PCA Based Robust Motion Data Recovery

ZHUORONG LI[1], HONGCHUAN YU[2], (Member, IEEE) , HAI DANG KIEU[3], TUNG LONG VUONG[3], AND JIAN JUN ZHANG[2],

[1]School of Computer and Computing Science, Zhejiang University City College, Hangzhou, 310015 China
[2]National Centre for Computer Animation, Bournemouth University, Bournemouth, BH12 5BB United Kingdom
[3]Faculty of Information and Technology, University of Engineering and Technology, Hanoi, Vietnam

Corresponding author: H. Yu (e-mail: hyu@bournemouth.ac.uk).

ABSTRACT Human motion tracking is a prevalent technique in many fields. A common difficulty encountered in motion tracking is the corrupted data is caused by detachment of markers in 3D motion data or occlusion in 2D tracking data. Most methods for missing markers problem may quickly become ineffective when gaps exist in the trajectories of multiple markers for an extended duration. In this paper, we propose the principal component eigenspace based gap filling methods that leverage a training sample set for estimation. The proposed method is especially beneficial in the scenario of motion data with less predictable or repeated movement patterns, and that of even missing entire frames within an interval of a sequence. To highlight algorithm robustness, we perform algorithms on twenty test samples for comparison. The experimental results show that our methods are numerical stable and fast to work.

INDEX TERMS Missing marker problem, MoCap data, 2D tracking data, principle component analysis

I. INTRODUCTION

Human motion tracking has always received wide attentions due to increasing demand in many applications, including human computer interaction (HCI), automatic surveillance, biomedical application (e.g. stroke rehabilitation), virtual reality, video games, animation and movie FX productions. The tracking techniques may be categorized into two groups, i.e. marker-based tracking (e.g. Vicon multiple camera system) and markerless tracking. However, human motion tracking remains challenging due to the motion complexity and highly articulated structure of human body. One of the well-known problems is of the self-occlusion between body parts. Although multiple camera systems may provide a constrained setting with minimal occlusion, the loss of tracking markers is a frequent challenge in 3D tracking from the current MoCap devices, as well as in 2D tracking from a monocular video [1]–[5]. For example, sport exercises and dancing usually twist and rotate body resulting in movements at high degree of freedom and the absence of markers. Particularly, some markers may be absent throughout (e.g. a marker dropping off) or multiple frames may be missed in some extreme scenarios (e.g. a blackout when all markers disappear at the same time [6]). The resultant incomplete motion data compromises

the accuracy of the motion analysis [4], and therefore may require a post-processing to fill missing data just as some commercial MoCap software packages have done, such as EVaRT (Motion Analysis Corporation, Santa Rosa CA, USA), Qualisys Track Manager (Qualisys AB, Gothenburg, Sweden), or Vicon (Oxford Metrics, Limited, Oxford, England). As raw MoCap data volume is always too large, the need for improving the efficiency of such data post-processing is of substantial interest in MoCap community [4], [7].

This paper aims at the "missing marker problem" and the extension of this problem, that is, the entire time frames are missed within an interval of a sequence. It may further be applied to motion editing, such as combining two motion type pieces into one piece.

The usual gap filling methods include linear, spline interpolation and reconstructing the marker trajectory in a local segment coordinate system, which are restricted to the gaps of short duration [8]. The recent low-rank matrix completion methods [9] make use of the sparsity of motion data rather than the traditional polynormal interpolations. The rising issues include that (1) performance tends to degrade noticeably when increasing the number of gaps or enlarging the length of gaps [6]; (2) interpolation will become difficult

if multiple markers are dropping off, particularly missing whole frames within an interval of a sequence; (3) it remains challenging when the motion sequence covers multiple motion types with less predictable or repeated movement patterns (e.g. walking and running are both involved in the same motion sequence) [10]. Most of the off-the-shelf methods tend to work on ONE piece in order to avoid a large training time cost. Consequently, a challenging issue is rising, i.e. it is unlikely for these methods to robustly deal with diverse types of motion.

To tackle the above challenges, our proposed methods employ the principal component analysis (PCA) technology to motion data recovery. While using a training set for good numerical stability, it has been shown that the proposed methods do not introduce large computational cost during training. The main contributions include,

1) a principal component eigenspace based learning mechanism is presented;
2) the proposed method work well both on the scenarios of missing entries in markers' trajectories and that of missing whole frames within an interval of a motion sequence;
3) the proposed method is fast to work almost in real-time.

II. RELATED WORK
Most of the gap filling methods can be roughly categorized into three groups as follows.

A. POLYNOMIAL INTERPOLATION
Polynomial interpolation is the standard gap filling procedure that is built in many commercial marker-based motion tracking systems. Linear interpolation, spline interpolation and monotone piecewise cubic interpolation [11] are well known examples of this kind of method. They usually work well for small gaps, typically 0.2 seconds [12], whereas they are unsuitable for big gaps. Hence, many new interpolation techniques are proposed to leverage the available temporal information or spatial characteristics. For instance, missing markers in a short sequence can be estimated using Kalman filters [13]–[15]. In addition, methods based on Gaussian process dynamic models [16] or linear dynamic systems (LDS) [17] are suited for real-time applications. Hidden Markov Model (HMM) has also been employed to model human motion [4]. Moreover, some algorithms are developed based on the principal components analysis (PCA). However, they lack a training sample set. For example, Federolf [8] applied the mapping between PCA spaces to motion data interpolation. However, this approach is unsatisfactory if gaps occur in multiple markers or the movements are less predictive or cyclic. This is because the number of principle component eigenvectors cannot be changed adaptively. To tackle this issue, Gløersen et al. [10] assigned the weights to the individual gaps and omitted either some whole frames or some markers' trajectories in a data-driven way in order to reduce overfitting problem

in least squares. These efforts obtained encouraging but not perfect results, particularly motion data with less predictable or repeated movement patterns. Liu and McMillan [18] filled the missing data through the projection onto PCA eigenspace and then refined the estimations through a local linear model that is yielded by the Random Forest classifier. However, although these methods can usually deal with small gaps effectively (typically less than 0.5 second for human full-body motion), they can be inadequate and fail when applied to larger gaps [6], which is analyzed in the section IV-B.

Moreover, [6] presents multiple regression models, such as Global linear regression (GLR), Local polynomial regression (LPR), Local generalized regression neural network (LGRNN) in order to unify these methods with their scheme of Probabilistic Model Averaging (PMA). While these models can be more or less effective, they are less robust due to lack of training data.

B. LOW-RANK DECOMPOSITION TECHNIQUES
Recently, methods based on the low-rank decomposition techniques have shown promising performance [7], [9], [19], [20]. The distinct merit is to preserve the spatial-temporal characteristics embedded in natural human motion and solve the out-of-sample problem compared to many data-driven methods [21], [22]. However, the low-rank regression is suitable to approximate linear structures while human motion is nonlinear and lies on a Riemannian manifold. To approximate nonlinear structures, [7] introduced truncated nuclear norm to make subspaces simultaneously optimized. Ref. [9] further showed that the representation of motion data in a high dimensional Hilbert space was of low-rank and presented a multiple kernel learning based low-rank matrix completion method. The key point of these methods is that the observed incomplete datasets are represented in a matrix form, which assumes that the dataset is redundant enough and thus the low rank matrix completion techniques can be applied accordingly. Herein the low-rank constraint must be satisfied. However, this rarely holds when the motion data lacks repeated movement patterns [9], [23].

