Research Article

Sports Action Recognition and Analysis Relying on Inertial Sensors

Yutong Liu,¹ Zunliang Zhou,¹ Xiaolong Qian,² and Jiaming Chen¹

¹Physical Education Department, Northeastern University, Shenyang, 110819 Liaoning, China
²College of Information Science and Engineering, Northeastern University, Shenyang, 110819 Liaoning, China

Correspondence should be addressed to Xiaolong Qian; 1971203@stu.neu.edu.cn and Jiaming Chen; 2001275@stu.neu.edu.cn

Received 7 March 2022; Revised 8 April 2022; Accepted 15 April 2022; Published 9 May 2022

Academic Editor: Wen Zeng

Copyright © 2022 Yutong Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The process of sports action recognition involves not only the change of motion speed but also the change of motion attitude, so it is necessary to carry out coordinate system transformation and attitude behavior recognition in combination with the actual situation, and inertial sensors play an important role in this process. Moreover, this paper combines inertial sensors to construct a sports action recognition and analysis system. In addition, based on the STA/LTA-AIC vibration wave picking method, an improved STA/LTA-AIC method based on wavelet packet decomposition is proposed to automatically identify the first arrival of vibration waves. Through the experimental research, it can be seen that the sports action recognition system based on an inertial sensor proposed in this paper has a good performance in sports action feature recognition.

1. Introduction

In a real environment or an indoor obstacle environment, the user’s walking speed is difficult to control. Regarding sensor signals, gait recognition by spatial angle sensor data is rarely involved in previous studies. The reason for this is that studies typically use sensors attached to the waist or smart handheld devices placed in a trouser pocket. The body part involved in this type of equipment is basically unable to measure the angle change of the leg, so it can only be analyzed through acceleration data. The study measured the angle change of each part of the whole leg. This method can indeed better reflect the gait stage, but it still remains at the laboratory level due to the need for a large number of customized sensor equipment. Moreover, it is still worth exploring using the built-in angle sensor of ordinary wearable smart devices to measure gait.

Special attention is required regarding the use of spatial angle sensor data in wearable devices. The space angle definition of these devices usually adopts the Euler angle system, which may cause the problem of Gimbal lock. The universal lock is a problem that occurs when the dynamic Euler angle is used to represent the rotation of a three-dimensional object; that is, when the axis is rotated to coincide with the axis, the degree of freedom is lost. Specifically, when an Euler angle is rotated beyond 180 degrees, it will jump to a position of -180 degrees. In recent years, the use of portable mobile devices for object motion behavior monitoring is one of the most popular researches in the field of mobile computing in recent years, and the corresponding motion behavior monitoring applications have also become a popular and popular category of various application software.

When the interactive behavior occurs in multiple different time and place contexts, by observing the interaction patterns in different contexts, the design content can be brought into the user’s different contexts [1]. When users move from one context to another, the multidimensional cognitive content of the context is generated, including context location, user subject, interaction behavior, interaction barriers, and interaction guidance, which are all extremely important concerns. Multidimensional context is a multangle analysis of the acquisition results of historical contexts. The information of these contexts can be obtained from observation or extracted from user interviews or contextual
storyboards [2]. By analyzing the content of different contextual dimensions, according to the relevant data of these historical contexts, the designers in the current context are provided with the context construction requirements [3].

This paper combines inertial sensors to construct a sports action recognition and analysis system and builds an intelligent system, which provides a reference for subsequent sports action recognition and analysis.

2. Related Work

When discussing various statistical properties of stochastic resonances, there are mainly the Langevin equation or the corresponding Fock-Planck equation in theory. Under the condition of adiabatic approximation, the literature [4] studied the SR effect of the two-state model of the bistable system and obtained the adiabatic approximation theory, but this theory has some shortcomings, it is only suitable for modulation signals with very small frequency and amplitude, and the transition probability formula it gives cannot predict the higher harmonics in the output spectrum. In order to better describe the output behavior of the bistable system, the literature [5] proposed the linear response theory of stochastic resonance, so that the SR effect can be better predicted theoretically. Reference [6] proposed the residence time distribution theory. The eigenvalue theory proposed in [7] is also known as the Flock theory. The eigenvalue theory does not require the assumptions made by the adiabatic approximation theory but regards the signal as a weak disturbance, not require the assumptions made by the adiabatic approximation theory, but regards the signal as a weak disturbance, not require the assumptions made by the adiabatic approximation theory...

