

Human Motion Enhancement via Joint Optimization of Kinematic and Anthropometric Constraints

Le Zhou¹, Nate Lannan¹, Guoliang Fan^{1,*}, and Jerome Hausselle²

¹School of Electrical and Computer Engineering

²School of Mechanical and Aerospace Engineering

Oklahoma State University, Stillwater, Oklahoma 74078, USA

Abstract

INTRODUCTION: Recently, low cost RGB-D depth sensors have emerged as a promising alternative for motion capture for clinical gait assessment. However, depth sensor based Mocap (D-Mocap) suffers from low accuracy and poor stability for 3D joint estimation due to noise, self-occlusion, interference, and other technical limitations, which prevents it from being widely used in health-related applications.

OBJECTIVES: The primary objective of this study is to integrate non-linear Kalman filters (KFs) and evolutionary algorithms to enhance the quality of D-Mocap data by jointly considering kinematic or anthropometric constraints at the joint-level and skeleton-level, respectively.

METHODS: We propose a hybrid approach to synergistically integrate the Tobit Kalman filter (TKF) and the Differential Evolutionary (DE) algorithm for human motion enhancement that is referred to as TKF-DE. Specifically, the joint-level TKF provides the predictive distribution of each joint that is kinematically admissible in time and probabilistically amenable in space for skeleton-level DE optimization in terms of all bone lengths. Two predictive distributions of the TKF, i.e., the Gaussian and uniform, are tested and compared in terms of their effectiveness in generating the initial DE population.

RESULTS: Two sets of motion capture data are used to validate the proposed TKF-DE methods, one simulated and one real-world. The first dataset is from the Carnegie Mellon University Database (CMU) which contains a multitude of various motions and is simplified to a 21-joint skeleton and corrupted with additive white Gaussian noise (AWGN). The second dataset was collected at two labs at Oklahoma State University (OSU). The Orbbec depth sensor and the NuiTrack SDK were used for D-Mocap data acquisition along with an optical Mocap system that was time-synchronized and skeleton-matched with the D-Mocap system as a reference for evaluation. The results confirm that the proposed TKF-DE algorithms significantly outperform other nonlinear KFs, including the extended KF (EKF), the unscented KF (UKF), and the TKF alone, with improved accuracy and stability of the estimation of joint positions, bone lengths, and joint angles. It is also shown that the Gaussian-based predictive distribution is better than the uniform one, further validating the efficacy and synergy of the two key components in the TKF-DE algorithm.

CONCLUSION: Our research synergistically integrates the TKF and DE algorithms in one framework referred to as the TKF-DE, that takes advantage of kinematic and anthropometric constraints for human motion enhancement. The experimental results on two D-Mocap datasets show that the proposed TKF-DE method significantly improves the quality of D-Mocap data in terms of joint positions, bone lengths, and joint angles. This study takes one step closer to bringing the D-Mocap technology to possible health-related applications.

Received on 18 February 2021; accepted on 03 April 2021; published on 09 April 2021

Keywords: *D-Mocap*, Mocap, Depth sensor, Kalman filters, extended Kalman filters, differential evolution, joint positions, bone length, joint angles, gait analysis

Copyright © 2021 Le Zhou *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.9-4-2021.169181

*Corresponding author: Guoliang Fan, guoliang.fan@okstate.edu

1. Introduction

Since the last decade, low-cost RGB-D or depth sensors have become popular and prevalent in many computer vision applications. There are numerous works in the fields of surveillance and entertainment where depth sensors are used for human motion capture (Mocap). Particularly, in medicine and health care, depth sensors have been used for gait analysis as an assessment, diagnostic, or even as a predictive tool [1–5]. Traditionally kinematic gait measurements are obtained by optical Mocap systems, which are considered the gold standard but are impractical in a clinical setting. There is a practical need for low-cost and user-friendly Mocap tools [6]. Recent progress has been made towards the development of markerless Mocap systems using an Inertial Measurement Unit (IMU) which integrates an accelerometer and a gyroscope to measure the specific force and angular velocity of the human body segments. IMU sensors have gained popularity in rehabilitation and other motion tracking applications [6, 7]. On the other hand, as a more affordable and practical alternative, depth-based Mocap (D-Mocap) has some advantages for markerless motion capture in a normal setting [5, 8]. However, D-Mocap suffers from low accuracy due to several technical limitations: the limited imaging range [9], self-occlusions during motion [10], and possible interference [11].

The intent of D-Mocap refinement is to improve the accuracy of the human motion data so that it can be used in a broad collection of applications. This allows for the use of depth sensors, a much more convenient and affordable technology than optical Mocap. There are several computational approaches to the enhancement of D-Mocap. Kalman filters (KF) and particle filters (PF) use a recursive algorithm and use a kinematic model for human joint motion to smooth data and remove irregularities [12]. However, human motion is a nonlinear system especially in the case of D-Mocap which suffers from distortion and outliers. Researchers have explored the use of the extended KF (EKF) [13] and the unscented KF (UKF) [14] in order to compensate for the nonlinear nature of D-Mocap data. However, these nonlinear methods are still limited to handle the non-Gaussian and biased noise inherent in the D-Mocap data. The Tobit KF (TKF) that involves an observation model for censored data [15] was applied to D-Mocap with noticeable improvement in [16]. In addition, learning-based methods like evolutionary algorithms (EA) can be used to search for the candidate with the best fitness in an iterative process. Some examples of learning methods in D-Mocap enhancement are, Differential Evolution (DE) used in [17] and Genetic

Algorithms (GA) used in [18], and in the case of deep learning-based methods, the improvement of D-Mocap data through a learned motion manifold in [19]. There have been a few attempt to integrate kinematic and anthropometric constraints with the purpose of improving joint trajectories over time and the skeleton structure in space [17, 18, 20].

Our work aims to develop a new hybrid method that integrates TKF-based filtering and DE-based learning into a computational flow, namely TKF-DE. Specifically, our approach provides a harmonious way to incorporate the kinematic and anthropometric constraints for D-Mocap enhancement. The TKF takes advantage of a kinematic model in the dynamic system and it provides a probabilistic predictive distribution for population initialization in the DE. Then the DE is applied to reduce the bone-length variation in space and makes bone lengths anthropometrically consistent. To demonstrate the validity of our approaches, several nonlinear Kalman filtering methods (EKF, UKF and TKF) are applied to the same D-Mocap data for quantitative evaluation and comparisons. In order to have a more comprehensive comparison, we implement a method which combines TKF and DE directly, which we refer to TKF+DE, which is similar in spirit to the weighted average method used in [18]. In order to illustrate the robustness of our proposed method, two datasets were used in our experimentation. The first being simulated simulated D-Mocap data created from the Carnegie Mellon University Mocap data [21] and second, real-world D-Mocap data collected by two laboratories at Oklahoma State University (OSU) which were collected side-by-side with an optical Mocap system to serve as the ground-truth reference.

2. Related Work

We briefly review the related work on improving D-Mocap data from three methodological perspectives, *filter-based*, *learning-based* and *hybrid* approaches, followed by a short summary about how the proposed approach is different from previous ones.

In filter-based approaches, researchers provide a B-spline wavelet-based filtering method in [22] to remove noise from noisy motion data as well as improve the smoothness of the motion. With regard to other approaches, the KF is widely used in human motion capture and analysis for denoising [12]. To achieve higher accuracy in the D-Mocap data, in [23] five Kinect sensors are established around the tracking target, that cover 180-degree range of its perimeter, forming an average angle of approximately 45 degrees between two adjacent Kinect sensors. In addition, non-linear KFs provide another idea to solve the low-accuracy problems in D-Mocap. In [13], Researchers use the sound localization capabilities of the Kinect sensor and

joints extracted from the SDK to obtain two sets of joint data of the subject's head joint respectively, where an EKF is applied for smoothing and correcting the noise data. Furthermore, the application of the TKF is explored in the human motion enhancement, in [16], where a the TKF was applied to reduce the noise and find the censored data which is caused by occlusions. Compared to traditional KFs, the TKF can achieve higher accuracy.