C. MODEL-BASED METHODS
The skeleton model-based methods have been widely studied, which usually impose the skeleton constraints to existing methods. For example, Li et al. [14] proposed a skeleton-constrained method, BoLeRO, which enforces the bone-length constraints in a linear dynamical system. Tan et al. [24] combined the skeleton constraints with Singular Value Thresholding (SVT) [25] while Peng et al. [26] applied Nonnegative Matrix Factorization (NMF) to a block-based skeleton model. Similarly, the constraint of fixing bone length is employed to drive natural looking reconstruction [9], [14], [24] through such skeleton constraints usually result in high computational complexity due to iterative optimization procedures. Moreover, such methods are usually defined for a special skeleton model based on a pre-defined

marker set. To remove pre-processing, [15] [27] presented the individual automatic methods to estimate a skeleton structure in MoCap data in order to scale the whole dataset to fit with the targeted actor.

III. METHOD

The basic idea of our proposed methods is to apply a set of training samples to missing marker problem [15]. A motion sequence is usually represented in a matrix form, i.e. $A \in R^{m \times n}$, where n denotes the number of the markers' coordinates and m denotes the frame number. Missing markers usually results in some gaps in a motion matrix. In some extrenal scenarios, it is possible that all the frames within a short interval are accidentally missed from the motion matrix. A training sample set contains K motion sequences, i.e. $\{A_i\}, i = 1, ..., K$ which may involve a variety of short motion sequences to cater for diverse motions.

A. FILLING GAPS

Consider the scenario of multiple gaps distributing over a motion matrix $B \in R^{m \times n}$. We can create a 0/1 template to indicate the locations of gaps in the incomplete matrix B and apply it to each training sample to form a set of sample pairs, i.e. complete-incomplete matrix pairs $\left\{A_i; A_i^{(0)}\right\}_{i=1}^{K}$, where the gaps are indicated by zero entries in the incomplete matrix $A_i^{(0)}$.

Consider that the training set is composed of K pieces, which are from a long motion sequence. It can be expressed in a matrix form, $A = \begin{pmatrix} A_1 \\ ... \\ A_K \end{pmatrix}$. In terms of PCA paradigm, we can apply Singular Value Decomposition (SVD) to the whole training set A and every sample $A_i^{(0)}$ respectively to span the individual eigenspaces,

$$\begin{cases} \left(A^T A\right) = U \Sigma U^T \\ \left(A_i^{(0)^T} A_i^{(0)}\right) = U_i^{(0)} \Sigma_i U_i^{(0)^T} \end{cases} \quad (1)$$

and assume that,

$$A_i U = A_i^{(0)} U_i^{(0)} T, i = 1...K \quad (2)$$

up to a transformation matrix T which facilitates the transformation from the incomplete matrix complete matrix. This is because U spans the common eigenspace of all the samples while $U_i^{(0)}$ spans the individual eigenspace for each sample. The projection of the sample A_i onto the common U may be equivalent to that of the incomplete $A_i^{(0)}$ onto the individual $U_i^{(0)}$ up to a factor. Due to singularity issue, we retain the first r eigenvectors of U and $U_i^{(0)}$ according to rank. Thus T is a $r \times r$ matrix. We can solve it in a least squares sense through K matrix equations in Eq.(2),

$$T = \left(\sum_{i=1}^{K} U_i^{(0)^T} A_i^{(0)^T} A_i U^T\right) \left(\sum_{i=1}^{K} U_i^{(0)^T} A_i^{(0)^T} A_i^{(0)} U_i^{(0)}\right)^{-1} \quad (3)$$

For an incomplete matrix B, we can yield an estimation of complete matrix of B as,

$$B^* = B U_B^{(0)} T U^T \quad (4)$$

where $U_B^{(0)}$ contains the first r eigenvectors of the incomplete B. In practice, the estimated complete matrix can be formed through filling the missing gaps of B by B^*. (The estimated complete matrix is still denoted as B^* thereafter.)

Moreover, let's consider the other scenario, i.e. all the samples are from different sources. As there are no temporal continuity in-between samples, the training set may be represented as, $A = (A_1, ..., A_K)$. Applying SVD to the whole training set A and every sample $A_i^{(0)}$ respectively yields,

$$\begin{cases} \left(AA^T\right) = V \Sigma V^T \\ \left(A_i^{(0)} A_i^{(0)^T}\right) = V_i^{(0)} \Sigma_i V_i^{(0)^T} \end{cases} \quad (5)$$

and assume that,

$$V^T A_i = F V_i^{(0)^T} A_i^{(0)}, i = 1...K \quad (6)$$

up to a transformation matrix F which facilitates the transformation from the incomplete matrix to the complete matrix. We have the same explanation on this assumption as in Eq.(2), and select the first r eigenvectors according to rank. We can solve F in a least squares sense through K matrix equations in Eq.(6),

$$F = \left(\sum_{i=1}^{K} V^T A_i A_i^{(0)^T} V_i^{(0)}\right) \left(\sum_{i=1}^{K} V_i^{(0)^T} A_i^{(0)} A_i^{(0)^T} V_i^{(0)}\right)^{-1} \quad (7)$$

To recover the complete matrix from an incomplete matrix B, an estimated complete matrix of B can be expressed as,

$$B^* = V F V_B^{(0)^T} B \quad (8)$$

where $V_B^{(0)}$ contains the first r eigenvectors of the incomplete B. Similarly, the estimated complete matrix is formed through filling the missing gaps of B by B^*.

We present two interpolation methods Eq.(4) and Eq.(8), in order to deal with two practical scenarios, one is that a training dataset is of a long motion sequence; and the other is that a training dataset consists of many short pieces that may be from different sources, in which the motion data is usually scaled to the same level in advance. We hope to point out that mathematically speaking, Eq.(4) and Eq.(8) have no essential difference except the form of input matrix A. Thus, there is no distinct difference on their performance.

B. WHOLE TIME FRAMES ARE MISSED

Consider the scenario of missing all the frames within an interval of a sequence. The incomplete motion matrix $B \in R^{m \times n}$ may be reduced as $\tilde{B} \in R^{j \times n}$ by removing the missing columns from B. The removed columns are grouped in $\check{B} \in R^{(m-j) \times n}$ that will be estimated. We can reduce all the training samples, $A_i, i = 1...K$, likewise and obtain two kinds of submatrices for each sample, i.e. reduced version $\tilde{A}_i$ and removed part $\check{A}_i$.

The training set is expressed as $A = (A_1, ..., A_K)$. Applying SVD to the training set A and each sample respectively yields,

$$\left(AA^T\right) = V\Sigma V^T \qquad (9)$$

Due to the singularity issue, we select the first r eigenvectors of V according to rank, that is, V is an $m \times r$ eigenmatrix and $m > j > r$. In terms of the row indices of the pair $(\tilde{B}, \check{B})$ in the original B, the V can be partitioned into two parts; one is the group $\tilde{V}$ with the same row indices of $\tilde{B}$ and the other is the group $\check{V}$ with the same row indices of $\check{B}$.

