2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Vision-based human joint angular velocity estimation during squat and walking on a treadmill actions

Konki Sravan Kumar[1], Ankhzaya Jamsarndorj[2], Dawoon Jung[1], Daehyun Lee[3], Jinwook Kim[1] and Kyung-Ryoul Mun[1]

Abstract— **Elderly health monitoring, rehabilitation training, and sport supervision could benefit from continuous assessment of joint angle, and angular velocity to identify the joint movement patterns. However, most of the measurement systems are designed based on special kinematic sensors to estimate angular velocities. The study aims to measure the lower limb joint angular velocity based on a 2D vision camera system during squat and walking on treadmill action using deep convolution neural network (CNN) architecture. Experiments were conducted on 12 healthy adults, and six digital cameras were used to capture the videos of the participant actions in lateral and frontal view. The normalized cross-correlation (Cc$_{norm}$) analysis was performed to obtain a degree of symmetry of the ground truth and estimated angular velocity waveform patterns. Mean Cc$_{norm}$ for angular velocity estimation by deep CNN model has higher than 0.90 in walking on the treadmill and 0.89 in squat action. Furthermore, joint-wise angular velocities at the hip, knee, and ankle joints were observed and compared. The proposed system gets higher estimation performance under the lateral view and the frontal view of the camera. This study potentially eliminates the requirement of wearable sensors and proves the applicability of using video-based system to measure joint angular velocities during squat and walking on a treadmill actions.**

I. INTRODUCTION

Estimation of lower limb Kinematics such as joint angular velocity is a significant clinical parameter that can help to assist elderly daily walking [1, 2] determine the progress during the rehabilitation of patients [3], improve athletes performance [4, 5], enhance human health, develop the sports gear equipment, diagnose early human joint movement disorders and recognize human activity [6]. Wearable sensor-based systems typically measure joint angular velocity. Wearable sensor-based systems employ sensors attached to human joints to extract joint movement information. Inertial measurement units (IMU) such as accelerometers, gyroscopes and magnetometers are effective, practical and powerful tools for human angle estimation and angular velocity measurements, drawing much attention from researchers [7–10]. The IMU sensors have been widely used as an alternative to the optoelectronic motion capture system, a gold standard for human joint kinematics measurements and showed sufficient reliability. However, one IMU sensor can only measure one body segment's motion, so usually, seven IMU sensors are required just for a lower limb motion such as pelvis, ankle, knee, and hip of each left and right side. When many IMU sensors are used, complex preprocessing processes such as sensor-to-segment calibration, subject calibration, and human kinematic chains should be considered to measure accurate joint kinematics. In addition, the use of motion tracking systems like IMU sensors hinder the natural movements of the subjects due to tight straps for the sensor attachment, and it is still expensive, complex handling of equipment, power consumption, participant cooperation and controlled laboratory restrictions [11]. To mitigate the limitations of wearable sensor-based measurement systems, novel vision-based systems were used to estimate human low limb kinematics. Recently emerged deep learning frameworks had opened a new possibility of human joint angle estimation from images captured by RGB-D cameras [12]. However, the Kinect depth image database suffers from sparse and noise. A 2D single-camera gait kinematics analysis system has been proposed in [13] to measure the knee angle. Estimate the joint angle from extracted center coordinates of the marker. But, marker tracker can only track the makers on camera facing lower limbs. It may not track the marker on the other views. Conversely, a marker-less system based on the OpenPose library was introduced for gait analysis [14, 15]. However, joint angles were extracted from the OpenPose algorithm based on the 2D keypoint datasets that are manually annotated and annotated keypoints are not center of the joint. Therefore, it may not be reliable to estimate angular velocity based on the Openpose library. The marker-less systems, which were mentioned above, can estimate the joint angles, but these were designed to track the participants only in side view of a camera, unable to distinguish between the lateral and frontal view (not view-invariant) and lacks in the adequate accuracy in joint angle estimation. Therefore, the motivation of the study was to design a view-invariant system for accurate joint angular velocity estimation.