Learning-based approaches have also shown promise in human motion enhancement of D-Mocap data. For instance, a deep recurrent neural network was used in [24] trained on D-Mocap data obtained from the Microsoft Kinect skeleton tracker. Different networks are used to refine the velocity and position separately and get two sets of results. This paper explores three methods of integrating the results of the two networks to improve the joint positions of D-Mocap data. An autoencoder-based deep learning framework was proposed in [19] and [25] for human motion enhancement, synthesis, and style transfer. The key component is the motion manifold represented in a latent space learned from the high-quality and diverse CMU Mocap dataset [21] which can be used to improve the quality of D-Mocap by projecting the erroneous or erratic data onto the latent space. In [26], researchers used a perceptual-based 3D skeleton motion data refinement network (BRA-P) to improve the refined motion data and further suppresses the bone-length variation as well as smooth the joint trajectories.

Hybrid approaches combine both learning-based and filtering-based ideas. In [24], researchers used the KF as one method to integrate the results of two separate neural networks. The recurrent neural networks (RNN) were trained separately on joint positions and joint velocities, and the results were used as the measurement model and process model for a Kalman filter. In [17], a multiple objective GA algorithm is embedded into PF to support importance sampling with the aim to improve accuracy of both bone-lengths as well as joint position. A constraint KF method was developed in [18] which combine DE and the KF to refine the bone-lengths fluctuations with an attempt to improve the smoothness of joint trajectories. In addition, in [20], traditional filtering is used to create a target for optimization over the latent space of a convolutional autoencoder. The authors use the KF and deep learning to recover corrupted Mocap or low quality D-Mocap data.

Inspired by previous studies, our approach is unique in several ways. First, we incorporate the probabilistic prediction of the TKF and the iterative correction of DE jointly in a synergistic way in our proposed method. Specifically, two different predictive distributions based on the TKF outcome are tested to show their effectiveness and robustness. Furthermore,

various approaches are employed for comparison to show our proposed method outperforms some filtering methods and hybrids methods mentioned in this paper by taking advantage of kinematic and anthropometric constraints from two techniques. Second, we explore the process for quantifying estimation of the D-Mocap data with regard to joint trajectories and bone-lengths. To do so, we collected both optical Mocap and D-Mocap data simultaneously. With the implementation of some software tools to overcome the mismatch problem between Mocap marker and D-Mocap joints, the comparison of quantitative results becomes more accurate and reliable. Third, we investigate the several lower body joint angles of both legs including hip angles and knee angles which are critical in gait assessment for rehabilitation [27, 28]. We are able to improve the error of joint angles from $8^\circ \sim 10^\circ$ to $4^\circ \sim 5^\circ$. Our long term goal is to achieve accuracy comparable to that of the IMU technology, the gold standard in clinical applications, where the angle errors range from $1.4^\circ \sim 4.3^\circ$ in rehabilitation applications [6].

3. Methodology

3.1. Kalman Filters for human kinematics

The Kalman filter (KF) is one of the most popular algorithms applied in human motion capture applications, such as human motion synthesis [29] and real-time human motion tracking. The system modeling involves a dynamical model and a Gaussian noise model for denoising where the observation value of the current state is corrected by prediction of previous state depend on a recursive formula [30]. In KF, the state transition model F will predict the state value at each time step k based on the previous instant. The state transition model will be contaminated by the process noise w_k , described by:

$$X_k = FX_{k-1} + w_k, w_k \sim N(0, Q_X), \quad (1)$$

where k is the frame index, and w_k is the process noise which is assumed to be a multivariate normal distribution with zero mean and covariance Q_X . Since the filter is applied to all joint positions separately, the joint index (j) is omitted in this section. Therefore, the state of the Kalman filter applied in the human kinematics at the frame k is given by:

$$X_k = [p_k^{(x)} \ p_k^{(y)} \ p_k^{(z)} \ v_k^{(x)} \ v_k^{(y)} \ v_k^{(z)}]^T \quad (2)$$

where $p_k^{(x)}$, $p_k^{(y)}$ and $p_k^{(z)}$ represent the position of the joint in Cartesian coordinate system, and $v_k^{(x)}$, $v_k^{(y)}$, and $v_k^{(z)}$ are corresponding velocities at the same frame for each axis. From (1), the updating of the state is dependent on the state-transition model F . In the

human motion system, this evolution is represented by the physics equation of motion considering each time step to be $\Delta t = 1/fs$, where fs is the frame rate of the depth sensor. We assume the speed of a joint remains constant along the direction of the joint's movement, thus the constant velocity model is used:

$$X_k = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} X_{k-1}. \quad (3)$$

The measurement Y_k in the KF model is defined as:

$$Y_k = [q_k^{(x)} \ q_k^{(y)} \ q_k^{(z)}]^T, \quad (4)$$

which is generated from the state X_k represented by:

$$Y_k = HX_k + u_k, u_k \sim N(0, Q_Y) \quad (5)$$

where u_k is observation noise with zero mean and covariance Q_Y . H is observation model which is defined by:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \quad (6)$$

The statistical property of the KF is based on a linear system. However, most of the systems in human motion capture are nonlinear. Therefore, non-linear KFs such as the EKF and UKF are employed to achieve a higher accuracy of D-Mocap data. The EKF uses Taylor decomposition to linearize nonlinear system [23]. However, EKF only provides a sub-optimal solution, because if the linearization is not accurate, its performance will be undermined. UKF estimates accurate state values by using unscented transformation (UT) [31][32]. Theoretically, the performance of UKF is better than EKF, and the accuracy of UKF can reach the second order or higher of EKF[33]. Gaussian-distributed samples are selected to estimate the true state in the UKF [34]. However, the real-world D-Mocap data are more anomalous with biased measurements or censoring data and non-Gaussian noise. In addition, the problem of censored measurements may be caused by occlusion or device limitation. It is difficult to solve these problems with the EKF and UKF. In order to overcome these problems, we employ the TKF to build a novel observation model for human kinematics.

3.2. Tobit Kalman Filter (TKF)

The TKF is a novel non-linear KF which integrates the property of the Tobit model in censored data estimation with the Kalman filter to provide an anticipation

for actual state value of a dynamic system in which parameters are partially known [15, 35]. The Tobit model is a truncated model which is widely used in economics because it can compensate for the impact of personal privacy or government secrecy policies on economic statistics [36, 37]. The Tobit model employed in the TKF which is applied to the D-Mocap data is depicted in Fig. 1.

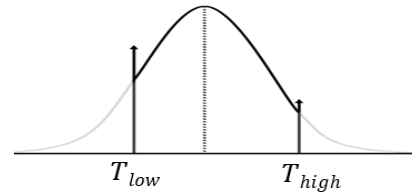


Figure 1. Tobit model

Based on the left censoring and right censoring of the Tobit model [34], two thresholds are defined in the TKF: the upper threshold T_{low} for left censoring and the lower threshold T_{high} for right censoring. For the left censoring, cases with a value no less than the upper threshold are censored, the censoring will take place above [38]. Likewise, for the situation of right censoring, the values that no more than the lower threshold are censored, the censoring will take place below. In the real world D-Mocap data, joint positions might be missing due to device limitations or occlusion. Therefore, a portion of parameters might hide in latent variables in which the situation meets the assumption of the TKF. The Tobit model can help to compensate for the impact of censoring joint data caused by occlusion or device limitation while the dynamic model providing a prediction for correcting noisy data. To apply TKF in the D-Mocap system, the upper and lower thresholds are selected along each axis of the current state depending on the previous state value and absolute values of maximum velocities in a small sequence (e.i. 50 frames) which is represented by $|v_{max}^x|$, $|v_{max}^y|$, and $|v_{max}^z|$. Therefore, thresholds for the k^{th} joint position are based on $|v_{max}|\Delta t$ and the $(k-1)^{th}$ step joint position in a small time window with a center at the $(k-1)^{th}$ step. For x -axis, thresholds are expressed by:

$$TH_k^{(x)} = p_{k-1}^{(x)} + |v_{max}^{(x)}|\Delta t \quad (7)$$

$$TL_k^{(x)} = p_{k-1}^{(x)} - |v_{max}^{(x)}|\Delta t, \quad (8)$$

where $TL^{(x)}$ and $TH^{(x)}$ are the lower and upper thresholds for x -axis at frame k , respectively. The same logic is applicable to other two axes (y and z). Hence,

the measurement model for x -axis is represented by:

$$\rho_k^{(x)} = \begin{cases} q_k^{(x)} & TL_k^{(x)} < q_k^{(x)} < TH_k^{(x)} \\ TL_k^{(x)} & q_k^{(x)} \leq TL_k^{(x)} \\ TH_k^{(x)} & q_k^{(x)} \geq TH_k^{(x)} \end{cases} \quad (9)$$

where $\rho_k^{(x)}$ is a latent variable for the x -axis [15]. $\rho^{(y)}$ and $\rho^{(z)}$ are defined in a similar way. Therefore, the 3D higher threshold vector is represented by:

$$\mathbf{TH}_k = [TH_k^{(x)} \quad TH_k^{(y)} \quad TH_k^{(z)}]^T, \quad (10)$$

This logic is equally applicable to the lower threshold vector \mathbf{TL}_k and 3D latent vector $\rho_k = [\rho_k^{(x)} \quad \rho_k^{(y)} \quad \rho_k^{(z)}]^T$. After being divided by two thresholds, there will be three parts in the measurement data: a portion censored equal to or above the upper threshold, a portion censored equal to or below the lower threshold, and an uncensored portion. And the expectation of the measurements is used as the observation value, which is expressed as:

$$E(Y_k) = \mathbf{p}_k^{(uc)} \odot \rho_k + \mathbf{p}_k^{(low)} \odot \mathbf{TL}_k + \mathbf{p}_k^{(high)} \odot \mathbf{TH}_k, \quad (11)$$

where \odot is the Hadamard product, $\mathbf{p}_k^{(uc)}$ represents the probability vector of uncensored measurement which is defined as:

$$\mathbf{p}_k^{(uc)} = [t_k^{(x)} \quad t_k^{(y)} \quad t_k^{(z)}]^T, \quad (12)$$

where $t_k^{(x)}$, $t_k^{(y)}$, and $t_k^{(z)}$ are uncensored probabilities of three coordinates. The same logic is equally applicable to $\mathbf{p}_k^{(high)}$ and $\mathbf{p}_k^{(low)}$, which represent the probability vectors of three coordinates being censored from above and below, respectively. The values of probabilities above are estimated based on the difference between latent measured variables in (9) and thresholds in (10), which is detailed in [15]. The Tobit Kalman gain is defined by:

$$K_k = R_{XY_k} R_{YY_k}^{-1}, \quad (13)$$

where R_{XY_k} is the cross-covariance between the observation and the state, and R_{YY_k} is the variance of the observation. For the details about R_{XY_k} and R_{YY_k} , refer to [15]. Thus, the complete TKF process applied in human kinematics is represented by the following equations:

$$X_{k|k-1} = FX_{k-1|k-1} \quad (14)$$

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q_X \quad (15)$$

$$X_{k|k} = X_{k|k-1} + K_k (\rho_k - E(Y_k)) \quad (16)$$

$$P_{k|k} = (I - K_k \mathbf{p}_k^{(uc)} H) P_{k|k-1}. \quad (17)$$

To make it easier to relate the results in context, we re-add the joint index to the final output of the TKF. Thus, the resultant joint positions which make up the filtered result in the k^{th} frame is given by:

$$\mathbf{x}_k^{(TKF)} = [p_{k,1}^{(x)} \quad p_{k,1}^{(y)} \quad p_{k,1}^{(z)} \quad \dots \quad p_{k,J}^{(x)} \quad p_{k,J}^{(y)} \quad p_{k,J}^{(z)}]^T \quad (18)$$

where J is the number of joints. The TKF differs from the traditional non-linear KFs like EKF and UKF in that it not only smooths the joint trajectories, but also utilizes a regression model for estimation of censoring measurement which makes state updating more reliable in the D-Mocap system. By selecting thresholds for censoring above and below, the measurements are divided into three parts. It helps the TKF to achieve an accurate estimation of the real state value while the measurement is replaced by the expectation. Using expected value can compensate for censored or biased data which makes the estimation close to the latent variables so that TKF yields a better result in human motion capture. Although the joint positions have been improved by denoising of nonlinear-KFs, the smoothness of the trajectory may cause changes in the skeleton structure. In order to limit the length of the bones in motion, we employed a differential evolution algorithm.

3.3. Differential Evolutionary (DE) Algorithm

Differential evolution (DE) is a population-based meta-heuristic search algorithm that optimizes a problem by iteratively driving a candidate solution toward better fitness regarding a given quality metric. It is executed by keeping a population of candidate solutions and creating new candidate solutions by mutation and crossover operations between exciting candidates, where the candidate solution with the best fitness will be kept. To reduce the bone length variation, a DE algorithm is employed to optimize the joint position in this research. If $(p_{m_i}^{(x)}, p_{m_i}^{(y)}, p_{m_i}^{(z)})$ and $(p_{m_j}^{(x)}, p_{m_j}^{(y)}, p_{m_j}^{(z)})$ are known as the 3-D joint positions at the endpoints of the

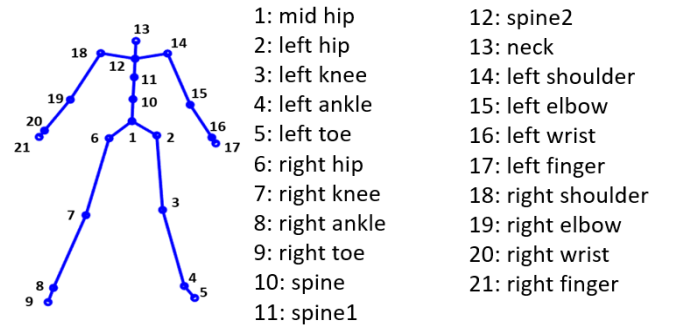


Figure 2. The 21-joint skeleton model used in our experiment.

m^{th} bone, its bone-length can be defined by:

$$L_m = \sqrt{(p_{m_i}^{(x)} - p_{m_j}^{(x)})^2 + (p_{m_i}^{(y)} - p_{m_j}^{(y)})^2 + (p_{m_i}^{(z)} - p_{m_j}^{(z)})^2}. \quad (19)$$

where m_i and m_j are two joint indices, $m_i, m_j \in [1, J]$. For the reason that DE is used for optimizing joint positions in each frame, the frame index (k) is omitted in this section. Joints positions are sampled in Cartesian coordinates (Fig. 2) to form a skeleton candidate, and multidimensional DE is used to maintain a high efficiency of the optimization method. Therefore, the joints are integrated into the skeleton candidate which is represented by:

$$C_n = [p_{1,n}^{(x)} p_{1,n}^{(y)} p_{1,n}^{(z)} p_{2,n}^{(x)} \dots p_{J,n}^{(x)} p_{J,n}^{(y)} p_{J,n}^{(z)}]^T, \quad (20)$$

where n is the index of the initial population. The search space of each parameters in the population is defined by a lower limit and an upper limit. The value of a parameter in the initial population is usually selected uniformly and randomly in the interval between the limit above and the limit below. For example, given that $C_{n,min}$ and $C_{n,max}$ represent the limit below and limit above of each parameter, the generation of other candidates in the initial population is represented by:

$$\hat{C}_n = C_n^{(min)} + \tau(C_n^{(max)} - C_n^{(min)}) \quad (21)$$

where τ is a number randomly selected from $[0, 1]$. Three parameters can be tuned in the mutation and crossover steps. First, the scaling factor F is used in the mutation step to generate a new skeleton candidate:

$$C_n^{(c)} = C_{n_3}^{(p)} + F * (C_{n_1}^{(p)} + C_{n_2}^{(p)}), \quad (22)$$

where n, n_1, n_2 , and n_3 are four positive integers randomly selected from numbers not larger than the population size, and the values of n_1, n_2 , and n_3 are different. $C^{(p)}$ and $C^{(c)}$ are the skeleton candidates for parent and offspring respectively. The scale factor $F \in (0, 2]$ is used to control the generation of the new population in the mutation step. The other two parameters, mutation rate, and crossover rate prevent the algorithm from stalling in a local optimum. In order to constrain bone lengths in D-Mocap data, we reduce deviation from corresponding reference bone lengths, which can be expressed as:

$$S = \sum_{m=1}^M |L_m - \bar{L}_m|, \quad (23)$$

where L_m is the bone length measured from the m^{th} bone, and \bar{L}_m is the reference bone length obtained from an initial application of the TKF, and M is the total number of bones. Normalization is applied to the objective function to improve convergence speed

and accuracy. Consequently, the objective function is represented by:

$$FV = \frac{1}{S + 1}, \quad (24)$$

where FV represents the fitness value. Thus the fitness function is defined as:

$$\Delta FV = |FV^{(p)} - FV^{(c)}|, \quad (25)$$

where $FV^{(p)}$ and $FV^{(c)}$ represent the fitness value of the parent and child generations, respectively [18]. The iteration of the DE will stop once ΔFV achieves a value no larger than η , where η is a very small value.

3.4. Integration of TKF and DE

To further improve the accuracy of D-Mocap data and reduce bone length variation, a hybrid method is proposed which integrates TKF and DE. We employ two different methods to integrate the TKF and DE based on initialization: TKF_U-DE and TKF_G-DE. In the following sections we refer to these two methods collectively as TKF-DE. In the TKF-DE, the results of the TKF, including output states and respective covariance matrices are used to generate an initial population in DE. Furthermore, the constraint of the search space in mutation and crossover also depends on the initial TKF results. The algorithm flow of TKF-DE is depicted in Fig. 3 (b). For comparison, we also employed a method called TKF+DE that directly combines the TKF and DE outputs, similar to the one in [18]. The flow chart of TKF+DE is depicted in Fig. 3 (a).

TKF+DE. The TKF and DE are operated separately in this method, and the final output is a weighted sum of the results of the TKF and DE, as shown in Fig. 3(a). The upper bound and lower bound of the search space are defined by the standard deviation (STD) of the joint trajectories and the velocity of each joint, which is detailed in [18]. Candidates in the initial population are randomly generated from the original D-Mocap confined by the upper and lower bounds defined above [18]. This is done so that the initial population can effectively span all the parameters of the algorithm. After optimization of mutation and crossover, the candidate with minimum bone length fluctuation is selected as $\mathbf{x}_k^{(DE)}$ at the k^{th} frame. The combination of TKF and DE in TKF+DE is represented by:

$$\mathbf{x}_k^{(final)} = \zeta * \mathbf{x}_k^{(TKF)} + (1 - \zeta) * \mathbf{x}_k^{(DE)}, \quad (26)$$

where $\mathbf{x}_k^{(TKF)}$ is the filtering result of the TKF at k^{th} frame, $\mathbf{x}_k^{(DE)}$ is the result of DE, ζ is the weight value, and the value of ζ depends on the speed of the joints. When the speed is small, the value of ζ will be larger, otherwise, it will be small [18].

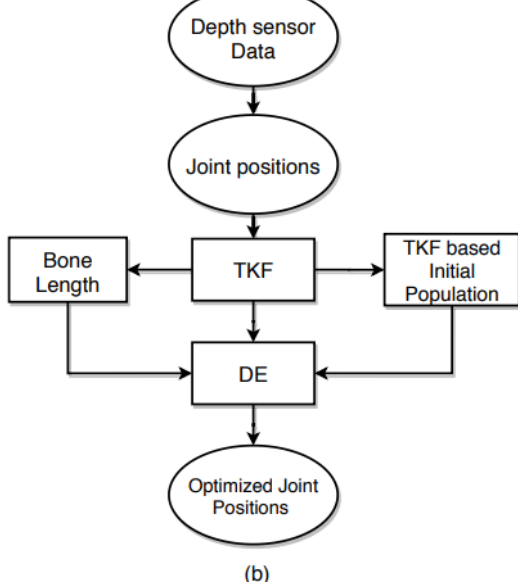
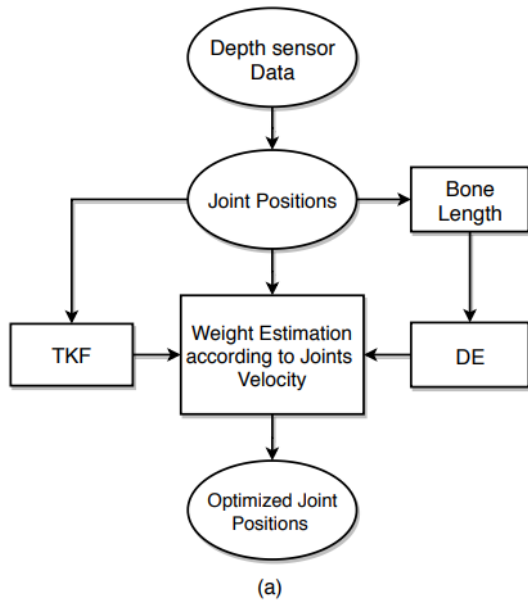


Figure 3. The flow charts of TKF+DE (a) and TKF-DE (b).

TKF_G-DE. In this method, the TKF and DE are integrated sequentially and probabilistically, as shown in Fig. 3(b). To improve the quality of initialization and to refine the search space, the initial population of DE is generated from the Gaussian-based probabilistic distribution of the TKF output, which is represented by the estimated joint position and its corresponding covariance matrix. This will allow the initial population to take advantage of both the mean and uncertainty from the TKF filtering result. The initial population is created joint-by-joint. For the j^{th} joint, a set of random 3D vectors, \mathbf{X}_j , is sampled from a 3D Gaussian distribution $N_3(\mu, \Sigma)$ as:

$$\mathbf{X}_j \sim N_3(\chi_j, P_j), \quad (27)$$

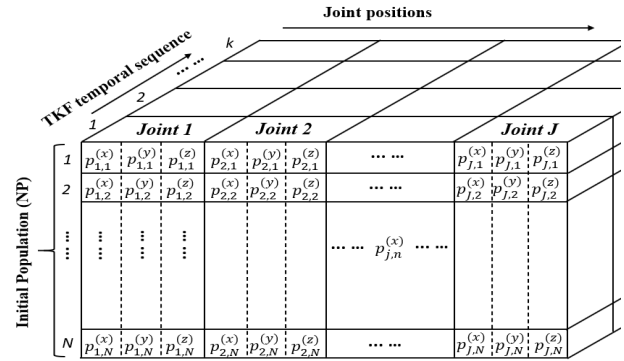


Figure 4. The initialization of population of the DE based on TKF is shown in this cube where N is the number of samples generated from each output joint position of each frame. $p_{j,n}^{(x)}$ represents coordinate value along x -axis of the n^{th} samples of j^{th} joint, $p_{j,n}^{(y)}$ and $p_{j,n}^{(z)}$ are defined in the same way.

where P_j is the covariance matrix of the j^{th} joint obtained from (17). Likewise, χ_j is the estimated joint position of the j^{th} joint obtained from TKF (16) where the frame index k is omitted for simplicity. \mathbf{X}_j has the same scale as the initial population and is reshaped into row vectors that reflect the original population in a single joint. This process is repeated for each joint one-by-one, after which the initial populations of the whole skeleton are accumulated to be used in the DE (Fig. 4), where the front layer denotes the initial population generated by the TKF output for all joints in the first frame. Then mutation, crossover, and selection are iteratively operated frame-by-frame to drive the population toward better candidates in the search space.