Projecting A_i onto V yields,

$$V^T A_i = \alpha_i \qquad (10)$$

which can be further split as,

$$\tilde{A}_i = \tilde{V}\alpha_i \qquad (11)$$

$$\check{A}_i = \check{V}\alpha_i \qquad (12)$$

where $V^T = (\tilde{V}^T, \check{V}^T)$ and $A_i = (\tilde{A}_i^T, \check{A}_i^T), i = 1...K$. They both share the same projection α_i. This is due to the following observation. Given an eigenvector matrix V and the projection α of some sample P onto V, the sample can be expressed as $P = V\alpha$. It can be noted that if dividing V into two parts, $V^T = (V_1^T, V_2^T)$, the P may be reconstructed by these two parts individually, $P = (V_1\alpha, V_2\alpha)$, in which the same projection α is shared by these two parts of V.

It can be noted that Eq.(11) provides an approximate solution for α_i in a least square sense. For all the training samples, we may apply the weighted least square method to it to further improve accuracy, that is,

$$\min_W \sum_{i=1}^{K} \left\| \left(\tilde{A}_i - \tilde{V}\alpha_i\right)^T W \left(\tilde{A}_i - \tilde{V}\alpha_i\right) \right\| \qquad (13)$$

where the weight is a $j \times j$ diagonal matrix, $W = diag(w_l), l = 1...j$. Then, we can update α_i by,

$$\alpha_i^* = \left(\tilde{V}^T W \tilde{V}\right)^{-1} \tilde{V}^T W \tilde{A}_i \qquad (14)$$

Substituting α_i^* to Eq.(12), the missing part $\check{A}_i$ can be expressed as,

$$\check{A}_i = \check{V}\left(\tilde{V}^T W \tilde{V}\right)^{-1} \tilde{V}^T W \tilde{A}_i \qquad (15)$$

Consequently, the missing rows $\check{B}$ may be estimated as,

$$\check{B} = \check{V}\left(\tilde{V}^T W \tilde{V}\right)^{-1} \tilde{V}^T \tilde{B} \qquad (16)$$

IV. EXPERIMENTS

A. EXPERIMENT SETUP

All the experiments are performed on the Southeast Asian traditional dance dataset.[1] This dataset covers a large variety of dancing actions, such as squat, sway, and stretching. Instead of cyclic motion such as walking on a treadmill or low-frequency motion like balancing in a one-leg stance [8], long sequences in this database usually involve multiple motion types, which are highly complicated and lowly predictable as shown in Fig.1. In 3D dancing datasets, every frame contains 15 markers (e.g., head, neck, shoulder etc.). Apart from 3D motion data, we perform our method on 2D tracking data as well. The 2D tracking data are obtained using the open source codes-OpenPose[2] and every frame contains 25 markers. It is natural that some markers are missed in 2D/3D motion datasets due to occlusions or marker dropping off. For robustness test, we use twenty pieces of 3D motion data as test samples which is different the current experiment setting, i.e. one test sample with multiple random masks [6]. It is not sufficient to justify algorithm robustness according to the comparison of testing algorithms on ONE sample.

The experiments focus on two scenarios, that is, there are multiple gaps over a sequence and a small set of whole time frames are missed in a sequence.

In the experiment of filling gaps, we make the gap evenly distributed over the whole motion data using a predefined 0/1 mask as the template shown in Fig.2, in which the successive gaps have the same interval in the column dimension. The main merit is that multiple test samples can be tested instead of one test sample available with different masks, which is suitable to evaluate numerical stability of algorithms.

Moreover, we also compare our methods with the state-of-the-art gap filling methods. Ref. [6] proposes the Probabilistic Model Averaging (PMA), which combines multiple methods together with probabilistic model averaging, such as weighted PCA-based method (WPCA) [10], Global Linear Regression (GLR) [6], Local Interpolation (LI) [12], Local Polynomial Regression (LPR) [6], and Local Generalized Regression Neural Network (LGRNN) [6]. The algorithms are evaluated in terms of (1) the length of the gaps, and (2) the number of the gaps.

For comparison, we use the Mean Absolute Error (MAE) [6] as the recovery error metric,

$$\epsilon = \frac{1}{g} \sum_{j=1}^{g} \frac{1}{n_{2j} - (n_{1j} + 1)} \sum_{n=n_{1j}}^{n_{2j}} \|\hat{m}_j(n) - m_j(n)\| \qquad (17)$$

where g denotes the number of gaps created, n_{1j} and n_{2j} signify the location (in frames) of the introduced gap j, $\hat{m}_j$ and m_j are respectively the recovered and the original trajectories. Additionally, our methods, Eq.(4) and Eq.(8),

---

[1] https://www.euh2020aniage.org/testthaidancedownload
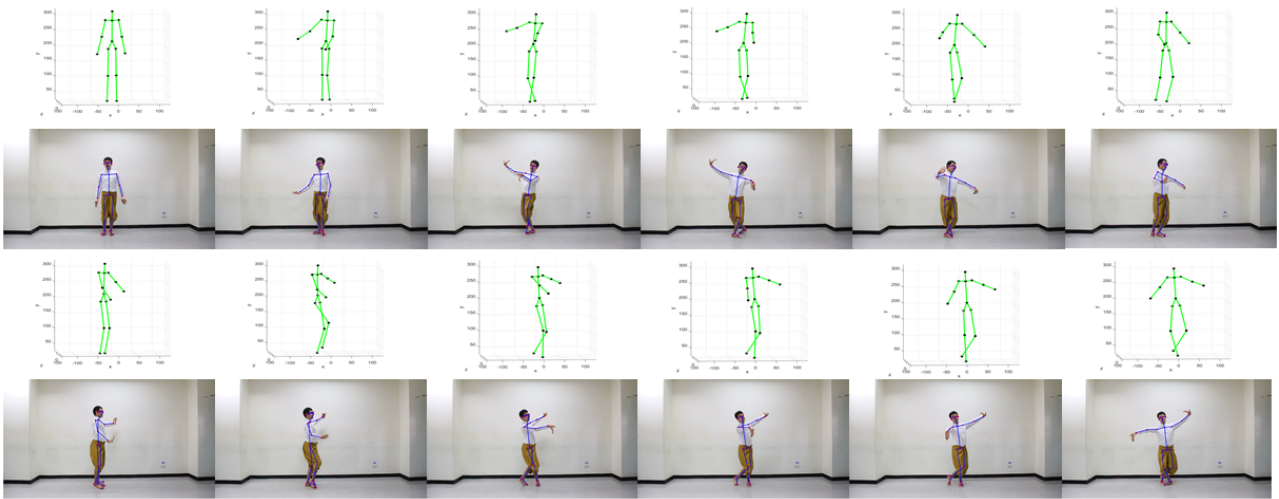[2] https://github.com/CMU-Perceptual-Computing-Lab/openpose

FIGURE 1: Examples of the 3D dancing data and the corresponding 2D tracking data of the Southeast Asian Dance Database.
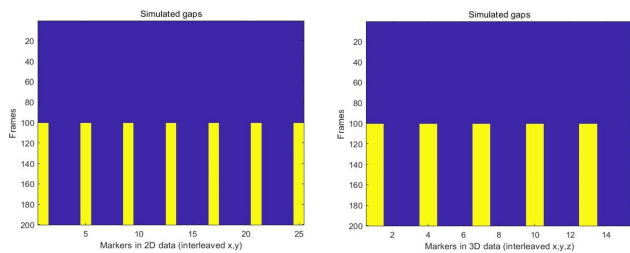

FIGURE 2: Illustration of the masks, where the yellow blocks represent the locations of gaps. Left: mask for artificially creating evenly distributed gaps on probe 2D tracking data. Right: mask for generating corrupted data on 3D motion data.

essentially have no difference since they both employ PCA technology. Thus, we only show the results by Eq.(4) in the following tests.

