

Constrained Particle Filter for Improving Kinect Based Measurements

Soumya Ranjan Tripathy¹, Kingshuk Chakravarty² and Aniruddha Sinha²

Abstract—Microsoft Kinect has been gaining popularity in home-based rehabilitation solution due to its affordability and ease of use. It is used as a marker less human skeleton tracking device. However, apart from the fact that the skeleton data are contaminated with high frequency noise, the major drawback lies in the inability to retain the antropometric properties, like the body segments' length, which varies with time during the tracking. In this paper, a particle filter based approach has been proposed to track the human skeleton data in the presence of high frequency noise and multi-objective genetic algorithm is employed to reduce the bone length variations. In our approach multiple segments in skeleton are filtered simultaneously and segments' lengths are preserved by considering their interconnection unlike other methods in available literature. The proposed algorithm has achieved MAPE of 3.44% in maintaining the body segment length close to the ground truth and outperforms state-of-the-art methods.

Index Terms—Kinect, Particle filter, NSGA, Multi objective optimization

I. INTRODUCTION

Over past few decades, several motion capturing technologies have been explored and applied to human monitoring for health care applications. In recent days, the demand for home based affordable rehabilitation has increased for movement analysis in gait, balance and postural stability. This is mainly to prevent fall and improve in the movement of body parts for people affected by stroke [1] and elderly population [2]. There are mainly two types of movement analysis namely, marker based and markerless. Marker based systems like Vicon are very much popular in human movement analysis for their reliability, precision and accuracy. However these systems are very expensive and require skilled personnel for operation [3]. On the other hand, marker less motion tracking solutions are mostly based on radar and ultrasonic technologies. As an example, a radar based system was proposed in [4] for gait monitoring. The applicability of these systems in real world is limited due to the problems like multipath fading and very low range of operation etc. The Microsoft Kinect Xbox One is a low cost markerless motion tracking device [5], which is a potential candidate for 24 × 7 monitoring device in home rehabilitation [6][7] solution. It consists of a RGB camera and infrared (IR) based depth sensor, which can track human skeleton joint positions in 3D space similar to other marker based systems

like Vicon [8] [9]. The major problem in Kinect is that the body segments' length (bone length) calculated from joint coordinates vary with time. Ideally, the bone lengths should remain constant through out the time irrespective of segment orientation in 3D space [10]. Malinowski et al. [11] thoroughly studied variation of eight bone lengths during walking and running in the treadmill. An average bone length disparity of 9-11 cm in Kinect sensor was reported in their work [11]. Moreover, Kinect shows significant variation in bone lengths compared to Vicon as discussed in [12]. In addition to these discrepancies, Kinect coordinates are also noisy due to IR interference, external lighting conditions and non-anthropometric skeleton model [13] etc. These issues make Kinect difficult to use in clinical applications like rehabilitation.

In order to reduce the noise in joint coordinates, many time domain filters like averaging, median, mean square filters etc. are proposed in [14]. In [15], constant Kalman filter and Weiner Process Acceleration (WPA) Kalman filter are used to smooth and track joint position simultaneously. However all these methods are mainly inclined on maintaining the latency and improving individual joint positions.

The focus on improving the bone length variation is still limited to few literature. In [16] a constrained Kalman filter based method was proposed to track the joint positions by preserving the bone length over time. Potential drawback of this method [16] is that it considers a single body segment at a particular time stamp for filtering without considering interconnection between body segments. So naturally filtering of one segment may negatively affect the other connected segments. For example elbow joint cannot be filtered efficiently by this method [16] without considering both the arm and forearm simultaneously. In [17] similar approach is taken to maintain the bone length constant using Kalman filter and Differential Evolution algorithm. Hence it [17] also suffers from the same problem as that of [16]. Specifically, the constraints formulated in these works have focused on preserving the bone length of single body segment at a time without being concerned about structural build of human skeleton.

