A Kalman Filter based Approach for Markerless Pose Tracking and Assessment

Xuqi Zhu, Issam Boukhennoufa, Bernard Liew, Klaus D McDonald-Maier and Xiaojun Zhai
University of Essex, Colchester, United Kingdom
{xz18173, ib20472, bl19622, kdm, xzhai}@essex.ac.uk

***Abstract*—The assessment and treatment of diseases that causes movement impairments typically rely upon clinical information obtained from self-reported rating scales and clinical observation from health professionals. Currently, objective clinical gait analysis requires the use of expensive cameras or wearable sensors, which may be too time-consuming for routine clinical usage. Therefore, an assessment system that can provide non-invasive gait analysis tool is desired. In this paper, we propose a cost-efficient assessment system that combines computer vision and artificial intelligence technology to analyse human gait, which can provide a basic clinical report for clinicians to evaluate the patients' recovery to facilitate clinical decision making. And a series of experiments is taken to improve this assessment system performance. Those experiments showed that when the visibility threshold (VT) is set to a relatively high level (VT = 0.4), the post-processing part, which includes a Kalman filter and a frequency domain filter, can improve the human pose detection model (BlazePose)'s joint angle prediction accuracy by 10%. This post-processing method can be applied to other human body detection models to achieve filtering and feature extraction from joint angle signals for clinical gait analysis.**

***Index Terms*—Gait analysis, Mediapipe, Assessment system.**

I. Introduction

Common to almost all neuromotor (e.g. stroke, Parkinson's disease) and lower limb diseases (e.g. arthritis) may cause degradation of motor control, which negatively impacts their life quality. According to Statistics from Global Burden of Disease Collaborators, in 2016, over 80 million people worldwide suffered a stroke, of which 5.5 million died as a result of a stroke [1]. And the disability rate of such disease may be as high as 75% [2]. Parkinson's disease affects an estimated 6.2 million people worldwide in 2015 and caused 117,000 deaths [3], while most Parkinson's victims suffer persistent symptoms of the disease, such as tremors, stiffness, slowness of movement and instability. In the United States, about 43 million individuals (1 in 6) have arthritis, which leads to an urgent need for advanced medical technology to cope with the increasing demand for arthritis-related hospitalization and treatment with rehabilitation [4]. Given that walking is fundamental to activities of daily living and exercise, the rehabilitation of this motor task is key to improving the quality of life and health of patients.

Disease recovery for these patients can be difficult and normally it has a long recovery process. Effective communication to assess the disease and customize personalized rehabilitation planning, combined with monitoring and feedback among clinicians, patients and their families, is the key to achieve a successful rehabilitation. To help healthcare professionals make timely recommendations, a home health assessment system using markerless gait detector is introduced in this paper, and it will focus on abnormal gait analysis, which is one of the significant pathological features of these diseases.

In order to reduce the noise of predicting biomechanical signals introduced by markerless gait detector, Kalman Filter (KF) and Discrete Fourier Transform (DFT) algorithms have been introduced to cope with the loss of target of the human pose detection model, and the extreme values of joint angle signal are smoothed [5], [6]. In addition to reducing noise in the biomechanical signals, the most important function of the evaluation system is to extract features from joint angle signals and use the extracted features to establish a normative sample model, which can be used to quantify the magnitude of movement impairment in patients. Furthermore, Principal Component Analysis (PCA) algorithm is used to achieve this requirement [7]. To evaluate the patients' health condition, the deviation for identifying normal gait and morbid gait is defined using Mahalanobis distance, which describes the distance between the target sample and the group sample [8]. These algorithms construct a special post-processing method for the proposed clinical gait analysis system.

This paper aims to design a rehabilitation assessment system for aiding clinicians in clinical-decision making, which provides an exploration of using human pose detection models in the clinical gait analysis field. In addition, a reusable post-processing algorithm and strategy for clinical gait analysis are also proposed. Moreover, to improve the joint angle signal accuracy and the assessment system performance, a series of experiments are performed to identify suitable hyper-parameter settings. Although the proposed assessment system is currently inadequate for a fully independent gait analysis and assessment that is completely free from professional involvement, the system can effectively assist healthcare professionals to learn patients' condition from a basic gait analysis without requiring special locations or time arrangements.

The rest of the paper is organized as follows. The overview of the assessment system is introduced in Section II. Section III illustrates the system architecture and discusses implementation details of the assessment system. The experimental

This work is supported by the UK Engineering and Physical Sciences Research Council through grants EP/V034111/1 and EP/V000462/1.

results and evaluation are provided in Section IV. In the last section, the conclusions are discussed.

