

# Gait Estimation and Analysis from Noisy Observations

Hafsa Ismail<sup>1,2</sup>, Ibrahim Radwan<sup>3</sup>, Hanna Suominen<sup>1-4</sup> and Roland Goecke<sup>1</sup>

**Abstract**—People’s walking style – their gait – can be an indicator of their health as it is affected by pain, illness, weakness, and aging. Gait analysis aims to detect gait variations. It is usually performed by an experienced observer with the help of different devices, such as cameras, sensors, and/or force plates. Frequent gait analysis, to observe changes over time, is costly and impractical. This paper initiates an inexpensive gait analysis based on recorded video. Our methodology first discusses estimating gait movements from predicted 2D joint locations that represent selected body parts from videos. Then, using a long-short-term memory (LSTM) regression model to predict 3D (Vicon) data, which was recorded simultaneously with the videos as ground truth. Feet movements estimated from video are highly correlated with the Vicon data, enabling gait analysis by measuring selected spatial gait parameters (step and cadence length, and walk base) from estimated movements. Using inexpensive and reliable cameras to record, estimate and analyse a person’s gait can be helpful; early detection of its changes facilitates early intervention.

## I. INTRODUCTION

Gait is the manner or style of walk. Gait analysis evaluates this style of locomotion [1]. While walking, one limb provides body support (stance phase for that leg), while the other leg swings forward (swing phase for that leg) in preparation for stance phase. The combination of stance and swing phases forms the gait cycle as shown in Fig. 1. In normal Walk, a symmetry occurred in the movements of both feet.

A person’s walk can be affected by pain, weakness, aging or injuries. Thus, gait abnormalities can be an indicator of a person’s health. Gait abnormalities can be detected by an experienced physiotherapist who observes the walking style and measures parameters to analyse the gait with the help of different devices. These gait parameters are: 1) Temporal and spatial, such as walking speed and lengths, and can be measured by a stop watch and a measuring tape. Floor sensors, accelerometers and/or wearable sensors can be added for more accurate measurements. 2) Kinematics and kinetics, such as joint movements and angles, and the forces involved to produce these movements. Specific devices are used to measure such parameters, *e.g.*, motion capture cameras and force plates. 3) Electromyography, which measures muscle activities during walk. Video recording is used widely in gait analysis to slow the motion of the walk action and the movement of the legs and feet for a detailed assessment and analysis.

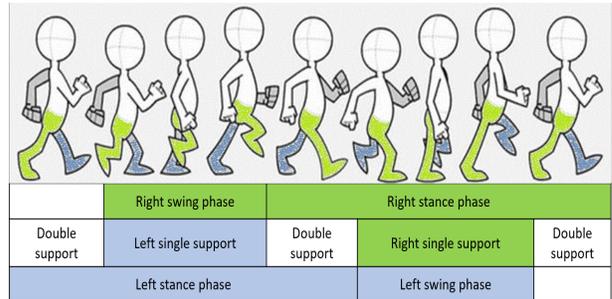


Fig. 1: Gait cycle consisting of stride and swing phases for each leg.

Human gait and sway has received much attention in computer vision in recent years for human recognition and identification (*e.g.*, [2]), for detecting falls (*e.g.*, [3]) using the general shape of the body and/or the movement patterns, or even for assessing the person’s balance with a view to reduce the reliance on expensive and/or impractical devices [13]

For the goal of using video cameras only to estimate and analyse human gait, spatial gait parameters that can be captured and measured from frontal and side views has been selected. In this paper, we estimate movements of selected body parts (feet, torso) from video using predicted 2D joint locations. The estimated movements are used to derive spatial gait parameters (step and cadence length, walk base). Estimated movements and gait parameters are compared with the ground truth Vicon data and parameters derived from it, respectively. Our long-term goal is to investigate the use of simpler and cheaper video technology as an alternative to costly specialist devices.

