



Markerless gait estimation and tracking for postural assessment

Chuan Zhi Tay¹ · King Hann Lim¹ · Jonathan Then Sien Phang¹

Received: 18 February 2021 / Revised: 2 June 2021 / Accepted: 3 January 2022 /

Published online: 21 February 2022

© The Author(s) 2022

Abstract

Postural assessment is crucial in the sports screening system to reduce the risk of severe injury. The capture of the athlete's posture using computer vision attracts huge attention in the sports community due to its markerless motion capture and less interference in the physical training. In this paper, a novel markerless gait estimation and tracking algorithm is proposed to locate human key-points in spatial-temporal sequences for gait analysis. First, human pose estimation using OpenPose network to detect 14 core key-points from the human body. The ratio of body joints is normalized with neck-to-pelvis distance to obtain camera invariant key-points. These key-points are subsequently used to generate a spatial-temporal sequences and it is fed into Long-Short-Term-Memory network for gait recognition. An indexed person is tracked for quick local pose estimation and postural analysis. This proposed algorithm can automate the capture of human joints for postural assessment to analyze the human motion. The proposed system is implemented on Intel Up Squared Board and it can achieve up to 9 frames-per-second with 95% accuracy of gait recognition.

Keywords Pose estimation · Gait recognition · Postural assessment · Markerless motion capture

1 Introduction

Postural assessments using computer vision have received substantial attention in the sports community for sports screening and injury prevention [3, 6]. Incorrect posture during sports

✉ Chuan Zhi Tay
tcz3103@gmail.com

King Hann Lim
glkhan@curtin.edu.my

Jonathan Then Sien Phang
jonathanpts@postgrad.curtin.edu.my

¹ Department of Electrical and Computer Engineering, Curtin University Malaysia, CDT 250, 98009, Miri, Sarawak, Malaysia

is one of the root causes of physical injuries that leads to joints inflammation and post-traumatic arthritis [39]. Despite the recovery from joint inflammation, a person may still possess the symptom of post-traumatic arthritis after 6 months and so considered as chronic pathological disease [32]. The statistic reported in [13] shows that 30 million people suffer from osteoarthritis (OA) in the United State, of whom more than half are under age 65. OA is a chronic condition that can affect any joints, such as knees, hips, lower back and neck, small joints of the fingers, the bases of the thumb and big toe [13]. Post-traumatic osteoarthritis [32] is recognized as a disabling condition as soon as 20 months after injury and there is no cure for it.

The capture of human motion can be classified into two types [41], i.e. (a) marker-based approach, and (b) markerless-based approach. Marker-based approach requires sensors or visual markers to be attached on human body for motion analysis. Marker-less motion approach uses passive cameras to detect human body visually and process with computer vision techniques. Reyes-Ortiz et al. [35] tracked human activity using the smartphone's inertial measurement unit (IMU) for human action recognition. Beravs et al. [4] proposed an IMU system with the validation algorithm to measure the joint angle of ankle and the motion of lower limb exoskeletons. However, the marker-based approaches require sensors to be attached to body and long setup preparation before any human motion analysis.

On the other hand, markerless-based motion analysis approach tracks the boundaries or features of human bodies without the mounted markers using camera. It retrieves the spatial-temporal dimension of salient objects such as human skeleton and motion [9, 22, 24, 34] in the video. Since it does not require the attachment of sensors and markers, the quality of kinematic data can be improved in the physical training for precise postural assessment in the sports field [26]. Hence, the sport community believe that the use of computer vision could improve the physical training efficiency to achieve high performance [10]. However, high computational cost and multiple cameras calibration setup are the limitations encountered in this approach in order to generate fast inference in real-time applications.

The information of the human posture is crucial cue in sports because it can be analyzed to achieve the optimal performance and reduce the risk of physical injury [15]. To achieve the low computational cost, this paper proposes a markerless gait estimation and tracking for postural assessment with edge computing to do the sports screening. This proposed system consists of four stages: (a) human joints estimation to extract key-points, (b) gait recognition for single human instance tracking, (c) pose tracking for processing time reduction, (d) postural assessment using joints information. The overall proposed pipeline architecture is shown in Fig. 1 and it is migrated to a low power embedded system, Intel Up Squared Board with single camera setup. It is hypothesized that the system enables the fast processing gait estimation with joints location invariant to camera orientation without marker attached to the human body for the postural assessment. The extension of this research can cover other areas of sports such as weight lifting, diving and cycling.

2 Related work

Human motion is a key information to analyze the human posture that receives huge attention in the the research of computer vision and artificial intelligence [2]. The recent computer vision research of human pose estimation associates with the deep learning approach to achieve significant improvement to the standard benchmark such as the accuracy of clustering body part and parsing human pose. Human pose estimation enables the

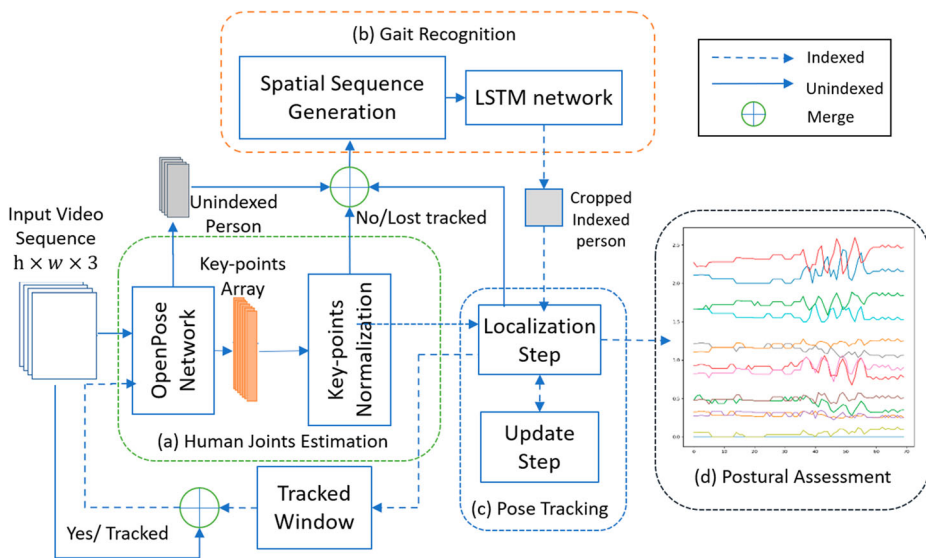


Fig. 1 The basic pipeline of markerless gait estimation and tracking for postural assessment consisting of 4 stages, i.e. (a) Pose estimation, (b) Gait recognition, (c) Pose Tracking and (d) Postural assessment

potential of computer vision to analyze the human motion and capture the invisible subtle postural information which are neglected by human eyes. Therefore, a real-time postural assessment system using markerless-based approach can be widely applied to the sport scene to reduce the risk of injury especially at the joints.

