Stabilizing Skeletal Pose Estimation using mmWave Radar via Dynamic Model and Filtering

Shuting Hu
Department of Electrical and Computer Engineering
The University of Arizona
Tucson, AZ USA
shutinghu@arizona.edu

Arindam Sengupta
Department of Electrical and Computer Engineering
The University of Arizona
Tucson, AZ USA
sengupta@arizona.edu

Siyang Cao
Department of Electrical and Computer Engineering
The University of Arizona
Tucson, AZ USA
caos@arizona.edu

***Abstract*—In this paper, we illustrate a method to stabilize the position estimation of human skeleton using mmWave radar. In our previous study, an optimized CNN architecture was used to extract the positions of human skeleton accurately. However, the position estimation of the joints vibrates over time. In the field of digital signal processing, filters are used to remove unwanted parts of signal and widely applied in noise reduction, radar, audio and video processing, etc. In this paper, three types of filters i.e. Elliptic, Savitzky-Golay, and Whittaker-Eilers are discussed and applied to both positions and angles of the human skeleton. This paper further presents a humanoid robotics dynamic model, specifically forward kinematics, to recalculate joint positions with improved stability. We define the root joint, a world coordinate system, and "T" pose, to get the subsequent joints' rotation matrix using kinematics chain of the skeleton, then compute the Euler angles. After the filtering, we compare the effect of different filters using a method of Standard Deviation (SD) of the angle slope. In addition, we analyze the change of localization accuracy after recalculating the positions using forward kinematics based on the current angle, root position, and bone length information. The data collection and experimental evaluation have shown a motion stability improvement of 54.05% compared to the CNN predicted value.**

***Index Terms*—mmWave radar, skeleton estimation, humanoid robotics dynamic model, digital signal filters**

I. Introduction

Recently, human behavior detection has become increasingly important in a variety of areas, such as preventive security monitoring, pedestrian monitoring for autonomous vehicles, and healthcare for the disabled and or elders. Commonly, optical or infrared sensors could be deployed for such applications, but privacy concerns have hindered its widespread adoption, especially in public settings. An alternate solution is to use wearable sensors. However, i) frequent battery charging, ii) inconvenience of attaching extra equipment to the human body, make this choice impractical and undesirable [1] in certain applications. Recently, lidar-based human behavior detection has garnered significant interest [2]. Lidar products could address the privacy concerns problems and are not affected by degraded lighting conditions, however, at the cost of an expensive module and continued maintenance. In recent years, small and low-cost on-chip 60 GHz and 77 GHz consumer and automotive mmWave radars have developed rapidly, and a series of applications have been derived from this. Zhang et al. [3] used TI AWR1642 mmWave radar to build a real-time human behavior detection and monitoring system. Singh et al. [4] proposed RADHAR, which performed human activity recognition using TI IWR1443BOOST mmWave radar. Sengupta et al. [5] proposed mm-Pose to detect and track human skeletal key points in real-time using two TI AWR1642 boost mmWave radars. Size et al. [6] propose MARS for smart health care using TI IWR1443 mmWave radar to reconstruct 19 critical human skeletal key points. To further improve accuracy, Sengupta et al. [7] presented mmPose-NLP, a Natural Language Processing (NLP) inspired approach, which could precisely estimate up to 25 skeletal key points with an average Mean Absolute Error (MAE) of 2.39 cm. While the aforementioned methods' primary objective was to accurately detect and localize key points in each frame, they do not address the inherent jitter and non-smooth skeletal-pose transitions between frames, which is extremely desired in real-world applications. To address this, we present a novel approach to stabilizing inter-frame skeletal-pose motion using digital signal filters and a forward kinematics dynamic model.

The remaining sections of this paper are organized as follows. Section II discusses the methodology, including the estimation of human skeleton, dynamics modeling of a humanoid robot, and digital signal filters. Section III demonstrates the experiment results of the proposed approach. Finally, Section IV concludes the study.

