Kalman Filter-Enhanced MediaPipe Pose Estimation for Quantitative Assessment of Taekwondo Training Strategies

Hsin-Yi Zheng
Department of Computer Science
University of Taipei
Taipei, Taiwan
a0983344722@gmail.com

YI-ZHEN YANG
Department of Physical Education
University of Taipei
Taipei, Taiwan
ok131055@utaipei.edu.tw

Jui-Chung Hung
Department of Computer Science
University of Taipei
Taipei, Taiwan
juichung@gmail.com

***Abstract*—In this paper address quantitative Assessment of Taekwondo coaching strategies. Based on MediaPipe Pose estimation, the system extracts precise body landmark coordinates from video recordings of fast and continuous Taekwondo actions. Due to the high-speed and dynamic nature of these movements, pose estimation often suffers from detection errors. To mitigate such inaccuracies, we propose a Kalman filter-based correction method, supplemented by an anomaly detection technique based on the Interquartile Range (IQR), to improve temporal coherence of the landmark data. Experimental evaluation shows that our approach corrects approximately 4.107% of erroneous MediaPipe detections, and further analysis reveals a 37.21% enhancement in movement stability metrics for the athletes.**

Keywords—MediaPipe, Kalman Filter, Taekwondo, Interquartile Range (IQR), Spectrum Analysis

I. INTRODUCTION

In recent years, the field of Human Action Recognition (HAR) technologies, the field has evolved from early reliance on wearable sensors and traditional machine learning methods to the use of deep learning for visual-based action analysis, resulting in significant improvements in both the accuracy and real-time performance of motion tracking[1]. Although object detection models such as YOLO can rapidly localize the human body, they only provide bounding box information and are unable to precisely analyze joint movements[2]. The development of OpenPose has enabled real-time localization and tracking of multiple human keypoints, driving progress in sports science and interactive systems[3]. However, the computational demands of OpenPose, particularl its relance on GPU resources, pose challenges on edge devices such as smartphones and wearables. MediaPipe Pose is a real-time pose estimation model based on the BlazePose architecture, integrating deep learning with efficient image processing techniques. This model is capable of detecting 33 human keypoints, with particular accuracy in capturing subtle movements of the face, hands, and lower limbs[4]. In sports analysis applications, MediaPipe Pose can quickly and accurately localize human keypoints on mobile devices, demonstrating excellent accuracy and real-time performance across various sports scenarios[5]. Therefore, this study adopts MediaPipe Pose as the tool for human pose data acquisition.

In the application of human pose detection and motion analysis, although MediaPipe's keypoint detection technology can provide real-time skeletal coordinates, practical deployment often encounters issues such as sensor noise, occlusion of body parts, and rapid movement changes. These may result in discontinuities or noticeable jitter in keypoint localization, leading to errors in keypoint estimation[6]. To enhance the stability and quality of keypoint data, this study introduces the Kalman Filter for correction. The Kalman Filter leverages the system's dynamic model, combining previous and current data to predict the next state, effectively smoothing continuous data and reducing errors caused by noise, occlusion, or abrupt motion changes[7], thereby producing more stable pose estimation results[8].

Furthermore, in this paper, the quality of motion tracking is evaluated through both time-domain and frequency-domain analyses to comprehensively examine the stability and consistency of athletes' movements. Time-domain analysis focuses on changes in motion over time and can be used to quantify features such as velocity, acceleration, mean, and standard deviation[9]. Frequency-domain analysis, through the use of Fast Fourier Transform (FFT), converts time series data into the frequency domain, revealing the periodicity of athletic movements[10].