2.1 Deep learning approach in human pose estimation

The markerless-based approach [19] are mainly classified based on its scanning techniques, i.e. (a) Top-Bottom method and (b) Bottom-Up method. Top-Bottom method detects the human location first and implements the pose estimation within location result. DensePose with Mask-Region-Convolutional Neural Networks baseline [1] was developed by using deep learning top-down method to locate the person on the image and predicted the bounding boxes to crop the detected person become Region of Interest (ROI) pooling. After that, the ROI was fed into the fully-convolutional network architecture in order to predict the coordinates of the body key-points. Based on the set of coordinates, it identified the body part and aligned them to estimate the pose for individual person [14, 17]. Ning et al. [27] used ResNet151 as backbone to detect the human candidates and crop the detected person in tracked window. A Cascaded Pyramid Network which was the single-person human pose estimator implemented within the tracked window. Then, a flow-based pose tracker was implemented to track the pose ID for persons who was in the image area. Top-bottom method reduces the false key-point prediction but the early commitment could affect the process if the person detector fails in the early stage.

Bottom-Up method estimates the body part nodes on the image and associates them as pairwise part connections using grouping algorithm to the belonging person without the process of person detection. Leonid Pishchulin et al. [30] proposed a bottom-up method, “DeepCut” for multi-person human pose estimation. It initialized all the possible body part candidates from the image and labelled them corresponding to the body part to classify

them to each possible person. Cao et al. [8] proposed a method to use CNN approach with two branches of the same sequential prediction process to train the part confidence map for body part and Part Affinity Field (PAF) for the association degree between the limbs. Then, a greedy parsing algorithm was implemented which performed a fast connecting process to form a high-quality parses of human pose estimation. Bottom-up method predicts the body parts in the image and implements the grouping algorithm to assign the body part belonging to the particular person. Meanwhile, the grouping algorithm could be complicated as the fully connected graph is a non-deterministic polynomial-time hardness problem [30].

2.2 Postural assessment using markerless-based approach

The recent developments of postural assessment in sports often use markerless-based approach with deep learning architecture such as CNN. Tian et al. [23] proposed a hierarchical CNN model to assign the social roles to the players which were defender and attacker. After that, the scene of the human activity was categorized based on the movement and action of the defender and attacker. Furthermore, Joshi et al. [20] proposed to use Inception-V3 neural network to classify the robust sport images to six classes such as basketball, tennis, badminton, cricket, rugby and volley. It used the steps feature descriptor for classification based on the human activity and surrounding environment to do the sport categorization.

Kim et al. [21] proposed to use the human motion analysis to predict the play evolution in the dynamic sport scenes in soccer field. Based on the human motion, the next ground level movements of the players could be predicted. This method was applied on the automatic camera control and the sports visualization for analysis. Bialkowski et al. [5] investigated the raw player detection and the strategy analysis from the noisy data in the hockey game. Based on the occupancy maps of the players which was predicted by histogram of gradient orientation, the specific team formation was retrieved and wrong prediction of strategy could be removed.

Hayasahi et al. [16] proposed a CNN model visual imagery analyzer to recognize the upper body pose features and use the head and pelvis to get the spine orientation and estimate body orientation to estimate the random decision of human movement. Jian et al. [40] developed an AI coach camera system to collect the 'great' pose of athlete in the spatial-temporal sequence. This system used the 'great' pose as benchmark to detect the 'bad' poses of the athletes in the next few moves. Nuttachot and Sajjaporn [31] proposed a model for practice badminton basic skills which used the pose estimation to collect the human joints in spatial-temporal sequence and compared it with the world-class players' posture as reference. It extracted the posture embedding by triplet loss technique and fed the embedding posture to One-shot Network to find the similarity score between the input posture and reference posture. In a nutshell, the markerless-based approaches have the potential to benefit the sports community [10].

3 Proposed postural assessment system

A markerless gait estimation and tracking for postural assessment system is proposed to detect and track human gait information to evaluate the posture of sport man. This proposed pipeline is implemented in a real-time manner with the input video stream sequence. The proposed pipeline consists of four stages: (a) human joints estimation to extract key-points,

(b) gait recognition for single human instance tracking, (c) pose tracking for processing time reduction, (d) postural assessment using joints information [12, 36, 37]. The benchmark measurements for the postural assessment are introduced in this research such as i.e. (a) step rate, (b) running gait, (c) angle of the elbow.

3.1 Human joints estimation

The detection of human joints information can be achieved using OpenPose network [8]. In Fig. 2, the stage-0 of the architecture uses ten layers of Very Deep Convolutional Neural Network (VGG-19 network) to analyze the color image with size $h \times w \times 3$ and generate a set of features maps. At stage-1, the feature maps from VGG-19 network are parsed into two branches, part confidence map and PAF with multi-stage manners. The first branch predicts a set of part confidence maps to locate key-points of human body, and the second branch generates PAF to encode the association degree and direction between the key-points across multi-stages. Subsequently, the greedy parsing algorithm is applied to assemble the part association according to the part confidence map location and PAF score to the particular person. The estimated human pose with the labeling is displayed in Fig. 3 regardless of the person physical appearance or wearing and it consists of 14 key-points in the human body. The coordinates (X_K, Y_K) of each points are generated from the first branch part confidence map and the associations between each key-points are calculated from the second branch PAF.