B. RESULTS

1) Filling gaps

The first comparison is undertaken when the length of gaps is varying. Our recovery method, Eq.(4), is performed on both the 3D motion data and 2D tracking data. The results are compared with [6] based on our 3D dancing datasets. Unlike [6], we randomly select 20 test samples from the 3D motion datasets and apply the predefined 0/1 mask in Fig.2 to these test samples. We iterate our methods and the methods mentioned in [6] on these test samples and separately average the recovery errors of each method over all the iterations.

Table 1 shows the mean recovery errors along with the different gap lengths, which are intuitively shown in Fig.3 as well. It can be noted that our method significantly outperforms the others. Although LGRNN is slightly better

than our method when the duration of gap is very short, e.g., 20 frames, it becomes gradually worse as the gap length increasing. Moreover, the performances of the methods mentioned in [6] are sensitive to the length of gaps. As PMA is the averaging of several methods, its quality of reconstruction relies on the individual methods. It is better than the worst (i.e. LPR in this test) but also inevitably weaker than the strongest (i.e. LGRNN). We hope to point out that our method takes advantage of a training dataset and is robust to deal with various test samples. In contrast, the state of the art methods [6] only exploits the current test sample for interpolation. It is natural that they perform numerically unstable under multiple sample tests.
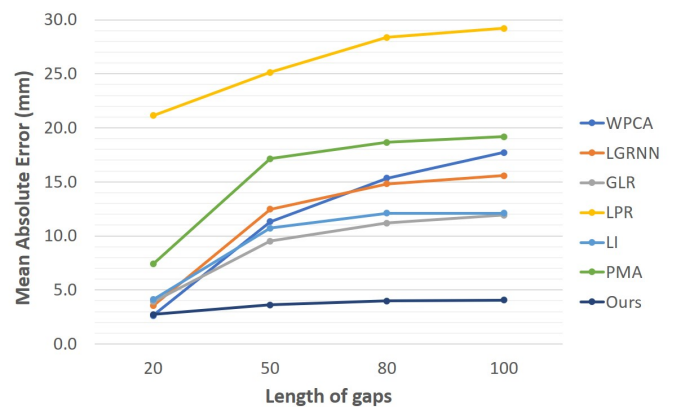

FIGURE 3: Mean recovery errors for different gap lengths. 5 evenly distributed gaps are placed on the sequences with varying gap length, i.e. the duration of 20 frames, 50 frames, 80 frames and 100 frames.

We further show the comparison of WPCA's and our method's performance in Fig.4 to illustrate the algorithm robustness. It can be noted that due to the motion complexity

TABLE 1: Mean recovery errors for different gap lengths. All values are in units of (mm).

Length of gap (frames)	20	50	80	100
WPCA	2.6125	11.3295	15.3370	17.7271
LGRNN	3.5361	12.4654	14.8164	15.5710
GLR	3.9057	9.5267	11.1776	11.9146
LPR	21.1284	25.1179	28.3632	29.1967
LI	4.1168	10.7123	12.1057	12.1030
PMA	7.4004	17.1423	18.6547	19.1778
Ours	2.7312	3.6110	3.9816	4.0601

of these 20 samples, the MAEs look visibly undulant. Compared with WPCA, our method noticeably tends to stable. Comparing with the others in [6], we can conclude the similar result.

The 2nd comparison is undertaken when the gap number is changed. We further test our method on the 20 test samples with different gap numbers, i.e. 1 gap, 2 gaps, 3 gaps, 5 gaps and 8 gaps are placed on the 3D motion data respectively, and each gap contains 100 frames. Table 2 shows the results of 3D motion data, and Table 3 shows the results of 2D tracking data. It can be noted that the average recovery error increases very slow. Even when gaps appear in every other marker, with missing rate up to 26%, the average recovery error is less than 6 mm on 3D motion data (see Table 2) and is merely 5 pixels on 2D tracking data (see Table 3). The error level is acceptable in practice. For intuitiveness, Fig.5 further shows the estimated trajectories of the missing markers in all the 20 test samples when there are 5 gaps. It can be noted that almost all the reconstructed trajectories are closely fitting the ground truth. Regardless of 2D or 3D data, our method yields relatively stable results which justifies the importance of using training data again.

TABLE 2: Average recovery error of the proposed methods on 20 pieces of 3D motion data at different missing rate. Every testing sample is a 200-frame segment of 3D motion sequences. Gaps start from the 101st frame to the end. Missing rate is the percentage of missing entry number over the total entry number in a motion matrix. All values are in units of (mm).

Number of gaps (missing rate)	1 gap (3.33%)	2 gaps (6.67%)	3 gaps (10%)	5 gaps (16.67%)	8 gaps (26.67%)
Eq.(4)	1.2831	2.9079	3.1031	3.9479	5.7161

TABLE 3: Average recovery error of the proposed methods on 2D tracking data. All values are in units of (pixels).

Number of gaps (missing rate)	2 gaps (4%)	4 gaps (8%)	5 gaps (10%)	9 gaps (18%)	13 gaps (26%)
Eq.(4)	1.5246	2.0104	2.1327	3.0697	4.6719

The above comparisons show a good performance of our method against the start-of-the-art methods in [6]. Our advantage mainly relies on the training data, that is, using PCA learning mechanism is more robust to gap length or motion complexity than the existing gap filling methods.

2) Missing whole time frames

It remains challenging to estimate the missed whole time frames. Most of the existing methods, including the state-of-the-art methods in [6], cannot work in the case of the blackout or markers dropping off. They usually need at least 3 or 4 present markers as references for reconstruction. We focus on a short blank interval in a sequence and apply our method Eq.(16) to estimate it.

The recovery error metric Eq.17 is simplified as,

$$\epsilon = \frac{1}{n_2 - (n_1 + 1)} \sum_{n=n_1}^{n_2} \|\hat{m}(n) - m(n)\| \qquad (18)$$

where $m(n)$ denotes the n-th original frame and $\hat{m}(n)$ denotes the n-th recovered frame.

We preform our method Eq.16 on both 3D motion data and 2D tracking data. To illustrate the numerical performance of our method, we plot the MAE of every single marker on 3D and 2D data along with the increasing missing frames, as shown in Fig.6 and Fig.7 separately. In addition, the average MAEs of all markers is provided in Fig.8. It can be noted that the proposed method, Eq.16, can satisfactorily recovery the blank areas with missing rate less than 20% on 3D motion data, and even up to 40% on 2D tracking data.

Moreover, to illustrate the numerical performance of Eq.16, we compare our method Eq.4 and Eq.16 as follow. We set the gap duration as 30 frames and then miss markers from the first marker to the end. Before the end marker, we run our method Eq.4 on 20 test samples, while performing our method Eq.16 in the extremal case, i.e. missing all the markers (or missing whole time frames within the gap duration). The results are plotted in Fig.9. It is reasonable that the errors increase with increasing the number of missing markers since the available spatial information is getting less and less when missing markers more and more.

Furthermore, to evaluate the quality of the estimated frames, we show the worst estimated frames, i.e., the middle frame of a blank area, in Fig.10 and Fig.11. For comparison, we also show both end frames of the blank area since these two frames have very small deviations. It can be noted that the worst estimated frames are still acceptable when the frame missing rate is less than 15% on both 2D tracking data and 3D motion data.