II. System Design

The assessment system focuses on tackling the following four challenges. Firstly, detection of joints' locations. Secondly, prediction of walking speed. Thirdly, de-noise of raw joint angle signals. Finally, create feature model to assess patients' disease conditions. In this section, the proposed solutions for the above challenges are discussed respectively.

A. Pose Detection

The joint angle is one of the important kinematic parameters in human gait analysis [9]. In this paper, BlazePose is applied to predict key points of the human body, and the lower limb joint angle-time series signals were obtained by further processing those key points [10].

To reduce the background environment noise level introduced in joint prediction, the pose segmentation mask will be applied in the pre-processing stage. Fig. 1(b) illustrates the anatomical joint landmarks [11] used by BlazePose for human pose prediction and segmentation masking for a video frame in a public dataset [12]).

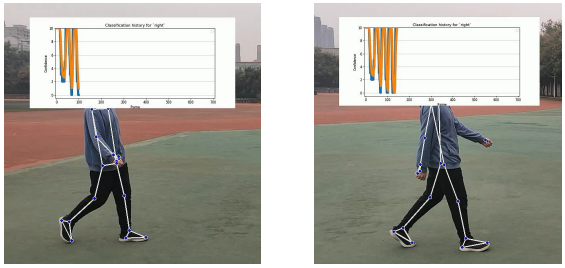
(a) Skeleton (b) Mask

Fig. 1. The key point landmark and joints prediction [12].

B. Pose Classification

The walking speed is one of important parameters for human gait analysis. The stride cycle can be roughly divided into two-step: left foot in the front (Posture A), and right foot in the front (Posture B). Thus, the counting strides task becomes the pose classification task. Since each state has a unique body posture, and those postures are described by a data structure called Pose Distance Embedding (PDE) containing a list of joint distances (signed value), the Mediapipe provides a solution to count the number of repeated movements that uses the K Nearest Neighbor (KNN) algorithm [11], [13] to identify the one state of the repeat action. Once the algorithm finds that the target is in the corresponding posture, the counter then updates.

To solve the problem that counter increases unexpectedly for the same posture in continuous video frames, exponential moving average (EMA) smoothing algorithm [11], [14] and two thresholds were applied respectively. The EMA algorithm can firstly reduce the undesired identification error, then entry threshold T_1 and exit threshold T_2 (e.g. $T_1 > T_2$) were set to avoid the cases when probability fluctuates around the threshold caused phantom counts.



(a) Posture A ('left') (b) Posture B ('right')

Fig. 2. Pose classification.

Fig. 2(a) and Fig. 2(b) shows two types of postures defined in this paper. The blue curves in Fig. 2(a), and Fig. 2(b) are the posture confidence before smoothing, and the orange curves in Fig. 2(a), and Fig. 2(b) are the posture confidence after smoothing. The walking speed can be then calculated from the number of strides divided by the frame interval of two strides and updated thereafter.

C. Post-processing

The raw joint signals predicted by BlazePose usually suffer from the pose detection failures and other noises caused by the extreme values. In this part, two post-processing algorithms were introduced to solve those issues.

1) Kalman Filter: Due to the errors of the prediction model as well as other overlapped body parts, it is inevitable to avoid detection failure while the target is moving. Most of the time, the target loss only occurs in a very short period. Therefore, a Kalman Filter (KF) can be designed to correct the targeted state from the prediction [15].

Since most of the time the changes of joint angle have steady velocity and acceleration, the joint angle time series signals can be thus considered as a simple linear model [15]. A KF can then automatically adjust parameters (error estimate covariance matrix P and Kalman gain K) to handle the influence of velocity and acceleration changes on the prediction results by combining the current measurement results.

$$\theta_t = \lim_{\Delta t \to 0} \theta_{t-1} + v_t \Delta t + \frac{1}{2} a_t \Delta t^2 + R \tag{1}$$

$$R = \sum_{n=3}^{\infty} \frac{1}{n!} v_t^{(n-1)} \Delta t^n \tag{2}$$

$$v_t = v_{t-1} + a_t \Delta t \tag{3}$$

$$a_t = a_{t-1} \tag{4}$$

The above classical kinematic model equation derived from the Maclaurin series can be used to describe the changes in joint angles over time [5], where θ_t represents the joint angle at time point t. Δt, is the reciprocal of frame per second (FPS), also known as the interval period of two consecutive frames. v is joint rotate speed, a is the angular acceleration,

and R is called remainder which is ignored because this value is extremely small with the growth of n. This kinematic model equation can be transformed into a matrix equation in KF.

