

Chapter 2

Video-Based Monitoring and Analytics of Human Gait for Companion Robot



Xinyi Liu, Md Imran Sarker, Mariofanna Milanova,
and Lawrence O’Gorman

Abstract Human gait is essential for long-term health monitoring as it reflects physical and neurological aspects of a person’s health status. In this paper, we propose a non-invasive video-based gait analysis system to detect abnormal gait, and record gait and postural parameters framework on a day-to-day basis. It takes videos captured from a single camera mounted on a robot as input. Open Pose, a deep learning-based 2D pose estimator is used to localize skeleton and joints in each frame. Angles of body parts form multivariate time series. Then, we employ time series analysis for normal and abnormal gait classification. Dynamic time warping (DTW)-based support vector machine (SVM)-based classification module is proposed and developed. We classify normal and abnormal gait by characterizing subjects’ gait pattern and measuring deviation from their normal gait. In the experiment, we capture videos of our volunteers showing normal gait as well as simulated abnormal gait to validate the proposed methods. From the gait and postural parameters, we observe a distinction between normal and abnormal gait groups. It shows that by recording and tracking these parameters, we can quantitatively analyze body posture. People can see on the display results of the evaluation after walking through a camera mounted on a companion robot.

X. Liu · M. I. Sarker · M. Milanova (✉)
University of Arkansas at Little Rock, Little Rock, AR, USA
e-mail: mgmilanova@ualr.edu

X. Liu
e-mail: xxliu8@ualr.edu

M. I. Sarker
e-mail: misarker@ualr.edu

L. O’Gorman
Nokia Bell Labs, Murray Hill, NJ, USA
e-mail: larry.o_gorman@nokia-bell-labs.com

2.1 Introduction

With the rise of aging populations, desire for independent living, high healthcare costs, non-invasive health monitoring, and smart personal and home communications, the time for “healthy-living” companion robots are approaching. Every 11 s, an older adult is treated in the emergency room for a fall; every 19 min, an older adult die from a fall. Healthcare cost of falls in the USA, including hospitalization, surgery, therapy, etc., is expected to increase as the population ages and may reach \$67.7 billion by 2020 [1, 2].

According to Selke [3], *lifelogging* is understood as different types of digital self-tracking and recording of everyday life. Another feature of *lifelogging* is that it is a continuous process that requires no user interaction. In the context of active-assisted living (AAL), sensors used for lifelogging can also be *ambient-installed* as opposed to wearable sensors, for instance, video surveillance or other cameras installed in nursing or smart homes to monitor and support older and fragile people [4–6]. Recently, there is a growing research on robots and their amplification in AAL Healthcare humanoid robots are designed and used by individuals at home or healthcare centers to analyze, treat and improve their medical conditions.

The quantitative gait analysis requires specific devices such as a 3D motion capture system, accelerometer, or force plate, which are time-, labor-, space-, and cost-consuming to use daily. Furthermore, monitoring gait in home using monocular camera without any annotations is interesting and not explored (practically unavailable for gait analyses).

The goal of this study is to develop a video-based marker-less system for non-invasive healthcare monitoring, using skeleton and joint location from pose estimation to extract gait features and generate alerts for abnormal gait, indicating needs for further medical attention. We propose use of Open Pose [7], a deep learning-based 2D keypoint framework that estimates the joint coordinates of persons in the image or videos obtained using a monocular camera, as it does not require external scales or markers. Using this estimator, we can automatically obtain the joint coordinates of persons in each image/movie recorded, thus enabling the calculation of joint angles or other spatial parameters useful for further gait analysis.

We first designed experiments to measure consistency of the system on a healthy population, then monitoring gait of subjects walking carrying the weight to replicate gait decay. The second set of experiments is “nudge” human posture monitoring. The system guides individuals in regular walking, freely assesses gait states and provides real-time personalized feedback to evaluate correct body posture. The robot will also connect individuals with family and friends through a virtual connection, and if needed, it will set up alarms.

Hypothesis Can we evaluate gait and motion dynamic integrating deep learning-based 2D skeleton estimator and time series analysis?

Research involves two complementary tasks:

1. Designing set of experiments and creating own dataset.

2. Analyzing gait changes using lower body skeleton representation and calculate falling risk.

Our contribution

1. We designed system to evaluate gait deviation and classification of normal and abnormal gait non-invasively over time.
2. Perform experiments on subject walking normally and the same subject walking carrying the weight.
3. Feature extraction and machine learning evaluation of human pose changes.

In Sect. 2.2, we describe evidence primarily from the medical and health-related literature on the correlation between gait and health. We also give an overview of other video gait recognition systems. In Sect. 2.3, we describe the methodology, which combines use of a deep learning gait and pose recognition engine, and classification module to determine significant health-related changes. Section 2.4 contains description of the experiment which simulated an attempt to understand how the system would work for people with changing gait or emotional state. Finally, in Sect. 2.5, we discuss results and current and future utility of this approach.

2.2 Related Work

In terms of data modalities, there are mainly two categories of gait analysis approaches: sensor based and vision based. Although sensor-based approach has shown its ability to reflect human kinematics, the requirement of certain sensors or devices and needs to be worn on human body for some approach has made it less convenient to be applied. Vision-based approaches are more unobtrusive and only requires cameras for data collection [8].

Recently, skeleton has been widely used [9–14]. Some researchers employed Microsoft Kinect camera to generate 3D skeleton using its camera SDK. Gait abnormality index is estimated using 3D skeleton in [11], joint coordinates are used as input of auto-encoders. Then, reconstruction errors from auto-coders are used to differentiate abnormal gaits. Jun et al. [12] proposed to extract features from 3D skeleton data using a two recurrent neural network-based auto-encoder for abnormal gait recognition and then evaluated the performance by feeding the features extracted to discriminative models. While Kinect RGB-D camera provided additional depth information, in our preliminary experiment of comparing skeleton output from Azure Kinect body tracking SDK and Open Pose, it is not as robust as Open Pose.