II. Methodology

A. Human Skeleton Estimation

Radar sensing uses multiple RF bands, and each band's application may vary across regions. The 77GHz band is common in automotive applications but has restrictions in regions that require human interaction. Frequency band 60-64GHz, which is used in this study, is expected to be allocated by FCC for monitoring vulnerable medical patients. The used antenna of mmWave radar consists of 3 transmit (TX) and 4 receive (RX) radio frequency (RF) components. The most basic principle in the radar system is the transmission and reception of electromagnetic wave signals. In mmWave radars, Frequency-Modulated Continuous-Wave (FMCW) signals are used, and the frequency increases linearly with time. FMCW

transmits a chirp signal and receives the signal that is transmitted and reflected by the target. By calculating the difference in frequency and phase between signals, and using MIMO technology, we can get 3D point cloud, each point consists of range, azimuth, elevation, radial velocity (i.e., Doppler), and signal-to-noise ratio (SNR). The 3D point cloud is in spherical coordinates, according to the trigonometric function, the spatial position of each point in Cartesian coordinates can be obtained. The intermediate process involves many techniques, such as range processing, static clutter removal, capon beamforming (BF), constant false alarm rate (CFAR) detection algorithm, radial velocity estimation, and so on. The first is range processing, which performs 1D FFT processing on chirp RF data to obtain radar cubes. Then static clutter removal was performed on the radar cube matrix by calculating the mean values of antenna symbols across the chirps per antenna, then subtracting the mean values from the symbols. In this way, stationary objects such as chairs and tables can be removed, leaving only the signal reflected from moving objects. The activities of the human body are captured by the front-end radar sensor and transmitted to the inside of the microcontrollers (MCUs) and digital signal processors (DSPs). After the range processing and static clutter removal steps, the capon beamformer is used to generate the 3D range-azimuth-elevation spectrum matrix (i.e., heatmap). Remove the cluster by a CFAR algorithm in the range domain and angle domain, respectively, to find detection points. Finally, to get the doppler information, it is necessary to operate in successive chirps, followed by a maximum peak search to estimate the radial velocity of the detected point. A comprehensive discussion on radar operation and the radar signal processing chain has been presented in our previous works [3], [5], [7], readers are referred to them, and here is only a brief overview.

Convolutional Neural Network (CNN) still uses the same principles that Multi Layer Perceptron (MLP) use, but it applies convolutional layers and is therefore more suitable for efficient processing of image and video information. In order to apply TensorFlow's Conv2D, 2D convolution layer, we operate the obtained point cloud to a uniform length of 100 by the method of zero padding. Then reshape it into a size of 10×10. The 5D (xyz position, velocity, SNR) information of the point cloud is used as the different channels of the convolution layer. As velocity, SNR, and 3-D spatial position data are independent variables, we use separate convolution layers for preliminary transformation, followed by concatenation at a later stage. To address the random initialization and subsequent local minima convergence of the CNN, we create 3 models to perform the training process at the same time and averaged the results. The optimized CNN architecture is shown in Fig. 1.

B. Dynamics Modeling of a Humanoid Robot

Dynamic models are very important in the field of robotics and play a role in stabilizing motion. Past approaches have tried to stabilize robot motion by setting constraints and optimal parameterization [8] [9]. Ding et al. used a kinematic constrained method to add multiple sets of loss functions to the radar point cloud during the training process to aid the neural network learn these relationships [10]. In this study, in order to explore the cause of jitter, we choose to use the most basic inverse and forward kinematic methods. Inverse kinematics find the robot joint angles needed to reach a particular position and produce the movement. Forward kinematics calculate the robot joint position in the next frame according to the joint rotation angles. The complete signal flow for the forward kinematic transform is shown in Fig. 2. Villegas et al. [11] proposed a recurrent neural network architecture with a forward kinematics layer and cycle consistency-based adversarial training objective for unsupervised motion retargeting. Shi et al. [12] introduced MotioNet, a deep neural network that could reconstruct the motion of a 3D human skeleton from a monocular video. Inspired by them, we obtain the Euler angles of the skeletal key points from the CNN outputs using inverse kinematics and then reconstruct the skeleton with forward kinematics.

Given the world positions of each joint, our first goal is to determine the joint angles. The midpoint of the left and right hip joints, referred to as *SpineBase*, is defined to be the root joint. We center the root joint to the origin of the world coordinate system, rotating to match the "T" pose, shown in Fig. 3. Each joint has its own coordinate system, and an offset value with respect to its parent joint.