$$\begin{bmatrix}\theta_t\\ v_t\\ a_t\end{bmatrix} = \begin{bmatrix}1 & \Delta t & \frac{\Delta t^2}{2}\\ 0 & 1 & \Delta t\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\theta_{t-1}\\ v_{t-1}\\ a_{t-1}\end{bmatrix} + Pnoise_t \tag{5}$$

$$\begin{bmatrix}\theta_t\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\end{bmatrix}\begin{bmatrix}\theta_{t-1}\\ v_{t-1}\\ a_{t-1}\end{bmatrix} + Mnoise_t \tag{6}$$

$$A = \begin{bmatrix}1 & \Delta t & \frac{\Delta t^2}{2}\\ 0 & 1 & \Delta t\\ 0 & 0 & 1\end{bmatrix}\ S_t = \begin{bmatrix}\theta_t\\ v_t\\ a_t\end{bmatrix}\ H = \begin{bmatrix}1 & 0 & 0\end{bmatrix} \tag{7}$$

The matrix equation (5) describes the calculation of (1), (2) and (3), which is the prediction step in KF [15]. The matrix A is a state-transition matrix, and $Pnoise_t$ in (5) is a processing noise obeying the Gaussian distribution $Pnoise_t \sim (0, Q)$, where Q is a process noise covariance matrix. The equation (6) extracts useful state information (joint angle θ), where the matrix S_t is a joint state at time point t, the matrix H is a measurement matrix, and the $Mnoise_t$ is measurement noise obeying the Gaussian distribution $Mnoise \sim (0, R)$, where R is measurement noise covariance matrix [16]. The above equations (5, 6, 7) only demonstrate the situation of one joint angle, in this paper, those matrices will contain four joint states (left knee, right knee, left hip, right hip), each of them has three parameters: joint angle θ, rotate speed v, and rotate acceleration. Thus, the state matrix is a (12, 1) matrix.

In addition to the kinematic model definition, the KF also requires initializing some basic matrix parameters. The process noise covariance matrix Q, the measurement noise covariance matrix R, and the error estimate covariance matrix P at $t = 0$, called P_0 [16]. Those matrix parameters need to be carefully adjusted in advance. In this paper, they are set as follows.

$$Q = \alpha\begin{bmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{bmatrix}\ R = \beta\begin{bmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{bmatrix} \tag{8}$$

$$P_0 = \begin{bmatrix}100 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 100\end{bmatrix} \tag{9}$$

The error estimate covariance matrix P_t is automatically adjusted per iteration. The matrix Q and matrix R reflect the confidence for the measured and predicted results respectively. To be more specific, a higher α or β means less confidence in the observations or predictions. By modifying coefficients α and β to adjust the performance of the KF to achieve a relatively ideal effect. The detail about α and β is discussed in section IV-B.

2) Frequency Domain Filter: Although the KF algorithm can solve joint angle loss when joint visibility is below the threshold, it has a limited effect on removing extreme values and waveform distortion. Thus, the time-series signal needs an extra filter to de-noise and smooth the signals. Usually, the extreme values and waveform distortion are caused by noise with relatively low energy in the energy density spectrum. To de-noise the joint angle signal, a frequency domain filter implemented by Discrete Fourier Transform (DFT) and Inverse Discrete Fourier Transform (IDFT) is applied in the post-processing part.

The DFT algorithm can compute the amplitude and phase of each frequency component. According to Parseval's theorem, the energy density of corresponding frequency components is $|Amplitude|^2/N$, where N is the sampling number of the signal. For the assessment system, the sampling number is the number of frames for input video. Except for the amplitude of each frequency component, the phase difference of the principal frequency component in the frequency domain of two legs is one of the features that can be used in future analysis.

Morgan K D and Noehren B [6], come up with a filter strategy that select out the frequency components with top K energy density to recover the corresponding time-domain signal. For easy to mark, we call it the Peak Top K Components strategy (PTKC). This method performs well for wearable sensor situations. Inspired by this method, the strategy to pick those frequency components that can recover a time-domain signal with the lowest Root Mean Square (RMS) error is proposed. For easy to mark, we call it the Lowest Root Means Square Error strategy (LRMSE). This strategy (LRMSE) is built on the assumption that the joint angle time series signals from BlazePose are close to the ground truth, otherwise, this strategy will make the recovered signal worse. The comparison between the experimental results for the two strategies is illustrated in Section IV-B.