Our motivation is to preserve the bone lengths of all the connected segments simultaneously by considering their dependencies on each other due to the presence of common joints between them. Hence, a novel Non-dominated Sorting Genetic Algorithm (NSGA) [18] based particle filtering method is developed so that particle filter can track multiple skeleton joints simultaneously whereas NSGA will be responsible for maintaining the bone lengths. Moreover, novelty of the proposed work also lies in the constraint

¹Soumya Ranjan Tripathy is currently with Tampere University of Technology (TUT), Finland. He did this work during his earlier association with TCS Research and Innovation. Email: st11@iitbbs.ac.in

²Kingshuk Chakravarty and Aniruddha Sinha are with Embedded Systems & Robotics, TCS Research and Innovation, Tata Consultancy Services (TCS), Kolkata, India. Email: kingshuk.chakravarty@tcs.com, aniruddha.s@tcs.com

formulation to ensure that the filtered coordinates will not drastically deviate from the original coordinates (which ensures smooth tracking). In order to achieve this, the constraint is mathematically formulated and applied on both the static and dynamic postures for performance evaluation. A NSGA based improvement of particle filter was done in [19] but our method enables to extend other multi-objective problems in conjunction with the particle filter.

The paper is organized as follows. Proposed methodology and database creation are presented in Section II and Section III respectively. Section IV contains the results along with its comparison with existing methods. Finally conclusion and discussion of future work are summarized in Section V.

II. PROPOSED METHOD

Kinect uses IR Projector, IR sensor and RGB camera to track human joint positions in 3D world co-ordinate system consisting of axes (a, b, c) . At any instance of time t it provides noisy co-ordinates of N joints (for Kinect Xbox One $N = 25$), $\chi_t = (a^i, b^i, c^i)$ where $i = 1, 2, \dots, N$. Unfortunately, the body segment lengths computed from $\chi_t \forall t = 1, 2, \dots, T$ vary with time which is quite unexpected in human physical structure. In order to remove this noise, a constraint is formulated based on two factors derived from human physical (skeleton) structure i.e. (I) bones (body segments) are interconnected to each other and (II) the length between any physically connected joints should remain constant over time. Specifically the constraint is defined in such a way that it will preserve the entire human skeleton structure and provide more realistic anthropometric measurements. Hence a dynamic filtering based approach with multiple bone length constraints is proposed to denoise χ_t obtained from Kinect. It is realized through the fusion of Particle filter algorithm and NSGA. In this context, the bone length constraint is defined as:

Definition 2.1: A pair of joints i and j are said to comply to bone length constraint if the two joints lie on a single bone and their coordinates should follow (1) for all time instances.

$$\|\chi_t^i - \chi_t^j\|^2 = L_{i,j}^2 \quad (1)$$

Where χ_t^i is the coordinate of i^{th} joint at time t and $L_{i,j}$ is the physical (actual) length of the bone present between those two joints.

In the proposed method, the state vector is formed from χ_t as $\mathbf{x} = [a_1, \dots, a_m, b_1, \dots, b_m, c_1, \dots, c_m]^T$, where $m (\leq N)$ is the total number of joints considered for filtering. The state vector contains all the connected body segments for e.g. 6-5, 5-4, 4-20, 20-8, 8-9, 9-10 are such six segments consisting of seven joints as shown in Fig. 1. Our particle filter based method models the joint motion trajectory as a Linear Dynamic Systems (LDS). In this LDS, the states are evolved as $\mathbf{x}_t = f(\mathbf{x}_{t-1}, \mathbf{u}_{t-1})$ where f is the state transition function and \mathbf{u}_t is process noise following i.i.d. (independent and identically distributed) $\mathcal{N}(0, \sigma_u)$. The joint position tracking is carried out recursively by estimating \mathbf{x}_t at each time step t depending on the Kinect based measurements $\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{n}_t)$ where h is the measurement function and

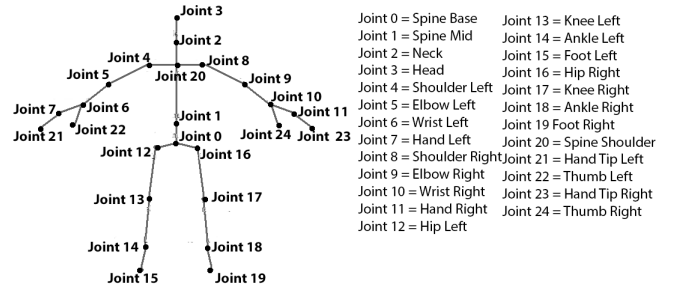


Fig. 1: Skeleton joints obtained from Kinect 2

\mathbf{n}_t is the measurement noise following i.i.d. $\mathcal{N}(0, \sigma_n)$. We have selected the covariance matrix of process noise and measurement noise to be diagonal with value 0.01. The whole process involves two steps called prediction of \mathbf{x}_t from measurements $\mathbf{y}_{1:t-1}$ and correction of \mathbf{x}_t given observed \mathbf{y}_t as shown in (2) and (3) respectively.

$$p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} \quad (2)$$

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1})} \quad (3)$$

The posterior probability density $p(\cdot)$ computed using (3) is approximated with the help of Monte Carlo (MC) simulations. MC simulations approximate the Probability Density Function (PDF) by a set of random samples and corresponding weights as given in (4).