In the next section related work is presented followed by information about the dataset, the proposed method, and the experiments and results.

## II. RELATED WORK

Early research on measuring and analysing human gait was entirely for medical purposes to distinguish normal gait pattern from a pathological one. Gait analysis was usually performed by an experienced physiotherapist who observed the walking style.

For accurately measuring and characterising human gait, different devices are used to measure body movements, mechanics, and muscle activity. These devices are either *wearable* sensors, such as accelerometers, gyroscopes, [9], [10], or *non-wearable* devices, such as cameras, force plates. Most of these are expensive and need to be installed and used by a specialist in a laboratory environment.

<sup>1</sup> University of Canberra, Bruce, Australia

<sup>2</sup> DATA61/CSIRO, Acton, Australia

<sup>3</sup> Australian National University, Acton, Australia

<sup>4</sup> University of Turku, Turku, Finland

With the advances in image processing algorithms, analysing gait became an interest for computer vision researchers, *e.g.*, to recognise and identify humans from their body movements. Gait is considered to be a biometric characteristic of an individual. Most computer vision studies on gait use the general shape of the body and/or the moving pattern for abnormal gait detection [5], [6], [7], human recognition [8], fall detection [3], and for safety and surveillance purposes.

Human recognition systems using gait have focused on image representation, feature dimensionality, and gait classification [8] to improve human recognition. On the other hand, for abnormal gait detection, body silhouettes have been used in [5] to detect and classify the observed walking pattern into normal or abnormal gait based on analysing recorded videos of seven subjects walking in normal and abnormal ways. In [6], the extracted silhouette, with frame-to-frame optical flow and motion metrics based on histogram representations of silhouette-masked flows, was used to determine different styles of walking and detect deviation from usual walking patterns using two separately recorded datasets, one for subjects simulated abnormal gait and the second dataset for professional actors performing pathological gaits.

Gait analysis, especially spatial-temporal parameters that can be noticed in videos, is receiving increasing attention in computer vision to provide an inexpensive tool to identify and detect gait abnormalities. Human gait has been classified into normal or abnormal in [7] by calculating some gait spatial-temporal parameters using vision data. The dataset consisted of 30 subjects (15 normally walking, 15 abnormally walking).

For the improvement of all aspects of life, new technologies and approaches are proposed to make the measurement and assessment of health conditions easier, faster, and more reliable. Although some progress has been made in using computer vision techniques for human gait analysis, a need for a technique that measures and analyses human gait in an inexpensive and reliable manner still exists. In this paper, we estimate the human gait from video and explore the utility of automated gait analysis by comparing the results with 3D Vicon data in a newly collected dataset.

### III. DATASET

After furnishing the proper ethics approvals and research permissions, a dataset that contains walk and balance actions was collected and used for this study. Existing datasets that include gait and balance activities have usually been clinical datasets, collected in health care settings and, hence, were confidential and unavailable to us. Moreover, video cameras or a motion capture system that we need for assessing the accuracy of the modeling were typically not included. This dataset is considered as a ground truth dataset to verify the proposed methodology in calculating the selected gait parameters. While this dataset contains both walking and balance activities, this paper investigates the walking (gait) data only.

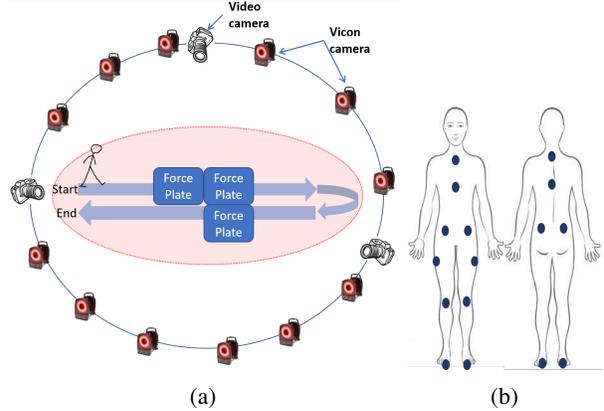


Fig. 2: Dataset recording (a) Lab installation and (b) Vicon marker locations on subjects' body