Similar to our work, Xue et al. [14] presented a system for senior care using gait analysis. They accurately calculated gait parameters, including gait speed, stride length, step length, step width, and swing time from 2D skeleton. Compare to their work, besides gait parameters, we also employed time series analysis techniques to characterize motion dynamics. In addition, our system gives classification result of normal or abnormal gait.

There are several gait-related public datasets. CASIA gait database [15] includes video files of the subject walking with variations in view angle, clothing, whether or not carrying a bag. However, it does not have samples of the subject walking with abnormal gait patterns. And the resolution is quite low, 320×240 , and captured at 25 fps. INIT Gait database [16] is designed for gait impairment research; it consists of one normal gait pattern and seven simulated abnormal gait styles. But only binaries silhouettes sequences were released. The walking gait dataset by Nguyen et al. is designed for abnormal gait detection [11]. It includes point cloud, skeleton, and frontal silhouette captured by Microsoft Kinect 2 camera. Nine subjects performed nine different gaits on a treadmill. Aside from one normal gait, they simulated abnormal gaits by padding a sole with three different levels of thickness and attaching a weight to ankle. However, it is captured in front view, while our method is designed on sagittal plane from the side view. Thus, we captured our own dataset for this research.

2.3 Methodology

2.3.1 Pipeline Overview

Our system is set up to capture people walking left-right or right-left through a camera view, termed an event. A camera is placed to capture the movement in sagittal plane of human body while walking (see Fig. 2.7).

The processing system contains the three modules as shown in Fig. 2.1. The system continually monitors a camera view, and when a person walks through the camera view, that event is detected and captured to a video clip. Pose estimation is performed on the clip to obtain a sequence of skeleton which contains location of body joints. In feature extraction and gait classification stage, we extracted gait and postural parameters and employed time series analysis to classify if the gait is normal or abnormal.

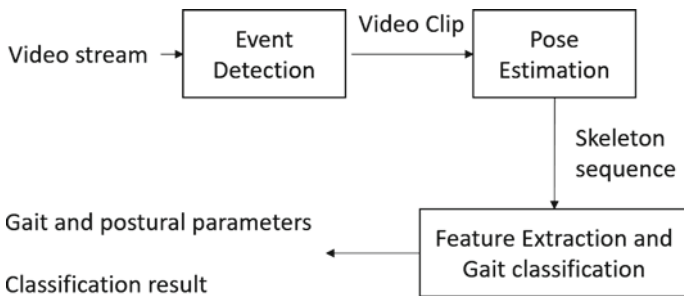


Fig. 2.1 Three processing modules of the system

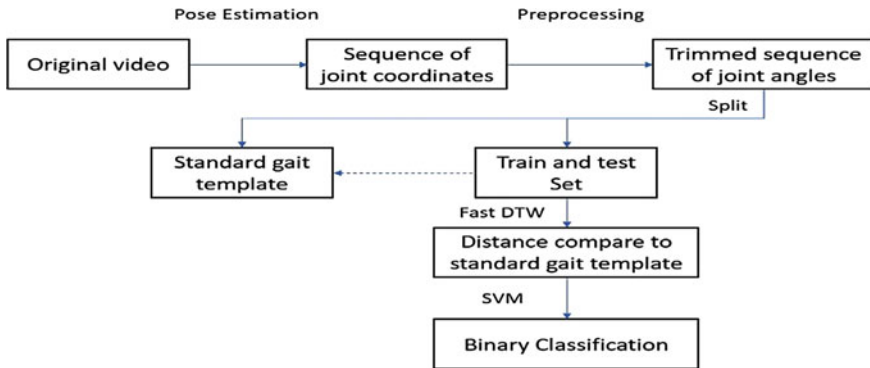


Fig. 2.2 Data processing flow

Event detection is performed in tripwire areas which located in both edges of the frame. When motion is detected in one side of the tripwire area, it starts recording an event; when motion is detected in another side, it ends the recording. We use a motion edge method (akin to optical flow) described in [17, 18].

After we captured the video clip of the subject walking perpendicular to the camera, we applied pose estimation on each frame of the video. The output of the pose estimation is a sequence of 25 anatomical joint coordinates in each frame. As shown in Fig. 2.2, in order to prepare the data for the binary classification, the following steps are performed: Trim the sequence to ensure all videos start from the same walking position. Keep a few samples of walking videos of normal gait as the standard gait template. Split the rest of the dataset into training and test set. We trained and evaluated two different algorithms of time series analysis to classify whether the gait in a video is normal or abnormal.

If the gait is classified as abnormal, it means the gait has deviated from their normal gait. To better analyze the gait, we also extract gait and posture features to help understand how the gait has changed.

2.3.2 Pose Estimation and Skeleton Extraction

Open Pose is an open-source real-time human 2D pose estimation deep learning model. It introduces a novel bottom-up approach to pose estimation using Part Affinity Fields (PAFs) to learn to associate body parts of a person in images or videos. Figure 2.3 shows Open Pose architecture (TensorFlow-based framework) [7].