Qin Guangrong, Hu Gang, and others have proved some mechanism, the research on signal detection and processing involving stochastic resonance is also vigorous. The monostable function given in the literature [15] is also known as the Flock theory. The eigenvalue theory does not require the assumptions made by the adiabatic approximation theory but regards the signal as a weak disturbance, not require the assumptions made by the adiabatic approximation theory...

In previous studies, step symmetry was used to measure gait symmetry (gait symmetry) [17]. Step symmetry is defined as the ratio of step regularity to stride regularity. By calculating the autocorrelation coefficient of the VT axis acceleration signal data, the time period difference with the most significant autocorrelation can be found [18]. If the sensor is fixed on the subject’s waist in the experiment, each peak corresponds to a step. At this time, since the step and the stride appear in a time series in a cross, the peak corresponding to the step regularity and the peak corresponding to the stride regularity also appear in a cross. By looking for the difference between the two peaks, the symmetry of left and right foot movements can be illustrated [19]. The calculation method of step regularity can also be obtained in the study, which is to find the peak value in the autocorrelation coefficient. Step regularity reflects whether the stride is regular and uniform in the whole walking cycle.

Although gait features can be extended in the gait stage, because the sensor data is susceptible to noise interference, research on the gait stage is mostly carried out with high-cost laboratory equipment, such as the motion capture device Vicon. Some studies have attempted to classify gait stages using acceleration sensors at the waist of the human body [20]. This is a very bold and worthwhile attempt, but due to the noise processing involved in filtering and the division of stages too detailed, the accuracy is still open to question. The research accuracy of acceleration sensor data for the gait stage tends to be low, mainly because the acceleration sensor captures the acceleration data of body buffering or shaking. The meaning of this data is not clear enough, the angle data is easier to capture, it is easier to find characteristic points in the signal, and it is more consistent with the various definitions of the gait cycle [21].
3. Sports Action Recognition and Analysis
Relying on Inertial Sensors

The Short-Term Average/Long-Term Average (STA/LTA) method is one of the commonly used automatic pick-up methods for vibration waves. The principle of the STA/LTA method is to use the ratio of STA (average value in a short time window of the signal) to LTA (average value in a long time window of the signal) to represent the change in the energy or amplitude of the sports signal and to predict the first arrival time of the sports vibration wave. Averages in short time windows are more sensitive to rapid fluctuations in time series amplitude, while averages in long time windows reflect only background noise. Therefore, when the sports signal comes, the variation of the average value in the short-term window (STA) will always be greater than the variation of the average value in the long-term window (LTA), and the STA/LTA also increases significantly. When the ratio of STA to LTA is greater than a preset threshold, it is determined that a sports event occurs. At this time, the mutation point of the ratio is the first arrival point of the sports signal, and the time point of the mutation is the first arrival time of the sports signal. The schematic diagram of the STA/LTA method is shown in Figure 1.

According to different calculation methods, the calculation methods of STA and LTA can be divided into two types: recursion and standard. The calculation formulas are shown in formulas (1) and (2), respectively:

Recursive STA/LTA are

\[
\begin{align*}
\text{STA}_t &= \text{STA}_{t-1} + \frac{f_c(t) - \text{STA}_{t-1}}{N_{\text{STA}}}, \\
\text{LTA}_t &= \text{LTA}_{t-1} + \frac{f_c(t) - N_{\text{LTA}} - 1}{N_{\text{LTA}}}. 
\end{align*}
\]

Standard STA/LTA are

\[
\begin{align*}
\text{STA}_t &= \text{STA}_{t-1} + \frac{f_c(t) - f(t - N_{\text{STA}})}{N_{\text{STA}}}, \\
\text{LTA}_t &= \text{LTA}_{t-1} + \frac{f_c(t - N_{\text{STA}} - 1) - f(t - N_{\text{LTA}} - N_{\text{STA}} - 1)}{N_{\text{LTA}}}. 
\end{align*}
\]

In the above formula, \(\text{STA}_t\) is the average value of the signal in the short time window at time \(t\), \(\text{LTA}_t\) is the average value of the signal in the long time window at time \(t\), and \(f_c(t)\) is the characteristic function value of the signal at time \(t\). \(N_{\text{STA}}\) and \(N_{\text{LTA}}\) are the number of recorded sample points included in the short-term average time window and the long-term average time window, respectively.