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) = \sum_{i=1}^S \mathbf{w}_t^i \delta(\mathbf{x}_t - \mathbf{x}_t^i) \quad (4)$$

Here, S is the number of particles obtained from importance sampling scheme [20] by defining the importance density function $q(\cdot)$ and in our case we have selected $S = 400$. The generated S particles undergo state transitions at each time instance as given in (2) and the corresponding weight \mathbf{w}_t^i of each particle is updated sequentially [20] using (5)

$$\mathbf{w}_t^i \propto \mathbf{w}_{t-1}^i \frac{p(\mathbf{y}_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, \mathbf{y}_t)} \quad (5)$$

The sequential update (5) is simplified [20] to (6) by taking $q(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i, \mathbf{y}_t) = p(\mathbf{x}_t^i | \mathbf{x}_{t-1}^i)$

$$\mathbf{w}_t^i \propto \mathbf{w}_{t-1}^i p(\mathbf{y}_t | \mathbf{x}_t^i) \quad (6)$$

As the state vector \mathbf{x} operates on multiple (m) joints simultaneously, it eventually preserves the interconnection between body segments. However, it is not responsible to keep the individual bone length constant. To achieve this and harness the power of particles, in the proposed filtering approach, a population based Genetic algorithm is employed. Even so, this algorithm has to deal with two objectives i.e. (I) the filtered coordinates should not deviate abruptly from the original coordinates obtained from Kinect and (II) minimize the bone lengths' variations over time. Mathematically the objective function for each particle i at t is formulated as

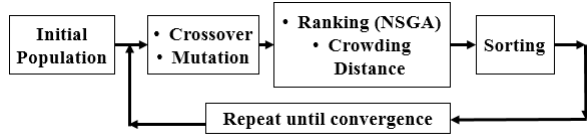


Fig. 2: Steps included in the NSGA

given in (7).

$$\max_{\mathbf{x}_t} \sum_{j=1}^{2*m} \mathbf{w}(j)_t^i \quad (7)$$

and $\min_{\mathbf{x}_t \forall \text{valid } i,j} \sum (\mathbf{l}(\mathbf{x}_t)_{i,j} - \mathbf{L}_{i,j})^2$

Where $\mathbf{l} = \{l_{i,j} | (i,j) \leq N \text{ and } i \neq j\}$ is bone length vector computed from S and $\mathbf{L} = \{L_{i,j} | (i,j) \leq N \text{ and } i \neq j\}$ is the actual lengths (ground truth) of the segments. According to (7) maximizing the sum of weights of all the state variables at a particular time step will ensure that the filtered coordinates are close to the measurement vector \mathbf{y}_t , as \mathbf{w} is modeled as likelihood function in (6). This type of multi-objective optimization is handled by NSGA [19]. S particles form the initial population vector of NSGA and undergo multiple steps as depicted in the Fig. 2 to provide optimized particles. Finally, the weighted mean of the particle distribution is taken as the final estimation of state vector at time t and it is expressed as $\hat{\mathbf{x}}_t = \sum_i \mathbf{w}_t \mathbf{x}_t$. $\hat{\mathbf{x}}_t$ is expected to provide corrected joint coordinates by satisfying both the constraints as mentioned earlier in the section. In this scenario, the NSGA helps the particle filter to find the multi-objective solutions by being used as an add-on to original particle filter algorithm. This method can also be extended to handle other multi-objective problems with the benefit of particle filters unlike the method given in [19]. It is to be noted that the degeneracy problem of particle filter i.e. most of the weights become insignificant, is avoided by re-sampling strategy as mentioned in [20]. All particles and corresponding weights are uniformly randomly initialized at time instance $t = 0$ and recursively estimated for the successive timestamps. The overall algorithm is explained in the Algorithm 1.