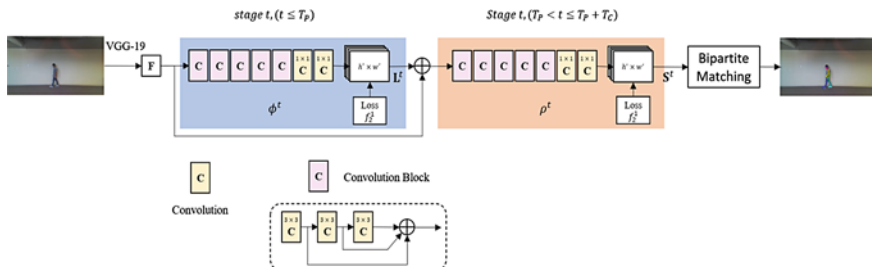


Fig. 2.3 Architecture of the multi-stage CNN

2.3.2.1 The Architecture of Open Pose

1. The system takes a color image as input, then analyzes it with a convolutional neural network (CNN) initialized and fine-tuned based on the first 10 layers of Visual Geometry Group-19 model (VGG-19).
2. A set of feature maps generated from CNN are feed into another multi-stage CNN.
3. The first set of stages predicts and refines PAFs, which is a set of 2D vector fields that encode the degree of association between body parts.
4. The last set of stages generates confidence maps of body part locations.
5. Finally, the confidence maps and the PAFs are parsed by bipartite matching to obtain 2D key points for each person in the image.

2.3.2.2 Part Affinity Fields for Part Association

PAFs contain location and orientation information across the region of support of the limb. It is a set of flow fields that encodes the unstructured pairwise relationship between body parts. Each pair of body parts have one PAF. PAFs are represented as set $\mathbf{L} = (\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_C)$, where $\mathbf{L}_c \in \mathbb{R}^{w \times h \times 2}$, $c \in \{1, \dots, C\}$. C denotes the number of pairs of body parts, $w \times h$ is the size of the input image. Each image location in \mathbf{L}_c encodes a 2D vector, if it lies on the limb c between body parts j_1 and j_2 , the value of PAF at that point is a unit vector that points from j_1 to j_2 ; otherwise, the vector is zero-valued. The ground truth PAF $\mathbf{L}_{c,k}^*$ at a point \mathbf{p} for person k as

$$\mathbf{L}_{c,k}^*(\mathbf{p}) = \begin{cases} \frac{\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}}{\|\mathbf{x}_{j_2,k} - \mathbf{x}_{j_1,k}\|_2} & \text{if } p \text{ on limb } c \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

where $\mathbf{x}_{j,k}$ is the ground truth position of the body part j of person k .

2.3.2.3 Confidence Map for Part Detection

Each confidence map is a 2D representation of the belief that a particular body part can be located in any given pixel. Each body part has one corresponding confidence map. Confidence maps are represented as set $\mathbf{S} = (\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_J)$, where $\mathbf{S}_j \in \mathbb{R}^{w \times h}$, $j \in \{1, \dots, J\}$ and J denotes the number of body parts. Individual confidence maps $\mathbf{S}_{j,k}^*$ for each person k is defined as

$$\mathbf{S}_{j,k}^*(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \mathbf{x}_{j,k}\|_2^2}{\sigma^2}\right) \quad (2.2)$$

where $\mathbf{x}_{j,k}$ is the ground truth position of the body part j of person k . The ground truth confidence map is an aggregation of individual confidence maps:

$$\mathbf{S}_j^*(\mathbf{p}) = \max_k \mathbf{S}_{j,k}^*(\mathbf{p}) \quad (2.3)$$

2.3.2.4 Multi-stage CNN

Stage $t = 1$: Given the feature maps \mathbf{F} generated from VGG-19, the network computes a set of part affinity fields, $\mathbf{L}^1 = \phi^1(\mathbf{F})$, where ϕ^1 refers to the CNN at stage 1.

Stage $2 \leq t \leq T_P$: Original feature maps \mathbf{F} and the PAF prediction from previous stage are concatenated to refine the prediction,

$$\mathbf{L}^t = \phi^t(\mathbf{F}, \mathbf{L}^{t-1}), \forall 2 \leq t \leq T_P \quad (2.4)$$

T_P refer to the number of PAF stages, ϕ^t is the CNN at stage t .

Stage $T_P < t \leq T_P + T_C$: After T_P iterations, starting from the most updated PAF prediction \mathbf{L}^{T_P} , the process is going to be repeated for T_C iterations to refine confidence map detection.

$$\mathbf{S}^{T_P} = \rho^t(\mathbf{F}, \mathbf{L}^{T_P}), \quad \forall t = T_P \quad (2.5)$$

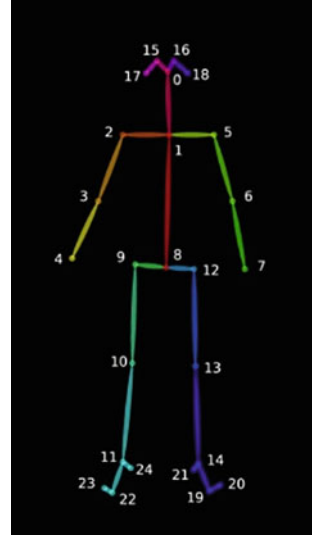
$$\mathbf{S}^t = \rho^t(\mathbf{F}, \mathbf{L}^{T_P}, \mathbf{S}^{t-1}), \quad \forall T_P < t \leq T_P + T_C \quad (2.6)$$

T_C refer to the number of confidence map stages, ρ^t is the CNN at stage t .