There are 14 key-points extracted to normalize the scale and view piont of human object in this system [29]. The key-points #1-#13 extracted from the OpenPose network and the pelvis (#14) measured at the middle point between #8 and #11 coordinates. The key-points array consists of 2×14 array of #1-#14 distance joint-to-neck presented in (X_K, Y_K) coordinates taking the top-left corner of an image as the origin. To differential the motion of each key-points corresponds to the relative person, the coordinate of each key-points are subtracted with the coordinate of neck (#1). The neck point is considered as the relative point of the person. To obtain invariant human key-points from any camera orientation, the joint-to-neck distance (D) is calculated as follows,

$$(d_{K,x}, d_{K,y}) = (X_K, Y_K) - (X_1, Y_1), \tag{1}$$

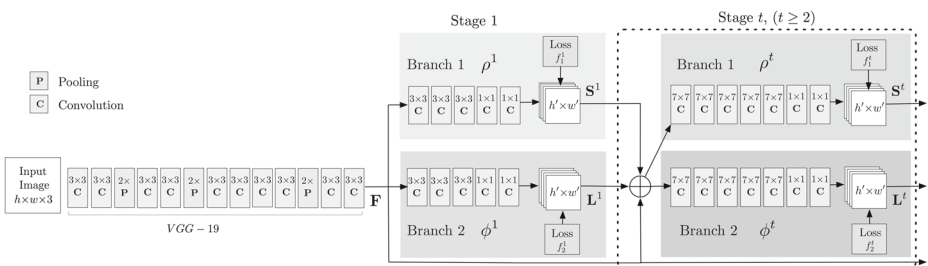


Fig. 2 The architecture of pose estimation based on Part Confidence Map and Part Affinity Field (PAF). Cao et al. [8]

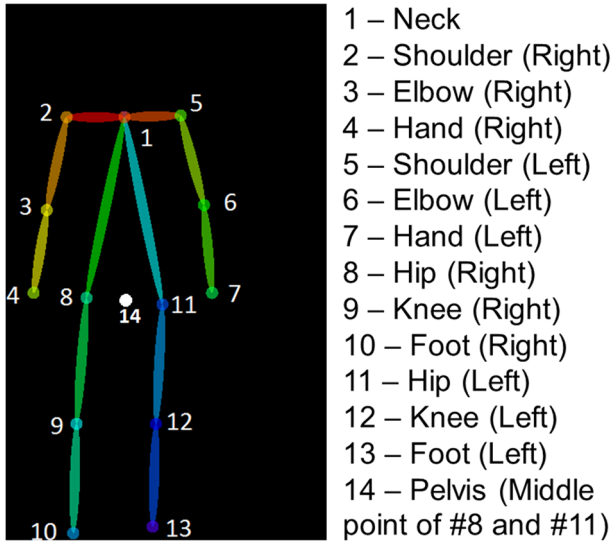


Fig. 3 The skeleton model with joint labeling generated

where the subscript K is the index of key-points, (X_1, Y_1) is the coordinate of neck and $(d_{K,x}, d_{K,y})$ is the distance from K -th joint to neck (#1). Instead of using coordinate of key-points, the distance array is arranged as the spatial information as follows,

$$D = \begin{pmatrix} (d_{1,y}, d_{1,y}) \\ \vdots \\ (d_{14,y}, d_{14,y}). \end{pmatrix} \tag{2}$$

The position of a person can be located at near or far field from the camera causing the variant of joint-to-neck distance. To normalize the joint-to-neck distances, Euclidean distance with relative neck to the pelvis distance is used to normalize all joints distance,

$$\hat{D} = \frac{D}{|S|}, \tag{3}$$

where $|S| = \sqrt{(X_1 - X_{14})^2 + (Y_1 - Y_{14})^2}$ is the Euclidean distance measured from neck (#1) to pelvis (#14) and \hat{D} is the normalized joint distance which is invariant to the distance between the person and camera. In this proposed pipeline, the neck (#1) to pelvis (#14) key-points must be co-existed conditionally to classify as a human object. Subsequently, the normalized joint-to-neck distance (\hat{D}) from each frame are stacked into a sequence of time-frame to generate spatial-temporal sequence of human motion.

3.2 Gait recognition

Gait recognition is used to identify a specified person based on the walking gait. The normalized key-points coordinates of a detected person are arranged in the spatial-temporal sequence for gait recognition. Long Short Term Memory (LSTM) network [18] is implemented as shown in Fig. 4 to process the spatial-temporal sequence and output the object index. LSTM has an internal hidden cell h_N used as the memory effect along with the

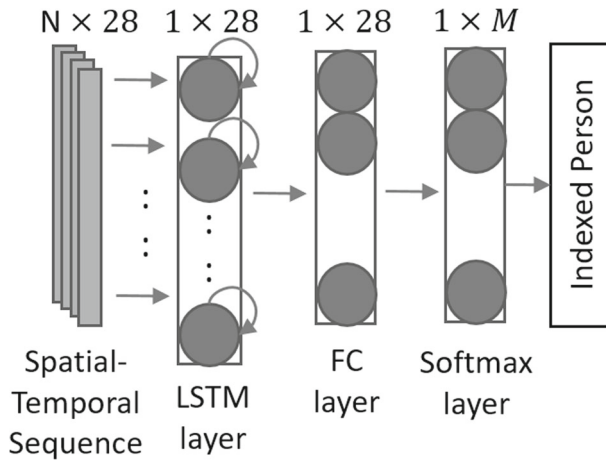


Fig. 4 Spatial-temporal sequences for gait recognition

time-series step N incorporating with an input gate’s activation vector i , forget gate f , state vector c and output gate o . The process is formulated as,

$$f_N = \sigma_g(W_f x_N + U_f h_{N-1} + b_f), \tag{4a}$$

$$i_N = \sigma_g(W_i x_N + U_i h_{N-1} + b_i), \tag{4b}$$

$$o_N = \sigma_g(W_o x_N + U_o h_{N-1} + b_o), \tag{4c}$$

$$\tilde{c}_N = \sigma_h(W_c x_N + U_c h_{N-1} + b_c), \tag{4d}$$

$$c_N = f_N \cdot c_{N-1} + i_N \cdot \tilde{c}_N, \tag{4e}$$

$$h_N = o_N \cdot \sigma_h(c_N), \tag{4f}$$

where σ_g is sigmoid activation function and σ_h is hyperbolic tangent activation function. b is the bias of the gates. W and U are the weight matrices of input vector x and recurrent connection h respectively. For each input $x = (x_1, x_2, x_3, \dots, x_N)$, the output feature of LSTM is denoted from the hidden cell h_N .