This paper proposes a video-based joint angular velocity estimation system using deep learning techniques. In particular, we focused on estimating six joint angular velocities of the three joints located on the right and the left side of human lower limbs. In joint angular velocity measurement during daily life actions, videos of a healthy adult were acquired using digital cameras during squat and walking on treadmill actions. Estimation results of the proposed architecture were analyzed and compared with ground truth angular velocities tracked by the motion analysis system.

[1]Center for Artificial intelligence, Korea Institute of science Technology, Seoul, Republic of Korea (e-mail: krmoon02@kist.re.kr).
[2]Department of Human Computer Interface Robotics Engineering, University of Science Technology, Daejeon, South Korea.
[3]Department of Biomedical Science and Technology, Graduate School, Kyung Hee University, Seoul, Korea.

II. METHODS

A. Experimental Configuration

To collect the dataset, experiments were performed in an experimental room installed with the motion analysis system, which has 8 optical cameras and a frequency of 120Hz. It tracks markers attached to the human body's lower limbs according to the Plug-in-gait marker set and carries out the accurate angular velocity of all joints during the experiments.The angular velocity is calculated by derivation of the angles. Fifteen markers were placed on the left and right side of the hip, knee and ankle joint to capture the lower body motion, as shown in Fig. 1. Xiaomi mijia 4K system has 6 digital cameras with a frame rate of 60Hz, and a resolution of 1920×1080 was used to capture videos of 12 healthy adults during squat and walking on treadmill actions within 30 seconds. There are four cameras placed in front, and two cameras were located on the side of a participant with the capture area has a size of 2.000 meters by 3.685 meters on the floor, as shown in Fig. 1. We designed a circuit to receive the flash signal from a flash device to synchronize digital cameras and optical cameras. When the flash fires, a photodiode will generate an electric current. An operational amplifier converts the small amount of current to the Transistor–transistor Logic (TTL) level signal. When the signal changes abruptly, the microcontroller (Arduino Nano) will generate pulses lasting in 100ms. The motion capture system which was used in this experiment is triggered a rising edge pulse. Furthermore, the digital cameras will capture the flashing moment. Hence, this signal was used to cut videos corresponding to actions and subjects. We calculated the average intensity of pixels in each frame, plotted them and find a threshold for flashing frames. Lastly, we down-sampled the frame rate of annotations from 120Hz to 60Hz because the frequency of optical cameras is twice the value for digital cameras.

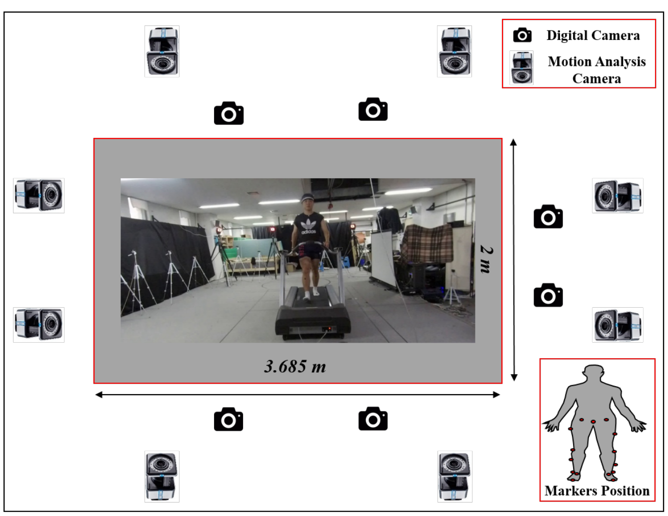

Fig. 1: In this figure, we show a scheme of the experimental setup. The healthy adults were walking on a treadmill at a comfortable speed. The digital cameras were located at the side and front of the treadmill. Six cameras are included to capture the subject during an action.