II. PROPOSED METHOD

A. *Keypoint Data Acquisition and Preprocessing*

Human pose estimation was performed using the MediaPipe Pose framework, which is based on the BlazePose architecture and enables real-time detection of 33 human keypoints, as illustrated in Fig. 1. Considering the characteristics of the Taekwondo spinning kick, all video data were sampled at a frame rate of 60 frames per second. Each frame was extracted using OpenCV, and the corresponding keypoint coordinates were obtained through the MediaPipe Holistic model. To minimize proportional errors caused by variations in camera distance or angle, this study adopted the Euclidean distance between the right shoulder p_{12} and right hip p_{24} as a reference for coordinate normalization. To further analyze the characteristics of the spinning kick, the velocity of the right ankle p_{28} keypoint, which is primarily engaged during the kick, were calculated based on the normalized keypoint coordinates. The velocity was determined by dividing the displacement in pixels of the keypoint between adjacent frames by the time interval between frames (1/60 seconds).

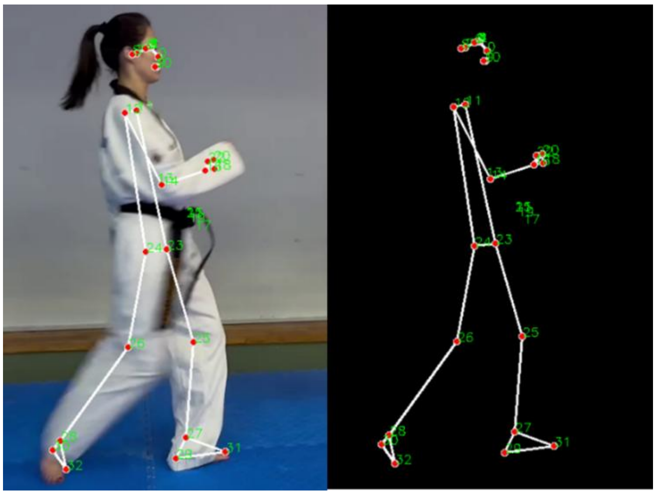

Fig. 1 Keypoints of the athlete performing the Taekwondo spinning kick were detected using MediaPipe Pose.

$$L_{ref} = \frac{1}{T}\sum_{t=1}^{T}\left\| p_{12}(t) - p_{24}(t) \right\| \tag{1}$$

$$\alpha(t) = \frac{L_{ref}}{\left\| p_{12}(t) - p_{24}(t) \right\|} \times H \tag{2}$$

$$\begin{aligned} &d_{p_{28}}(t) = \\ &\alpha(t)\cdot\sqrt{\left(x_{p_{28}}(t+1) - x_{p_{28}}(t)\right)^2 + \left(y_{p_{28}}(t+1) - y_{p_{28}}(t)\right)^2} \\ &, t = 1, 2, \ldots, S-1 \end{aligned} \tag{3}$$

$$v_{p_{28}}(t) = \frac{d_{p_{28}}(t)}{1/60}, t = 1, 2, \ldots, S-1 \tag{4}$$

where L_{ref} is the reference length value, T represents the number of frames during which the subject remains stationary prior to initiating the movement, and $\| p_{12}(t) - p_{24}(t) \|$ represents the pixel distance between the right shoulder and right hip, $\alpha(t)$ the ratio for converting pixel measurements to actual length, H represents the actual distance between the right shoulder and right hip , $d_{p_{28}}$ is the displacement of the right ankle keypoint between frame t and frame $t+1$, $v_{p_{28}}$ denotes the velocity, S is the total number of frames for each measurement.

B. *Anomaly Detection and Dynamic Correction*

Due to the rapid kicking motions of Taekwondo athletes, misidentification of keypoints can occur, as illustrated in Fig. 2. In the time domain, these errors are also evident as significant fluctuations in keypoint trajectories. Therefore, this study employs interquartile range (IQR) outlier detection and Kalman filter based dynamic correction to address keypoint anomalies.