The position of any joint can be expressed as

$$\overrightarrow{P_n} = \overrightarrow{P}_{parent(n)} + R_n\overrightarrow{j_n}, \tag{1}$$

here $\overrightarrow{P_n}$ is the coordinates of the n-th joints while $\overrightarrow{j_n}$ is the joint offset from the parent joint axes. Using *Knee* joint as an example

$$\overrightarrow{P}_{Knee} = (\overrightarrow{P}_{SpineBase} + R_{SpineBase}\overrightarrow{j}_{Hip}) + R_{SpineBase}R_{Hip}\overrightarrow{j}_{Knee}. \tag{2}$$

When calculating the angles between each of the joints, we subtract the root position from every joint position and center it at the world space origin. We start from the joints closest to the root joint and then get the subsequent joints' rotation matrix using the kinematics chain of the skeleton [13]. Using *Knee* joint as an example, if we know the positions of the *SpineBase* (root), *Hip*, and *Knee*, we could obtain the rotation matrix from

$$R_{Hip}\overrightarrow{j}_{Knee} = R_{root}^{-1}(\overrightarrow{P}_{Knee} - \overrightarrow{P}_{Hip}). \tag{3}$$

We then computed the Euler angles from the rotation matrix. A rotation of θ_x radians about the x-axis, θ_y radians about the y-axis, θ_z radians about the z-axis are defined as $R_x(\theta_x)$, $R_y(\theta_y)$, $R_z(\theta_z)$, correspondingly. As matrix multiplication is not commutative, the resulting rotation matrix depends on the order in which the three rotations are performed. We choose to use a zxy Euler angle rotation matrix

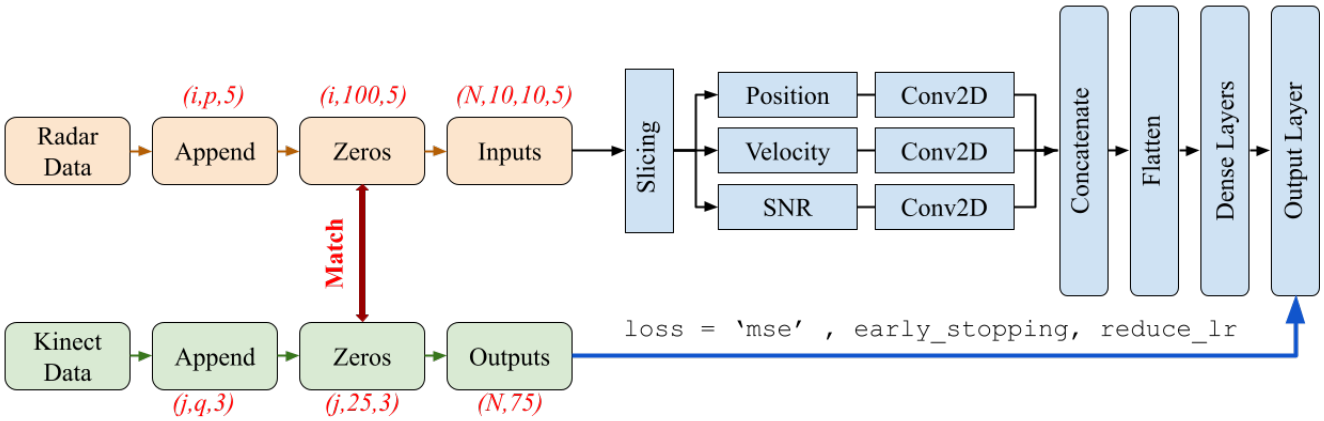

Fig. 1. CNN network architecture.