Eighteen subjects, twelve males and six females, performed normal walking for about 10m, including making a U-turn after 5m, as well as a balance test on an AMTI force plate. Each subject performed this sequence of activities three times in a row, with a minimal break in between. The activities were recorded by three video cameras: a front view, a side view, and a corner back view. Twelve T-series Vicon cameras were used to capture 3D motion from sixteen markers that were placed on the subject's body (Fig. 2). As the feet, legs, and trunk are the major motor reactors in walking, Vicon markers were placed on different joints related to these body parts to capture the body movements during the gait and balance test activities. Captured data from the Vicon cameras and the force plate were systemically synchronised, where the videos were manually aligned with the heel strike.

### IV. PROPOSED METHOD

Our goal is to provide a gait analysis method that is accurate enough to provide an alternative to highly accurate but expensive devices, such as force plate and motion capture systems. These devices also need a special setup or need to be worn most of the time. To achieve this goal, both frontal and side views in RGB video cameras are considered. Some parameters, such as walking base and sway, are more obvious from the frontal view while others, such as step and cadence length, are easier to measure from the side view.

To determine the accuracy of the calculated gait parameters from videos, Vicon data are used as an accurately captured ground truth for body movements. First, the body joint locations are estimated from the video frames. Second, joints of the limb related to the walking action are selected from both video and Vicon. Feet and legs represent the walking activity, whilst upper body movements reflect the body sway when walking or standing. Then, a regression method is used to predict the real measurements for the selected spatial gait parameters from the estimated joints from the video (Fig. 3). Next, we describe the methods for pre-processing, gait estimation and gait analysis.

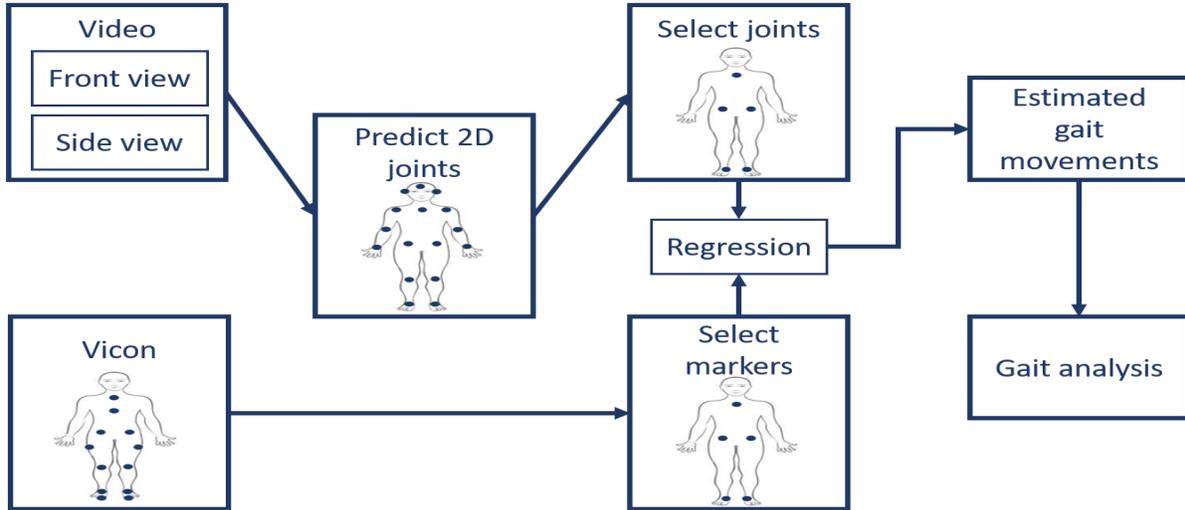


Fig. 3: Proposed method: Movements of selected body joint locations from video are regressed with movements from corresponding Vicon markers in order to estimate body part movements and use them in gait analysis. Note that while the Vicon and real life distances are in millimetres, the respective image metric uses pixels. The LSTM regression method is used to estimate joint movement measurements in millimetres for the gait parameter calculations.