An L_2 loss function is applied at the end of each stage; it is specially weighted to tackle the case when people in some images are not completely labeled. Loss function at PAF stages t_i is

$$f_{\mathbf{L}}^{t_i} = \sum_{c=1}^C \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{L}_c^{t_i}(\mathbf{p}) - \mathbf{L}_c^*(\mathbf{p})\|_2^2 \quad (2.7)$$

Fig. 2.4 BODY_25 skeleton output. *Source* [19]



where \mathbf{L}_c^* is the ground truth PAF, \mathbf{W} is a binary mask. If pixel \mathbf{p} is not labeled, $\mathbf{W}(\mathbf{p}) = 0$. Loss function at confidence map stages t_k is

$$f_S^{t_k} = \sum_{j=1}^J \sum_{\mathbf{p}} \mathbf{W}(\mathbf{p}) \cdot \|\mathbf{S}_j^{t_k}(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p})\|_2^2 \quad (2.8)$$

where \mathbf{S}_j^* is the ground truth part confidence map. The overall objective is to minimize the total loss.

$$f = \sum_{t=1}^{T_p} f_L^t + \sum_{t=T_p+1}^{T_p+T_c} f_S^t \quad (2.9)$$

The output of the Open Pose is BODY-25 output format as Fig. 2.4 shows; it consists of an (x, y) coordinate pair and confidence score for each of 25 joints [19].

2.3.3 Gait and Postural Feature Extraction

After getting pose estimation of each frame in the videos, we calculated angles of back and each lower leg respect to the vertical axis in each frame. The vertical axis in the coordinate system oriented downwards. Back angle is the angle of the vector start from “Neck” (key point 1) to “Mid Hip” (key point 8) respect to vertical. The angle of lower legs uses the vector point from knee to ankle, respect to vertical. Left lower leg: “LKnee” (key point 13) and “LAnkle” (key point 14); Right lower leg: “RKnee” (keypoint 10) and “RAnkle” (key point 11) (Table 2.1).

Table 2.1 List of body parts angles

Angle name	Joints
Left lower leg angle	Left knee (key point 13), left ankle (key point 14)
Right lower leg angle	Right knee (key point 10), right ankle (key point 11)
Back angle	Neck (key point 1), mid hip (key point 8)

Each video has three corresponding time series representing the angles mentioned above. And preprocessing is performed to clean up missing data, crop, and align time series.

In Fig. 2.5, left and right lower leg (knee-ankle) angles are colored in blue and orange, respectively. The first row (a–d) is a sample of normal gait, while the second row (e–h) is for abnormal gait. First column shows the original frame, the skeleton was extracted and displayed as shown in (b) and (f). (c) And (g) shows the vector, point from knee to ankle. The lower leg angle is the angle between the vector and vertical axis. (d) and (h) are the sequence of lower leg angle in the video, the horizontal axis of the plot is the frame index, and the vertical axis is angle in degree.

From the pose estimation result of the video, we also extracted gait and postural features listed below (see Figs. 2.6 and 2.7):

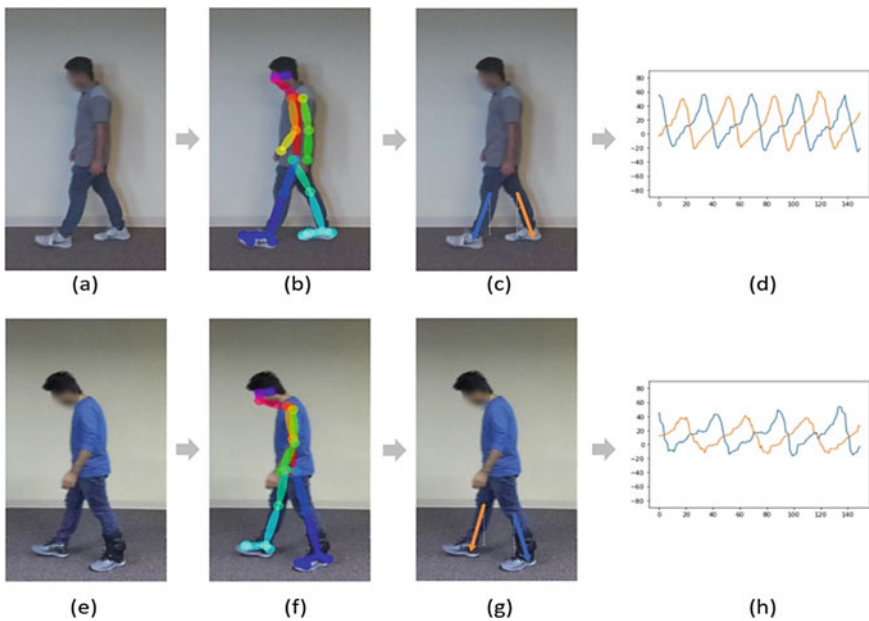


Fig. 2.5 Time series of lower leg angles

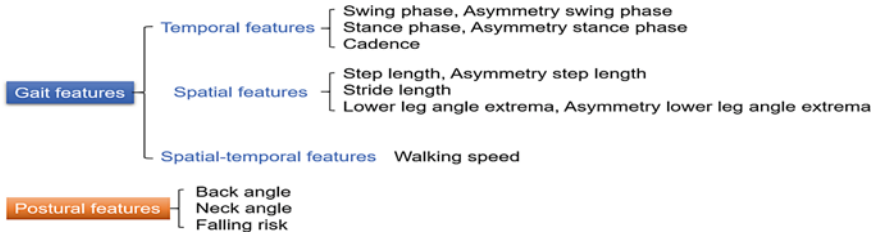


Fig. 2.6 Gait and postural feature

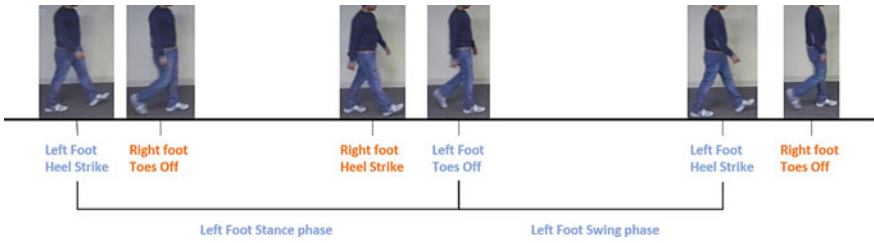


Fig. 2.7 Phases of the gait cycle

Stance Phase: The period starts from the heel strike, the heel of the same foot strike floor to toe-off, the foot is lifted from the floor. Swing Phase: The period that foot left the floor and swung forward in the air until the heel strikes the floor again [20].