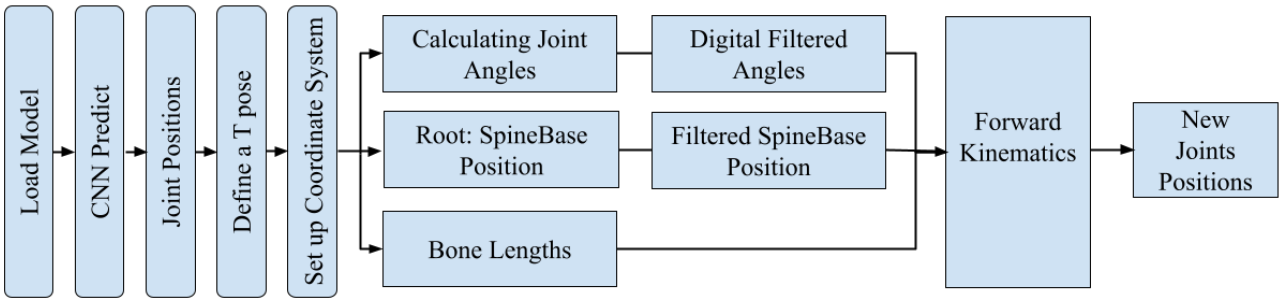

Fig. 2. Forward kinematic workflow.

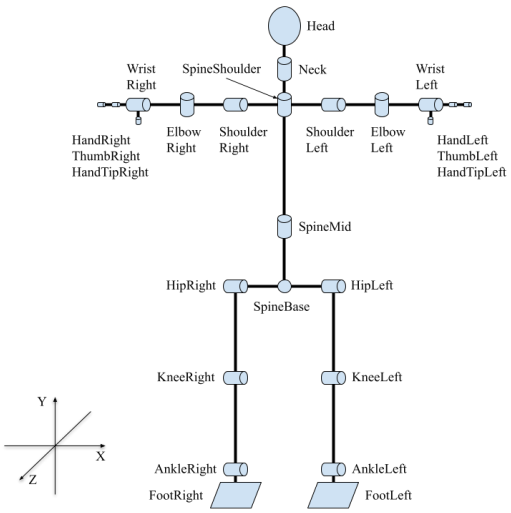

Fig. 3. T pose and coordinate system.

$$
\begin{aligned}
R &= R_z(\theta_z)R_x(\theta_x)R_y(\theta_y) \\
&= \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta_z\cos\theta_y - \sin\theta_z\sin\theta_x\sin\theta_y & -\sin\theta_z\cos\theta_x & -\cos\theta_z\sin\theta_y + \sin\theta_z\sin\theta_x\cos\theta_y \\ \sin\theta_z\cos\theta_y + \cos\theta_z\sin\theta_x\sin\theta_y & \cos\theta_z\cos\theta_x & \sin\theta_z\sin\theta_y - \cos\theta_z\sin\theta_x\cos\theta_y \\ -\cos\theta_x\sin\theta_y & \sin\theta_x & \cos\theta_x\cos\theta_y \end{bmatrix}.
\end{aligned} \tag{4}
$$

To simplify the expression, the above rotation matrix could be expressed as

$$
\begin{bmatrix} R_{00} & R_{01} & R_{02} \\ R_{10} & R_{11} & R_{12} \\ R_{20} & R_{21} & R_{22} \end{bmatrix}. \tag{5}
$$

We could get the joint Euler angles θ_x, θ_y and θ_z by

Case 1: If $\theta_x \in (-\pi/2, \pi/2)$,

$$
\theta_x = \arcsin(R_{21}),\ \theta_z = \arctan 2(-R_{01}, R_{11}),\ \theta_y = \arctan 2(-R_{20}, R_{22}). \tag{6}
$$

Case 2: If $\theta_x = \pi/2$,

$$
\theta_x = +\pi/2,\ \theta_y + \theta_z = \arctan 2(R_{02}, R_{00}). \tag{7}
$$

Case 3: If $\theta_x = -\pi/2$,

$$
\theta_x = -\pi/2,\ \theta_y - \theta_z = \arctan 2(R_{02}, R_{00}). \tag{8}
$$

C. Digital Signal Filters

In the field of digital signal processing, filters are used to remove unwanted parts of the signal and are widely used in applications such as noise reduction, radar, audio and video processing, etc. In this paper, three types of filters i.e., Elliptic, Savitzky-Golay, and Whittaker-Eilers have been discussed theoretically and experimentally.

Elliptic filter [14] can also be called as Cauer filters. Elliptic filters are equiripple in both the passband and stopband. The gain of a lowpass elliptic filter is given by:

$$G_n(\omega) = \frac{1}{\sqrt{1+\epsilon^2 R_n^2(\xi, \omega/\omega_0)}}, \tag{9}$$

where R_n is an n^{th} order elliptic rational function, which is set to 4 in our experiment. ω_0 is the cut-off frequency in half-cycles/sample - set to 0.125. ϵ is the ripple factor - set to 0.01 decibels. And ξ is the selectivity factor - set to 120 decibels.