III. DATABASE CREATION

Twenty six subjects (age: range 24-55 years, weight: 52kg-97kg and height: 1.42m-1.96m) with no symptoms of neuro-physiological or musculo-skeletal disorders, have been chosen for the study. These subjects belong to TCS Innovation lab and voluntarily participated in the data collection. The subjects performed active Range Of Motion (ROM) exercises - shoulder abduction/adduction or flexion/extension in front of the Kinect Xbox One sensor placed at a distance of 2 meter approximately. In the beginning of the exercise, the subjects are told to stand in a stationary posture for 30 seconds and then perform the exercise. Dataset comprises of 25 skeleton joint coordinates for both static and dynamic postures.

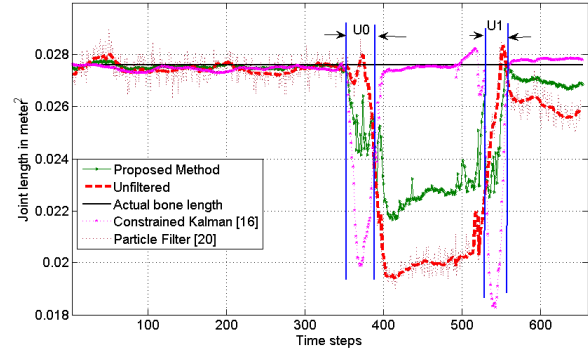


Fig. 3: Variation of right arm length (segment 20-8)

Algorithm 1 Proposed algorithm

Initialize:
 Random initialization of particles \mathbf{x}_0^i and weights \mathbf{w}_0^i at $t = 0$ for $i = 1, 2, \dots, S$

loop on t:

for each particle $i = 1$ to S **do**
 $\mathbf{x}_t^i \leftarrow \mathbf{x}_{t-1}^i + \mathbf{u}_t$: state Transition
 $\mathbf{w}_t^i \leftarrow p(\mathbf{y}_t | \mathbf{x}_t^i)$: weight allocation
 $C_1(i) = \sum_{j=1}^{2*m} \mathbf{w}(j)_t^i$: 1st Cost
 $C_2(i) = \sum (\mathbf{l}_{\forall \text{segments}}(\mathbf{x}_t^i) - \mathbf{L}_{\forall \text{segments}})^2$: 2nd Cost
end for
 $\mathbf{x}_t \leftarrow \text{NSGA}(\mathbf{x}_t, C_1, C_2)$: maximize(C_1), minimize(C_2)
 $\mathbf{w}_t^i \leftarrow \mathbf{w}_{t-1}^i * p(\mathbf{y}_t | \mathbf{x}_t^i)$, $\forall i$
 $\mathbf{w}_t \leftarrow \mathbf{w}_t / \sum \mathbf{w}_t$
if degeneracy in \mathbf{w}_t then
 $(\mathbf{x}_t, \mathbf{w}_t) \leftarrow \text{resample}(\mathbf{x}_t, \mathbf{w}_t)$.
end if
 $\hat{\mathbf{x}}_t = \sum_i \mathbf{w}_t \mathbf{x}_t$.
if $t < T$ then
goto loop.
else return
end if

IV. RESULTS AND DISCUSSIONS

Our algorithm is evaluated on the basis of its ability to minimize the bone lengths' variations computed from the filtered joint coordinates with respect to actual bone lengths. We have considered joints from both right and left hands simultaneously i.e. 6-5, 5-4, 4-20, 20-8, 8-9, 9-10 as shown in Fig. 1 for filtering. Fig. 3 (red dotted line) clearly depicts the variation in bone length between joint number 20 and 8 while performing the shoulder abduction/adduction exercise in the right hand (between time steps or frame numbers 380 to 540). Moreover, from the Fig. 4 it is quite clear that length of the left arm also varies even it is almost in static posture during the entire exercise. The observation holds true for all other body segment lengths. In order to demonstrate robustness of the proposed algorithm, it has been applied on the above mentioned seven joints simultaneously and green line in Fig. 3 and Fig. 4 portray how it is able to bring the associated segment lengths close to the actual ones. It is quite evident from the Fig. 5 that the bone length corrections don't come by randomly (abruptly) adjusting the joint coordinates but by closely following joint positions over