When using the STA/LTA method to process sports signals, the stability and accuracy of the results obtained will be affected by many factors, such as the selection of feature functions, the size of the time window, and the setting of the trigger threshold.

The selection of the characteristic function directly affects the accuracy of the vibration wave pickup. The long-short time window mean ratio method is used to identify the vibration wave and vibration wave. The more common characteristic functions are as follows:

\[
\begin{align*}
f_c(t) &= |Y(t)|, \\
f_c(t) &= Y(t)^2, \\
f_c(t) &= Y(t) - Y(t-1), \\
f_c(t) &= Y(t)^2 - Y(t-1)Y(t+1).
\end{align*}
\]

The calculation process of the mean ratio method of long and short time windows is fast and simple, and it can better carry out vibration wave pickup experiments in the environment of large-scale propagation paths and high signal-to-noise ratio. However, when the initial fluctuation is not obvious and the environmental noise is large, the pickup effect of the vibration wave is not good, and it is easy to produce large errors. The selection of the length of the time window also directly affects the accuracy of the automatic pickup of vibration waves, and the length of the short time window can be obtained from the above STA/LTA calculation formula.

In order to more clearly confirm the optimal threshold, this paper introduces the parameter \(Q\) to represent the perceptual quality of the early warning system, where \(Q\) is the ratio of the accuracy rate to the false alarm rate. According to the experimental results in Table 1, Figure 2(b) plots the numerical variation of the ratio \(Q\) of the accuracy rate to the false alarm rate under different thresholds. The larger the value of \(Q\), the higher the accuracy and the better the effect of the early warning system.

Figure 2(a) shows that the accuracy and false alarm rate of the STA/LTA method change with the threshold \(R\). It can be seen from this figure that the accuracy rate increases with the increase of the threshold and the false alarm rate decreases with the increase of the threshold \(R\). When the optimal threshold point is reached, the false alarm rate increases continuously with the increase of the threshold \(R\), which indicates that the threshold is set too high, and some vibration events that meet the requirements are ignored and not identified. Combining with Figures 2(a) and 2(b), we can find that in the experimental environment of this paper, when the threshold \(R\) is equal to 2, the vibration recognition accuracy is the highest, so the threshold size of the STA/LTA method involved in this paper is set to 2.

3.1. Setting of Long and Short Time Windows. The short time window (STA) is mainly used to obtain a time window for sports signals, so the shorter the short time window, the more advantageous it is for the acquisition of sports signals in a short period, and the long time window (LTA) is used to measure the time window. STA/LTA can automatically adjust the acuity of the sports signal according to the noise level in the background environment. It is necessary to find the optimal choice of long time window and short time window in this experimental environment. This paper uses a fixed long time window size to analyze the picking results corresponding to different short time windows and uses a...
fixed short time window size to analyze the picking results corresponding to different long time windows and determines the size of the long and short time windows. It is obtained that the accuracy rate reaches the highest when the short-time window size is 0.2 s and the long-time window size is 2 s. Therefore, the short-time window size in this paper is set to 0.2 s, and the long-time window size is set to 2 s.

When processing the data using the $|Y(t)|$ and $Y^2(t)$ characteristic functions, it only reflects the change in the amplitude of the vibration wave in sports, and cannot highlight the change in frequency. Although $Y(t) - Y(t - 1)$ can reflect the changes of amplitude and frequency, the recognition effect is poor in the background of low signal-to-noise ratio. Based on the low signal-to-noise ratio of the data collected by the inertial sensor, $Y(t)^2 - Y(t - 1)Y(t + 1)$ is used as the feature function of this paper to characterize the data, and the result after characterizing the data is shown in Figure 3.