Savitzky-Golay filter [15] is very popular for smoothing and interpolation of noisy data, based on local least-squares polynomial while maintaining the shape and height of waveform peaks [16]. For a sequence of samples $x[n]$, N is the total samples number, we obtain the coefficients of a polynomial

$$p(n) = \sum_{k=0}^{N} a_k n^k. \tag{10}$$

The minimum mean-squared approximation error could be calculated by

$$\varepsilon_N = \sum_{n=-M}^{M} \left(p(n) - x[n]\right)^2 = \sum_{n=-M}^{M} \left(\sum_{k=0}^{N} a_k n^k - x[n]\right)^2, \tag{11}$$

where M is the "half-width" of the approximation interval. In our experiment, we set the length of the filter window equal to 5, and the order of the polynomial to 2.

Whittaker presented an extended idea based on penalized least squares [17]. It has several advantages, such as automatic adaptability to boundaries, handling missing values automatically, extremely fast cross-validation with sparse matrices, etc. Suppose given a noisy digital signal Y of length M, assumed that the data have been sampled at equal intervals. We define the roughness of Z

$$R = \sum_i \left(Z_i - Z_{i-1}\right)^2. \tag{12}$$

The differences of signal and smooth series can be measured as

$$S = \sum_i \left(Y_i - Z_i\right)^2. \tag{13}$$

In order to keep both the fidelity to the data and the roughness of smooth series, after combination, we get the sum

$$Q = S + \lambda R, \tag{14}$$

where, we choose $\lambda = 10$. The larger λ is, the stronger the influence of R on the goal Q and the smoother Z will be (at the cost of the fit to the data getting worse).

III. Experiments and Results

A. Human Skeleton Estimation

In this study, we use IWR6843ISK-ODS 60 GHz mmWave sensor, with 120° azimuth field-of-view (FoV) and a 120° elevation FoV. The detailed radar system design parameters are listed in Table I.

TABLE I
Parameters of mmWave Radar

Parameter (Units)	Value
Start Frequency (GHz)	60.6
Slope (MHz/μs)	54.725
Samples per chirp	96
Chirps per frame	288
Frame duration (ms)	50
Sampling rate (MHz)	2.950
Bandwidth (MHz)	2249
Range resolution (m)	0.084
Max Unambiguous Range (m)	7.2
Max Radial Velocity (m/s)	8.38
Velocity resolution (m/s)	0.17
Angle resolution (deg)	29

The Microsoft Kinect v2 sensor is used to obtain the Ground Truth (GT) skeletal key points during this experiment. To recover the human skeleton from the radar point cloud, we first preprocess the collected data, extract the required information, and perform radar and GT frames matching according to timestamp. It should be noted that in the process of preprocessing, we analyze the number of points in the newly synthesized frames. Empirically, we observed that in a typical frame we obtained a maximum of 98 points in the radar point cloud distribution, we set the frame points to a constant value of 100, for a consistent input size to the neural network. For cases where the number of points N is less than 100, the remainder of $100 - N$ elements is zero-padded, as shown in Fig. 1. Each point in the mmWave radar point cloud data comprises of 5 parameters, viz. 3-D spatial location, radial velocity, and Signal-to-Noise Ratio (SNR). The purpose of this study is to provide a general approach to stabilizing skeletal poses, so only the author is involved in data collection. The radar and Kinect sensors collected data simultaneously and were mounted on a stabilizer frame to ensure a constant translation offset between them. The author does a continuous walking motion in front of the system. A Windows 10 laptop is used to collect and store the data (since both data are on the same computer, so there is no time synchronization problem), and the radar and ground truth frame data are matched by timestamp. After radar and GT frames matching, we have 24719 train frame pairs and 4805 test frame pairs. We split the 24719 frames into 19775 for training, and 4944 for validation. After training the model, the predicted accuracy of Test data could be seen in Table. III, marked as "CNN_predict" with an average MAE of 2.85 cm, which is comparable to our previous method that also uses CNN [5]. However, we do not overemphasize the value of the obtained localization errors in this study.