Cadence: Cadence is the number of steps taken in a given period of time, expressed in steps per minute.

Step length: Distance between the contact points of two heels.

Stride length: Distance between two consecutive heel contact points of the same leg [21].

Lower leg angle extrema: The maximum and minimum of the lower leg angle time series, which reflects knee flexion while walking.

Asymmetry measure: To represent feature’s asymmetry between left and right leg, we calculate the asymmetry measure of each feature. Let f denote feature, f_L and f_R denote feature extracted from left and right leg, respectively. Asymmetry measure A_f is defined as [22]:

$$A_f = \frac{|f_L - f_R|}{\max(f_L, f_R)} \tag{2.10}$$

Walking speed: Walking speed is calculated by dividing walking distance with the time period taken.

Back Angle: Back angle is between the mid-hip (key point 8) and bottom of the neck (key point 1)

Neck Angle: Neck angle is between the bottom of the neck and the nose (key point 0).

Falling risk: Body posture at left heel strike phase is used to calculate falling risk. Similar to [22], falling risk Fr is defined as:

$$Fr = \frac{\left| Nose_x - \frac{(LHeel_x + RHeel_x)}{2} \right|}{\frac{|LHeel_x - RToe_x|}{2}} \tag{2.11}$$

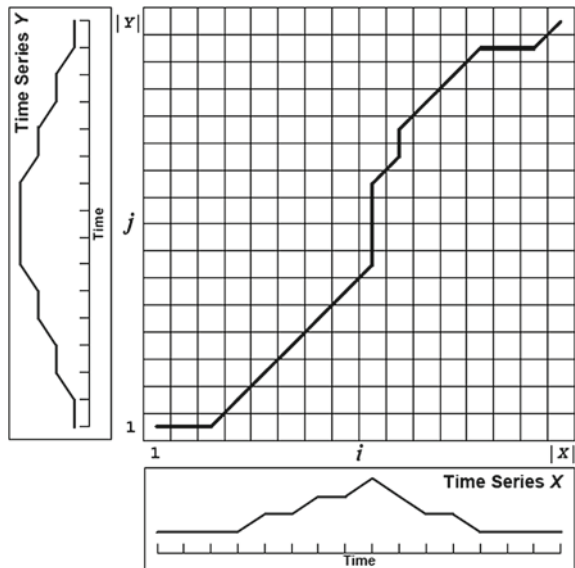
where $Nose_x$ is the x -coordinate of Nose (key point 0), $LHeel_x$, $RHeel_x$ and $RToe_x$ are x -coordinate of Left heel (keypoint 21), Right heel (keypoint 24), and Right big toe (keypoint 22).

2.3.4 Time Series Analysis

In order to measure the similarity between time series derived from pose estimation, we use fast dynamic time warping (DTW) [23] method to help calculate Euclidean distance with optimal alignment (see Fig. 2.8). And, support vector machine (SVM) is used for binary classification of normal and abnormal gait.

Due to the nature of the human walk, there might have shifts and distortions in gait data between each walk in the time axis, caused by subtle difference in walking speed or cadence. It is hard to have sequences aligned perfectly. Even between multiple samples of the normal walk, the slight time shift causes the distance to be considerably

Fig. 2.8 Cost matrix with the minimum-distance wrap path traced through it. Source [23]



large, making it unable to differentiate with abnormal gait, as their distances are both large. So directly calculating Euclidean distance is going to give poor results.

Dynamic time warping (DTW) [24] algorithms are commonly used to overcome shifts in the time dimension. Assume we have two time series X and Y . The value of a cell $D(i, j)$ in cost matrix D is the minimum-distance warp path of sequence $X' = x_1, \dots, x_i$ and $Y' = y_1, \dots, y_j$, $\text{Dist}(i, j)$ denote distance of two data points x_i and y_j .

$$D(i, j) = \text{Dist}(i, j) + \min[D(i-1, j), D(i, j-1), D(i-1, j-1)] \quad (2.12)$$

Fast DTW approximates DTW, using a multi-level approach to achieve linear time and space complexity, in contrast to quadratic time and space complexity in standard DTW algorithm. It first produces different lower resolutions of the time series, by taking an average of adjacent pairs of points. Then, project the minimum distance calculated from lower resolution to higher resolution as an initial guess. Finally, refine the wrap path by local adjustments.

After calculation of DTW distance for each time series, we used support vector machine (SVM) on a multidimensional DTW distance vector for classification.

2.4 Experiments and Results

2.4.1 Data Collection

In our experiment, we aim to simulate gait in different health states within the laboratory. Because it is difficult to change health status of our volunteers, we propose to use different levels of physical ankle weights in the experiment to help them demonstrate abnormal gait. Intuitively, when additional weights added on human body, gait will change accordingly as mobility and stability of walking are affected.

The adjustable ankle weights strap with removable sand packets were used because it can easily adjust the weight by adding or removing sand packets on the strap. Each strap has five slots to hold sand packets, and each packet weighs around 0.6 lbs.

We designed the experiment capturing the same person walking across the camera normally for 10 times, walking with three different levels of weights 10 times, respectively, and walking with the 3rd level of weights plus carrying a heavy box for 10 times (see Fig. 2.9).