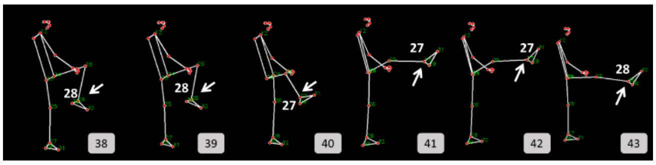

Fig. 2 Error detection of the right ankle keypoint

a. IQR anomaly detection

To detect anomalies, this study employs the IQR method to analyze outliers in the normalized velocity sequence of the right ankle keypoint, denoted as v_{28}. Specifically, the first quartile (Q_1, 25th percentile), the third quartile(Q_3, 75th percentile), and the interquartile range ($IQR = Q_3 - Q_1$) are used to measure the variability of the central 50% of the data.

To assess outliers within the dataset, the anomaly detection threshold is set at 1.5 times the IQR. This threshold is a classical statistical criterion for identifying outliers and is widely adopted in anomaly detection applications[11].

$$Threshold = Q_3 + 1.5 \times IQR \tag{5}$$

b. Kalman Filter Based Dynamic Correction

To correct the anomalies detected by the IQR method, this paper employs a first-order Kalman filter to construct a dynamic system model, using two-dimensional position and velocity as the state vector. This approach enables recursive estimation and correction of anomalies within the continuous time series data.

$$p(t+1)=\underbrace{\begin{bmatrix}1 & 0 & 1/60 & 0\\ 0 & 1 & 0 & 1/60\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}}_{A}p(t)+\mathbf{w}(t) \tag{6}$$

where $p(t)=\left[x(t)\quad y(t)\quad v_x(t)\quad v_y(t)\right]^t$ denotes the state vector at frame t, A is the state matrix, and $\mathrm{w}(t)$ represents Gaussian white noise with zero mean and unit variance.

In summary, this study detects keypoint anomalies using the IQR method. The Kalman filter, leveraging its ability to estimate future positions based on historical state information, is further applied for dynamic correction. This approach maintains the continuity of keypoint trajectories and ensures the stability of subsequent data.

C. *Time-Frequency Analysis*

To further investigate the changes in athletes' movements before and after Taekwondo training, this study evaluates the stability and consistency of the actions from different perspectives in both the time domain and the frequency domain.

In the time domain, the normalized displacement is calculated as shown in (3), and its mean and standard deviation are subsequently determined.

$$\bar{d}_{p_{28}}=\frac{1}{S}\sum_{t=1}^{S}d_{p_{28}}(t) \tag{7}$$

$$\sigma d_{p_{28}}=\sqrt{\frac{1}{S}\sum_{t=1}^{S}\left(d_{p_{28}}(t)-\bar{d}_{p_{28}}\right)^2} \tag{8}$$

where $d_{p_{28}}(t)$ denotes the displacement from frame t to frame $t+1$, and $\|\cdot\|$ represents the Euclidean distance, which is the distance between the coordinates of the right ankle keypoint at the current and subsequent frames. $\bar{d}_{p_{28}}$ indicates the mean displacement, while $\sigma d_{p_{28}}$ is the standard deviation of the displacement.

In the frequency domain, the y-axis coordinate of keypoint $p(t)$ is subjected to a Fast Fourier Transform (FFT), which is defined as follows:

$$P(f)=\sum_{t=0}^{S-1}p(t)\cdot e^{-j2\pi ft/S} \tag{9}$$

Here, $P(f)$ is the Discrete Fourier Transform of the sequence $p(t)$, S is the total number of frames. The analysis utilizes features such as the dominant frequency, peak power, spectral centroid, and spectral bandwidth.

$$f_{dominant}=f_{k_{max}},k_{max}=\arg\max_{k}\left|Y\left(f_k\right)\right| \tag{10}$$

where f_k is the k -th frequency component, $Y\left(f_k\right)$ represents the Fourier transform result at the k -th frequency, and $\left|Y\left(f_k\right)\right|$ is its magnitude.

$$P_{peak}=\left|Y\left(f_{k_{max}}\right)\right|^2 \tag{11}$$

P_{peak} represents the power at the dominant frequency component, with units in the square of the original signal unit (pixels squared).

$$C_{ratio}=\frac{\left|Y\left(f_{k_{max}}\right)\right|}{\sum_{k=0}^{S-1}\left|Y\left(f_k\right)\right|} \tag{12}$$