In this gait recognition, the 14 normalized vectors of X and Y axis are parsed into LSTM cells with N time steps, where N is the number of time frame. The LSTM network consists of 28 LSTM cells input size and 28 fully connected cells with ReLu activation function. Subsequently, the output of fully connected layer is fed into softmax layer which is used to classify the gait pattern according to trained gait sequence, where M is the number of indexing classes. The gait recognition can index a person from the global pose estimation for local pose tracking using smaller tracked window. This is to assist the sports screening system for a targeted personal performance monitoring.

3.3 Pose tracking

For the pose tracking, the result of person identification from LSTM network is necessary and is cropped into a smaller window size to create ROI. Then, it is parsed into a single tracking iteration using discriminative correlation filter [25] for pose tracking. There are two steps of pose tracking, i.e. localization step and update step. Localization step identifies

the new target location (p_t) by finding the position of the maximum in correlation between h_{t-1} and image patch features f using,

$$\tilde{g}(h_d) = \sum_{d=1}^{N_c} f_d * h_d \cdot \tilde{w}_d, \quad (5)$$

where $\tilde{g}(h_d)$ is the location of maximum of correlation responses, the symbol ‘*’ is the circular correlation, N_c is the number of channel, f_d is the set of N_c channel features and h_d is the constraint correlation filter. The channel reliability can be computed using element-wise product as,

$$\tilde{w}_d = \tilde{w}_d^{(update)} \cdot \tilde{w}_d^{(det)}, \quad (6)$$

where it is normalized s.t. $\sum_d \tilde{w}_d = 1$, $\tilde{w}_d^{(update)}$ is the learned channel filter obtained from the update step and $\tilde{w}_d^{(det)}$ is the channel detection reliability measured on the ratio between the second and first highest non-adjacent peaks in the channel response map [7]. The two largest peaks in the response map are obtained as two largest values after a 3×3 non-maximum suppression. The detection reliability is estimated from per-channel responses. The new scale (s_t) is derived from the new position based on the channel responses.

In the update step, the ROI of foreground histogram (c^f) and background histogram (c^b) are extracted for region analysis. The foreground histogram is computed using Epanechnikov kernel within the estimated object bounding box and the background histogram is computed from the neighborhood twice the ROI size [25]. The foreground and background histograms can be updated using the exponential moving average with the histogram learning rate (η_c) as follows,

$$c_t^f = (1 - \eta_c)c_{t-1}^f + \eta_c c^f, \quad (7)$$

$$c_t^b = (1 - \eta_c)c_{t-1}^b + \eta_c c^b. \quad (8)$$

where c_t is the histogram of foreground and background, η is the correlation filter learning rate. The foreground and background histograms are optimized to construct the spatial reliability map (m) to identify pixels in the training region which likely belong to the target. The constraint correlation filter (\tilde{h}_t) is estimated using the spatial reliability map, that identifies pixels which is set to zero in the learned filter. The per-channel learning reliability weights are measured from the maximum response value of a learned channel filter. Subsequently, the learned weights and detection weights are calculated in (6). The single iteration update step of constraint correlation filter (h_t) and channel reliability (w_t) are updated independently in every input frame with the correlation filter learning rate (η) as follows,

$$h_t = (1 - \eta)h_{t-1} + \eta \tilde{h}_t. \quad (9)$$

$$w_t = (1 - \eta)w_{t-1} + \eta \tilde{w}. \quad (10)$$

The tracking algorithm is implemented in every frame after the gait recognition to locate the person with less computational cost and time reduction. It leads this system to become a real-time application.

3.4 Postural assessment in sports

Once the human object is tracked, the cropped image of the individual is fed into OpenPose network each time for local joint estimation to extract the normalized joints vectors. Once those joints information are obtained, the postural assessment can be enabled. Since the postural assessment is enabled, the more activities or postural analysis could be added in

the future. In this paper, running action is the main demonstration because it is an important action in the most of the high performance sports. Three measurements are performed in this proposed pipeline. i.e. (a) step rate, (b) gait measurement, and (c) angle of elbow. Step rate is defined as the frequency of the foot landing and it always associates with human muscle activity [12]. Higher step rate reduces the foot ground inclination angle that decreases the peak hip abduction angle as well as moment during stance phase. The higher step rate increases the muscle activity that improves the joint moment and energy absorption. In order to obtain the step of each foot, the number of step (n) of right (#10) and left (#13) are counted separately based on the trough amount of the sinusoidal waveform in the Y-axis of joint-to-neck distance. The average step of recreational runner is 150 to 170 steps per minute and elite runner can obtain up to 180 steps per minute [33]. The step rate of each foot is the fraction between the number of step (n) and the total time (t_{total}) which is formulated as,

$$S_{min} = \frac{n_{steps}}{t_{total}}. \quad (11)$$

The inverse of step rate in second is the time taken between the first contact of two consecutive footsteps of the same foot and it is expressed as,

$$T_{sec} = \frac{60}{S_{min}}, \quad (12)$$

where T_{sec} is known as stride time where it is the time in second of one stride. Step rate and stride time manipulate the muscle activity in anticipation of foot-ground contact. The time of one stride is useful to evaluate the symmetry of stride time [36]. The energy absorption on knee and hip is corresponding to the symmetry of stride using left and right stride. If both stride time has a big gap, the unsymmetrical gait causes the muscle be imbalanced and the uneconomical movement pattern which the wrong posture wastes the unnecessary power that will over-strain the joints and muscles increasing the risk of injury [38].