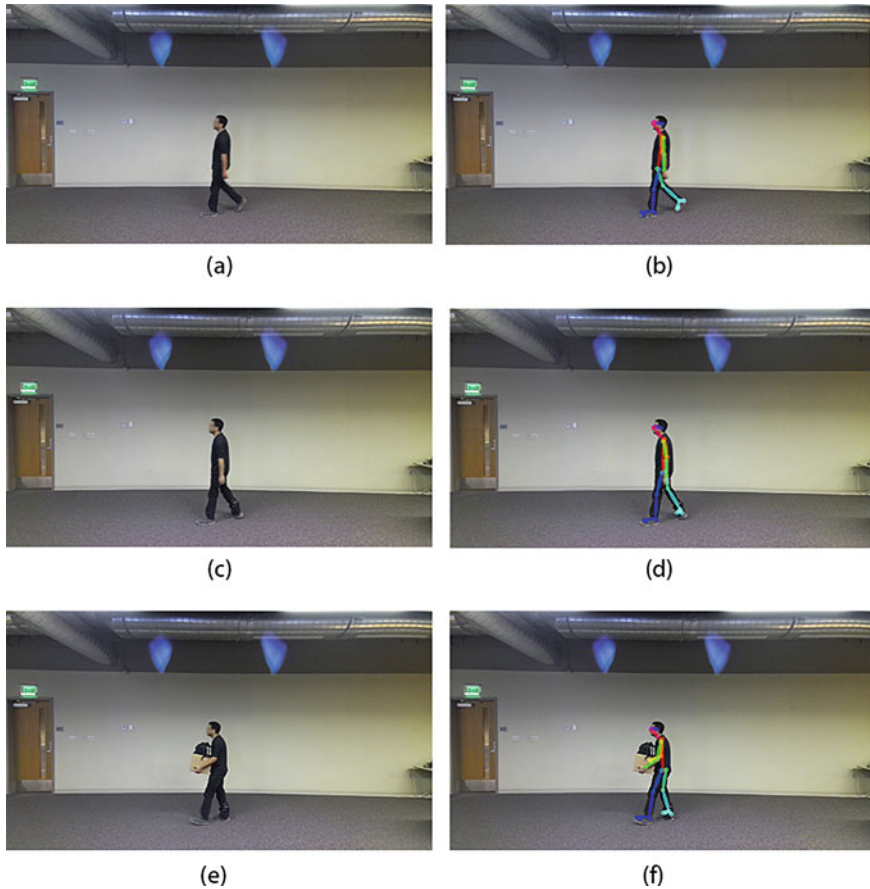


Fig. 2.9 Sample frames in the dataset: **a** original frame in video of normal walk; **b** normal walk frame with skeleton overlay; **c** original frame in video of walk with 3rd level of weights; **d** frame of 3rd level of weights experiment with skeleton overlay; **e** original frame in video of walk with 3rd level of weights and carrying a heavy box; **f** frame of 3rd level of weights and a heavy box with skeleton overlay

2.4.2 Experiment Setting

Videos we captured are with resolution of 1920×1080 and frame rate of 30 frames per second. In pose estimation stage, Open Pose was also processed at 30 FPS. We dropped frames with incomplete or low confidence joints to filter out the low-quality frames. Failure of pose estimation in these frames usually caused by motion blur in the frame, especially at lower leg and foot area where the amplitude of motion is largest.

In the preprocessing stage, we first performed imputation using interpolation to fill the missing data. Then, we cropped all the sequences to start from the first peak

(maximum) point of the right lower leg angle’s time series, so that all of the sequences should start from nearly the same gait phase which minimized the noise from data misalignment.

In our experiment, for each person, we set aside three videos of normal gait as standard template gait data to be compared with. The rest of the videos were split into training and test set, and for each one of them, we calculate the distance with three standard template gait data using fast DTW then take an average to measure how close it is comparing to normal gait. Linear kernel is used in SVM for classification.

2.4.3 Results for Gait Classification

2.4.3.1 Evaluation Metrics

Gait classification is a binary classification task to predict if an unlabeled video shows normal or abnormal gait. Detection of abnormal gait is defined as “positive.” If the ground truth label of the data matches detection result, it is defined as “true,” otherwise it is “false.”

We employed accuracy rate, precision rate, recall rate and F_1 score in our experiment.

2.4.3.2 Intra-subject Cross Validation

To evaluate the performance of our proposed methods, leave-one-out cross validation is performed within each subject. As Fig. 2.10 shows, for each subject’s data, we reserve one as test set and use the rest of this subject’s data to train the algorithms. The same process is performed on each of four subjects.

Results of leave-one-out cross validation are listed as below (Table 2.2).

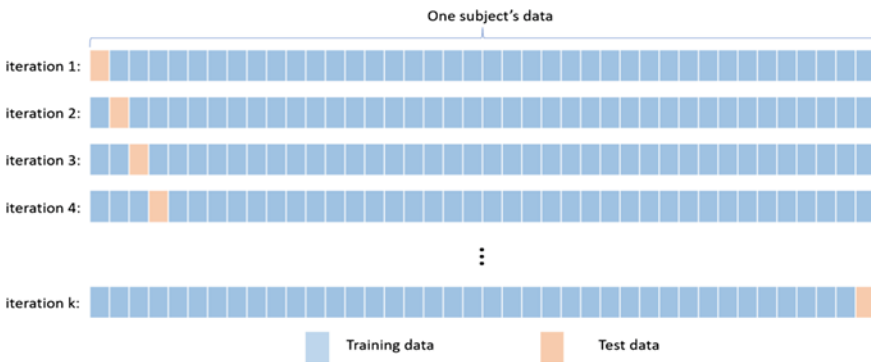


Fig. 2.10 Leave-one-out cross validation

Table 2.2 Results of leave-one-out cross validation

Subject	Precision	DTW-SVM Recall	Accuracy	F_1 score
1	1.000	1.000	1.000	1.000
2	0.974	0.925	0.915	0.949
3	1.000	0.975	0.979	0.987
4	1.000	1.000	1.000	1.000
Average	0.993	0.975	0.973	0.984

2.4.3.3 Inter-subject Cross Validation

We also performed leave-one subject-out cross validation to validate the performance when applied on a different subject that is not included in training set. As Fig. 2.11 shows, in each iteration, one subject's data is used as test set. All other subjects' data is used for training.