#### A. Pre-processing

The following steps are taken to prepare the variety of data, produced by the Vicon system and video cameras, for gait estimation and analysis.

*Joint Extraction:* The DeepPose method [4] that estimates human pose is applied to extract 2D joint locations of 18 different body parts. 2D joints related to feet, hips, and upper body (torso) are selected for subsequent use in gait estimation and analysis. In the DeepPose method, a generic convolutional deep neural network (DNN) is learned. Using a cascade of DNN-based pose predictors allows for increased precision of joint localisation. Starting with an initial pose estimation, based on the full image, DNN-based regressors are learned to refine the joint predictions by using higher resolution sub-images. The method is used on both side and frontal views.

*Joint Normalisation – Frontal View:* Although the extracted joints from the frontal view are clear and represent the body movements over the frames, the apparent movements decay when being too far from the camera. To avoid this distortion, predicted joint locations are normalised to the biggest bounding box that surrounds the body over all video frames as shown in Fig. 4.

*Data Re-sampling:* The video and Vicon data are recorded at different frame rates. The video frame rate is 25Hz, while the Vicon system captures the markers’ movements at 100Hz. The higher Vicon sampling rate results in a noisy signal. Since the signal of interest (markers’ movements) is characterised by low frequencies, the well-known Butterworth filter [11] is applied as a low-pass filter on the Vicon data. It cuts off the high frequencies in order to reduce the noise and make the data better suited for gait analysis through regression with an aim to learn the pattern of movements.

The output of the pre-processing steps are signals that

represent the selected joints movements over the video frames and the movements from the corresponding Vicon markers over time. In this paper, we work on the feet markers to calculate the gait parameters, namely step and stride length and walk base.

#### B. Gait Estimation and Analysis

The proposed method for gait estimation is based on regression. To predict the gait signal in the scale of the Vicon data space, two bi-directional long-short term memory (LSTM) units are used. The output signal at each time  $t$  is estimated by

$$y(t) = \sigma(\vec{h}_t, \overleftarrow{h}_t) \quad (1)$$

where  $\sigma$  is the non-linear activation function of the concatenation or averaging of the forward passing,  $\vec{h}_t$ , from  $t - n$  to  $t - 1$  and the backward passing,  $\overleftarrow{h}_t$ , from  $t - 1$  to  $t - n$ , where  $n$  is the history size of LSTM unit [12].

More precisely, the prediction task considers the relationships between the Vicon data and the video signals. The

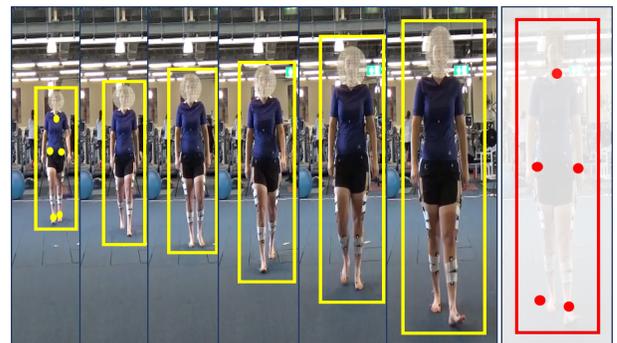


Fig. 4: Predicted 2D joints are normalised to the biggest bounding box surrounding the body over the video frames