The running gait can be split into two phases, i.e. stance and swing phase. Stance phase dominates approximately 40% before toe-off while swing phase covers around 60% of the gait movement. In between 50%-60% and 90%-100% in the swing phase, the runner enter a floating stage where both feet are lifted away from the ground in a very short period. As a result, a running cycle consists of 20% floating time, 40% of stance and 40% of swing [11]. In this system, only stance and swing phases are measured to create a clear indication.

In a running postural assessment [37], the elbow angle is suggested to be less than 90 degree during running and the shoulders can be relaxed and facing forward to smooth the breathe pulse. This elbow angle analysis can be viewed from the side view with the camera setup as shown in Fig. 5. The elbow angle (θ) is calculated using cosine rule as follows:

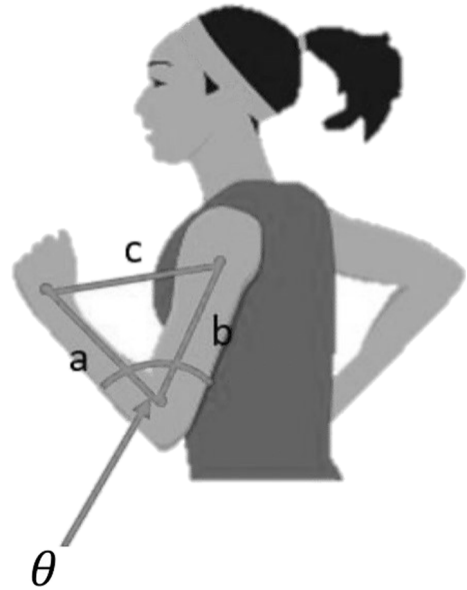
$$\theta = \cos^{-1}\left(\frac{a^2 + b^2 - c^2}{2ab}\right). \quad (13)$$

where a , b and c are the distances in between the wrist, elbow and shoulder.

4 Result and discussion of the proposed system

The proposed markerless gait estimation and tracking for postural assessment system is implemented on a microcomputer, Intel Up Squared board with Intel Movidius Neural Compute Stick 2 to achieve edge computing and low power implementation. The specification of the board consists of Intel Celeron™ N3350, 2GB ram. This postural assessment system

Fig. 5 Angle measurement at elbow using cosine rule



is equipped with single camera. In this section, the result of gait analysis computation and validation of postural assessment are covered.

4.1 Computation of gait analysis

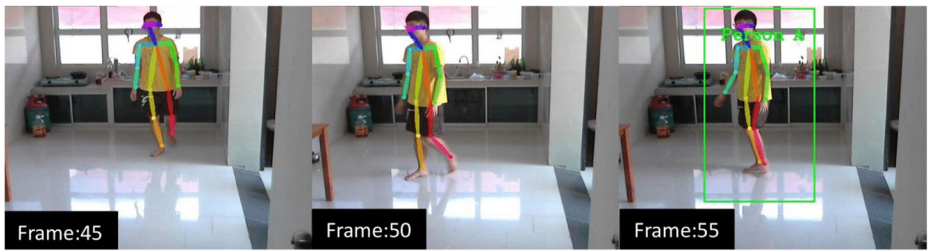
For the human pose estimation model, OpenPose is optimized using Intel OpenVino to integrate to the Intel Processor. The proposed system aims to retrieve 2×14 key-points arranging in a series of spatial-temporal features. Subsequently, these features are fed into LSTM network for gait recognition. In the gait recognition experiment, two person's walking sequences are recorded to train the gait recognition and M is set to 2. The human joints estimation is performed during the video recording with 5 frame-per-second (FPS) in 1280×720 resolution. There are 20 training set and 10 validation set of each person walking action. LSTM network was implemented with cross-entropy loss and Adam optimizer with learning rate set to 10^{-4} .

Figure 6a and b shows the pipeline classifies a person based on the walking action from "Person A" and "Person B" using the joints trajectory of walking in the continuous manner in Fig. 6c. In the gait recognition pipeline, the N time step is tested in 10, 15, and 30 of the 2×14 key-points spatial-temporal sequence. The different N time step setup of the gait recognition pipeline and the validation accuracy is calculated through (14).

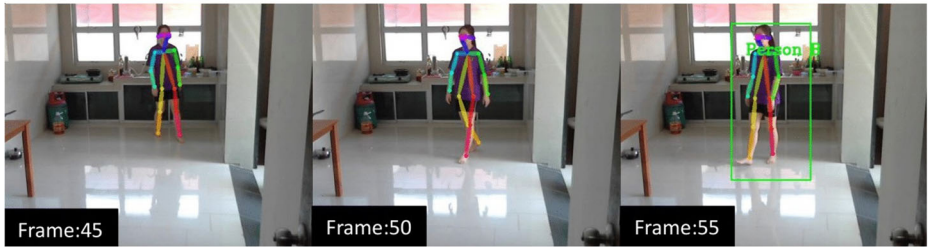
$$Accuracy = \frac{TP}{N_v} \times 100\%, \quad (14)$$

where TP is true positive and N_v is the total number of validation set.