B. Dynamics Modeling and Filters

We first compared the difference between GT and CNN output joint angles and observed that the angle fluctuation of

TABLE II
STANDARD DEVIATION COMPARISON OF THE JOINT ANGLE SLOPE

	SD of the angle slope (radius)				
Joint	GT	CNN_predict	Elliptic	Savitzky-Golay	Whittaker-Eilers
SpineMid	0.0044	0.0051	0.0021	0.0029	0.0015
SpineShoulder	0.0104	0.0135	0.0065	0.0082	0.0059
Neck	0.0091	0.0160	0.0058	0.0085	0.0047
ShoulderLeft	0.1022	0.1569	0.0844	0.1006	0.0733
ElbowLeft	0.0669	0.0527	0.0268	0.0329	0.0220
WristLeft	0.6891	0.0714	0.0292	0.0411	0.0238
ShoulderRight	0.1264	0.1725	0.0925	0.1103	0.0810
ElbowRight	0.1007	0.0657	0.0327	0.0401	0.0282
WristRight	0.9750	0.2457	0.1312	0.1574	0.1123
HipLeft	0.0440	0.0629	0.0348	0.0409	0.0304
KneeLeft	0.0539	0.0708	0.0367	0.0444	0.0316
AnkleLeft	0.2183	0.0532	0.0293	0.0347	0.0236
HipRight	0.0431	0.0659	0.0350	0.0422	0.0307
KneeRight	0.0512	0.0729	0.0388	0.0466	0.0333
AnkleRight	0.0923	0.0585	0.0250	0.0333	0.0209
Average SD	0.0494	0.0707	0.0374	0.0449	**0.0325**
Improvement	-	-	47.12%	36.45%	**54.05%**

the skeletal-pose motion recovered from CNN is significantly larger than that of the GT, which is shown in Fig. 4.

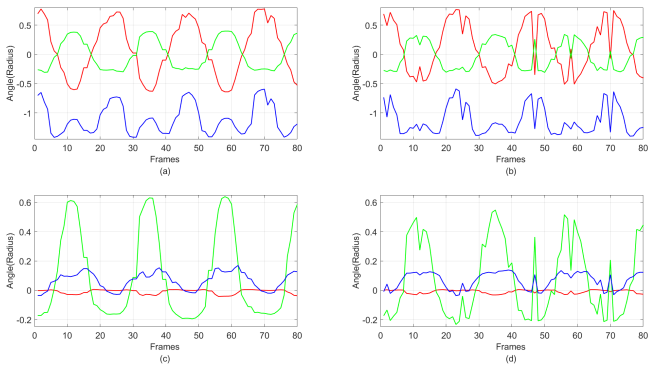

Fig. 4. Comparison between GT and CNN predict joint angles. Red: θ_x, Green: θ_y, Blue: θ_z. (a)(c): angle of GT (a) *ShoulderRight*, (c) *HipRight*; (b)(d): angle of CNN predict (b) *ShoulderRight*, (d) *HipRight*.

This phenomenon gives us the motivation to remove jitter by filtering. After the selected filters, we compared their effect of them on the same signal, e.g. Y-axis angle of *ShoulderRight*, as shown in Fig. 5 (a), (c), and (e). It can be seen that after filtering, some instantaneous jitter can be effectively attenuated. To evaluate the stabilization or smoothness, we use Standard Deviation (SD) of the angle slope and provide the plot of the frequency distribution, as shown in Fig. 5 (b), (d), and (f). From the figure, we can intuitively observe the distribution concentration: Whittaker-Eilers $>$ Elliptic $>$ Savitzky-Golay.