D. Feature Model

The final object of this work is to build a system to assess the patients' disease condition. To achieve this goal, the Principal Components Analysis (PCA) is used to create a model to obtain a group of sample features [17].

$$S_{(n,d_1)} = \overbrace{\begin{bmatrix}Sample_{1,1} & \cdots & Sample_{1,d_1}\\ \vdots & \ddots & \vdots\\ Sample_{n,1} & \cdots & Sample_{n,d_1}\end{bmatrix}}^{Features} \tag{10}$$

$$Cov_{(d_1\times d_1)} = S^T_{(n\times d_1)}S_{(n\times d_1)} \tag{11}$$

$$Cov_{(d_1\times d_1)}V_{i,(d_1,1)} = \lambda_i V_{i,(d_1,1)} \quad (i \in [1, d_2]) \tag{12}$$

$$LoadingMatrix_{(d_1,d_2)} = \begin{bmatrix}V_1 & \cdots & V_{d_2}\end{bmatrix} \tag{13}$$

$$EigenValue_{(1,d_2)} = \begin{bmatrix}\lambda_1 & \cdots & \lambda_{d_2}\end{bmatrix} \tag{14}$$

$$Reduction_{(n,d_2)} = S_{(n,d_1)}LoadingMatrix_{(d_1,d_2)} \tag{15}$$

Equations (10-15) illustrate how the PCA can be used to generate a feature model. Suppose d_1 is the original dimension size, d_2 is the size of the reduced dimensions, and n is the number of samples. In addition to dimension reduction, eigenvectors (loading matrix is constructed by eigenvectors) and

eigenvalues generated by PCA can be used to build a feature model. The eigenvalues indicate which axis (eigenvectors) are important at the new scale.

The value located on (i, j) in the covariance matrix Cov shows the internal relationship between the $i-th$ scale (e.g. height) and $j-th$ scale (e.g. age). If this value is positive, then those two scales have a positive correlation. If this value is negative, then those have a negative correlation. The loading matrix is a matrix combining a list of eigenvectors, the $i-th$ eigenvector is sorted out according to its score, where the score is $\lambda_i/sum(EigenValue)$ (the $i-th$ eigenvalue is the variances of the $i-th$ eigenvector) [8], [17]. In this paper, the PCA will create a feature model with reduced dimensions that can explain 90% of samples' variance, which require a loading matrix to select out a set of eigenvectors (also be known as principal components, PCs) who satisfies $\sum_i^{d_2} \lambda_i/sum(EigenValue) > 90\%$. The $Reduction$ matrix is the final feature model for n samples with reduced dimensions.

III. System Design

In this sections, the details about how the system processes video input and generates reports will be discussed.

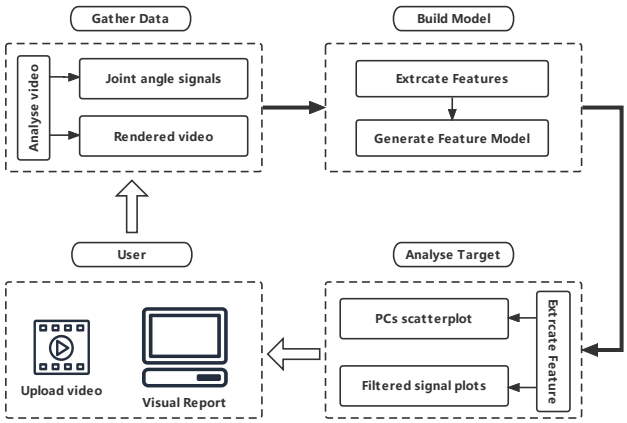


Fig. 3. System work flow.

Fig. 3 summarises the system working flow. The process of the assessment system can be mainly divided into three steps: 1) Gather data; 2) Build a feature model; 3) Analyse the target.

In the first step, the users need to provide a video input that meet the system requirements (e.g. horizontal and stabilized view, around 30 FPS, only one target in the camera, suitable resolution and a clear silhouette of the human body), then the BlazePose will return a 33 predicted joint locations and corresponding visibility scores. In this paper, the hip and knee angles are the main characteristics that the assessment system focuses on, while other joints are adopted to help pose classification function design in stride counter module. The joint angles in each frame are combined into a discrete-time domain sequence $[\theta_0, \theta_1, \cdots, \theta_t]$ as the raw joint angle signal, where the joint angles are computed by using the cosine law $cos(\theta) = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}$. Then KF algorithm will replace loss value in the raw joint angle signals. After that, the no-loss signals will be transformed into the frequency domain by using the DFT algorithm. A strategy called Least Root Mean Square Error (LRMSE) is adopted as the filter strategy to select principal frequency components, and then the system temporarily saves the recovered signal for rendering video and following processing.

In the second step, the samples were divided into different groups according to their age and gender, given that female and male participants have different lower limb skeleton structures and the gait parameters are sensitive to the physiological age [18]. Then a set of uniform gait parameters were gathered from samples: age, gender, height (cm), weight (kg), hip phase different, knee phase different, left knee max angle, left knee min angle, right knee max angle, right knee min angle, left knee flexion angle, right knee flexion angle, left hip max angle, left hip min angle, right hip max angle, right hip min angle, left hip flexion angle and right hip flexion angle. Those discrete gait parameters was be standardized and uploaded as a feature matrix for the respective age group, to generate a feature model using a PCA algorithm. The system directly used the feature model to analyze the target after it has been generated.

