	0.0578	0.0619	0.0601	0.0582
	90°	0.0769	0.0534	0.0484	0.0552	0.0535	0.0567	0.0555
	120°	0.0958	0.0552	0.0519	0.0568	0.0535	0.0608	0.0554
drink	60°	0.0393	0.0350	0.0329	0.0343	0.0323	0.0347	0.0357
	90°	0.0292	0.0330	0.0282	0.0240	0.0281	0.0284	0.0280
	120°	0.0298	0.0403	0.0296	0.0328	0.0339	0.0331	0.0331
pickup	60°	0.1306	0.1881	0.1292	0.1228	0.1235	0.1369	0.1260
	90°	0.2294	0.2514	0.1570	0.1243	0.1475	0.1754	0.0917
	120°	0.1602	0.1827	0.1587	0.1384	0.4062	0.1468	0.1055
stretch	60°	0.0951	0.0564	0.0622	0.0559	0.0667	0.0659	0.0761
	90°	0.0974	0.0679	0.0534	0.0479	0.0485	0.0606	0.0537
	120°	0.0811	0.0686	0.0743	0.0687	0.0740	0.0732	0.0789
yoga	60°	0.1499	0.1255	0.1179	0.1269	0.1373	0.1310	0.1205
	90°	0.1460	0.1550	0.1575	0.1766	0.1880	0.1639	0.1503
	120°	0.1081	0.0981	0.1128	0.1147	0.1237	0.1111	0.0963

IV. Conclusions

In this paper, an effective approach is proposed for non-rigid structure from motion to deal with the image sequence with a complex deformation. The strategy of rank selection in each local trajectory groups is verified to be effective to improve the estimation accuracy. The experimental results based on several widely used sequences demonstrated that the proposed method can decrease the estimation error of sequences with complex deformations, or with large rigid motion.

References

[1] X. Tao, Y. Liu, C. Jiang, Z. Wang, and X. Qin, "QoE-oriented multimedia assessment: A facial expression recognition approach," *IEEE MultiMedia*, vol. 26, no. 1, pp. 41-50, 2019.

[2] D. Kirovski, H. Malvar, and Y. Yacobi, "A dual watermark-fingerprint system," *IEEE MultiMedia*, vol. 11, no. 3, pp. 59-73, 2004.

[3] L. Kong and R. Dai, "Object-detection-based video compression for wireless surveillance systems," *IEEE MultiMedia*, vol. 24, no. 2, pp. 76-85, 2017.

[4] G. Srivastava and R. Srivastava, "Modification of gradient vector flow using directional contrast for salient object detection," *IEEE MultiMedia*, vol. 26, no. 4, pp. 7-16, 2019.

[5] 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*, 2000, pp. 690-696.

[6] 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, 2011.

[7] 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.

[8] 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*, 2011, pp. 3065-3072.

[9] P. F. U. Gotardo and A. M. Martinez, "Kernel non-rigid structure from motion," *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 802-809.

[10] 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*, 2012, pp. 260-273.

[11] 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.

[12] M. Lee, J. Cho, C. H. Choi, and S. Oh, "Procrustean normal distribution for non-rigid structure from motion", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1280-1287.

[13] M. Lee, C. H. Choi, and S. Oh, "A procrustean Markov process for non-rigid structure recovery," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1550-1557.

[14] A. Agudo and F. Moreno-Noguer, "Global model with local interpretation for dynamic shape reconstruction," *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017, pp. 264-272.

[15] M. Lee, J. Cho, and S. Oh, "Consensus of non-rigid reconstructions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4670-4678.

[16] G. Baudat and F. Anouar, "Feature vector selection and projection using kernels," *Neurocomputing*, vol. 55, no. 1-2, pp. 21-38, 2003.

[17] R. White, K. Crane, and D. Forsyth, "Capturing and animating occluded cloth," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 34, 2007.

[18] 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," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947-954.

[19] M. Lee, J. Cho, and S. Oh, "Procrustean normal distribution for non-rigid structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, pp. 1388-1400, 2017.