The PCA learning mechanism through using a training sample set in our methods is to cover diverse motion types to cater for various "missing marker problems". In general performance will be improved through enlarging training datasets. A rising issue is that a large training dataset usually results in an exponential training time. In contrast, applying a large training dataset to our algorithms does not result in a high time cost, which is shown in Table 4. Since Eq.(4) and Eq.(8) have no essential difference, Table 4 only shows the results by Eq.(4) here. However, it can be noted that the running time of our method is comparable with those of the existing methods.
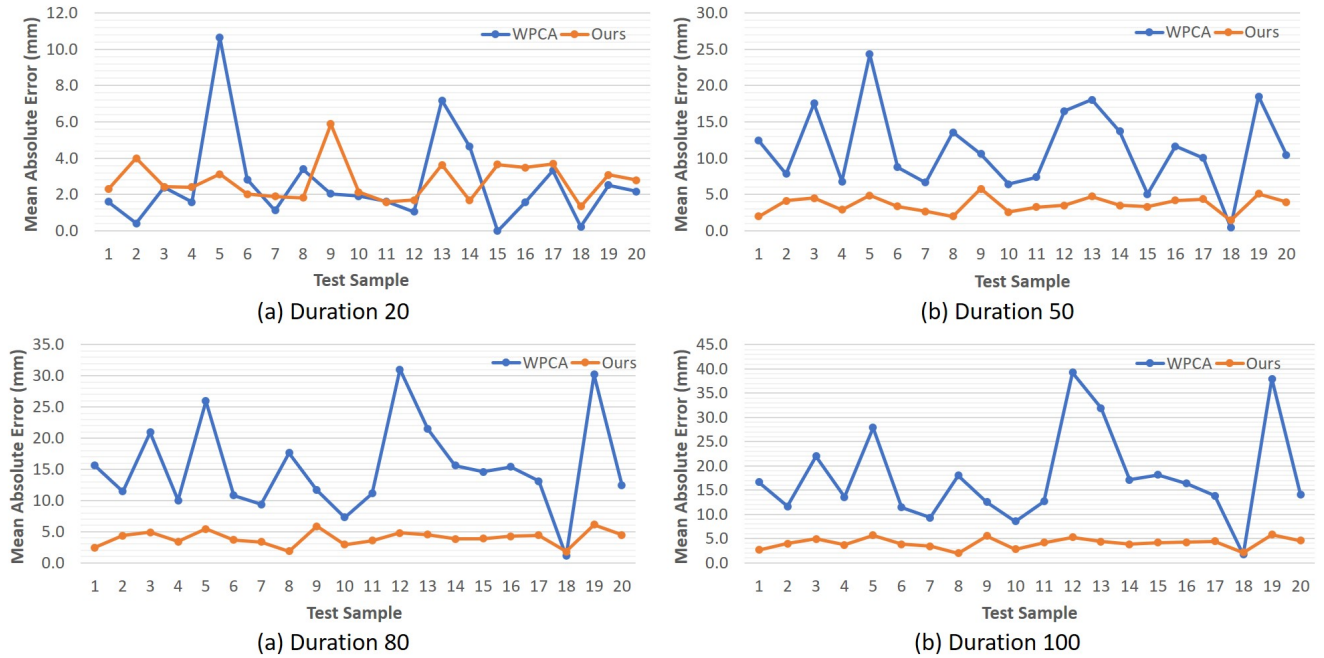
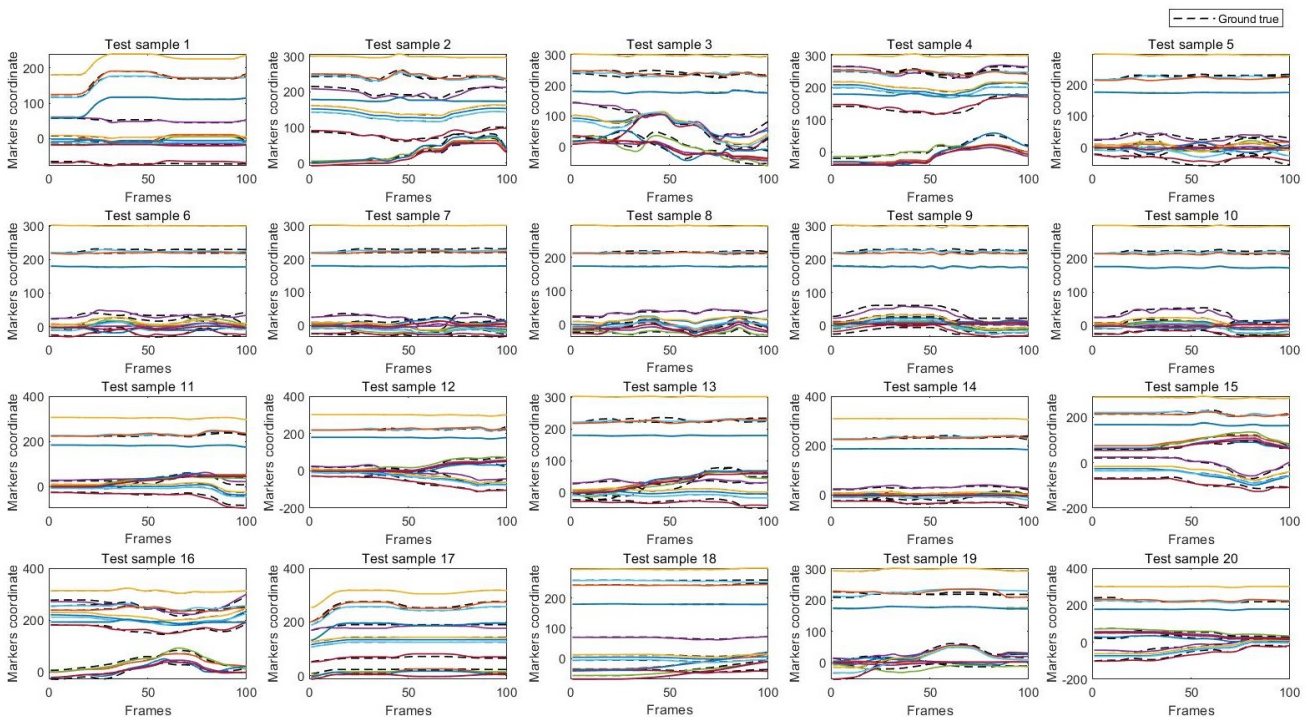FIGURE 4: Comparison of WPCA and our method's performance.