It represents the proportion of the dominant frequency magnitude to the sum of magnitudes across all frequencies.

$$f_{centroid}=\frac{\sum_{k=0}^{S/2-1}f_k\cdot\left|Y\left(f_k\right)\right|}{\sum_{k=0}^{S/2-1}\left|Y\left(f_k\right)\right|} \tag{13}$$

First, the spectral centroid is calculated as the weighted mean of the frequencies, where f_k denotes the k -th frequency component and $\left|Y\left(f_k\right)\right|$ represents the magnitude at the k -th frequency. Subsequently, the spectral bandwidth is computed as the weighted standard deviation of the frequencies relative to the centroid.

$$BW=\sqrt{\frac{\sum_{k=0}^{S/2-1}\left(f_k-f_{centroid}\right)^2\cdot\left|Y\left(f_k\right)\right|}{\sum_{k=0}^{S/2-1}\left|Y\left(f_k\right)\right|}} \tag{14}$$

where. Since the Fourier transform of a real-valued signal yields a conjugate symmetric spectrum, this study analyzes only the spectrum from 0 to the Nyquist frequency[12].

By integrating time-domain and frequency-domain analyses, the results and effects of coaching on Taekwondo athletes can be reflected both temporally and spectrally.

III. EXPERIMENTAL RESULT

This study conducted a three-month kicking training follow-up on two Taekwondo athletes: Student A, a third-year university student with a height of 179 cm, a weight of 49 kg, and a best competition result of third place at the National High School Games; and Student B, a first-year university student with a height of 174 cm, a weight of 55 kg, and a best competition result also of third place at the National High School Games.

For the correction of keypoint anomalies, Fig. 3 illustrates the temporal variation of Student A's right ankle during the second training session. In the figure, the gray line represents the raw data, the orange line indicates the data after correction by the Kalman filter, and the red crosses mark the anomalies detected by the IQR method. In Fig. 4, 1.62% of the data points are identified as anomalies, while 4.13% of the keypoints in Student B's total training data were corrected.

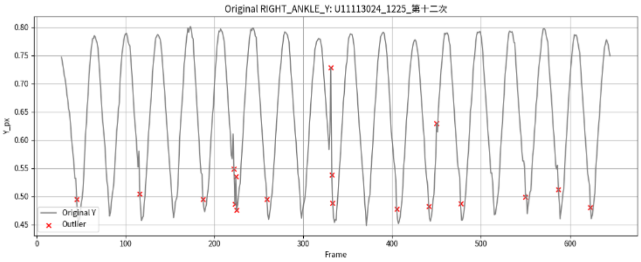

Fig. 3 Time-Domain Plot Before Correction

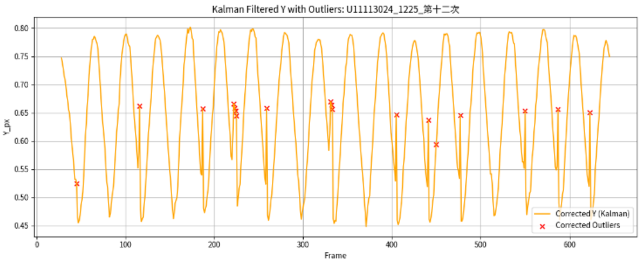

Fig. 4 Time-Domain Plot After Correction by the Kalman Filter

In Fig. 3, the red crosses indicate the original erroneous values of the detected anomalies, while Fig. 4 presents the values of these anomalies after correction.

In the time-domain analysis, the temporal characteristics of the kicking motion—including average kicking height and standard deviation—were evaluated after one month and three months of training. After one month, Student A's average kicking height increased by 15.9%, and the standard deviation decreased by 11.7%. For Student B, the average kicking height increased by 14.4%, and the standard deviation decreased by 13.1%, indicating improved movement stability. After three months, Student A's average kicking height decreased by 1.8%, with the standard deviation remaining unchanged, while Student B's average kicking height increased by 6.7%, and the standard deviation decreased by 69.6%. The significant reduction in standard deviation for both athletes after three months reflects a trend toward more consistent performance.