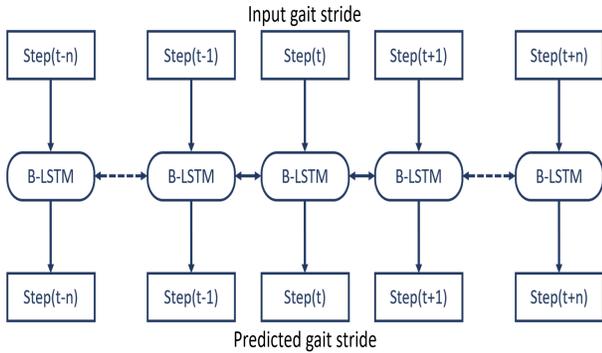


Fig. 5: The proposed LSTM model

LSTM model is designed based on a many-to-many structure, as illustrated in Fig. 5 where the input and the output are of the same size. The sequential connection in its middle layer shows the process of encoding the temporal relationships between the subsequent frames. The input of the model is the signal of one view for a typical joint of the detected pose, which is passed to the two bi-directional LSTM units as an input. Then, the output is passed to a fully connected layer to predict the final output signal.

To analyse gait, we consider the following three spatial parameters as described in Fig. 6: *stride length*, *step length*, and *walk base*. The stride length is the distance between successive points of heel strike of the same foot. In normal gait, the stride length is twice the step length, which is the distance between corresponding successive points of heel strike of one foot and the other foot. The walk base is the perpendicular distance between the two lines that pass through the heel strike points, respectively, for the two feet.

## V. EXPERIMENTS AND RESULTS

In this section, we present our experiment for applying the proposed method on the walking part of the dataset. First, we estimate feet movements from video and measure the selected spatial gait parameters from these movements. Second, we compare both estimated movements and measured parameters with data from the Vicon markers and the gait parameters derived from them. Vicon data is considered as an accurate reference for the collected data and hence, used as the ground truth or gold standard in this study.

Fig. 7 illustrates the Vicon space directions —  $X$ ,  $Y$ ,

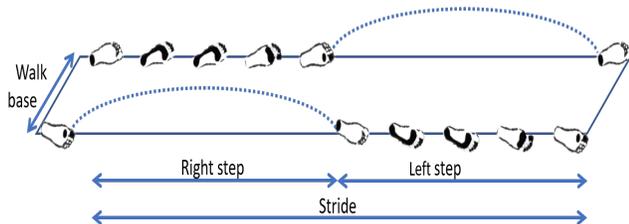


Fig. 6: Gait parameters considered in this study: stride and step length and walk base

and  $Z$  — and shows the corresponding directions from both video views, frontal and side.

### A. Gait Estimation from Video

Synchronised video and Vicon data are cut into single strides (from heel strike to next heel strike for a single foot) as a learning unit to be fed into the proposed LSTM network. An example of predicted gait movements for one stride using the proposed regression method and the corresponding ground truth is shown in Fig. 8.

The predicted feet movements in the  $X$ -direction from side view and  $Z$ -direction from frontal view (heel contact and off) are correlated with the corresponding Vicon movements with an average of 90% and 95%, respectively, over all subjects as shown in Fig. 11. Results for a sequence of strides for one foot are shown in Fig. 9 for the  $X$ -direction and the  $Z$ -direction. Movements in the  $X$ -direction are estimated from the side view where the subject crosses the scene, which contains some background objects (*e.g.* a camera tripod) that can lead to the 2D prediction drifting to them, which produces outlier movements in the signal, thus affecting the predicted gait movements. The smaller correlation in the  $X$ -direction is explained by errors multiplying from the movement extraction phase. Relying on the high correlation of estimated movements with the Vicon data, vision-based data can be used to measure and analyse human gait.

### B. Spatial Parameter Measurements

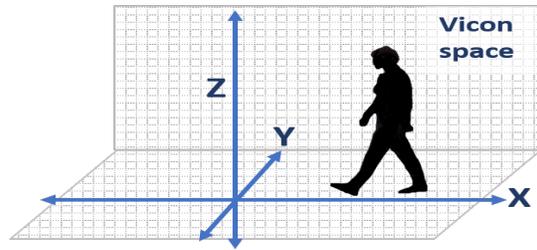
Since working with 2D joint representations, stride and step lengths are measured from predicted feet movements in the side view, while step base is measured from predicted feet movements in the frontal view.