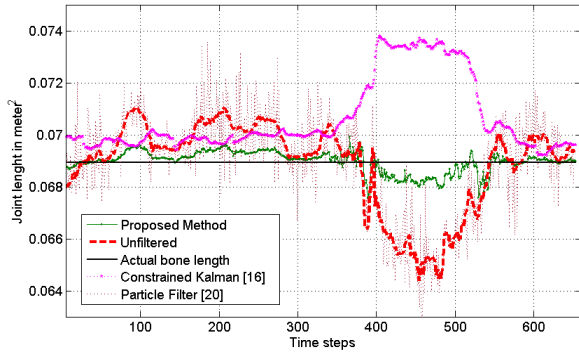


Fig. 4: Variation of left arm length (segment 4-5)

time. On the contrary, in constrained Kalman filter [16], the joint coordinates greatly deviates from actual trajectory to satisfy the bone length constraint. Moreover in a very small duration (zone $U1$ of Fig. 5), the joint coordinate corrected by the algorithm [16] changes drastically which is quite unexpected. In addition to that, Fig. 3 and Fig. 4 also depict the performance comparison of our proposed method with respect to state-of-the-art algorithms [16] and [20] for right and left arm lengths. The outcome of constrained Kalman filter [16] is inferior to our approach for both dynamic (right arm) and static body segments (left arm) because the formulation doesn't include the interconnection between the joints. As shown in Fig. 3, constrained Kalman filter also fails to adopt the changes in posture as it varies abruptly in zone $U1$ and $U2$ (where the right arm starts moving) whereas it performs well in between $U1-U2$ (where the right arm is at rest). As the formulation of our proposed particle filter accommodates the inter relationships between body segments, it resists abrupt changes in the bone length and maintains the variation of bone lengths minimum during the transition of body segments from resting state to dynamic state. Moreover, other constraint less methods ([20]) closely follow the unfiltered bone length and performance is not satisfactory compared to the constrained approaches (i.e. [16] and proposed one). The overall performance is evaluated based on the parameter Mean Absolute Percentage Error (MAPE) between the bone length obtained from filtered signal and original signal for all time instances. MAPE (in %) is mathematically defined as (8). It is desirable to have minimum MAPE to ensure better performance.

$$MAPE = \frac{100}{T} \sum_{t=1}^T \left| \frac{\mathbf{L} - \mathbf{I}(\mathbf{x}_t)}{\mathbf{L}} \right| \quad (8)$$

Table I presents the comparison between our method and state-of-the-art algorithms mentioned in [16], [15] and [20]. The evaluation is carried out per bone length based on the average of MAPE over all subjects. Here the constant velocity models are considered for Kalman filter based methods. It is clear from the Table I that our algorithm is able to minimize the MAPE for all segments over all subjects whereas performance of other ones are not so satisfactory. Finally, our particle filter and NSGA based

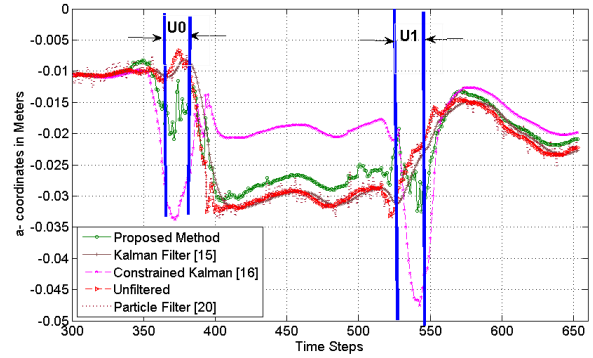


Fig. 5: Temporal variation of coordinate "a" for joint no. 8

TABLE I: Comparison of four methods in terms of MAPE

Segments bet. joints	Kalman Filter[15]	Constrained Kalman[16]	Particle Filter[20]	Proposed
20-8	15.60	13.92	17.66	3.82
8-9	16.60	14.25	16.77	4.22
9-10	22.01	19.57	22.52	5.69
20-4	17.17	17.17	18.39	2.71
4-5	13.99	13.42	14.20	1.74
5-6	23.97	23.71	24.36	2.46

algorithm outperforms them and is able to achieve a MAPE of 3.44% over all the subjects and all the joints in comparison to constrained Kalman filter [16] with MAPE \approx 17%.