FIGURE 5: The predicted trajectories of the missing markers in all the test samples when there are 5 gaps. There are 5 missing markers and 15 trajectories of x-y-z coordinates. Dash lines denote the ground-truth trajectories and color lines denote the estimated trajectories.
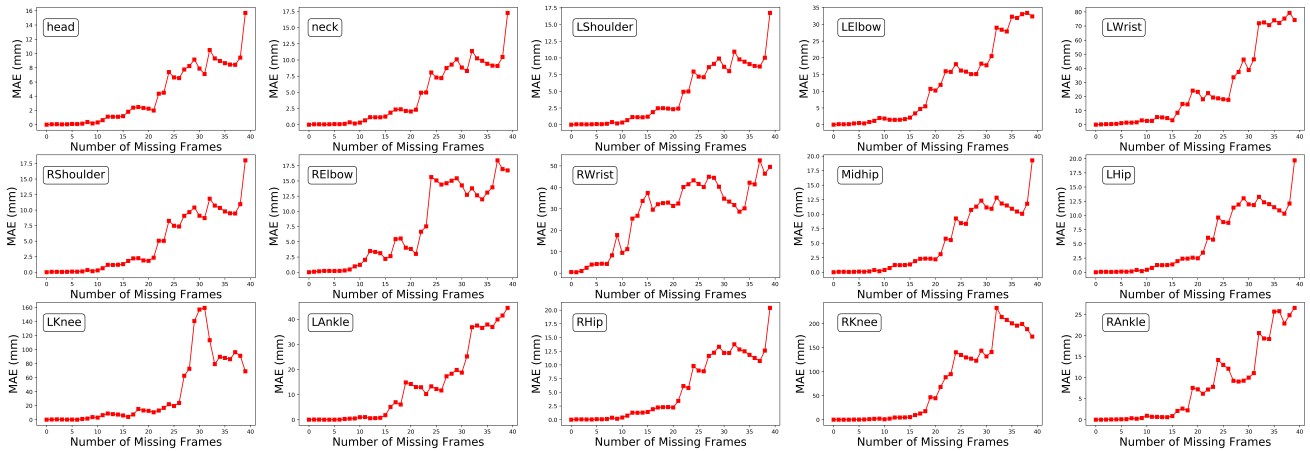
FIGURE 6: The MAE of every single joint. Method Eq.16 is tested on 3D motion data with different numbers of missing frames.
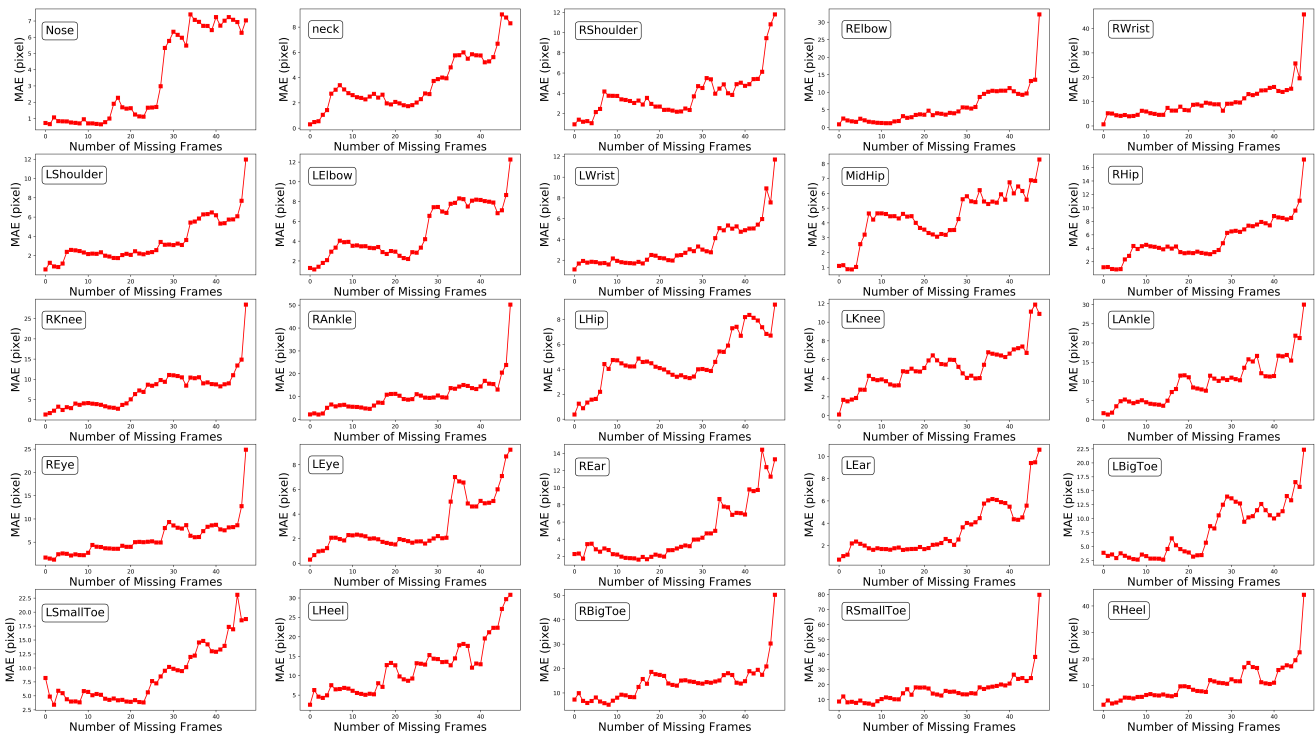

FIGURE 7: The MAE of every single joint. Method Eq.16 is tested on 2D tracking data with different numbers of missing frames.
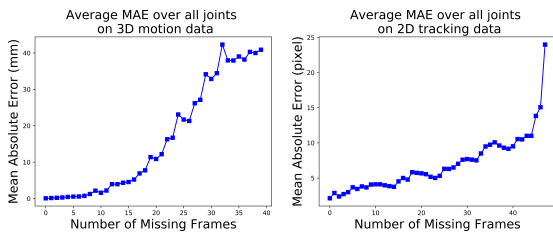
FIGURE 8: The MAE of all joints in the extreme scenario. Method Eq.16 is tested with increasing missing frames on 3D motion data (left) and 2D tracking data (right) respectively.
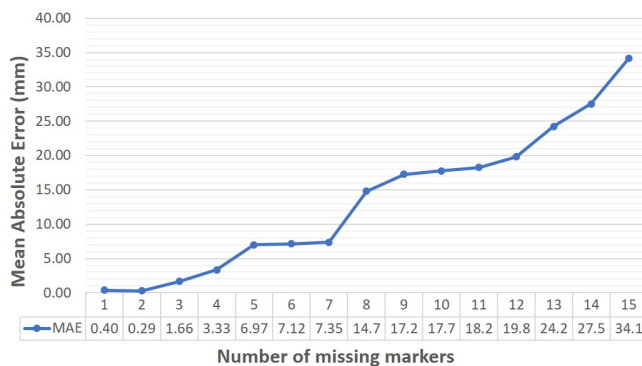

FIGURE 9: . MAEs of missing markers from 1 to all. The gap duration is set as 30 frames.

Compared Fig.6, 7 and 8 with Table 1, 2 and 3, it can be noted that the performance of Eq.(4) is noticeably better than that of Eq.(16). This is because for filling gaps, our method, Eq.(4), does not remove the whole time frames with gaps. For missing whole time frames, our method, Eq.(16), has to remove them from the motion matrix. In contrast, the existing methods [8], [10], [18] all remove the whole-time frames with gaps when filling gaps. It is inevitable that many spatial information is discarded as well. Therefore, for large gap duration, their performance is noticeable worse than our method, Eq.(4). Fig.9 further justifies that removing whole

TABLE 4: The running times.

Training data size (frames) in Eq.4	Running times (seconds) of Eq.4
500	0.1466
800	0.1511
1000	0.1777
1500	0.2281
2000	0.2575
2500	0.2732
Other methods without training data	Running times (seconds)
PCA	0.0670
GLR	0.1055
LGRNN	0.2618
LI	0.0825
LPR	0.0535
PMA	0.4522

time frames results in performance degradation as well.

Moreover, recalling our methods, Eq.(4), Eq.(8) and Eq.(16), it can be noted that the eigenspaces are pre-estimated. In essential, our methods are linear time algorithms.