TKF_U-DE. For comparison, we also introduce an alternative TKF-DE integration method called TKF_U-DE that involves a uniform distribution to create the initial population for the 3D position of each joint. Sharing the same algorithm flow as TKF_G-DE Fig. 3(b), TKF_U-DE creates the initial population for each joint as:

$$\hat{p}_{1:N} = p_{1:N}^{(min)} + \tau(p_{1:N}^{(max)} - p_{1:N}^{(min)}) \quad (28)$$

where τ is a number randomly selected from $[0, 1]$. $p_{1:N}^{(max)}$ and $p_{1:N}^{(min)}$ are the maximum and minimum thresholds that restrict the search space with regard to each of three Cartesian coordinates (x , y and z). In the interest of simplicity, the joint (j^{th}) index is left out in this section. Mutation, crossover, and selection are executed after initialization, and the process repeats itself until the candidate with the best fitness value is found.

4. Experiment Results

In order to test our methods, we used two datasets. The first dataset is from the Carnegie Mellon University Database (CMU) [21] which contains a multitude of different motions such as jumping and shaking, shaking hands. The CMU data is simplified to a 21 joint skeleton and corrupted with additive Gaussian white noise (AGWN). The AGWN for this corruption is 0 mean with a standard deviation of 7 cm, and along with the original CMU data forms our set of simulated data. An example of motion in the CMU simulated dataset is presented in Fig. 5. The second dataset is OSU D-Mocap data which was collected by the Visual Computing and Image Processing Lab (VCIPL) and Biomechanical Analysis and Musculoskeletal Modeling (BAMM) Lab at Oklahoma State University (OSU), where the Orbbec depth sensor and the NuiTrack SDK were used in the data acquisition. OSU D-Mocap data only contains 6 lower joints including left/right hip, left/right knee, and left/right ankle. The OSU D-Mocap data contains 3 actions performed by a male adult subject including normal walking motion represented in Fig. 6, elongated stepping motion depicted in Fig. 7, and the motion of stepping in place shown in Fig. 8. In addition, an optical Mocap system was time-synchronized with the D-Mocap system as a reference for evaluation.

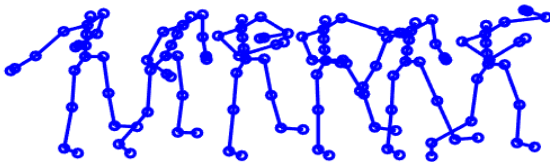


Figure 5. The motions in the CMU data

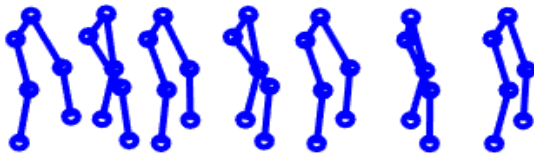


Figure 6. The normal walk motion in the OSU data

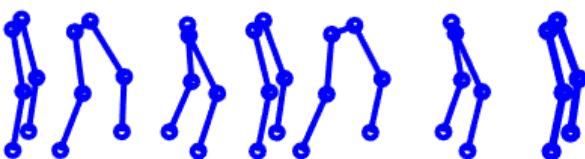


Figure 7. The long step motion in the OSU data

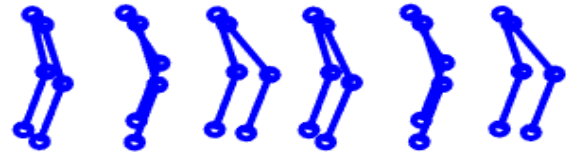


Figure 8. The Step in Place motion in the OSU data

4.1. Reference generation of OSU Data

Due to the differing methods of data collection, the joints estimated by the NuiTrack SDK and the Mocap marker data are not congruent. The markers of the Mocap system are placed on the surface of the human body, and the change in the relative position between the markers and if used to calculate bone length directly results in excessive variation. To match the corresponding joint positions of D-Mocap and Mocap data, MotionBuilder and OpenSim are employed to generate reference data from Mocap data. First, an OptiTrack marker-based Mocap system is used in the data acquisition where the markersets for the lower body is depicted in Fig. 9 (a). The Mocap system and D-Mocap system are executed simultaneously, where Mocap markers adhere to the certain positions of the lower half of the body, and the trajectories of these markers are recorded by the Mocap system. Then the recorded Mocap data is imported into OpenSim where the Mocap markers from a single frame are depicted as magenta dots attached to the lower body in Fig. 9 (b). The reference joint data is generated by forming a rigid body from select markers and calculating the rigid body's centroid. The generated reference joints in a single frame are depicted as red pots in Fig. 9 (c). We see that the generated reference joints more accurately correspond with the skeleton model of the NuiTrack SDK (Fig. 9 (d)). In order to validate the accuracy of the generated reference data, we calculated the bone lengths from reference joints generated by our method, and the bone length obtained from markers defined in Fig. 9(a). Here, LASIS and RASIS are used as the left and right hips, markers LTT and RTT are used as the left and right knees, and markers LLM and RLM are used as the left and right ankles. The STD of the bone length obtained from the Mocap markers is 10.34cm while the result of the reference data is 1.02cm. The resultant bone lengths are depicted in Fig. 10 which shows the consistency and robustness of the reference bone lengths for future algorithm evaluation.

4.2. Quantitative Evaluation

We present results comparing the performance of the different approaches applied to both simulated data and real-world D-Mocap data in this section. The accuracy assessment of the joint positions for all

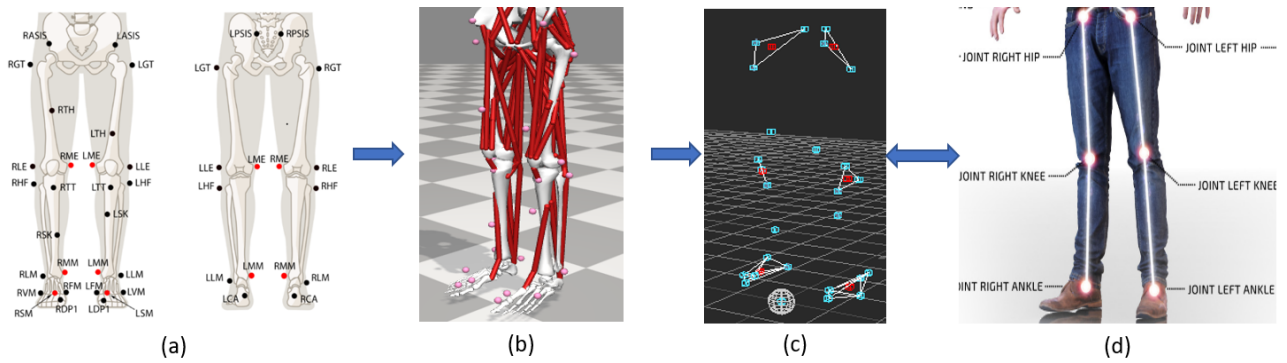


Figure 9. The process of reference data generation of OSU data

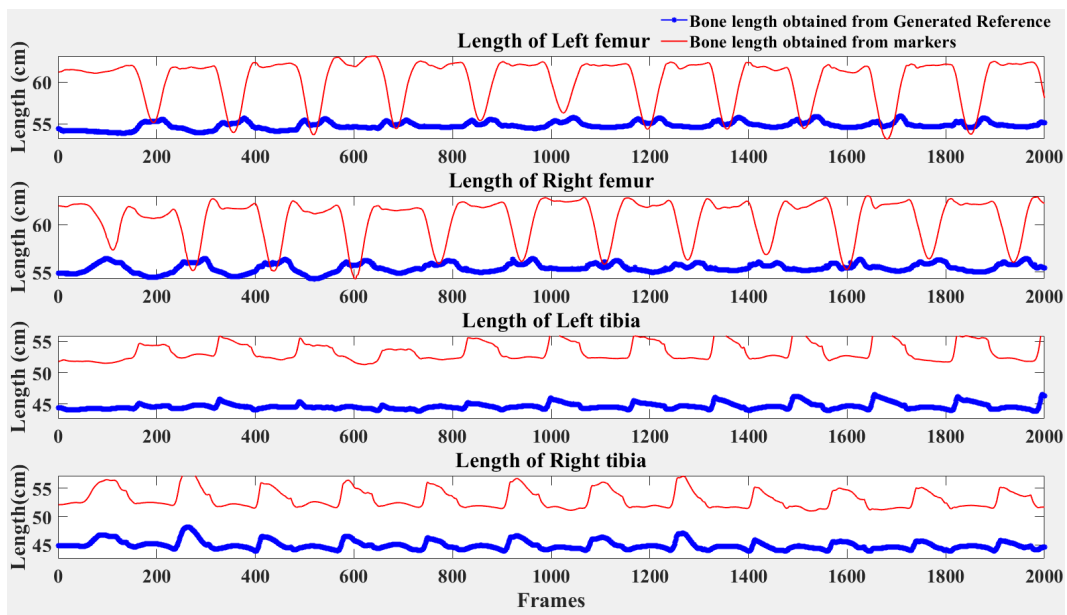


Figure 10. Comparison between the bone length obtained from markers and generated reference