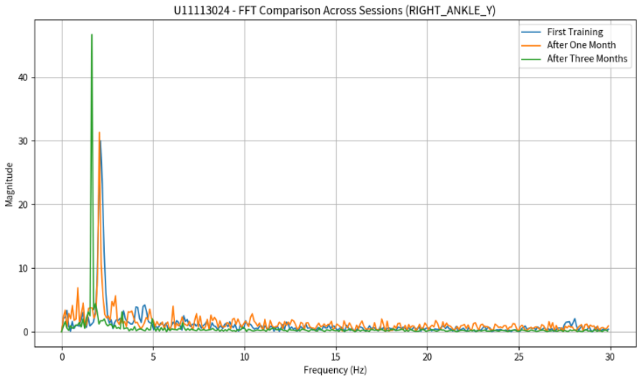

Fig. 5 Spectrogram of Student A's Training Over Three Months

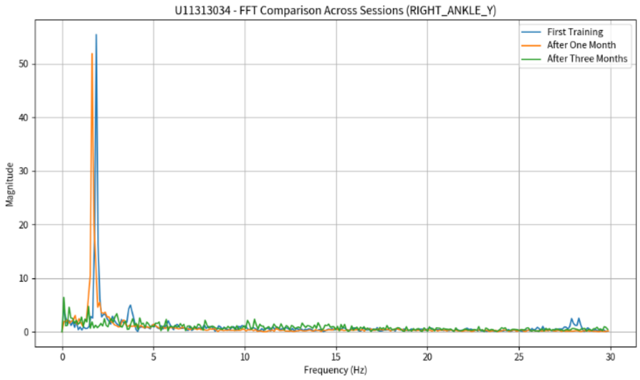

Fig. 6 Spectrogram of Student B's Training Over Three Months

In the frequency-domain analysis, the frequency-domain characteristics of the kicking motion—including dominant frequency, peak power, spectral centroid, and spectral bandwidth—were compared at the beginning, after one month, and after three months of training. After one month, Student A's dominant frequency decreased from 2.13 Hz to 2.06 Hz, peak power increased from 901.25 to 981.62, spectral centroid decreased from 0.087 to 0.077, and spectral bandwidth increased from 8.33 Hz to 8.6 Hz. For Student B, the dominant frequency decreased from 1.87 Hz to 1.64 Hz, peak power decreased from 3075.71 to 2693.65, spectral centroid increased from 0.183 to 0.206, and spectral bandwidth decreased from 8.58 Hz to 6.94 Hz.

After three months, Student A's dominant frequency decreased from 2.13 Hz to 1.65 Hz, peak power increased from 901.25 to 2181.18, spectral centroid increased from 0.087 to 0.283, and spectral bandwidth decreased from 8.33 Hz to 7.85 Hz. For Student B, the dominant frequency decreased from 1.87 Hz to 0.1 Hz, peak power decreased from 3075.71 to 41.21, spectral centroid decreased from 0.183 to 0.025, and spectral bandwidth decreased from 8.53 Hz to 8.33 Hz.

Based on the above results, it can be concluded that the dominant frequency for the male student continuously decreased, indicating a longer kicking cycle and a slower rhythm. The significant increase in peak power suggests improvements in both movement stability and explosiveness. The spectral centroid initially decreased and then increased substantially, reflecting enhanced coordination and consistency of movements after three months of training. The spectral bandwidth first increased and then decreased, with the final energy distribution becoming more concentrated around the dominant frequency, indicating more precise movement control.

In contrast, the female student exhibited a substantial decrease in dominant frequency after three months, along with marked reductions in peak power and spectral centroid. This may reflect poor adaptation to training or the influence of other external factors, resulting in decreased movement performance and stability.