B. Joint angular velocity estimation based on deep learning architecture

In this section, we presented in detail the joint angular velocity estimation system. The framework of the proposed architecture is shown in Fig. 2. We built a joint angular velocity estimator to simultaneously estimate the joint angular velocities at hips, knees, and ankles on both the right and left sides. A sequence of 9 lateral or frontal-view video frames during the action as features and the corresponding target joint angular velocities are fed to the deep-leaning networks in a supervised manner. Initially, we employed a 4^{th} version of the YOLO (You Only Look Once) object detection algorithm to detect the human in the action videos [16]. Subsequently, the images from videos were cropped based on resultant human bounding box detection by the YOLO algorithm and then padded depending on the bounding box's size to obtain a uniform square image in action videos. Finally, the padded images were resized to a size of 448 × 448. In this study, we employ three deep convolutional neural networks modified from Densenet121, ResNet18, ResNet50 architectures [17] to address the problem of joint angular velocity estimation. To make the network appropriate for our problem, we replaced the last fully connected (FC) layer with six neurons. The six neurons correspond to all the six joint angular velocities such as left/ right hip flexion/extension, knee flexion/extension, and ankle dorsiflexion/plantarflexion.

C. Network implementation and training scheme

The proposed deep CNN networks were implemented using the PyTorch deep learning open-source library in Python with the mean absolute error (MAE) as the loss function and Adaptive Moment Estimation (Adam) for optimization on a PC configured with an Intel® Core™ i7-9700 CPU at 3.00 GHz x 8, using Ubuntu 18.04.5 LTS. The Densenet121 model was first trained on the ImageNet dataset [18] and then utilized the pre-trained model to initialize the joint angular velocity estimation network. We choose a learning rate of 0.0001, batch size of 8, dropout probability at the last layer as 0.5 and trained a network using a stack of 9 consequent video frames with corresponding ground truth angular velocities in 20 epochs. There are videos of six healthy subjects during a squat and walking on a treadmill, intended for training, videos of three healthy subjects for validation and three healthy subjects for testing. This new dataset contains 203160 frames, 106920 for training, 48120 for validation and 48120 for testing.

III. EXPERIMENTAL RESULTS

This section demonstrates detailed explanations of experimental results to estimate joint angular velocity using deep learning architectures. The ground-truth and the estimated joint angular velocities at the hip, knee, and ankle joint by the different deep CNN architectures as shown in Fig. 3. The estimated results for joint angular velocities at the hip, knee, and ankle joint during squat action are displayed in Fig. 4. From Fig. 3. and Fig. 4., we have noticed that proposed deep CNN models show good angular velocity estimation.

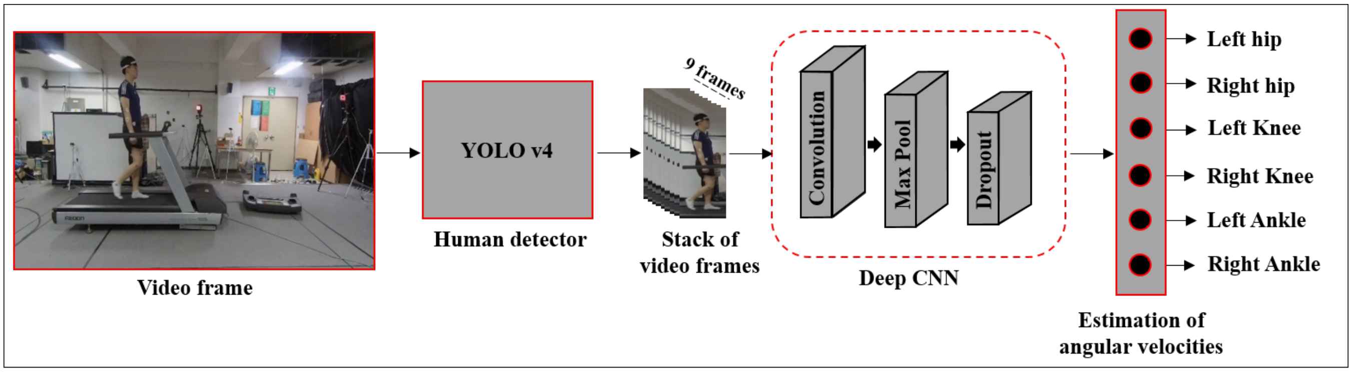

Fig. 2: The framework of the proposed angular velocity estimator: YOLO algorithm detects the human in the action videos and deep CNN extracts 2D convolutional features and estimate the joint angular velocities.