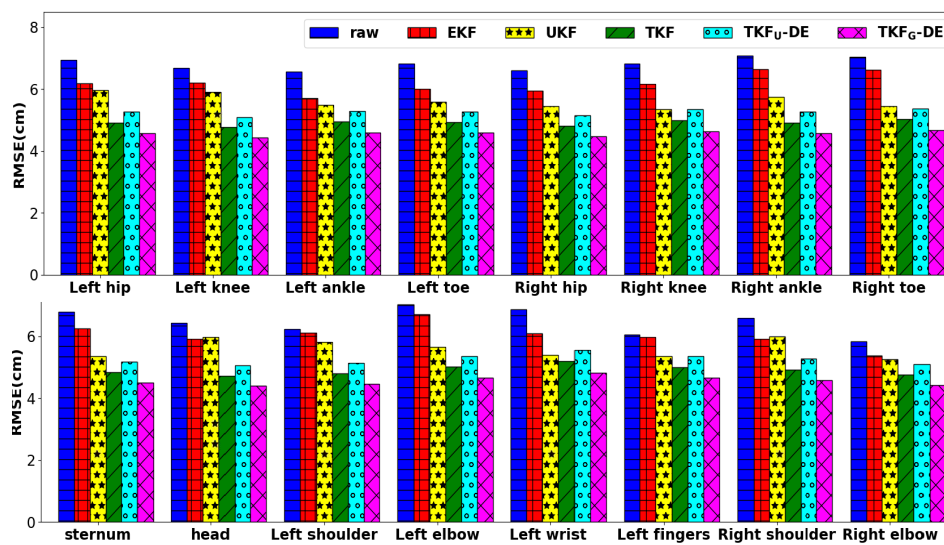


Figure 11. The RMSE of joint positions of CMU simulated data.

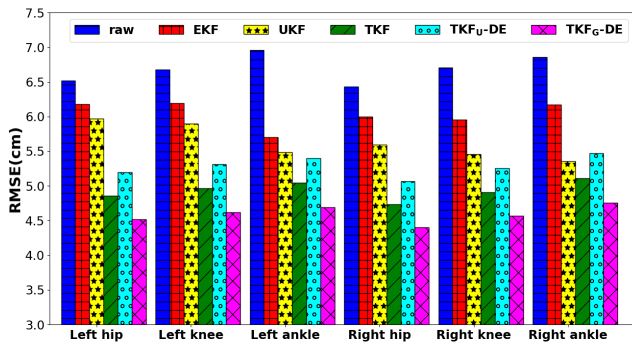


Figure 12. The error of joint positions of OSU real-world data.

different methods is performed using the root mean square error (RMSE) metric which is represented as:

$$RMSE = \sqrt{\frac{1}{J * K} \sum_{k=1}^K \sum_{j=1}^J (\tilde{X}_{j,k} - \hat{X}_{j,k})^2}, \quad (29)$$

where $\tilde{X}_{j,k}$ and $\hat{X}_{j,k}$ represent the j^{th} joint positions of the D-Mocap data and reference data at the k^{th} frame where J represents the number of joints and K is re-defined as the number of frames. The overall performance of bone-lengths and joint angles is evaluated by using the same metric. Figs. 11 and 12 joint position error for each method on the simulated dataset and real-world D-Mocap dataset, respectively. The overall performance of bone length including the RMSE and standard deviations of the D-Mocap data are depicted in Figs. 13 and 14, where the former displays the result of simulated data and later represents real-world data. In addition, Tables 1 and 2 show the numerical results of the different approaches directly, including the bone length error and RMSE of the joint positions. We also included three versions of TKF+DE, i.e., TKF+DE_{raw}, TKF+DE_{TKF}, and TKF+DE_{ref} where the bone lengths are estimated by using the raw D-Mocap data, the TKF filtering output, and the reference data. In addition, four versions of TKF-DE, i.e., TKF_U - DE_{TKF}, TKF_U - DE_{ref}, TKF_G - DE_{TKF}, TKF_G - DE_{ref} with different DE population initialization and bone length constraints.

Our methods are further verified by calculating relevant joint angles from joint positions and calculating the error with regard to reference values. Five joint angles are analyzed in our experiments, including left/right knee flexion angle (LKF/RKF), hip flexion angle (LHF/RHF), hip extension angle (LEF/REF), hip abduction angle (LHAB/RHAB), and hip adduction angle (LHAD/RHAD). The definitions of these joint angles are represented in Fig. 15. The hip extension angle and hip flexion angle are calculated together for simplicity, where the result is referred to as LHF/RHF in Table 3 and Fig. 16. The hip abduction angle and

Table 1. CMU simulated data results

	Bone length (cm)		Joint Positions	
	RMSE	Δ %	RMSE	Δ %
Raw data	7.2	–	6.7	–
EKF	6.6	8.2%	5.9	11.4%
UKF	6.4	10.5%	5.5	18.3%
TKF	6.0	16.7%	5.0	26.1%
TKF + DE _{raw}	6.4	11.1%	6.1	8.7%
TKF + DE _{TKF}	6.1	15.2%	5.9	12.1%
TKF + DE _{ref}	5.9	18.0%	5.6	17.3%
TKF _U - DE _{TKF}	5.5	24.0%	5.3	22.1%
TKF _U - DE _{ref}	5.6	21.7%	5.1	24.7%
TKF _G - DE _{TKF}	5.5	23.2%	4.8	29.4%
TKF _G - DE _{ref}	5.2	27.4%	4.5	33.1%

Table 2. OSU real-world data results

	Bone length (cm)		Joint Positions	
	RMSE	Δ %	RMSE	Δ %
Raw data	5.2	–	6.4	–
EKF	5.1	1.7%	5.8	9.4%
UKF	4.9	4.4%	5.2	18.7%
TKF	4.2	17.6%	4.4	31.3%
TKF + DE _{raw}	4.4	13.4%	5.7	11.3%
TKF + DE _{TKF}	4.2	18.3%	5.4	15.6%
TKF + DE _{ref}	4.0	22.2%	5.1	19.8%
TKF _U - DE _{TKF}	3.9	22.8%	4.8	25.4%
TKF _U - DE _{ref}	3.7	25.2%	4.6	28.1%
TKF _G - DE _{TKF}	4.1	21.4%	4.2	34.3%
TKF _G - DE _{ref}	3.8	26.2%	4.1	36.7%

Table 3. Error of joint angles in the OSU real-world data

	Raw Data	EKF	UKF	TKF	TKF+DE	TKF-DE
LKF	9.76	8.98	8.14	5.84	8.11	6.17
LHF	8.74	7.37	6.31	6.51	5.97	5.74
LHA	8.17	7.31	6.61	6.32	6.14	5.86
RKF	9.42	5.41	5.61	4.70	5.82	5.19
RHF	3.54	2.38	2.19	1.95	2.47	1.73
RHA	9.51	7.53	6.26	6.51	6.73	5.32
AVG	8.19	6.50	5.78	5.30	5.87	5.00

hip adduction angle are combined by the same token, where LHA/RHA are used as the name of the combination. The error of joint angles in OSU real-world data are shown in the Table 3. In our experiment, the average of the joint angle error of OSU real-world data is about 5°. Although this accuracy is still moderately lower than that of IMU sensors (e.g., 2° - 3° error in