When recalculating the positions using forward kinematics, we also filtered the position of *SpineBase* joint, because the remaining skeletal keypoint locations are based on the position of the root node. We observed that the inter-frame stability significantly improved at the cost of a marginal decrease in localization accuracy. For example, the MAE increased from 2.85 cm to 3.34 cm after the Whittaker-Eilers filter. But as mentioned earlier, an increase of about 0.5cm in average

TABLE III
COMPARISON OF MAE OVER 25 JOINTS

MAE over 25 joints (cm)				
Data	Azimuth(X)	Elevation(Y)	Depth(Z)	Average
GT	-	-	-	-
CNN_predict	2.33	1.73	4.49	2.85
Elliptic	2.85	2.74	4.96	3.51
Savitzky-Golay	2.66	2.77	4.91	3.45
Whittaker-Eilers	2.67	2.70	4.65	**3.34**

localization errors for 25 joints is acceptable. Using this Formula,

$$I\% = \frac{|\overline{SD}_{Filter} - \overline{SD}_{CNN\_predict}|}{\overline{SD}_{CNN\_predict}} \tag{15}$$

three filters achieved 47.12%, 36.45%, and 54.05% improvement of smoothness ($I\%$), as shown in Table. II. Note that we list 15 main joints here, which we are more concerned with, including the joints on the torso, knee, and elbow joints. The recalculated MAE after filters of the CNN predicted data are 3.51 cm, 3.45 cm, and 3.34 cm in Table. III, with the Whittaker-Eilers filter offering the most optimum SD-MAE trade-off, for our study. Besides, we noticed that the average SD values after the filters is even smaller than the GT. We argue that the wrist and ankle joints from Kinect have their own abnormal jitter [18], so the average SD was enlarged.

IV. CONCLUSION

We introduced an approach to stabilize inter-frame skeletal-pose motion using mmWave radar via dynamic model and filtering. A CNN architecture with three separate convolutional layers to extract spatial position, velocity, and SNR features respectively, followed by a concatenation at a later stage, is implemented. A method that could evaluate the smoothness of motion, the SD of angle slope, was proposed with a motion stability improvement of 54.05% compared to the CNN predicted value. The improvement of motion stability plays a catalytic role in the application of radar-based activity recognition and real-time clinical health-care monitoring. In the future, we will use the obtained stable skeleton to model the live posture/behavior of patient to detect gait, postural transition, and physical frailty parameters. It provides the possibility for mmWave radar products to be applied in the field of tele-medicine and remote diagnosis. Furthermore, we will use the transition pattern between activities to determine and predict the forthcoming behavior of the patient. In addition, we hope that on another level, to perform fall risk analysis and implement proactive strategies to prevent falls.

REFERENCES

[1] J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, and J. Wang, "Wearable sensors: modalities, challenges, and prospects," *Lab on a Chip*, vol. 18, no. 2, pp. 217–248, 2018.

[2] R. Opromolla, G. Fasano, G. Rufino, and M. Grassi, "Uncooperative pose estimation with a lidar-based system," *Acta Astronautica*, vol. 110, pp. 287–297, 2015.