IV. CONCLUSION

This study developed a Taekwondo motion analysis workflow that integrates MediaPipe keypoint extraction, IQR-based anomaly detection, and data refinement using a Kalman filter. The workflow effectively rerefining the quality of motion data, supporting accurate downstream analysis.. The training outcomes of athletes were quantitatively assessed through both time-domain and frequency-domain analyses. Experimental results indicate that, on average, approximately 4.107% of keypoint data required correction, which contributed to improved accuracy in subsequent analyses. By further combining time-domain and frequency-domain analyses, the training effectiveness of the athletes was quantitatively evaluated.

After training, Student A demonstrated improvements in both kicking height and stability, with energy distribution becoming more concentrated and movement performance more consistent. Student B exhibited a continuous increase in kicking height and a marked enhancement in stability; however, after three months, a noticeable decrease in kicking frequency was observed, and the energy distribution became more dispersed. Overall, both students showed progress in movement stability and quality following training, though the trends in their performance varied. Therefore, coaches can design customized training programs tailored to the specific characteristics and needs of individual athletes.

This study utilized MediaPipe to detect Taekwondo athletes' kicking motions, enabling real-time acquisition of keypoint data and subsequent motion analysis to assist coaches in dynamically correcting athletes' movements. Through quantitative data analysis, coaches can more accurately assess athletes' performance and use this information to adjust training programs, thereby enhancing training efficiency and movement quality. Furthermore, this automated analytical approach facilitates long-term tracking of athletes' technical progress, providing a scientific basis for Taekwondo training.

REFERENCE

[1] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-based Activity Recognition: A Survey," 2017/07/12, doi: 10.48550/arXiv.1707.03502.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2015/06/08, doi: 10.48550/arXiv.1506.02640.

[3] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," 2016/11/24, doi: 10.48550/arXiv.1611.08050.

[4] Y. Lin, X. Jiao, L. Zhao, Y. Lin, X. Jiao, and L. Zhao, "Detection of 3D Human Posture Based on Improved Mediapipe," *Journal of Computer and Communications*, vol. 11, no. 2, 2023–02–15, doi: 10.4236/jcc.2023.112008.

[5] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines," 2019/06/14, doi: 10.48550/arXiv.1906.08172.

[6] C. S. T. Hii *et al.*, "Automated Gait Analysis Based on a Marker-Free Pose Estimation Model," *Sensors (Basel, Switzerland)*, vol. 23, no. 14, 2023 Jul 18, doi: 10.3390/s23146489.

[7] C. Urrea and R. Agramonte, "Kalman Filter: Historical Overview and Review of Its Use in Robotics 60 Years after Its Creation," *Journal of Sensors*, vol. 2021, no. 1, 2021/01/01, doi: 10.1155/2021/9674015.

[8] O. Taheri, H. Salarieh, and A. Alasty, "Human Leg Motion Tracking by Fusing IMUs and RGB Camera Data Using Extended Kalman Filter," 2020/11/01, doi: 10.48550/arXiv.2011.00574.

[9] Z. Sha, Z. Zhou, and B. Dai, "Analyses of Countermovement Jump Performance in Time and Frequency Domains," *Journal of Human Kinetics*, vol. 78, no. 1, 2021 Mar 31, doi: 10.2478/hukin-2021-0028.

[10] W. S. Gan, "Fast Fourier Transform," *Signal Processing and Image Processing for Acoustical Imaging*, 2020, doi: 10.1007/978-981-10-5550-8_5.

[11] H. Beyer, "Tukey, John W.: Exploratory Data Analysis. Addison-Wesley Publishing Company Reading, Mass. — Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 S.," *Biometrical Journal*, vol. 23, no. 4, 1981/01/01, doi: 10.1002/bimj.4710230408.

[12] C. L. Farrow, M. Shaw, H. Kim, P. Juhás, and S. J. L. Billinge, "Nyquist-Shannon sampling theorem applied to refinements of the atomic pair distribution function," *Physical Review B*, vol. 84, no. 13, 2011–10–18, doi: 10.1103/PhysRevB.84.134105.