After the sample data is processed by the feature function, the abnormal vibration waveform features can be significantly amplified under the condition that the signal-to-noise ratio is relatively low. Therefore, after the sample data is characterized by the feature function $Y(t)^2 - Y(t - 1)Y(t + 1)$, the pick-up rate of the STA/LTA algorithm can be effectively improved.

The Akaike Information Criterion (AIC), also known as the minimum information criterion, is used to measure the estimated model complexity and the goodness of the model's

![Figure 1: Vibration wave of STA/LTA method.](image)

| Number | Sports recognition | Number | Sports recognition | Number | Sports recognition |
|--------|-------------------|--------|-------------------|--------|-------------------|
| 1      | 92.19             | 17     | 92.24             | 33     | 93.76             |
| 2      | 88.46             | 18     | 93.99             | 34     | 90.78             |
| 3      | 88.47             | 19     | 93.47             | 35     | 92.30             |
| 4      | 90.29             | 20     | 90.51             | 36     | 93.83             |
| 5      | 88.91             | 21     | 89.51             | 37     | 92.92             |
| 6      | 90.54             | 22     | 91.91             | 38     | 92.85             |
| 7      | 92.38             | 23     | 90.03             | 39     | 88.22             |
| 8      | 91.98             | 24     | 88.06             | 40     | 92.23             |
| 9      | 93.66             | 25     | 90.74             | 41     | 92.90             |
| 10     | 89.41             | 26     | 91.55             | 42     | 87.30             |
| 11     | 93.15             | 27     | 88.00             | 43     | 88.11             |
| 12     | 93.89             | 28     | 88.36             | 44     | 90.15             |
| 13     | 90.56             | 29     | 89.34             | 45     | 89.40             |
| 14     | 90.78             | 30     | 87.02             | 46     | 92.22             |
| 15     | 90.76             | 31     | 89.12             | 47     | 88.98             |
| 16     | 87.95             | 32     | 87.26             | 48     | 89.61             |

Table 1: Recognition effect of sports action feature of sports action recognition system based on inertial sensor.
fitting data. The basic expression of AIC is shown in

$$AIC = 2k - 2 \ln(\text{likelihood function})$$ \hspace{1cm} (5)

Among them, $k$ is the number of parameters.

The basic principle of using the AIC method to identify sports vibration waves is that the point corresponding to the minimum point of the AIC curve is the optimal dividing point between the sports signal and the background environmental noise. The AIC curve of the sports signal is solved in
the corresponding time window, and the minimum point corresponding to the curve is the point where the vibration wave reaches. The basic schematic diagram of the AIC algorithm is shown in Figure 4.

Its expression is shown in

\[
\text{AIC} = C - 2 \log (L).
\]

Among them, \(C\) is a constant and \(L\) is the maximum likelihood function.

The sports signal recording can be divided into two steady-state process sequences (sports signal and noise), and the first arrival of the vibration wave is used as the dividing line to distinguish the two steady-state processes of the sports signal and noise, which means that the minimum point of AIC is the vibration wave. At that time, in the boundary point of the two steady-state processes, the arithmetic expression of AR-AIC is shown in

\[
\begin{align*}
\text{AIC}(K) &= (K - N) \log (\sigma_{1,\text{max}}^2) \\
&\quad + (M - N - K) \log (\sigma_{2,\text{max}}^2) + C.
\end{align*}
\]

In the above formula, \(N\) is the order of fitting the autoregressive model data, \(M\) is the length of the sports data, \(\sigma_{1,\text{max}}^2\) and \(\sigma_{2,\text{max}}^2\) are the variance of the data fitting in the two time intervals, and \(C\) is a constant.

The expression of VAR-AIC is shown in

\[
\begin{align*}
\text{AIC}(K) &= K \times \log \{\text{var} (x[1,K])\} \\
&\quad + (M - K - 1) \log \{\text{var} (x[K+1,M])\}.
\end{align*}
\]

Among them, \(K\) includes all the sample points of the input sports signal \(x\), which is the corresponding serial number of the sample points in the selected time window, and \(\text{var} (x[1,K])\) and \(\text{var} (x[1,K+1])\) are the variances of two different time series. At this moment, the minimum value of the AIC curve is the first arrival point of the vibration wave.