From the result of leave-one-subject-out cross validation, we can see that both methods successfully classified normal and abnormal gaits. The DTW-SVM-based method achieved 0.982 in F_1 score (Table 2.3).

**Fig. 2.11** Leave-one-subject-out cross validation**Table 2.3** Results of leave-one subject-out cross validation

Subject	Precision	DTW-SVM Recall	Accuracy	F_1 score
1	1.000	1.000	1.000	1.000
2	0.907	0.975	0.894	0.940
3	1.000	0.975	0.979	0.987
4	1.000	1.000	1.000	1.000
Average	0.977	0.988	0.968	0.982

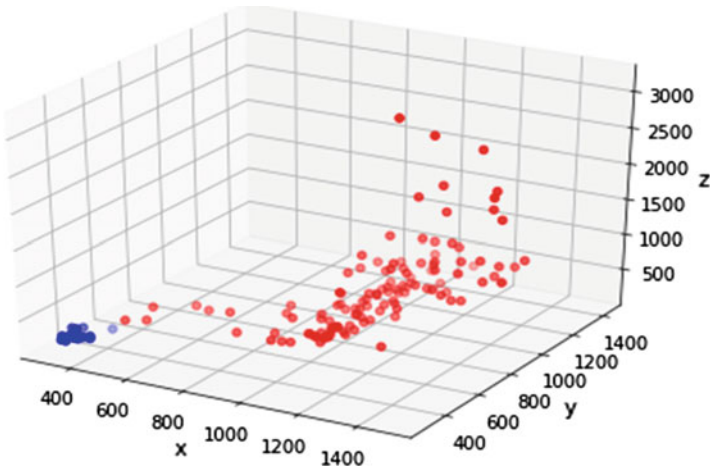


Fig. 2.12 DTW distance to standard gait template; data points of normal gait were marked with blue dots; data points of abnormal gait were marked with red dots

2.4.4 Discussion

The proposed DTW-SVM-based method performed classification by measuring the deviation from the standard gait template. As Fig. 2.12 shows, the x -, y -, z -axis denotes the DTW distance of left and right lower leg angle, back angle compares to standard gait template, respectively. Normal gait's data is marked with blue while abnormal gait's data is marked with red. Intuitively, normal gait is closer to the standard gait template, yet the distance should be small, so the data points are clustered near origin of the coordinate system.

As DTW-SVM-based method measures deviation, so it has better performance while applied on inter-subject prediction. It only requires a few samples as standard gait template to be compared with.

2.4.5 Graphical User Interface and Companion Robot

The gait monitoring system based on body posture and walking speed is implemented as a module in companion robot. On a robot's monitor directly in line with the walking trajectory (see Fig. 2.7), we displayed a graphical "reward" of the person's gait with respect to average values among the population (see Fig. 2.13). So, people can see the results of the evaluation after they walked through the room. We have heard from users that this system motivates them to improve body posture.

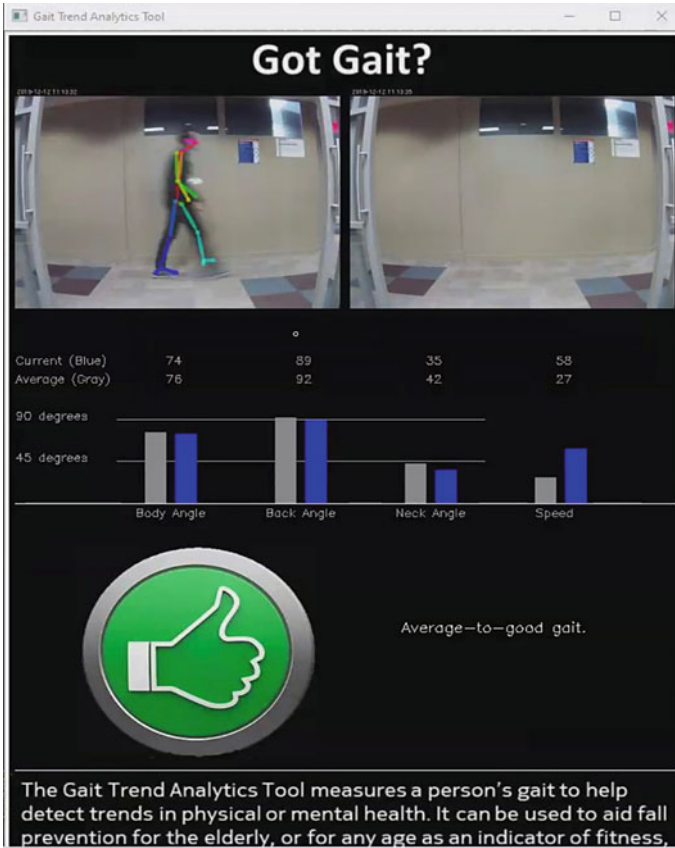


Fig. 2.13 Graphical user interface (GUI) of the gait analysis system

2.5 Conclusions and Future Work

This paper proposed a video-based gait analysis system to detect abnormal gait and capture gait features non-invasively over time. Gait analysis and gait abnormality detection allows early intervention and treatment to prevent underlying conditions develops and cause a fall. It can also evaluate the recovery progress of the physical therapy.