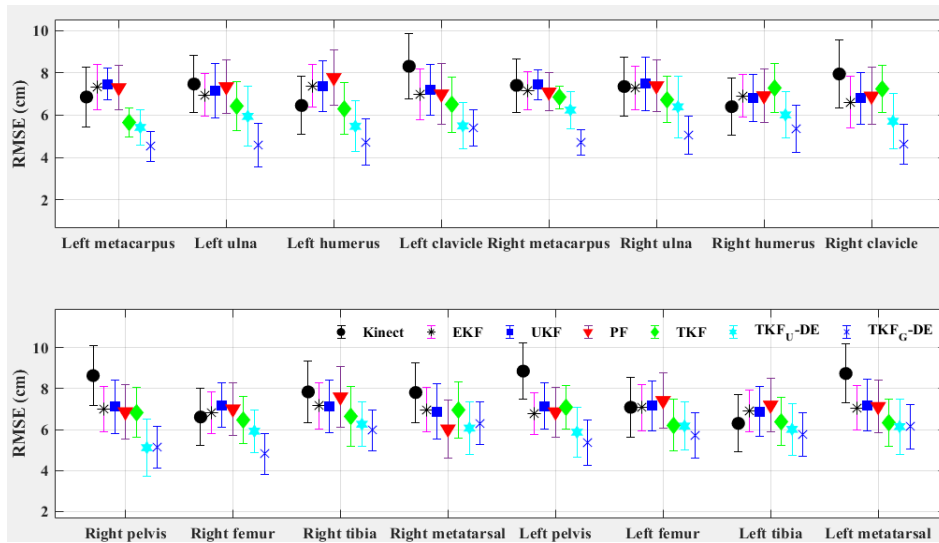


Figure 13. The bone length RMSE and standard deviations of simulated data.

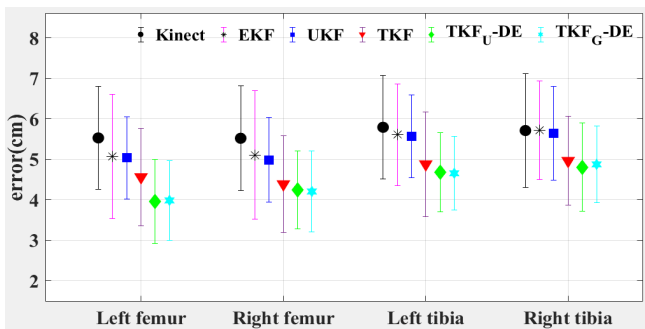


Figure 14. Bone length error and standard deviations of OSU real-world data.

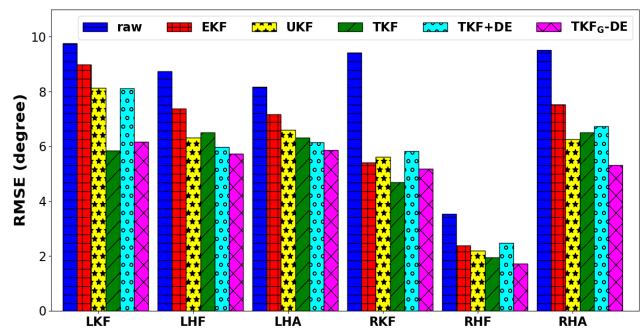


Figure 16. The error of joint angles of OSU real-world data.

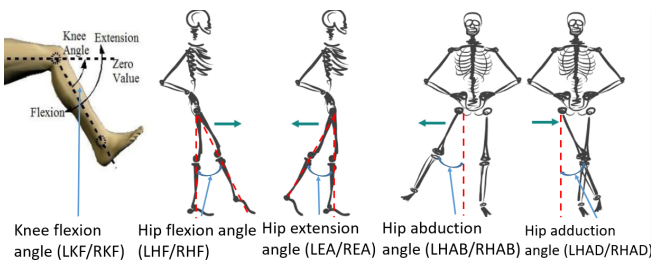


Figure 15. The definitions of the different joint angles.

[6]), D-Mocap does offer advantages in low-cost and efficient motion capture. In addition, there is still much potential to be further enhanced with advanced deep learning approaches [26].

4.3. Evaluation of Nonlinear KFs

From all results above, we confirm that the EKF and UKF significantly improve the accuracy of joint positions and joint angles for both simulated and real-world motion data. They also demonstrate

moderate improvements in the accuracy of the bone length estimation. Moreover, we note that the TKF significantly outperforms the EKF and UKF by a considerable margin in all three metrics, showing the usefulness and robustness of the Tobit model for the censored measurements in the D-Mocap data.

4.4. Analysis of TKF-DE integration

Essentially, the TKF+DE is a weighted average of the TKF and DE results. Both joint positions and bone lengths are improved in this method. Although the results of three versions of TKF+DE show that the more accurate bone length constraint leads to improved results, TKF+DE algorithms do not show much advantages over the TKF alone, implying that the advantages of TKF and DE may not be fully exploited to their full potentials. On the other hand, four TKF-DE methods show significant improvements over the three TKF+DE methods for both the simulated CMU data and the real-world OSU data. Specifically, $TKF_{G-DE_{TKF}}$ outperforms $TKF_{U-DE_{TKF}}$ with regard

to the accuracy of joint positions and angles with similar performance on bone length estimation. This is because the Gaussian predictive distribution from the TKF is more informative and useful than the uniform to initialize the DE population. More importantly, as shown in Tables 1, 2, and 3, all three aspects of D-Mocap have been significantly improved by $\text{TKF}_{G-DE_{TKF}}$ and $\text{TKF}_{G-DE_{ref}}$ over the TKF only. It is also understandable that a more accurate bone length constraint leads to better overall performance.

5. Conclusion

The relevant nature of our work lies in the fact that Mocap systems, although highly accurate, are expensive and inconvenient to use. If the accuracy of D-Mocap can be improved, depth sensor systems could provide an economical and convenient alternative to Mocap. We have introduced a novel hybrid approach to D-Mocap refinement synergistically combining the TKF and DE algorithms. Through the use of the TKF, we capitalize on kinematic prediction and a model for censored data. TKF joint trajectories are used to initialize DE, which optimizes this data to maintain accurate and consistent anthropometric measurements. This combined approach proves harmonious and improves upon other traditional approaches. To validate our work we utilized two sets of data, one set of real-world D-Mocap, and one set of Mocap data corrupted with AGWN. Using the metrics of joint position, bone length, and joint angle we find that our methods show more improvement than the TKF alone, as well as the EKF and UKF. Our future work will involve deep learning approaches to further enhance the accuracy and sensitivity of D-Mocap data for clinical and rehabilitation applications.

Acknowledgement

This work is supported in part by the US National Institutes of Health (NIH) Grant R15 AG061833 and the Oklahoma Center for the Advancement of Science and Technology (OCAST) Health Research Grant HR18-069.