Through the analysis of the experimental results of using the STA/LTA method and the AIC method alone to pick up the first arrival time of the vibration wave, the AIC method alone has a large deviation from the real arrival time of the vibration wave. The AIC method has a
smaller deviation but also has a larger pickup error. At present, the STA/LTA-AIC comprehensive method is widely used. The specific steps of using this method in this paper are as follows:

(Step 1) It uses the characteristic function determined.

(Step 2) It uses the STA/LTA calculation formula to calculate the value of STA/LTA, and the obtained ratio is compared with the previously set threshold $R$. When the ratio is larger than the threshold $R$, the algorithm proceeds to the next step, and when it is smaller than the threshold $R$, the algorithm ends.

(Step 3) The algorithm uses the AIC criterion within the time window of 2 s before and after to calculate the corresponding AIC curve.

The minimum value of the curve is the arrival time of the vibration wave.

Figure 5 is the flow chart of the STA/LTA-AIC integrated method to pick up the vibration wave when the vibration wave arrives.

Traditional vibration signal analysis and processing generally use windowed Fourier analysis. This method is an analysis method in which the window function does not change, so it cannot explain the characteristics of vibration signals such as short duration, frequency domain time
domain localization, and nonstationarity. Wavelet analysis is a time-domain localized analysis method with variable shape but constant window area. In this method, only the low-frequency signal is decomposed and decomposed again, and the high-frequency signal is not decomposed again, and the frequency resolution will decrease as the frequency increases. The transformation formula is as follows:

\[
\begin{align*}
\mathcal{C}_k^{n+1} &= I_0 (\mathcal{C}_n^k), \\
\mathcal{C}_k^{n+2} &= I_1 (\mathcal{C}_n^k),
\end{align*}
\]

Among them, \(\mathcal{C}_n^k \in \{ \mathcal{C}_n^k \}_{k \in \mathbb{Z}}, \mathcal{C}_k^{n+1} \in \{ \mathcal{C}_k^{n+1} \}_{k \in \mathbb{Z}}, \mathcal{C}_k^{n+2} \in \{ \mathcal{C}_k^{n+2} \}_{k \in \mathbb{Z}},\) and the expressions of operators \(I_0\) and \(I_1\) such as formula (10) are

\[
\begin{align*}
I_0 (S_k^j)(j) &= \sum_{k \in \mathbb{Z}} S_k \mathcal{D}_{k-2j}, \\
I_1 (S_k^j)(j) &= \sum_{k \in \mathbb{Z}} S_k \mathcal{G}_{k-2j}.
\end{align*}
\]

Wavelet packet analysis divides the frequency bands at multiple levels and decomposes each frequency band again after decomposing and obtains more detailed signal decomposition than wavelet transform. In the first decomposition of the wavelet packet, two parts of high frequency and low frequency are obtained, and the two parts are decomposed simultaneously in the second decomposition. Moreover, two sequences are obtained after each decomposition. The wavelet packet decomposition tree is shown in Figure 6.

The original sports signal is regarded as the decomposition transformation \(\mathcal{C}_0^0\) of the scale zero node \((0, 0)\), that is, \(\{ \mathcal{C}_0^0 \}_{j \in \mathbb{Z}}\). The algorithm performs wavelet packet decomposition transformation on the zero node \((0, 0)\) through formula (10) to obtain two nodes \((1, 0)\) and \((1, 1)\) with a scale of 1. The corresponding coefficients are \(c_0^1\) and \(c_1^1\), which, respectively, include the low-frequency and high-frequency parts of the original sports signal. Following this rule, the coefficient corresponding to a node of scale \(k\) is denoted \(c_k^0, c_k^1, \ldots, c_k^{2k-1}\).

The wavelet packet avoids the disadvantage that the local performance of the spectrum is deteriorated due to the increase of the scale of the wavelet transform and can analyze the sports signal more accurately. Therefore, this paper proposes the method of wavelet packet decomposition and reconstruction to decompose and reconstruct abnormal vibration signals.