To perform the quantitative analysis, we calculated Normalized cross-correlation [19] for angular velocity estimation using deep CNN models during squat and walking on a treadmill as presented. Cc_{norm} analysis for the joint-wise angular velocities is used to obtain a degree of symmetry of the ground truth and estimated angular velocity waveform patterns. Average Cc_{norm} of angular velocities using three deep CNN models as presented in Table I. DenseNet121 model produces higher Cc_{norm} of 0.89 and 0.91 during squat and walking on a treadmill, respectively.

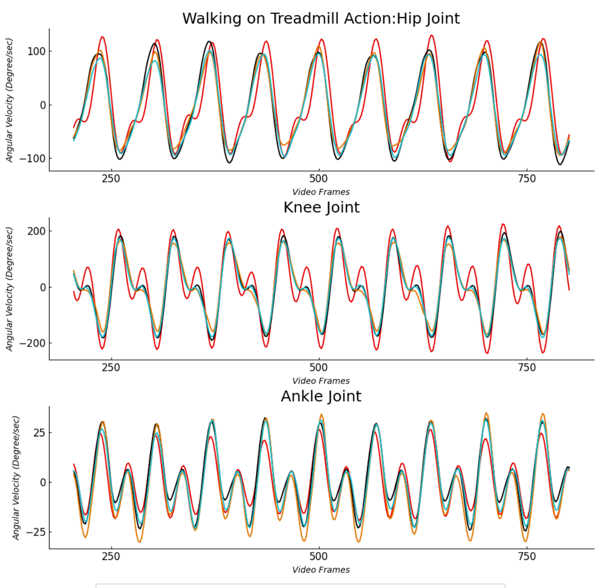

Fig. 3: This figure reports an estimated angular velocity at the hip, knee, and ankle joint during walking on a treadmill action. Redline represents ground truth angular velocities, the black line with ResNet18, the orange line with ResNet50 and the cyan line is estimated using the DenseNet121 model.

TABLE I: An average Cc_{norm} for estimated angular velocities during squat and walking on a treadmill by deep CNN architectures.

Model	Squat	Walking on treadmill
ResNet18	0.87	0.90
ResNet50	0.88	0.90
DenseNet121	0.89	0.91

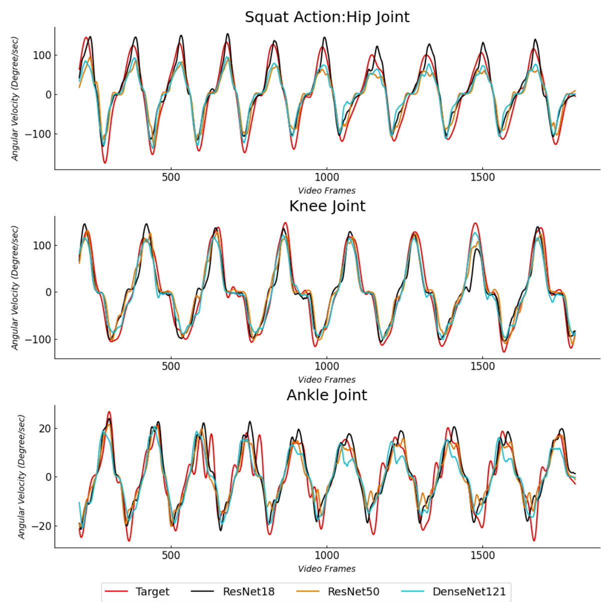

Fig. 4: This figure shows an estimated angular velocity at the hip, knee, and ankle joint during squat action. Redline represents ground truth angular velocities, black line with ResNet18, orange line with ResNet50 and the cyan line is estimated using the DenseNet121 model.

An average Cc_{norm} for the joint-wise angular velocity at hip, knee and ankle during walking on a treadmill was estimated using different deep CNN architectures as shown in Fig. 5 and in Table II. DenseNet121 model produces higher Cc_{norm} of 0.87, 0.95 and 0.86 at hip, knee and ankle joint, respectively. Fig. 6 and Tables III illustrates an average Cc_{norm} of angular velocities at hip, knee and ankle joint during squat were estimated using different deep CNN architectures. The mean Cc_{norm} for angular velocity estimation under different camera views is demonstrated in Table IV. An average Cc_{norm} for side and front view are 0.90 and 0.87, respectively.

TABLE II: An average Cc_{norm} analysis of joint wise angular velocities at hip, knee and ankle joint during walking on a treadmill.