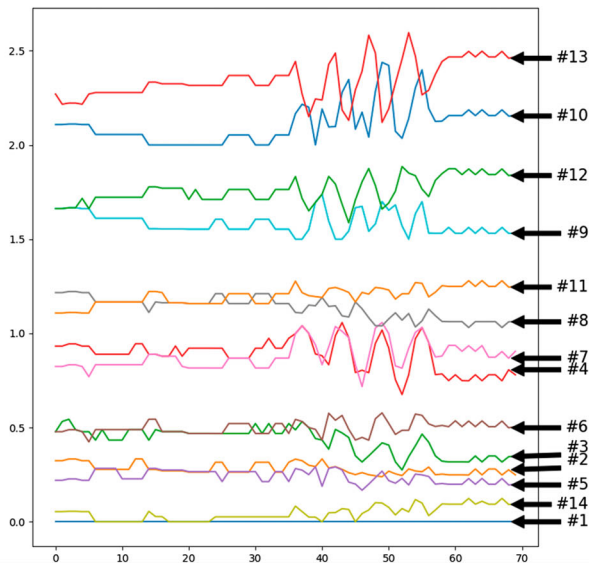
Table 1 shows the comparison between the different N time step setup of the gait recognition pipeline and the validation accuracy is calculated through (14). The result shows that 15 time step setup achieves the highest validation accuracy, 95.83% and it is chosen to implement in this system. Meanwhile, it simulates the scenario of multi-person are indexed in stadium. Every trainees are tracked in the stadium while they are entering the stadium



(a)



(b)



(c)

Fig. 6 Gait recognition for (a) Person A, (b) Person B and (c) Joints trajectory of Person A in the continuous manner

based on the walking action. Then, their sports posture information will be transferred to the sport analysis function.

Figure 6c demonstrates an example plot of the normalized joint-to-neck distance in spatial-temporal sequence of walking action of Person A. Despite the noise was obtained

Table 1 Comparison of different N time step in accuracy and loss

Time steps (N)	Validation loss	Validation Accuracy (%)
10	6e-4	88.89
15	7e-4	95.83
30	9e-4	91.67

due to the some occlusions or inaccuracy of pose estimation, the sequence could be analyzed in periodic gait movement. Apparently, the fluctuation of each sinusoidal plot are unique from every individuals. Once a human subject is indexed, the discriminative correlation filter locates the indexed person using a tracked window with less computational cost. Table 2 tabulates the comparison of proposed system in frame-per-second (FPS) using the setup on Intel Up Squared board. A raw machine setup was tested that the result was around 5.0 FPS [28] while decreasing the extracting point that only improved slightly up to 5.65 FPS [29]. The proposed method is tested and the computation time is reduced and the FPS can be achieved up to 9.

4.2 Validation of postural assessment

Step rate, running gait and the elbow angle are measured in this proposed pipeline. For the step rate, Fig. 7 is using the dot line for indication purpose to demonstrate the joint-to-neck distance plot of #10 and #13 as the joint motion in Y-axis. The trough of the joint-to-foot distance cycle represents the foot landing that generates the number of steps (n_{steps}). It demonstrates two significant troughs in Fig. 7 for both foot where there are two steps for both foot. The step counting function is mainly utilized for step rate and running gait measurement. The comparison between the actual walking step by counting manually and the measured step amount of walking with random route within a minute obtained an average error of the three testing sets is 4.29% that is lower than 5% as shown in Table 3.

Figure 7 shows the left feet (#13) is main relative point with a complete running cycle. The intersection points represent both foot are lifted away from the ground. Thus, the swing, stance phase and floating time are predicted respectively. Subsequently, the athlete can refer the result of the running gait measurement to adjust the running gait and avoid the incorrect gait. The assessment of the elbow angle is using the side view of the sport man. The plot of the elbow angle degree in time series during running is visualized in Fig. 8. As a result, the peak of the sinusoidal signal represents the position of wrist is at waist level while the trough shows the wrist is above the waist level. It does not require the wrist have to be in any form but the elbow angle is suggested to be lower than 90 degree to help maintain a fast cadence and get the shoulder relaxing to reduce the muscle stress.

Figure 9 demonstrates the real time implementation to do postural assessment. At the first stage in Fig. 9a, the system identifies the user according to the walking gait. After

Table 2 Comparison of the Frame-Per-Second (FPS) of different setup

Experimental pipeline setup	Frame-Per-Second (FPS)
Extract 18 joints (w/o optimization)[28]	5.0
Extract 14 joints (w/o Pose Tracking)[29]	5.65
Proposed method: Extract 14 joints with Pose Tracking	9.0

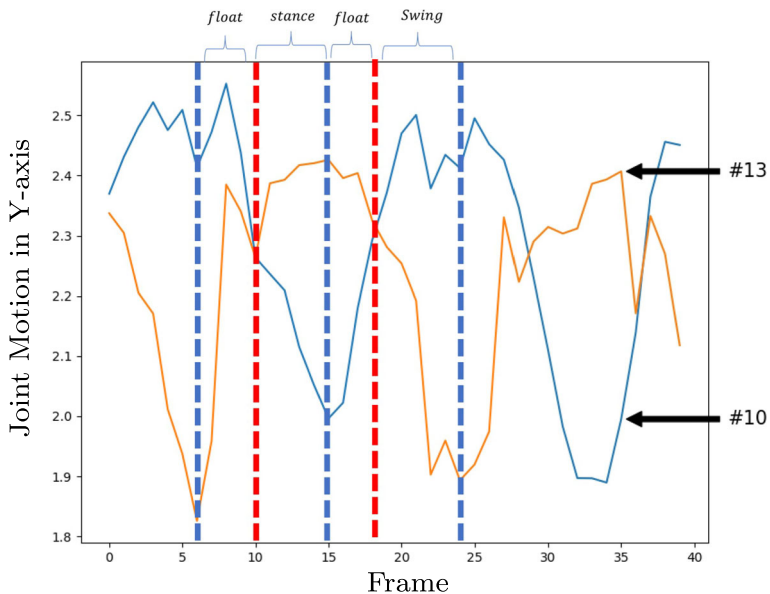


Fig. 7 Running gait evaluation, #13 is the main relative feet

Table 3 Actual and measured step amount of the walking

Test No.	Actual value	Measured value	Error rate (%)
1	72	71	1.38
2	69	73	5.79
3	70	74	5.71
		average error	4.29