Finally, a standard score was used to assess the disease state. In this paper, a value called Mahalanobis distance [17], also known as T^2 score, was applied to measure the deviation between individuals and entity samples. Suppose the target feature matrix is $T_{(1,d_1)}$ and the λ_i is the i^{th} PC's eigenvalue, and following equations show the T^2 score definition.

$$T^2 = z^t D^{-1} z \tag{16}$$

$$z = T_{(1,d_1)} LoadingMatrix_{(d_1,d_2)} \tag{17}$$

$$D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{d_2} \end{bmatrix} \tag{18}$$

Instead of Euclidean distance, the T^2 score takes the variances of PCs (eigenvalue) in reduced dimensions as a factor, a higher eigenvalue the more information the corresponding PC contains [17].

IV. Experiment

A. Public Datasets

Ghorbani S and Mahdaviani K provides a set of video datasets [12] with 30 FPS and 800 $\times$ 600 resolution processed by Visual3D and corresponding joints' location. The volunteers in the video samples were required to wear normal clothing to complete a series of actions and sports movements. Those actions and sports movements were recorded by video cameras and motion capture cameras in a space of approximately 3 by 5 meters. Moreover, since the system focused on walking gait parameters, the selected video samples are clipped and only save the walking period. In this paper, 9 male samples aged 21 - 30 years, height 160 - 180 cm, weight 60 - 80 kg were selected to generate feature models, while 8 male

and female samples were used to evaluate the recovered signal accuracy.

B. Hyper-parameters Setting

To improve the system performance, a series of work was performed to determine the optimal set of hyper-parameters. The error rate [5] was defined to quantify the effects of hyper-parameters in the results.

$$error = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{J}\sum_{j=1}^{J}\frac{\sqrt{\frac{1}{F}\sum_{t=1}^{F}(q_{s,j}(t)-q_{gt,s,j}(t))^2}}{max(q_{gt,s,j})-min(q_{gt,s,j})} \quad (19)$$

where S represents the number of samples, The parameter J represents the number of joint angles. The parameter F represents the number of frames for the corresponding sample. The function $q_{s,j}$ is the joint angle time series for the sample s joint j. The 'None' value in raw signal (joint visibility is below the threshold) is replaced by '0'.

According to equation (8), the value of α and β have a great influence on the KF behaviours. To be more specific, a higher α means less confidence in measurement, while a higher β means less confidence in the prediction.

TABLE I
COMPARE KF PERFORMANCE WITH DIFFERENT α AND β
(BLAZEPOSE SEGMENTATION MASK = TRUE)

	Raw signal	α and β value	After KF
Error (%)	33.02	α = 0.0001 β = 0.001	27.67
		α = 0.001 β = 0.001	**23.54**
		α = 0.0005 β = 0.001	24.25
		α = 0.001 β = 0.01	27.70

Table I illustrates that the KF algorithm performed relatively better when $\alpha = 0.001$ and $\beta = 0.001$. Although the error rate of the signal after KF is decreased by 10%, there were undesired noises in the result. Thus, the frequency domain filter strategy LRMSE was applied to pick principal frequency components to get a lower error rate signal. In Section II-C2, two strategies were introduced to implement the frequency domain filter. A comparison experiment was performed to evaluate the performance of two filter strategies for the frequency domain filter of de-noising joint angle signal task.

TABLE II
COMPARE PERFORMANCE OF TWO FILTER STRATEGIES
(BLAZEPOSE SEGMENTATION MASK = TRUE, $\alpha = 0.001$, $\beta = 0.001$)

	After KF	N components	LRMSE	PTKC
Error (%)	23.54	n = 2	24.53	24.53
		n = 3	23.84	23.84
		n = 4	23.61	23.61
		n = 5	**23.21**	24.53
		n = 6	23.19	23.19
		n = 7	**23.17**	23.84
		n = 8	23.25	23.25

Table II demonstrates that LRMSE can improve the accuracy slightly (In the case picking 5 or 7 frequency components), but the behaviour of LRMSE is the same as PTKC in most cases since the frequency components with a higher energy density usually determine the overall trend of the time domain waveform. If there is a demand for low time consumption, the PTKC is then the better option.

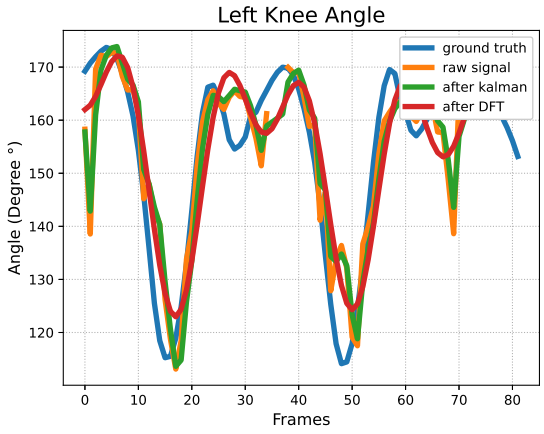


Fig. 4. Recovered signal v.s. raw signal (Subject_24, LRMSE, 5 components).