V. CONCLUSION

We present PCA based simple and robust methods to solve the missing marker problems. As the PCA learning mechanism requires training sample sets, we propose three algorithms with respect to three practical scenarios, including that (1) the training set may be of a long motion sequence, (2) the dataset may be a set of short pieces from different sequences, and (3) the gap may be of a blank area of missing entire time frames in a piece. The proposed algorithms are performed on 3D motion data and 2D tracking data respectively. The experiments show that our methods are numerical stable and low time complexity, and outperform the state of the art methods. In particular, due to the PCA learning mechanism, our methods can effectively deal with the motion sequences with less predictable or repeated movement patterns or a blank area in which all the time frames are accidentally missed.

However, our methods have some limitations. For example, one marker is missed throughout the whole sequence. It is very challenging to interpolate the trajectory of the missed marker since there is no mean vector available for the missed trajectories. Moreover, adding more training samples will increase the likelihood of overtraining our models, thereby limiting the ability to generalize. We will aim at these challenges in our future work.

REFERENCES

[1] D. J. Fleet, "Motion models for people tracking," in Visual Analysis of Humans. Springer, 2011, pp. 171–198.

[2] H. Sidenbladh, M. J. Black, and D. J. Fleet, "Stochastic tracking of 3d human figures using 2d image motion," in European conference on computer vision. Springer, 2000, pp. 702–718.

[3] S. Saini, D. R. B. A. Rambli, S. B. Sulaiman, M. N. B. Zakaria, and S. Rohkmah, "Markerless multi-view human motion tracking using manifold model learning by charting," Procedia Engineering, vol. 41, pp. 664–670, 2012.

[4] C.-L. Huang and C.-Y. Chung, "A real-time model-based human motion tracking and analysis for human-computer interface systems," EURASIP Journal on Advances in Signal Processing, vol. 2004, no. 11, p. 616891, 2004.

[5] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," arXiv preprint arXiv:1812.08008, 2018.

[6] M. Tits, J. Tilmanne, and T. Dutoit, "Robust and automatic motion-capture data recovery using soft skeleton constraints and model averaging," PloS one, vol. 13, no. 7, p. e0199744, 2018.

[7] W. Hu, Z. Wang, S. Liu, X. Yang, G. Yu, and J. J. Zhang, "Motion capture data completion via truncated nuclear norm regularization," IEEE Signal Processing Letters, vol. 25, no. 2, pp. 258–262, 2017.

[8] P. A. Federolf, "A novel approach to solve the "missing marker problem" in marker-based motion analysis that exploits the segment coordination patterns in multi-limb motion data," PloS one, vol. 8, no. 10, p. e78689, 2013.

[9] G. Xia, H. Sun, B. Chen, Q. Liu, L. Feng, G. Zhang, and R. Hang, "Nonlinear low-rank matrix completion for human motion recovery," IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 3011–3024, 2018.
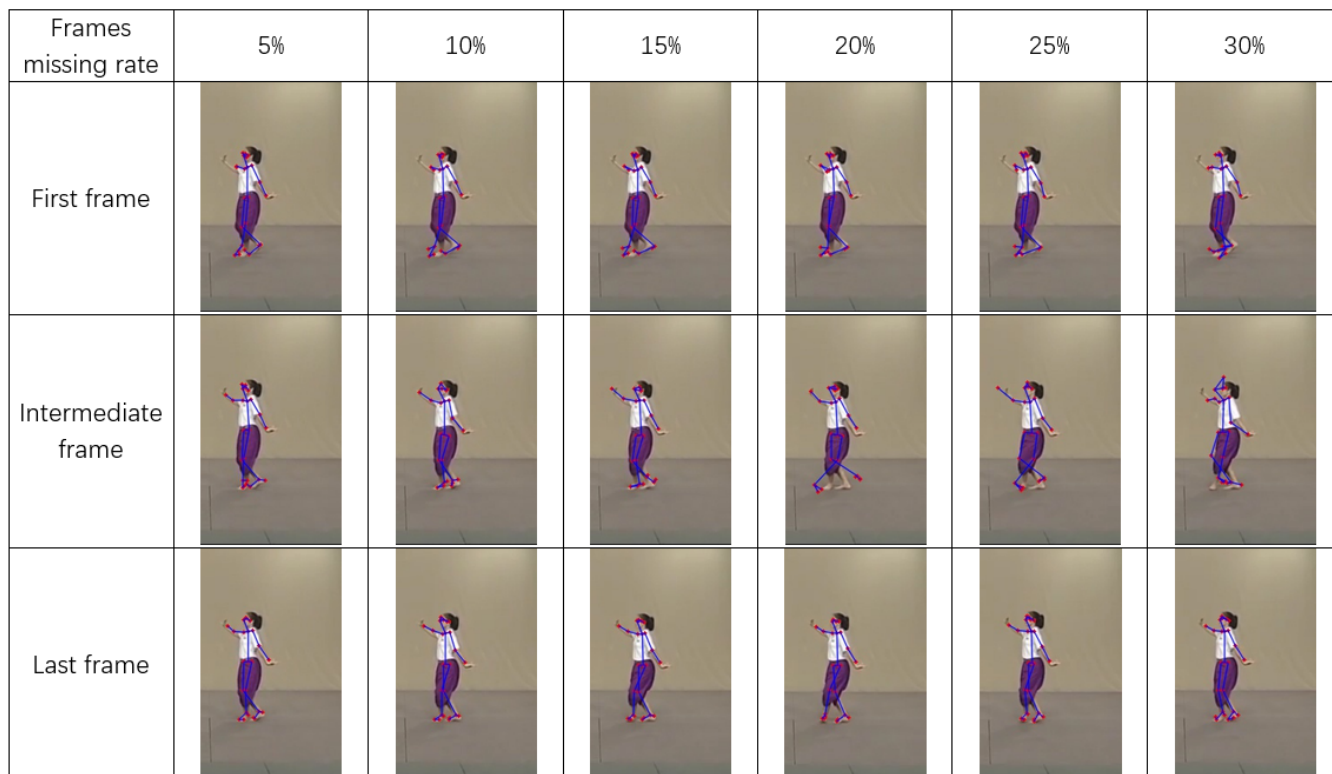
Frames missing rate	5%	10%	15%	20%	25%	30%
First frame						
Intermediate frame						
Last frame						

FIGURE 10: Rendered results of the whole frames recovery for 2D tracking data with different frame missing rates.