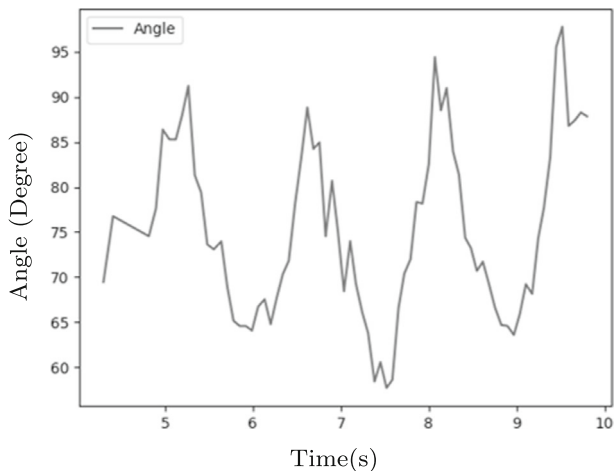
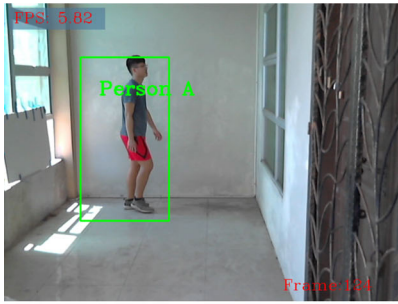


Fig. 8 The analysis of the elbow angle trajectory



(a)



(b)



(c)



(d)

Fig. 9 Real-time implementation, (a) Person is identified by the gait recognition, (b) Postural assessment, (c) Correct posture result, (d) Incorrect posture result

the user is indexed, a red bounding box tracks the identified user for postural assessment shown in Fig. 9b. The new tracked window is used to predict the position of a subject in the next frame. Subsequently, the gait measurement results are transferred to cloud database through a lightweight publish-subscribe network protocol, MQTT. User can access to a dashboard which is demonstrated in Fig. 9c and d as on the cloud remotely to observe the past postural assessment with relative to time sequence. Figure 9c shows a correct result of running gait. The step rate is shown on the top left of the dashboard with a green indication where the elbow angle remains 90 degree during running. The pie chart at the bottom right of the dashboard indicates that the running gait consists of approximately half of the swing and stance phases respectively. Figure 9d shows an incorrect posture with the step rate highlighted in red color and the angle of the running elbow is below a threshold angle. The

pie chart shows that the swing phase and stance phase are not in a synchronize manner. The unsymmetrical running gait is causing the risk of injury at a long term scenario.

5 Discussion and implications

The proposed pipeline comprises of the human joints estimation, gait recognition, pose tracking using discriminate correlation filter for the application of postural assessment. The key-points coordinate from the human joints estimation are firstly normalized to obtain the invariant key-points to camera's orientation. The 14 normalized vectors are combined to generate the human motion in the frame series. It leads that the resulting position of the human object that does not degrade the performance of gait recognition and postural assessment. With this proposed computer vision processing, it does not require to attach sensors to the body during athlete training and it could avoid the physical interference causing an unnatural posture to reduce the risk of sports injury.

To reduce the computational cost of the proposed pipeline, this gait estimation algorithm is optimized by reducing the cycles of estimating part confidence maps and part affinity field. Subsequently, the discriminate correlation filter locates the human object in every frame after the gait recognition to further reduce the computational cost. Only the region of interest within the tracked window is cropped and fed into OpenPose network to extract the joints information. The capturing rate in frame-per-second is significantly increased to enable the real-time application using a small-scale standalone module. As a result, when the capturing rate increases, it increases the human joints data samples on the motion plot to improve the signal quality for accurate postural assessment.

The development of real-time processing unit using Intel Up Squared board enables the edge computing system. Intel OpenVino facilitates the optimization of the deep learning model and optimizes the human pose estimation model to accelerate the model implementation on Intel CPU. Consequently, this sports screening system is optimized to reduce the latency for edge computing. The edge computing brings the computation closer to the athlete during the physical training. Essentially, it saves the transmission time and bandwidth usage to minimize the need for long distance of the communication between the server and client. Thus, the overall proposed gait estimation and tracking pipeline enables the fast processing computation to elevate the performance of postural assessment in sports applications.

6 Conclusion

A markerless gait estimation and tracking system with edge computing for postural assessment is proposed in this paper. Human joints are extracted from pose estimation with 14 joint-to-neck distances, which are normalized with relative to neck-to-pelvis distance to obtain invariant features from any camera's orientation. These features are used to generate the spatial-temporal sequence and passed to LSTM network for gait recognition. The indexed person is tracked for postural assessment. The automation postural assessment system can achieve up to 9 FPS on Intel Up Squared board with 95% recognition accuracy and the purpose of the low power consumption and latency reduction are achieved. The pipeline introduced the gait estimation as a significant measure to body posture in sports analysis based on the bio-mechanics of human body. The body orientation of the athlete creates the hard-occlusions and elbow angle only applied with side view which affects the result of

sports analysis. In the future works, to filter out the noise of the spatial-temporal sequence caused by hard-occlusions, 3D geometry pose estimation will be used to simulate the joint which has the potential to eliminate zero dropping caused by the occlusions to improve the gait recognition, sports analysis and also the quality of dataset collection.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. We would like to gratefully acknowledge the support of NVIDIA Corporation with the donation of the the Quadro P6000 GPU used for this research.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alp Güler R, Neverova N, Kokkinos I (2018) Densepose: Dense human pose estimation in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7297–7306
2. Andriluka M, Pishchulin L, Gehler P, Schiele B (2014) 2d human pose estimation: New benchmark and state of the art analysis. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3686–3693
3. Barris S, Button C (2008) A review of vision-based motion analysis in sport. *Sports Med* 38(12):1025–1043
4. Beravs T, Reberšek P, Novak D, Podobnik J, Munih M (2011) Development and validation of a wearable inertial measurement system for use with lower limb exoskeletons. In: 2011 11th IEEE-RAS international conference on humanoid robots. IEEE, pp 212–217
5. Bialkowski A, Lucey P, Carr P, Denman S, Matthews I, Sridharan S (2013) Recognising team activities from noisy data. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 984–990
6. Blanchard N, Skinner K, Kemp A, Scheirer W, Flynn P (2019) keep me in, coach!: A computer vision perspective on assessing acl injury risk in female athletes. In: 2019 IEEE Winter conference on applications of computer vision (WACV), pp 1366–1374
7. Bolme DS, Beveridge JR, Draper BA, Lui YM (2010) Visual object tracking using adaptive correlation filters. In: 2010 IEEE Computer society conference on computer vision and pattern recognition, pp 2544–2550
8. Cao Z, Simon T, Wei SE, Sheikh Y (2017) Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7291–7299
9. Chen C, Wang G, Peng C, Fang Y, Zhang D, Qin H (2021) Exploring rich and efficient spatial temporal interactions for real-time video salient object detection. *IEEE Trans Image Process* 30:3995–4007
10. Colyer SL, Evans M, Cosker DP, Salo AI (2018) 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* 4(1):24
11. Dugan SA, Bhat KP (2005) Biomechanics and analysis of running gait. *Physical Medicine and Rehabilitation Clinics* 16(3):603–621
12. ES C, CM W, MP M, BC H (2012) Changes in muscle activation patterns when running step rate is increased. *Gait & Posture* 36(2):231–235
13. Foundation A (2019) Arthritis by the numbers pp 16–27