V. CONCLUSIONS

In this paper we have proposed a probabilistic framework for estimating the skeleton joint locations using particle filter based joint location estimation approach accompanied with the Non-dominated Sorting Genetic Algorithm to reduce the variation in bone length. This method can track multiple body segments and can reduce the variation in the segments' length by considering their interconnection in dynamic as well as static conditions. This filtering approach is able to preserve the skeleton structure in a more realistic manner. Experimental results on healthy subjects demonstrate remarkable reduction in the MAPE compared to the earlier reported methods including constrained Kalman [16] and standalone particle filter [20] approaches. In future we plan to experiment with the patient data and measure the improvement in accuracy on the clinically derived parameters like joint angles.

REFERENCES

- [1] A. Towfighi and J. L. Saver, "Stroke declines from third to fourth leading cause of death in the united states historical perspective and challenges ahead," *Stroke*, vol. 42, no. 8, pp. 2351–2355, 2011.
- [2] Z. Zimmer and S. McDaniel, "Global ageing in the twenty-first century: An introduction," *Global Ageing in the Twenty-First Century: Challenges, Opportunities and Implications*, p. 1, 2016.
- [3] B. Carse, B. Meadows, R. Bowers, and P. Rowe, "Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3d motion analysis system," *Physiotherapy*, vol. 99, no. 4, pp. 347–351, 2013.
- [4] M. G. A. Bijan G. Mobasseri, "A time-frequency classifier for human gait recognition," pp. 7306 – 7306 – 9, 2009. [Online]. Available: <http://dx.doi.org/10.1117/12.819060>
- [5] J. Sell and P. O'Connor, "The xbox one system on a chip and kinect sensor," *IEEE Micro*, vol. 34, no. 2, pp. 44–53, 2014.

- [6] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, "Full body gait analysis with kinect," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2012, pp. 1964–1967.
- [7] M. E. Nixon, A. M. Howard, and Y.-P. Chen, "Quantitative evaluation of the microsoft kinect tm for use in an upper extremity virtual rehabilitation environment," in *Virtual Rehabilitation (ICVR), 2013 International Conference on*. IEEE, 2013, pp. 222–228.
- [8] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, and L. Rochester, "Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease," *Gait & posture*, vol. 39, no. 4, pp. 1062–1068, 2014.
- [9] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis," *Journal of medical engineering & technology*, vol. 38, no. 5, pp. 274–280, 2014.
- [10] M. Cirstea and M. F. Levin, "Compensatory strategies for reaching in stroke," *Brain*, vol. 123, no. 5, pp. 940–953, 2000.
- [11] M. J. Malinowski and E. Matsinos, "On using the Microsoft Kinect™ sensors to determine the lengths of the arm and leg bones of a human subject in motion," *ArXiv e-prints*, May 2015.
- [12] B. Bonnechere, B. Jansen, P. Salvia, H. Bouzahouene, L. Omelina, J. Cornelis, M. Rooze, and S. Van Sint Jan, "What are the current limits of the kinect sensor," in *Ninth International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT)*, 2012, pp. 287–294.
- [13] B. Bonnechere, B. Jansen, P. Salvia, H. Bouzahouene, L. Omelina, F. Moiseev, V. Sholukha, J. Cornelis, M. Rooze, and S. V. S. Jan, "Validity and reliability of the kinect within functional assessment activities: comparison with standard stereophotogrammetry," *Gait & posture*, vol. 39, no. 1, pp. 593–598, 2014.
- [14] N. Yang, F. Duan, Y. Wei, C. Liu, J. T. C. Tan, B. Xu, and J. Zhang, "A study of the human-robot synchronous control system based on skeletal tracking technology," in *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2191–2196.
- [15] M. Edwards and R. Green, "Low-latency filtering of kinect skeleton data for video game control," in *Proceedings of the 29th International Conference on Image and Vision Computing New Zealand*. ACM, 2014, pp. 190–195.
- [16] S. R. Tripathy, K. Chakravarty, A. Sinha, D. Chatterjee, and S. K. Saha, "Constrained kalman filter for improving kinect based measurements," in *Circuits and Systems (ISCAS), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1–4.
- [17] P. Das, K. Chakravarty, D. Chatterjee, and A. Sinha, "Improvement in kinect based measurements using anthropometric constraints for rehabilitation," in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.
- [18] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.
- [19] S. M. K. Heris and H. Khaloozadeh, "Non-dominated sorting genetic filter a multi-objective evolutionary particle filter," in *Intelligent Systems (ICIS), 2014 Iranian Conference on*. IEEE, 2014, pp. 1–6.
- [20] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.