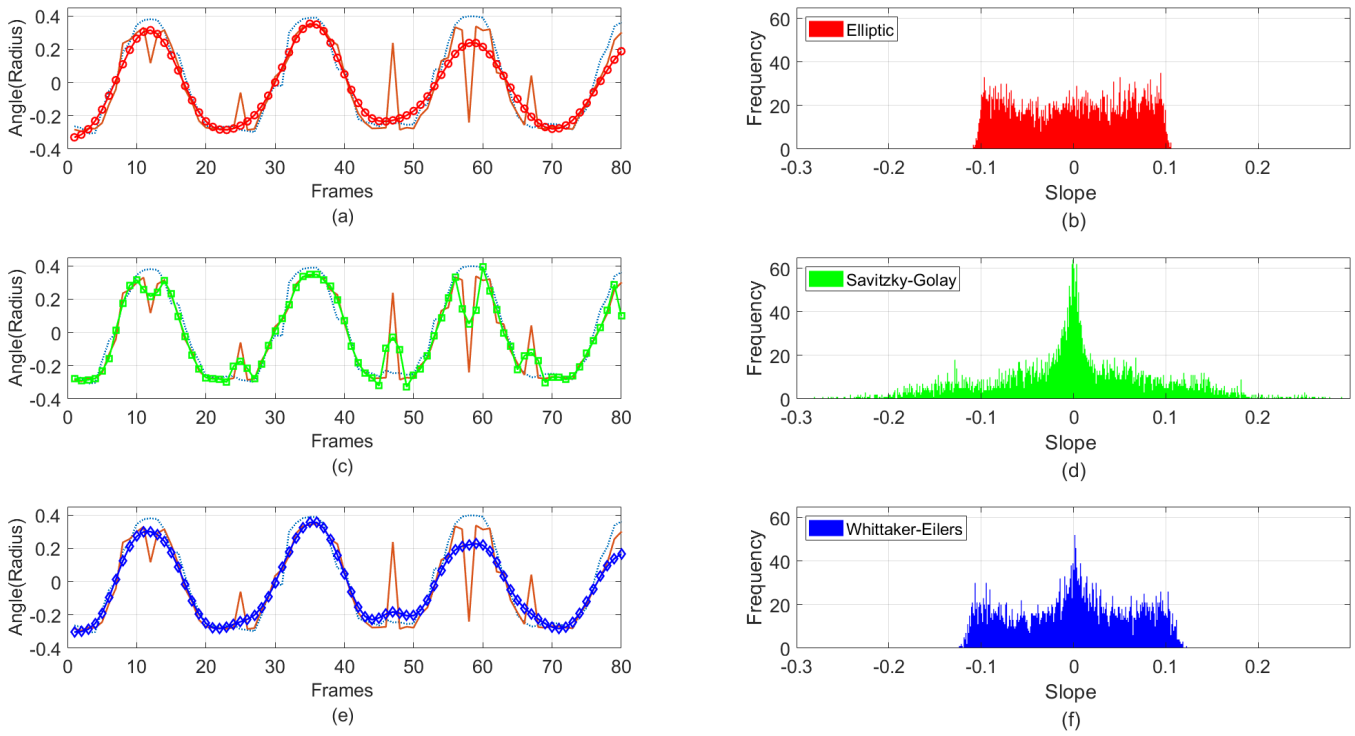

Fig. 5. Comparison of Filters and the Distribution of Angle Slope. Dotted line: GT, Solid line: CNN_predict, (a) Solid line with Circle: Elliptic, (c) Solid line with Square: Savitzky-Golay, (e) Solid line with Diamond: Whittaker-Eilers.

[3] R. Zhang and S. Cao, "Real-time human motion behavior detection via cnn using mmwave radar," *IEEE Sensors Letters*, vol. 3, no. 2, pp. 1–4, 2018.

[4] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," in *Proceedings of ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 51–56.

[5] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 032–10 044, 2020.

[6] S. An and U. Y. Ogras, "Mars: mmwave-based assistive rehabilitation system for smart healthcare," *ACM Transactions on Embedded Computing Systems*, vol. 20, no. 5s, pp. 1–22, 2021.

[7] A. Sengupta and S. Cao, "mmpose-nlp: A natural language processing approach to precise skeletal pose estimation using mmwave radars," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[8] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.

[9] S. Feng, E. Whitman, X. Xinjilefu, and C. G. Atkeson, "Optimization-based full body control for the darpa robotics challenge," *Journal of field robotics*, vol. 32, no. 2, pp. 293–312, 2015.

[10] W. Ding, Z. Cao, J. Zhang, R. Chen, X. Guo, and G. Wang, "Radar-based 3d human skeleton estimation by kinematic constrained learning," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23 174–23 184, 2021.

[11] R. Villegas, J. Yang, D. Ceylan, and H. Lee, "Neural kinematic networks for unsupervised motion retargetting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8639–8648.

[12] M. Shi, K. Aberman, A. Aristidou, T. Komura, D. Lischinski, D. Cohen-Or, and B. Chen, "Motionet: 3d human motion reconstruction from monocular video with skeleton consistency," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 1, pp. 1–15, 2020.

[13] T. Sugihara and M. Morisawa, "A survey: dynamics of humanoid robots," *Advanced Robotics*, vol. 34, no. 21-22, pp. 1338–1352, 2020.

[14] P. Podder, M. Hasan, M. Islam, and M. Sayeed, "Design and implementation of butterworth, chebyshev-i and elliptic filter for speech signal analysis," *arXiv preprint arXiv:2002.03130*, 2020.

[15] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[16] R. W. Schafer, "What is a savitzky-golay filter?" *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.

[17] P. H. Eilers, "A perfect smoother," *Analytical chemistry*, vol. 75, no. 14, pp. 3631–3636, 2003.

[18] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, "Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study," *Sensors*, vol. 20, no. 18, p. 5104, 2020.