Fig. 4 displays the joint angle signal in three stages. The KF algorithm complements the lost angles in the raw signal, and the signal after DFT demonstrated that the frequency domain filter smooths the waveform, but it is inevitable for information losing by applying filter which causes a gap between ground truth and recovered signal.

According to practice experiences, the BlazePose model and KF algorithm require an iteration to update internal variables to enter a stable stage which is, called the convergence period. The predicted joint angle signal during the convergence period has a higher error rate, which makes the frequency domain filter fail to extract the principal frequency components. To avoid a higher error signal part affecting the whole signal, the beginning of the signal is usually discarded.

TABLE III
THE ACCURACY COMPARISON FOR DIFFERENT Cut AND N $components$

Cut (%)	Error (%) Raw signal	Error (%) After KF	N components	Error (%) After DFT
0	33.02	23.54	n = 4	23.61
			n = 5	23.21
			n = 6	23.19
5	33.13	23.34	n = 4	23.19
			n = 5	23.13
			n = 6	23.04
10	33.35	23.17	n = 4	23.16
			n = 5	22.93
			n = 6	22.95
15	33.76	23.22	n = 4	23.20
			n = 5	22.94
			n = 6	22.96

Table III shows the relation between the clipping percentage and processed signal accuracy. In practice, the filtered signal

can get relatively lower error rates when $cut = 10\%$ and the number of frequency components is set to 5.

In addition to post-processing part hyper-parameters, the BlazePose provides segmentation mask (SM), min detection confidence (MDC), and min tracing confidence (MTC) as initialization parameters to adjust the performance and computing resource consumption for different situations. To improve BlazePose performance, some experiments were taken to find the suitable parameters setting. Except for those initialization parameters, the influence of visibility threshold (VT) for joint landmarks is discussed. The frame lost rate (FL), the number of detection failed frames divided by the total number of frames in a video, was used to assess the system performance.

Table IV shows that MTC and MDC seemed to have less effect on BlazePose performance. However, the VT and SM are key factors that can greatly affect the result's accuracy. Turning on the segmentation mask will improve the raw signal accuracy in lower VTs, but it decreases the raw signal accuracy in the higher VT. According to the increment in loss frames rate for $VT = 40\%$ when turning on the segmentation mask, a possible explanation is that the segmentation mask reduces the background noise but makes the BlazePose model difficult to detect the key points.

TABLE IV
SIGNAL ACCURACY FOR DIFFERENT BLAZEPOSE PARAMETERS

SM	VT (%)	MDC (%)	MTC (%)	Error (%) Raw signal	Error (%) After KF	Error (%) FL (%)
TRUE	20	50	30	22.87	23.57	0
	40	30	30	32.83	23.53	1.85
	40	30	50	32.83	23.53	1.85
	40	50	30	33.02	23.54	1.85
	40	50	50	33.02	23.54	1.85
FALSE	20	50	30	24.60	25.54	0
	40	30	30	27.67	25.47	0.46
	40	30	50	27.67	25.47	0.46
	40	50	30	28.10	25.51	0.46
	40	50	50	28.10	25.51	0.46

An interesting phenomenon is that the lower the VT, the higher the accuracy of the raw signal, but the signal accuracy after KF is even worse than the corresponding raw signal. One possible reason is that the lost joint angle (None) is replaced by an imprecise predicted value (extreme value). Because "None" is replaced by 0 when calculating the signal error, these extreme values are closer to the ground truth than 0, thus, the error rate at low VT is decreased. However, extreme values in the raw signal made the KF algorithm trust the measured values by mistake, resulting in a bias. The evidence is that the signal error after KF in $VT = 20\%$ is higher than that in $VT = 40\%$.

C. Evaluation of the Assessment System

In this section, an example of the proposed assessment system is given for assisting healthcare worker, where it helps the healthcare worker to learn the conditions of patient with arthritis. According to previous experimental results, the hyper-parameters are set as follow: $MDC = 50\%$, $MTC = 30\%$, $VT = 40\%$, $SM = TRUE$, $cut = 10\%$, $\alpha = 0.001$, $\beta = 0.001$, $N \quad components = 5$ and the filter strategy is LRMSE.

Three volunteers recruited from the university (e.g. age between 21 to 25, male, height between 170 to 180 cm, weight between 75 to 85 kg) were invited to participate in this experiment. They are labelled as target A, target B, and target C. Target A is a sample in which the volunteer does not have arthritis, while target B is a sample in which the patient has arthritis and target C is a sample in which the volunteer has recovered from arthritis.