Based on the STA/LTA-AIC vibration wave picking method, this paper proposes an improved STA/LTA-AIC method based on wavelet packet decomposition to automatically identify the first arrival of vibration waves. The basic idea of the method is as follows:

1. **Step 1** The algorithm uses the STA/LTA method to roughly pick up the picked original signal and compares the STA/LTA ratio with the threshold \(R\) set in Section 2. When the ratio is greater than the threshold \(R\), the algorithm proceeds to the next step, and when the ratio is less than the threshold \(R\), the algorithm ends the experimental step.

2. **Step 2** The algorithm decomposes the original vibration signal with three-scale wavelet packet and reconstructs the original vibration signal using the decomposed coefficients.

3. **Step 3** The algorithm then calculates the AIC curve of the reconstructed signal in the time window of \(t \pm 2s\) under the three scales according to formula (8) and then superimposes the AIC curves of the three scales, and the minimum value...
Figure 6: Wavelet packet decomposition tree.

Figure 7: Flow chart of picking up vibration wave at the first arrival based on wavelet packet.
obtained is the arrival time of the vibrational wave.

Figure 7 is the flow chart obtained when the STA/LTA-AIC method based on wavelet packet decomposition proposed in this paper is used to pick up the first arrival of the vibration wave.

4. Sports Action Recognition and Analysis System Based on an Inertial Sensor

The intelligent sports action recognition system uses motion sensor (including acceleration sensor and angular velocity sensor) data for learning and judgment. Action sensors are susceptible to noise, especially shaking when the body is in motion. For the operation of sports recognition, recognition using angle sensor data is another path. The data from the angle sensor is easier to process and characterize. In addition, another way to improve sports recognition is by adding a filter model of the user’s motion state to the data processing model. Figure 8 shows an inertial sensor-based sports motion recognition technique. The sports action recognition technology directly performs gesture recognition based on the data in the data buffer. Through the additional process, the data in the buffer is first processed, the motion cycle data is extracted, and the motion data is modeled. Moreover, by means of signal subtraction or filtering, the interference of the motion state to gesture recognition can be reduced, and the recognition accuracy can be improved.

On the basis of the above analysis, the sports action recognition system based on inertial sensor proposed in this paper is verified, the sports action recognition effect is counted, combined with multiple sets of data for comparative analysis, and the statistical test results are shown in the following Table 1 and Figure 9.

From the above research, it can be seen that the sports action recognition system based on inertial sensor proposed in this paper has a good performance in sports action feature recognition.
5. Conclusion

Previous studies mostly used speed as a variable to analyze the impact of movement speed on the use of smart devices. However, the real situation is more detailed and specific. When a user moves in a certain motion state, his body shakes, and visual and cognitive resources are scattered. In particular, motion causes instability of the device screen, making the use of smartwatches difficult. Furthermore, they refine the study of specific gait conditions and the impact of specific phases of the stride on performance. In addition, in previous studies, the preferred walking speed was used as the setting to control the movement state. This method can quantitatively measure the movement state, but it is only applicable to the experimental environment on the treadmill. In this paper, the inertial sensor is used to construct a sports action recognition and analysis system, and an intelligent system is constructed. Through experimental research, it can be seen that the sports action recognition system based on inertial sensors proposed in this paper has a good performance in sports action feature recognition.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding authors upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This study is sponsored by Northeastern University.

References

[1] A. Martínez-González, M. Villamizar, O. Canèvet, and J. M. Odobez, "Efficient convolutional neural networks for depth-based multi-person pose estimation,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 11, pp. 4207–4221, 2019.

[2] M. Li, Z. Zhou, and X. Liu, “Multi-person pose estimation using bounding box constraint and LSTM,” IEEE Transactions on Multimedia, vol. 21, no. 10, pp. 2653–2663, 2019.

[3] J. Xu, K. Tasaka, and M. Yamaguchi, “Fast and accurate whole-body pose estimation in the wild and its applications,” ITE Transactions on Media Technology and Applications, vol. 9, no. 1, pp. 63–70, 2021.