Model	Hip Joint	Knee Joint	Ankle Joint
ResNet18	0.80	0.94	0.84
ResNet50	0.85	0.94	0.84
DenseNet121	0.87	0.95	0.86

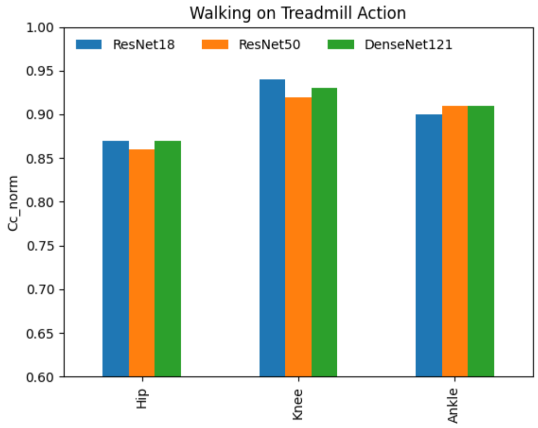

Fig. 5: An average Cc_{norm} analysis of the joint-wise angular velocities at hip, knee and ankle joint during walking on a treadmill action. Blue bar with ResNet18, orange bar with ResNet50 and green bar is estimated using the DenseNet121 model.

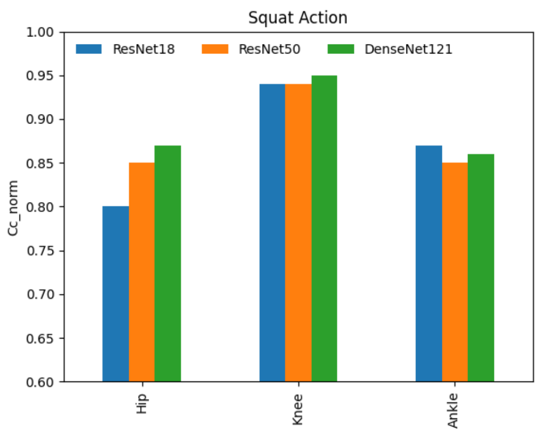

Fig. 6: An average Cc_{norm} analysis of the joint-wise angular velocities at hip, knee and ankle joint during squat action. Blue bar with ResNet18, orange bar with ResNet50 and green bar is estimated using the DenseNet121 model.

To evaluate the accuracy of the proposed models, we compare the results in terms of MAE for joint-wise angular velocity during squat and walking on a treadmill as shown in Table V and Table VI. The MAE was used to compare the performance of the models for predicting the error between the estimated and ground-truth angular velocities. From Table V, DenseNet121 model gives MAE of 37.45, 29.44 and 5.93 at hip, knee and ankle joint during a squat action, respectively. Table 6, DenseNet121 model gives MAE of hip, knee and ankle joint during walking on a treadmill are 27.14, 37.36 and 6.06.

TABLE III: An average Cc_{norm} analysis of joint wise angular velocities at hip, knee and ankle joint during a squat action.

Model	Hip Joint	Knee Joint	Ankle Joint
ResNet18	0.87	0.94	0.90
ResNet50	0.86	0.92	0.91
DenseNet121	0.87	0.93	0.91

TABLE IV: An average Cc_{norm} for estimated angular velocities of all healthy subjects compared with under different camera views.

Subject	Side view	Front view
S18	0.89	0.90
S30	0.91	0.82
S31	0.90	0.91
Average	0.90	0.87

TABLE V: The MAE for estimated joint-wise angular velocities of all three deep CNN was compared during a squat action.

Model	Hip Joint	Knee Joint	Ankle Joint
ResNet18	43.04	30.08	6.07
ResNet50	41.26	33.40	6.28
DenseNet121	37.45	29.44	5.93

TABLE VI: The MAE for estimated joint-wise angular velocities of all three deep CNN compared during walking on a treadmill.