The system consists of three stages; first, the event detection module records the gait event video clip while a person walks through the camera. Second, pose estimation is applied to extract sequence of skeleton and joints in the video. Then, feature extraction and gait classification is performed to calculate gait and postural parameters and to classify if the gait is normal or abnormal.

In gait classification task, we proposed DTW-SVM-based method, using time series of lower leg angle and back angle. Experiment results shows that DTW-SVM-based method achieved higher accuracy in inter-subject classification. Gait and postural parameters extracted shows distinction between normal and abnormal gait. With tracking the parameters on a day-to-day basis, we can quantitatively monitor the gait changes in long term.

In the future, this work can be extended by adding another camera, located in front view, to analyze the motion and gait asymmetry in the coronal plane or frontal plane of the human body. The system can be integrated with a face recognition module to fit the need of a multi-person household or facility so that it can automatically recognize the identity of the person in the video and then add the extracted gait features to their record. When the system detected gait abnormality of a person, an alert will be generated for medical attention.

Acknowledgements This work was supported by the National Science Fund of Bulgaria: KP-06-H27/16 *Development of efficient methods and algorithms for tensor-based processing and analysis of multidimensional images with application in interdisciplinary areas* and NOKIA Corporate University Donation number NSN FI (85) 1198342 MCA.

References

1. NCOA: Falls Prevention Facts. Retrieved from: <https://www.ncoa.org/news/resources-for-reporters/get-the-facts/falls-prevention-facts/>. 12 May 2020
2. MedlinePlus: Walking Problems. <https://medlineplus.gov/walkingproblems.html>. Accessed 12 May 2020
3. Selke, Stefan: Book Lifelogging. Springer, Heidelberg (2016)
4. Jalal, A., Kamal, S., Kim, D.: Depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* **14**(7), 11735–11759 (2014)
5. Climent-Pérez, P., Spinsante, S., Mihailidis, A., Florez-Revuelta, S.: A review on video-based active and assisted living technologies for automated lifelogging. *Expert Syst. Appl. J.* **139** (2020)
6. Liu, X., Milanova, M.: An image captioning method for infant sleeping environment diagnosis. In: IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction, pp. 18–26. Springer, Cham (2018)
7. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y.: Open Pose: real-time multi-person 2D pose estimation using Part Affinity Fields. arXiv preprint [arXiv:1812.08008](https://arxiv.org/abs/1812.08008) (2018)
8. Muro-De-La-Herran, A., Garcia-Zapirain, B., Mendez-Zorrilla, A.: Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors* **14**(2), 3362–3394 (2014)
9. Meng, M., Drira, H., Daoudi, M., Boonaert, J.: Detection of abnormal gait from skeleton data. In: 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016), vol. 3: VISAPP, pp. 133–139 (2016)
10. Khokhlova, M., Migniot, C., Morozov, A., Sushkova, O., Dipanda, A.: Normal and pathological gait classification LSTM model. *Artif. Intell. Med.* **94**, 54–66 (2019)
11. Nguyen, T.N., Huynh, H.H., Meunier, J.: Estimating skeleton-based gait abnormality index by sparse deep auto-encoder. In: 2018 IEEE Seventh International Conference on Communications and Electronics (ICCE), pp. 311–315 (2018)

12. Jun, K., Lee, D.W., Lee, K., Lee, S., Kim, M.S.: Feature extraction using an RNN autoencoder for skeleton-based abnormal gait recognition. *IEEE Access* **8**, 19196–19207 (2020)
13. Dolatabadi, E., Zhi, Y.X., Flint, A.J., Mansfield, A., Iaboni, A., Taati, B.: The feasibility of a vision-based sensor for longitudinal monitoring of mobility in older adults with dementia. *Arch. Gerontol. Geriatr.* **82**, 200–206 (2019)
14. Xue, D., Sayana, A., Darke, E., Shen, K., Hsieh, J.T., Luo, Z., Li, L.-J., Lance Downing, N., Milstein, A., Fei-Fei, L.: Vision-based gait analysis for senior care. arXiv preprint [arXiv:1812.00169](https://arxiv.org/abs/1812.00169) (2018)
15. CASIA Gait Database. <https://medlineplus.gov/walkingproblems.html>. Accessed 12 May 2020
16. INIT Gait Database homepage. <https://www.vision.uji.es/gaitDB/>. Accessed 12 May 2020
17. O’Gorman, L., Yin, Y., Ho, T.K.: Motion feature filtering for event detection in crowded scenes. *Pattern Recogn. Lett.* **44**, 80–87 (2014)
18. O’Gorman, L., Yang, G.: Orthographic perspective mappings for consistent wide-area motion feature maps from multiple cameras. *IEEE Trans. Image Process.* **25**(6), 2817–2832 (2016)
19. OpenPose Output Format. Retrieved from: <https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md>. 12 May 2020
20. Dale, R.B.: Clinical gait assessment. *Physical Rehabilitation of the Injured Athlete*, pp. 464–479 (2012). <https://doi.org/10.1016/b978-1-4377-2411-0.00021-6>
21. Richards, J., Chohan, A., Erande, R.: Biomechanics. *Tidy’s Physiotherapy*, pp. 331–368 (2013). <https://doi.org/10.1016/b978-0-7020-4344-4.00015-8>
22. Ortells, J., Herrero-Ezquerro, M.T., Mollineda, R.A.: Vision-based gait impairment analysis for aided diagnosis. *Med. Biol. Eng. Comput.* **56**(9), 1553–1564 (2018)
23. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.* **11**(5), 561–580 (2007)
24. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: *KDD Workshop*, vol. 10, No. 16, pp. 359–370 (1994)