References

- [1] CLARK, R.A., PUA, Y.H., FORTIN, K., RITCHIE, C., WEBSTER, K.E., DENEHY, L. and BRYANT, A.L. (2012) Validity of the microsoft kinect for assessment of postural control. *Gait & posture* 36(3): 372–377.
- [2] GALNA, B., BARRY, G., JACKSON, D., MHIRIPIRI, D., OLIVIER, P. and ROCHESTER, L. (2014) Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait & posture* 39(4): 1062–1068.
- [3] CLARK, R.A., VERNON, S., MENTIPLAY, B.F., MILLER, K.J., MCGINLEY, J.L., PUA, Y.H., PATERSON, K. et al. (2015) Instrumenting gait assessment using the kinect in people living with stroke: reliability and association with balance tests. *Journal of neuroengineering and rehabilitation* 12(1): 1–9.
- [4] ABREU, J., REBELO, S., PAREDES, H., BARROSO, J., MARTINS, P., REIS, A., AMORIM, E.V. et al. (2017) Assessment of microsoft kinect in the monitoring and rehabilitation of stroke patients. In *World Conference on Information Systems and Technologies* (Springer): 167–174.
- [5] DAWE, R.J., YU, L., LEURGANS, S.E., TRUTY, T., CURRAN, T., HAUSDORFF, J.M., WIMMER, M.A. et al. (2019) Expanding instrumented gait testing in the community setting: A portable, depth-sensing camera captures joint motion in older adults. *PloS one* 14(5): e0215995.
- [6] BONNET, V., JOUKOV, V., KULIĆ, D., FRAISSE, P., RAMDANI, N. and VENTURE, G. (2015) Monitoring of hip and knee joint angles using a single inertial measurement unit during lower limb rehabilitation. *IEEE Sensors Journal* 16(6): 1557–1564.
- [7] PEREIRA, A., GUIMARÃES, V. and SOUSA, I. (2017) Joint angles tracking for rehabilitation at home using inertial sensors: a feasibility study. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*: 146–154.
- [8] AHMED, M.H., SABIR, A.T. and MAGHDID, H.S. (2018) Khgi: Kinect-based human gait identification using statistical moments. In *1st International Conference on Advanced Research in Engineering Sciences (ARES)* (IEEE): 1–6.
- [9] BONNECHERE, B., JANSEN, B., SALVIA, P., BOUZAHOUENE, H., OMELINA, L., CORNELIS, J., ROOZE, M. et al. (2012) What are the current limits of the Kinect sensor. In *Ninth International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT)*: 287–294.
- [10] CHOI, W., PANTOFARU, C. and SAVARESE, S. (2011) Detecting and tracking people using an RGB-D camera via multiple detector fusion. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*: 1076–1083.
- [11] MAIMONE, A. and FUCHS, H. (2012) Reducing interference between multiple structured light depth sensors using motion. In *IEEE Virtual Reality Workshops (VRW)*: 51–54.
- [12] FLEET, D.J. (2011) Motion models for people tracking. In *Visual Analysis of Humans* (Springer), 171–198.
- [13] SHU, J., HAMANO, F. and ANGUS, J. (2014) Application of Extended Kalman filter for improving the accuracy and smoothness of Kinect skeleton-joint estimates. *Journal of Engineering Mathematics* 88(1): 161–175.
- [14] LARSEN, A.B.L., HAUBERG, S. and PEDERSEN, K.S. (2011) Unscented kalman filtering for articulated human tracking. In *Scandinavian conference on Image analysis* (Springer): 228–237.
- [15] ALLIK, B., MILLER, C., PIOVOSO, M.J. and ZURAKOWSKI, R. (2015) The Tobit Kalman filter: an estimator for censored measurements. *IEEE Transactions on Control Systems Technology* 24(1): 365–371.
- [16] LOUMPONIAS, K., VRETOS, N., DARAS, P. and TSAKLIDIS, G. (2016) Using kalman filter and Tobit Kalman filter in order to improve the motion recorded by kinect sensor ii. In *Proceedings of the 29th Panhellenic Statistics Conference*, 1: 2.

- [17] TRIPATHY, S.R., CHAKRAVARTY, K., SINHA, A., CHATTERJEE, D. and SAHA, S.K. (2017) Constrained particle filter for improving Kinect based measurements. In *IEEE International Symposium on Circuits and Systems (ISCAS)* (IEEE): 1–4.
- [18] DAS, P., CHAKRAVARTY, K., CHATTERJEE, D. and SINHA, A. (2017) Improvement in kinect based measurements using anthropometric constraints for rehabilitation. In *IEEE International Conference on Communications (ICC)* (IEEE): 1–6.
- [19] HOLDEN, D., SAITO, J. and KOMURA, T. (2016) A deep learning framework for character motion synthesis and editing. *ACM Trans. Graph.* **35**(4): 138:1–138:11. doi:10.1145/2897824.2925975, URL <http://doi.acm.org/10.1145/2897824.2925975>.
- [20] LANNAN, N. and FAN, G. (2019) Filter guided manifold optimization in the autoencoder latent space. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [21] UNIVERSITY, C.M. (2019), Carnegie Mellon mocap database. Data retrieved from Carnegie Mellon Mocap Database, <http://mocap.cs.cmu.edu/>.
- [22] HSIEH, C.C. and KUO, P.L. (2008) An impulsive noise reduction agent for rigid body motion data using b-spline wavelets. *Expert Systems with Applications* **34**(3): 1733–1741.
- [23] MOON, S., PARK, Y., KO, D.W. and SUH, I.H. (2016) Multiple Kinect sensor fusion for human skeleton tracking using Kalman filtering. *International Journal of Advanced Robotic Systems* **13**(2): 65.
- [24] PARK, Y., MOON, S. and SUH, I.H. (2016) Tracking human-like natural motion using deep recurrent neural networks. *arXiv preprint arXiv:1604.04528*.
- [25] HOLDEN, D., SAITO, J., KOMURA, T. and JOYCE, T. (2015) Learning motion manifolds with convolutional autoencoders. In *SIGGRAPH Asia 2015 Technical Briefs*, SA '15 (New York, NY, USA: ACM): 18:1–18:4. doi:10.1145/2820903.2820918, URL <http://doi.acm.org/10.1145/2820903.2820918>.
- [26] LI, S.J., ZHU, H.S., ZHENG, L.P. and LI, L. (2020) A perceptual-based noise-agnostic 3d skeleton motion data refinement network. *IEEE Access* **8**: 52927–52940.
- [27] BARTON, C.J., LEVINGER, P., MENZ, H.B. and WEBSTER, K.E. (2009) Kinematic gait characteristics associated with patellofemoral pain syndrome: a systematic review. *Gait & posture* **30**(4): 405–416.
- [28] MILLS, K., HETTINGA, B.A., POHL, M.B. and FERBER, R. (2013) Between-limb kinematic asymmetry during gait in unilateral and bilateral mild to moderate knee osteoarthritis. *Archives of Physical Medicine and Rehabilitation* **94**(11): 2241–2247.
- [29] SUL, C., JUNG, S. and WOHN, K. (1998) Synthesis of human motion using kalman filter. In *International Workshop on Capture Techniques for Virtual Environments* (Springer): 100–112.
- [30] MALAGUTI, A., CARRARO, M., GUIDOLIN, M., TAGLIAPIETRA, L., MENEGATTI, E. and GHIDONI, S. (2019) Real-time tracking-by-detection of human motion in RGB-D camera networks. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*: 3198–3204.
- [31] LAVIOLA, J.J. (2003) A comparison of Unscented and Extended Kalman filtering for estimating quaternion motion. In *Proceedings of the 2003 American Control Conference, 2003.* (IEEE), **3**: 2435–2440.
- [32] KHAZRAJ, H., FARIA DA SILVA, F. and BAK, C.L. (2016) A performance comparison between extended kalman filter and unscented kalman filter in power system dynamic state estimation. In *2016 51st International Universities Power Engineering Conference (UPEC)*: 1–6. doi:10.1109/UPEC.2016.8114125.
- [33] JULIER, S.J. and UHLMANN, J.K. (2004) Unscented filtering and nonlinear estimation. *Proceedings of the IEEE* **92**(3): 401–422.
- [34] ALLIK, B., MILLER, C., PIOVOSO, M.J. and ZURAKOWSKI, R. (2015) Nonlinear estimators for censored data: a comparison of the EKF, the UKF and the Tobit Kalman filter. In *2015 American Control Conference (ACC)* (IEEE): 5146–5151.
- [35] TOBIN, J. (1958) Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society* : 24–36.
- [36] DAHLBERG, M. and JOHANSSON, E. (2002) On the vote-purchasing behavior of incumbent governments. *American political Science review* : 27–40.
- [37] DUBOIS, H.F. and FATTORE, G. (2011) Public fund assignment through project evaluation. *Regional & Federal Studies* **21**(3): 355–374.
- [38] LOUMPONIAS, K., DIMOU, A., VRETOS, N. and DARAS, P. (2018) Adaptive tobit kalman-based tracking. In *2018 14th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*: 70–76. doi:10.1109/SITIS.2018.00021.