14. Girshick R, Radosavovic I, Gkioxari G, Dollár P, He K (2018) Detectron <https://github.com/facebookresearch/detectron>
15. Hamer P, Bloomfield J (2009) Posture
16. Hayashi M, Oshima K, Tanabiki M, Aoki Y (2015) Upper body pose estimation for team sports videos using a poselet-regressor of spine pose and body orientation classifiers conditioned by the spine angle prior. *IPSJ Trans Comput Vision Appl* 7:121–137
17. He K, Gkioxari G, Dollár P, Girshick RB (2017) Mask R-CNN. arXiv:1703.06870
18. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Computation* 9:1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>
19. Jin S, Ma X, Han Z, Wu Y, Yang W, Liu W, Qian C, Ouyang W (2017) Towards multi-person pose tracking: Bottom-up and top-down methods. In: *ICCV Posetrack workshop, vol 2*, p 7
20. Joshi K, Tripathi V, Bose C, Bhardwaj C (2020) Robust sports image classification using inceptionv3 and neural networks. *Procedia Comput Sci* 167:2374–2381
21. Kim K, Grundmann M, Shamir A, Matthews I, Hodgins J, Essa I (2010) Motion fields to predict play evolution in dynamic sport scenes. In: *2010 IEEE Computer society conference on computer vision and pattern recognition*. IEEE, pp 840–847
22. Konstantinou E, Lasenby J, Brilakis I (2019) Adaptive computer vision-based 2d tracking of workers in complex environments. *Automation in Construction*
23. Lan T, Sigal L, Mori G (2012) Social roles in hierarchical models for human activity recognition. In: *2012 IEEE Conference on computer vision and pattern recognition*. IEEE, pp 1354–1361
24. Li Y, Li S, Chen C, Hao A, Qin H (2020) A plug-and-play scheme to adapt image saliency deep model for video data. *IEEE Transactions on Circuits and Systems for Video Technology*
25. Lukezic A, Vojir T, Cehovin Zajc L, Matas J, Kristan M (2017) Discriminative correlation filter with channel and spatial reliability. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*
26. Mündermann L, Corazza S, Andriacchi TP (2006) The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of Neuroengineering and Rehabilitation* 3(1):6
27. Ning G, Liu P, Fan X, Zhang C (2018) A top-down approach to articulated human pose estimation and tracking. In: *Proceedings of the European Conference on Computer Vision (ECCV)*
28. Osokin D (2018) Real-time 2d multi-person pose estimation on cpu: Lightweight openpose. arXiv:1811.12004
29. Phang JTS, Lim KH (2019) Real-time multi-camera multi-person action recognition using pose estimation. In: *Proceedings of the 3rd international conference on machine learning and soft computing*, pp 175–180
30. Pishchulin L, Insafuldinov E, Tang S (2015) Deepcut: Joint subset partition and labeling for multi person pose estimation. arXiv:1511.06645
31. Promrit N, Waijanya S (2019) Model for practice badminton basic skills by using motion posture detection from video posture embedding and one-shot learning technique. In: *Proceedings of the 2019 2nd artificial intelligence and cloud computing conference*, pp 117–124
32. Punzi L, Galozzi P, Luisetto R, Favero M, Ramonda R, Oliviero F, Scanu A (2016) Post-traumatic arthritis: overview on pathogenic mechanisms and role of inflammation. *RMD open* 2(2)
33. Quinn TJ, Dempsey SL, LaRoche DP, Mackenzie AM, Cook SB (2019) Step frequency training improves running economy in well-trained female runners. *Journal of Strength and Conditioning Research*
34. Raju JP, Reddy YC, Reddy P (2019) Smart posture detection and correction system using skeletal points extraction. In: *International conference on e-business and telecommunications*. Springer, pp 177–181
35. Reyes-Ortiz JL, Anguita D, Ghio A, Parra X (2012) Human activity recognition using smartphones data set. *UCI Machine Learning Repository*; University of California, Irvine, School of Information and Computer Sciences: irvine, CA USA
36. Richards, Ambreen, Chohan, Renuka, Erande (2013) Biomechanics. In: *Tidy's physiotherapy*. Churchill livingstone, pp 331–368
37. Romanov N, Robson JDr (2002) Nicholas Romanov's Pose Method of Running: A New Paradigm of Running. *Dr. Romanov's Sport Education*. PoseTech
38. Sweeting K, Mock M (2007) Gait and posture - assessment in general practice. *Australian Family Phys* 36:398–401, 404
39. Valdes AM, Doherty SA, Muir KR, Wheeler M, Maciewicz RA, Zhang W, Doherty M (2013) The genetic contribution to severe post-traumatic osteoarthritis. *Ann Rheum Dis* 72(10):1687–1690

40. Wang J, Qiu K, Peng H, Fu J, Zhu J (2019) Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance. In: Proceedings of the 27th ACM international conference on multimedia, MM '19. association for computing machinery, USA, pp 374–382
41. Zhou H, Hu H (2008) Human motion tracking for rehabilitation—a survey. *Biomed Signal Process Control* 3(1):1–18

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.