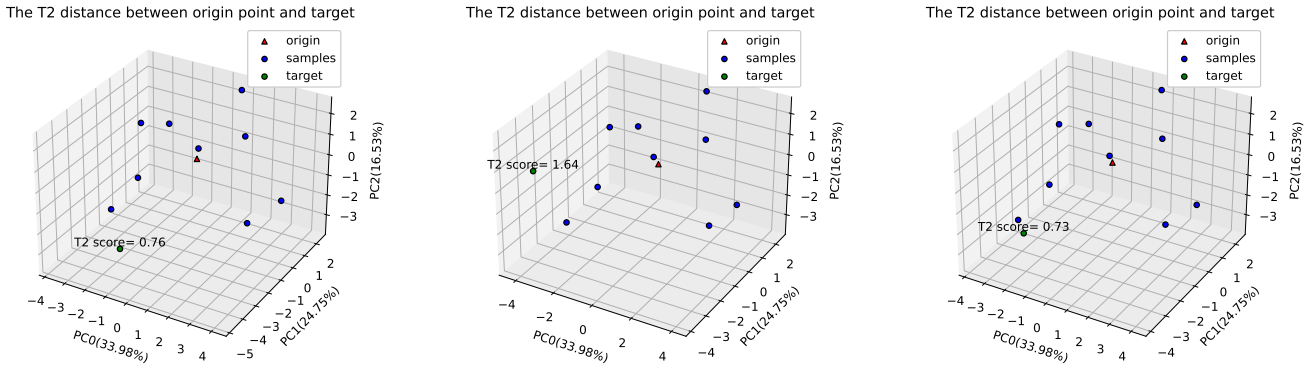
The signals at the front of the red dash lines in Fig. 6 are discarded. According to the comparison between joint angle signal plots in Fig. 6, target B's left knee flexion declines sharply from normal gait. The target C has a disease on his left leg.

The scatter plots in Fig. 5 demonstrate the deviation between normal samples and targets. Nine Samples (blue points) from public dataset are scattered in the plots, and the origin point (red triangle) is the geometrical center of those samples, while the T^2 score describes a kind of distance between target (green point) and origin point. Three PCs (PC0, PC1, PC2) can explain around 75% of the samples' variance, and the T^2 score takes account of the PCs having 90% variance explanation totally. Target B's gait gets a higher T^2 score (1.64) than target A (0.76) and target C (0.73), which means the lower limbs diseases result in abnormal gait.

To summarize, the assessment system can help medical workers to get essential knowledge about patients' diseases and demonstrate visualized reports for further clinical diagnosis. The system can be used as a convenient kit for healthcare staff to keep track of their patients.