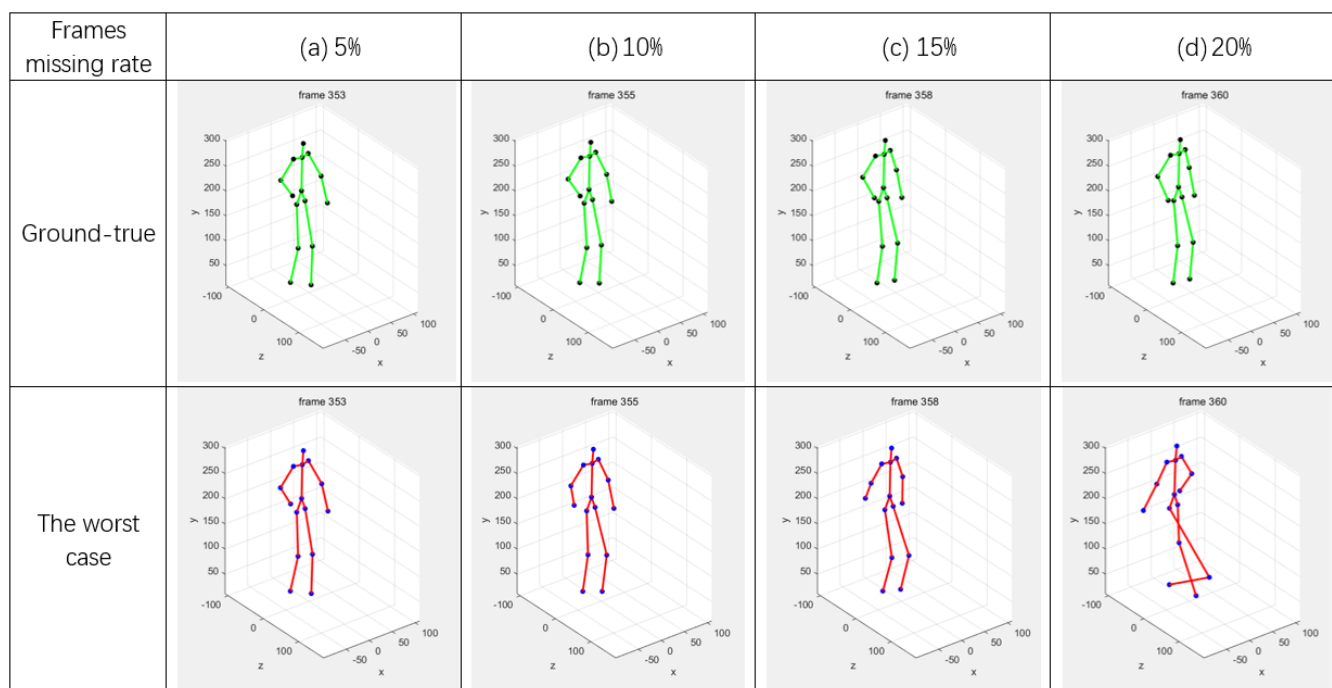| Frames missing rate	(a) 5%	(b) 10%	(c) 15%	(d) 20%
Ground-true	frame 353	frame 355	frame 358	frame 360
The worst case	frame 353	frame 355	frame 358	frame 360

FIGURE 11: The worst cases of our method on 3D motion data with increasing whole frame missing rates.

[10] Ø. Gløersen and P. Federolf, "Predicting missing marker trajectories in human motion data using marker intercorrelations," PloS one, vol. 11, no. 3, p. e0152616, 2016.

[11] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," SIAM Journal on Numerical Analysis, vol. 17, no. 2, pp. 238–246, 1980.

[12] S. J. Howarth and J. P. Callaghan, "Quantitative assessment of the accuracy for three interpolation techniques in kinematic analysis of human movement," Computer methods in biomechanics and biomedical engineering, vol. 13, no. 6, pp. 847–855, 2010.

[13] K. Dorfmüller-Ulhaas, "Robust optical user motion tracking using a kalman filter," Institut für Informatik, Tech. Rep. 2003-06, 2007.

[14] L. Li, J. McCann, N. Pollard, and C. Faloutsos, "Bolero: a principled technique for including bone length constraints in motion capture occlusion filling," in Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Eurographics Association, 2010, pp. 179–188.

[15] Q. Wu and P. Boulanger, "Real-time estimation of missing markers for reconstruction of human motion," in 2011 XIII Symposium on Virtual Reality. IEEE, 2011, pp. 161–168.

[16] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," IEEE transactions on pattern analysis and machine intelligence, vol. 30, no. 2, pp. 283–298, 2007.

[17] L. Li, J. McCann, N. S. Pollard, and C. Faloutsos, "Dynammo: Mining and summarization of coevolving sequences with missing values," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009, pp. 507–516.

[18] G. Liu and L. McMillan, "Estimation of missing markers in human motion capture," The Visual Computer, vol. 22, no. 9-11, pp. 721–728, 2006.

[19] C.-H. Tan, J. Hou, and L.-P. Chau, "Human motion capture data recovery using trajectory-based matrix completion," Electronics Letters, vol. 49, no. 12, pp. 752–754, 2013.

[20] X. Liu, Y.-m. Cheung, S.-J. Peng, Z. Cui, B. Zhong, and J.-X. Du, "Automatic motion capture data denoising via filtered subspace clustering and low rank matrix approximation," Signal Processing, vol. 105, pp. 350–362, 2014.

[21] G. Xia, H. Sun, G. Zhang, and L. Feng, "Human motion recovery jointly utilizing statistical and kinematic information," Information Sciences, vol. 339, pp. 189–205, 2016.

[22] H. Lou and J. Chai, "Example-based human motion denoising," IEEE Transactions on Visualization and Computer Graphics, vol. 16, no. 5, pp. 870–879, 2010.

[23] J. Hou, L.-P. Chau, Y. He, J. Chen, and N. Magnenat-Thalmann, "Human motion capture data recovery via trajectory-based sparse representation," in 2013 IEEE International Conference on Image Processing. IEEE, 2013, pp. 709–713.

[24] C.-H. Tan, J. Hou, and L.-P. Chau, "Motion capture data recovery using skeleton constrained singular value thresholding," The Visual Computer, vol. 31, no. 11, pp. 1521–1532, 2015.

[25] R. Y. Lai, P. C. Yuen, and K. K. Lee, "Motion capture data completion and denoising by singular value thresholding." in Eurographics (Short Papers), 2011, pp. 45–48.

[26] S.-J. Peng, G.-F. He, X. Liu, and H.-Z. Wang, "Hierarchical block-based incomplete human mocap data recovery using adaptive nonnegative matrix factorization," Computers & Graphics, vol. 49, pp. 10–23, 2015.

[27] A. G. Kirk, J. F. O'Brien, and D. A. Forsyth, "Skeletal parameter estimation from optical motion capture data," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2. IEEE, 2005, pp. 782–788.

HONGCHUAN YU is a Principal Academic of computer graphics in National Centre for Computer Animation, Bournemouth University. His specialties include geometry, graphics and image processing. He received his PhD in Computer Vision, Inst. of Intelligent Machine, Chinese Academy of Sciences, in 2000. He is a Member of IEEE and a fellow of High Education of Academy United Kingdom.

HAI DANG KIEU received the degree of Bachelor in Computer Science in 2018 at the University of Engineering and Technology (UET). 2018 - 2019, working as a researcher and teaching assistant at UET. 2019 until now, researcher at Bournemouth University follows a program sponsored by Aniage Project.

TUNG LONG VUONG received the B.S. Degree in Computer Science from VNU University of Engineering and Technology, Vietnam in 2016. From 2016 to 2019, he worked as researcher at Human-Machine laboratory, VNU University of Engineering and Technology. From 2019 to now, he currently is research fellow at Bournemouth University. His research interest includes the reinforcement learning and deep learning.

JIANJUN ZHANG is Professor of Computer Graphics at the National Centre for Computer Animation (NCCA), Bournemouth University where he leads the National Research Centre. His research focuses on a number of topics relating to 3D Computer Animation, including virtual human modelling and simulation, geometric modelling, motion synthesis, soft body deformation and physics-based animation.

Prof Zhang has published over 250 peer reviewed journal and conference publications and has given over 30 invited/keynote talks around the world. He has chaired over 30 international conferences and symposia, and serves on a number of editorial boards.

• • •

ZHUORONG LI received the PhD degree in Control Science and Engineering from Zhejiang University of Technology. She is currently a lecturer at the Zhejiang University City College. Her research interests include computational vision, deep learning and corresponding applications, machine learning.