Model	Hip Joint	Knee Joint	Ankle Joint
ResNet18	32.92	42.85	6.06
ResNet50	28.66	40.29	6.88
DenseNet121	27.14	37.36	6.06

IV. DISCUSSION

This study presented a video-based system using deep learning techniques to accurately measure the lower limb joint angular velocities during a squat and walking on the treadmill action. The deep CNN models are trained by action videos frames features and corresponding target angular velocity of joints. This proposed systems can estimate the joint angular velocity of the hip flexion/extension, the knee flexion/extension, and the ankle dorsiflexion/plantarflexion without the use of any wearable sensors. The estimated joint angular velocities and ground truths were graphically compared to investigate the estimation performance of deep CNN models as shown in Fig. 3 and Fig. 4. Estimated angular velocities well follow its ground truth profile during both actions. Estimation accuracies of all deep CNN models were evaluated using the Cc_{norm} and MAE. Cc_{norm} measures the degree of symmetry of the ground truth, and estimated angular velocities. Cc_{norm} computation generates the values range from 0 to 1, and a Cc_{norm} value close to 1 shows that both ground truth and estimated angular velocity have strong symmetry. Estimated angular velocities of all deep CNN models were highly correlated with ground truth. However, DenseNet121 model reports better estimation with more than 0.90 average Cc_{norm} in walking on a treadmill and 0.89 during squat actions as presented in Table I. The performance of the proposed models was further investigated based on the hip, knee, and ankle joint angular velocity. As shown in Fig. 5 and Table II, DenseNet121 model depicts a better average Cc_{norm} of three joints during walking on a treadmill action. Furthermore, the knee joint shows a higher correlation of 0.95. Similarly, DenseNet121 model gets higher accuracies than other models and knee joint reported better estimation with 0.93 during squat action as shown in Fig. 6 and Table III. It was indicated that the knee joint gets higher estimation than hip and ankle joints. The knee joint has more stable ground truth angular velocity, ankle joint having a higher degree of freedom, the hip joint estimation accuracy may affect by variations in human body size of subjects, and motion artifacts due to the fluctuations of the marker at the hip joint. As reported in Table IV, the mean Cc_{norm} in lateral and frontal view is about 0.90, indicating that the proposed system maintains relatively higher estimation accuracy in both camera views. Table V and Table VI demonstrate the MAE of all three joint-wise angular velocities estimated by ResNet18, ResNet50 and DenseNet121. We can notice that MAE for DenseNet121 exhibits lesser than ResNet18, ResNet50. However, all these deep CNN methods still showing higher error due to time delay in the estimation results, and an offset value error can be generated from the imprecise positions of makers on the human body, and the human body size and shape variation within the subjects. In our future work, the proposed pre-trained CNN model will be extended to use of complex recurrent, attention networks to obtain a temporal information for more stable angular velocity estimation over time. Furthermore, a larger dataset with numerous daily life actions and different age groups of subjects with gait pathologies is needed to evaluate the practical performance.

V. CONCLUSIONS

This study presents a vision-based angular velocity estimation system using deep learning techniques to estimate the lower limb joint angular velocities during a squat and walking on a treadmill action. The main contribution of this paper is to employ a modified deep CNN architecture to measure the joint angular velocity of the hip, knee, and ankle joints. Additionally, we constructed a new dataset of walking on a treadmill and squat action. The deep CNN models are trained by action videos frames features and corresponding target angular velocity of joints. This proposed estimation system offers good estimation accuracy under different camera view angles. Our preliminary study has proven the applicability of using action video to estimate angular velocity based on CNN. However, this vision-based system is still showing higher MAE errors due to time delays in estimations. Therefore, the proposed pre-trained CNN model will be extended to use of complex recurrent, attention networks to obtain a temporal information for more stable angular velocity estimation over time.

ACKNOWLEDGMENT

This research was supported by the Korea Institute of Science and Technology Institutional Program (Project No. 2E31051), and the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 1711139131).

REFERENCES

[1] J. P. Batista, L. M. Facci, A. R. M. Pelegrinelli, D. C. D. Souza, M. Taglietti, J. M. Dias, C. Disselhorst-Klug, and J. R. Cardoso, "Joint angle and movement velocity effects on muscle activity of elderly with knee osteoarthritis categorized and probabilistic analysis," *J Electromyogr Kinesiol*, vol. 41, pp. 58–59, Aug. 2018.

[2] L. Jin and M. E. Hahn, "Comparison of lower extremity joint mechanics between healthy active young and middle age people in walking and running gait," *Scientific Reports*, vol. 9, no. 1, Apr 2019.