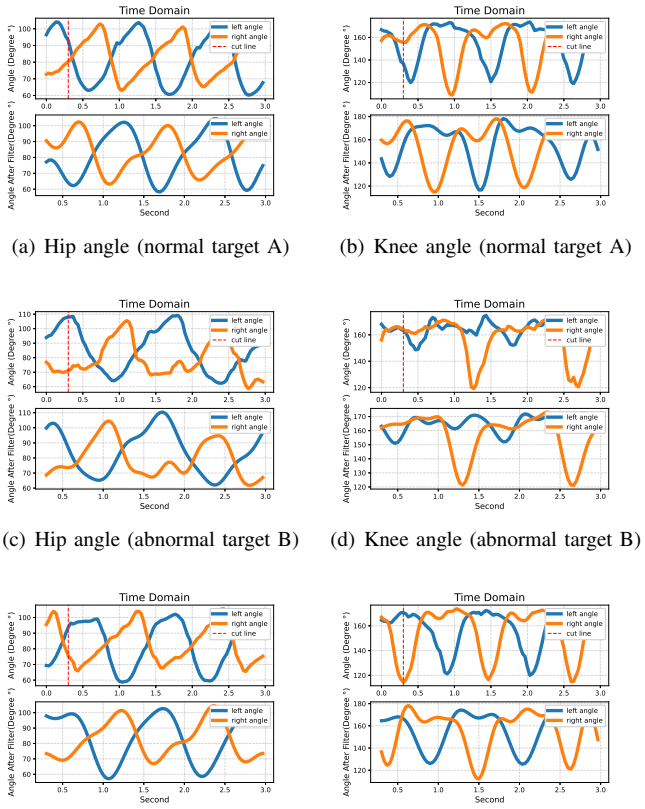
V. CONCLUSION

This paper focuses on the work of clinical gait analysis utilising a human posture detection model, and it also proposes a post-processing scheme for clinical gait analysis based on Kalman Filter and Frequency Domain Filters. This paper also explains how the PCA technique can be used to build feature models from noisy biomechanical signals. The experiment results suggest that the proposed post-processing can improve 10% BlazePose's joint angle prediction accuracy. Moreover, this post-processing method can also be used to filter and extract features of joint angle signals for other human body detection algorithms. Although the accuracy of the assessment system is currently insufficient to accomplish a fully independent clinical analysis task, the system could assist the expert in clinical diagnosis by analysing the patient's gait and returning the visualized reports. Compared with traditional gait analysis technology, it does not need bulky and expensive equipment and specific data collection sites, while the replacement is simple and inexpensive equipment (mobile phones and personal computers) and user-friendly operation mode (simple program commands). But, the feature model used in this paper may not sufficient due to the fact that only joint angle signals are used.



(a) Normal target A (b) Abnormal target B (c) Normal target C

Fig. 5. The scatterplots for three targets.



(a) Hip angle (normal target A) (b) Knee angle (normal target A)

(c) Hip angle (abnormal target B) (d) Knee angle (abnormal target B)

(e) Hip angle (normal target C) (f) Knee angle (normal target C)

Fig. 6. The joint angle signals for three targets.

Thus, we will focus on extending the feature model and try to combine other gait analysis techniques to obtain more effective information about patients' gait in the future.

REFERENCES

[1] V. L. Feigin, E. Nichols, T. Alam, M. S. Bannick, E. Beghi, N. Blake, W. J. Culpepper, E. R. Dorsey, A. Elbaz, R. G. Ellenbogen *et al.*, "Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016," *The Lancet Neurology*, vol. 18, no. 5, pp. 459–480, 2019.

[2] E. H. Lo, T. Dalkara, and M. A. Moskowitz, "Mechanisms, challenges and opportunities in stroke," *Nature reviews neuroscience*, vol. 4, no. 5, pp. 399–414, 2003.

[3] C. A. Davie, "A review of parkinson's disease," *British medical bulletin*, vol. 86, no. 1, pp. 109–127, 2008.

[4] M. J. Elders, "The increasing impact of arthritis on public health." *The Journal of Rheumatology. Supplement*, vol. 60, pp. 6–8, 2000.

[5] F. De Groote, T. De Laet, I. Jonkers, and J. De Schutter, "Kalman smoothing improves the estimation of joint kinematics and kinetics in marker-based human gait analysis," *Journal of biomechanics*, vol. 41, no. 16, pp. 3390–3398, 2008.

[6] K. D. Morgan and B. Noehren, "Identification of knee gait waveform pattern alterations in individuals with patellofemoral pain using fast fourier transform," *PloS one*, vol. 13, no. 12, p. e0209015, 2018.

[7] K. Deluzio and J. Astephen, "Biomechanical features of gait waveform data associated with knee osteoarthritis: an application of principal component analysis," *Gait & posture*, vol. 25, no. 1, pp. 86–93, 2007.

[8] K. J. Deluzio, U. P. Wyss, B. Zee, P. A. Costigan, and C. Serbie, "Principal component models of knee kinematics and kinetics: normal vs. pathological gait patterns," *Human Movement Science*, vol. 16, no. 2-3, pp. 201–217, 1997.

[9] C. Prakash, R. Kumar, and N. Mittal, "Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges," *Artificial Intelligence Review*, vol. 49, no. 1, pp. 1–40, 2018.

[10] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *arXiv preprint arXiv:2006.10204*, 2020.

[11] Google.com, "MediaPipe pose solution." [Online]. Available: https://google.github.io/mediapipe/solutions/pose.html

[12] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje, "Movi: A large multi-purpose human motion and video dataset," *Plos one*, vol. 16, no. 6, p. e0253157, 2021.

[13] O. Sutton, "Introduction to k nearest neighbour classification and condensed nearest neighbour data reduction," *University lectures, University of Leicester*, vol. 1, 2012.

[14] F. Klinker, "Exponential moving average versus moving exponential average," *Mathematische Semesterberichte*, vol. 58, pp. 97–107, 2011.

[15] G. Bishop, G. Welch *et al.*, "An introduction to the kalman filter," *Proc of SIGGRAPH, Course*, vol. 8, no. 27599-23175, p. 41, 2001.

[16] R. J. Meinhold and N. D. Singpurwalla, "Understanding the kalman filter," *The American Statistician*, vol. 37, no. 2, pp. 123–127, 1983.

[17] K. Dunn, "Process Improvement Using Data." [Online]. Available: https://learnche.org/

[18] M.-J. Chung and M.-J. J. Wang, "The change of gait parameters during walking at different percentage of preferred walking speed for healthy adults aged 20–60 years," *Gait & posture*, vol. 31, no. 1, pp. 131–135, 2010.