[4] G. Szücs and B. Tamás, “Body part extraction and pose estimation method in rowing videos,” Journal of Computing and Information Technology, vol. 26, no. 1, pp. 29–43, 2018.

[5] R. Gu, G. Wang, Z. Jiang, and J. N. Hwang, “Multi-person hierarchical 3d pose estimation in natural videos,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 11, pp. 4245–4257, 2019.

[6] M. Nasr, H. Ayman, N. Ebrahimi, R. Osama, N. Mosaad, and A. Mounir, “Realtime multi-person 2D pose estimation,” International Journal of Advanced Networking and Applications, vol. 11, no. 6, pp. 4501–4508, 2020.

[7] N. T. Thành và P. T. Công, “An evaluation of pose estimation in video of traditional martial arts presentation,” Journal of Research and Development on Information and Communication Technology, vol. 2019, no. 2, pp. 114–126, 2019.

[8] I. Petrov, V. Shakhuro, and A. Konushin, “Deep probabilistic human pose estimation,” IET Computer Vision, vol. 12, no. 5, pp. 578–585, 2018.

[9] G. Hua, L. Li, and S. Liu, “Multipath affino stacked—hourglass networks for human pose estimation,” Frontiers of Computer Science, vol. 14, no. 4, pp. 1–12, 2020.

[10] K. Aso, D. H. Hwang, and H. Koike, “Portable 3D human pose estimation for human-human interaction using a chest-mounted fisheye camera,” in Augmented Humans Conference 2021, pp. 116–120, Finland, Feb 2021.

[11] D. Mehta, S. Sridhar, O. Sotnychenko et al., “Vnect: real-time 3d human pose estimation with a single rgb camera,” ACM Transactions on Graphics (TOG), vol. 36, no. 4, pp. 1–14, 2017.

[12] L. Liu, Y. Li, and G. Hua, “Human pose estimation in video via structured space learning and halfway temporal evaluation,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 7, pp. 2029–2038, 2019.

[13] S. Ershadi-Nasab, E. Noury, S. Kasaei, and E. Sanaei, “Multiple human 3d pose estimation from multiview images,” Multimedia Tools and Applications, vol. 77, no. 12, pp. 15573–15601, 2018.

[14] X. Nie, J. Feng, J. Xing, S. Xiao, and S. Yan, “Hierarchical contextual refinement networks for human pose estimation,” IEEE Transactions on Image Processing, vol. 28, no. 2, pp. 924–936, 2019.

[15] Y. Nie, J. Lee, S. Yoon, and D. S. Park, “A multi-stage convolution machine with scaling and dilation for human pose estimation,” KSII Transactions on Internet and Information Systems (TIIS), vol. 13, no. 6, pp. 3182–3198, 2019.

[16] A. Zarkeshev and C. Csiszár, “Rescue method based on V2X communication and human pose estimation,” Periodica Polytechnica Civil Engineering, vol. 63, no. 4, pp. 1139–1146, 2015.

[17] W. McNally, A. Wong, and J. McPhee, “Action recognition using deep convolutional neural networks and compressed spatio-temporal pose encodings,” Journal of Computational Vision and Imaging Systems, vol. 4, no. 1, pp. 3–3, 2018.

[18] R. G. Díaz, F. Laamarti, and A. El Saddik, “DiTCoach: your digital twin coach on the edge during COVID-19 and beyond,” IEEE Instrumentation & Measurement Magazine, vol. 24, no. 6, pp. 22–28, 2021.

[19] A. Bakshi, D. Sheikh, Y. Ansari, C. Sharma, and H. Naik, “Pose estimate based yoga instructor,” International Journal of Recent Advances in Multidisciplinary Topics, vol. 2, no. 2, pp. 70–73, 2021.

[20] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. Salo, “A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system,” Sports Medicine-Open, vol. 4, no. 1, pp. 1–15, 2018.

[21] I. Sárhidi, T. Linder, K. O. Arras, and B. Leibe, “Metrabs: metric-scale truncation-robust heatmaps for absolute 3d human pose estimation,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 1, pp. 16–30, 2021.