[3] Y. Huang, Y. Liu, W. Hsu, L. Lai, and M. S. Lee, "Monitoring and assessment of rehabilitation progress on range of motion after total knee replacement by sensor-based system," *Sensors*, vol. 20, no. 6, p. 1703, Mar 2020.

[4] A. Struzik, G. Konieczny, M. Stawarz, K. Grzesik, S. Winiarski, and A. Rokita, "Relationship between lower limb angular kinematic variables and the effectiveness of sprinting during the acceleration phase," *Applied Bionics and Biomechanics*, vol. 2016, pp. 1–9, Jul 2016.

[5] D. Hahn, W. Herzog, and A. Schwirtz, "Interdependence of torque, joint angle, angular velocity and muscle action during human multi-joint leg extension," *European Journal of Applied Physiology*, vol. 114, no. 8, pp. 1691–1702, Aug 2014.

[6] F. Ince, I. F. Ince, M. E. Yıldırım, J. S. Park, J. K. Song, and B. W. Yoon, "Human activity recognition with analysis of angles between skeletal joints using a rgb-depth sensor," *ETRI Journal*, vol. 42, no. 1, pp. 78–89, Nov 2019.

[7] R. Williamson and B. J. Andrews, "Detecting absolute human knee angle and angular velocity using accelerometers and rate gyroscopes," *Medical Biological Engineering Computing*, vol. 39, no. 3, pp. 294–302, May 2001.

[8] A. Mannini, V. Genovese, and A. M. Sabatin, "Online decoding of hidden markov models for gait event detection using foot-mounted gyroscopes," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1122–1130, Jul 2014.

[9] M. Versteyhe, H. D. Vroey, F. Debrouwere, H. Hallez, and K. Claeys, "A novel method to estimate the full knee joint kinematics using low cost imu sensors for easy to implement low cost diagnostics," *Sensors*, vol. 20, no. 6, p. 1683, Jan 2020.

[10] M. Mundt, A. Koeppe, S. David, T. Witter, F. Bamer, W. Potthast, and B. Markert, "Estimation of gait mechanics based on simulated and measured imu data using an artificial neural network," *Frontiers in Bioengineering and Biotechnology*, vol. 8, no. 41, Feb 2020.

[11] M. Schepers, M. Giuberti, and G. Bellusci, "Xsens mvn: Consistent tracking of human motion using inertial sensing," *Xsens Technologies*, pp. 1–8, Mar 2018.

[12] A. Guler, N. Kardaris, S. Chandra, V. Pitsikalis, C. Werner, K. Hauer, C. Tzafestas, P. Maragos, and I. Kokkinos, "Human joint angle estimation and gesture recognition for assistive robotic vision," in *2016 Lecture Notes in Computer Science Computer Vision – ECCV 2016 Workshops*, Nov 2016, pp. 415–431.

[13] C. Yang, U. C. Ugbolue, A. Kerr, V. Stankovic, L. Stankovic, B. Carse, K. T. Kaliarntas, and P. J. Rowe, "Autonomous gait event detection with portable single-camera gait kinematics analysis system," *Journal of Sensors*, vol. 2016, pp. 1–8, Jan 2016.

[14] A. Schmitz, R. S. M. Ye, R. Yang, and B. Noehren, "Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system," *Journal of Biomechanics*, vol. 47, no. 2, pp. 587–591, Nov 2014.

[15] E. D'Antonio, J. Taborri, E. Palermo, S. Rossi, and F. Patanè, "A markerless system for gait analysis based on openpose library," in *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2020, pp. 1–6.

[16] A. Bochkovskiy, C. Wang, and H. Liao, "Yolov4: Optimal speed and accuracy of object detection," *ArXiv*, vol. abs/2004.10934, 2020.

[17] Y. Que and H. J. Lee, "Densely connected convolutional networks for multi-exposure fusion," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018, pp. 417–420.

[18] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[19] W. Sheng, W. Guo, F. Zha, Z. Jiang, X. Wang, and H. Zhang, "The effectiveness of gait event detection based on absolute shank angular velocity in turning," in *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2019, pp. 899–904.