The stride length calculated from estimated feet movements shows an average of 1.17m for the right foot with a mean square error (MSE) of 0.036m compared to the Vicon measurements (1.30m). An average of 1.12m for the stride length for the left foot with an MSE of 0.044m to the Vicon measurements (1.30m) (Table I). The step length derived from estimated feet movements shows an average of 0.585m and 0.563m for the right and left foot, respectively, while the average step length for both right and left foot from the Vicon data is 0.65m.

Although the gait parameters measured from predicted feet movements from video and the Vicon markers are close, the slight difference can be traced back to how accurate the extracted 2D feet joints represent the actual feet movements. In the Vicon data, the heel strike is well defined because

TABLE I: Performance evaluation (Mean and MSE measures) for the stride and step predictions for the right and left foot on the test and validation data

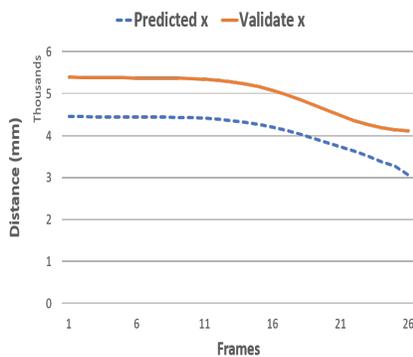
		Stride (metre)		Step (metre)	
		Mean	MSE	Mean	MSE
Right	Test	1.17	0.036	0.585	0.040
	Validation	1.30		0.650	
Left	Test	1.12	0.044	0.563	0.047
	Validation	1.30		0.650	



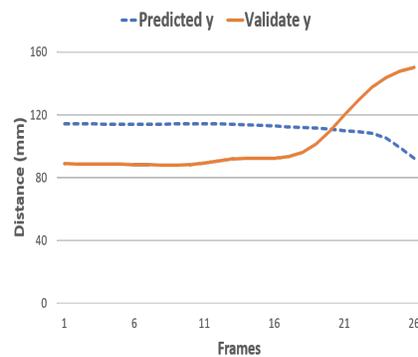
	Front view	Side view
X-direction	Y-Vicon	X-Vicon
Y-direction	Z-Vicon	Z-Vicon



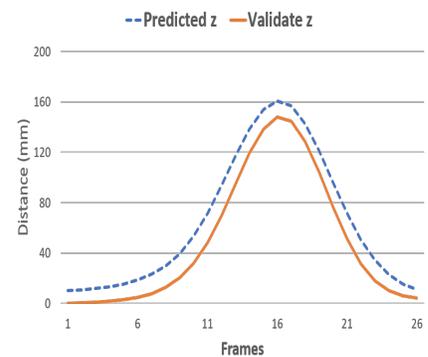
Fig. 7: Vicon and video coordinate systems and their correspondence



(a)

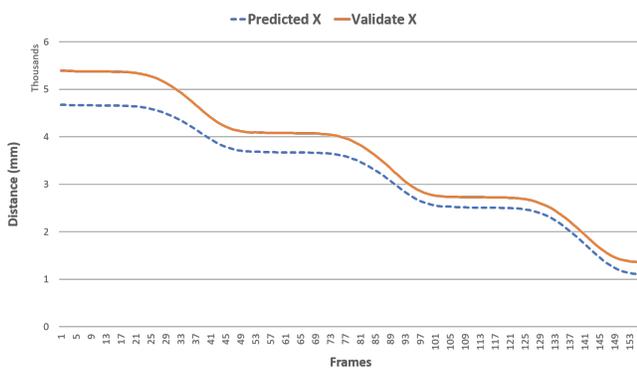


(b)

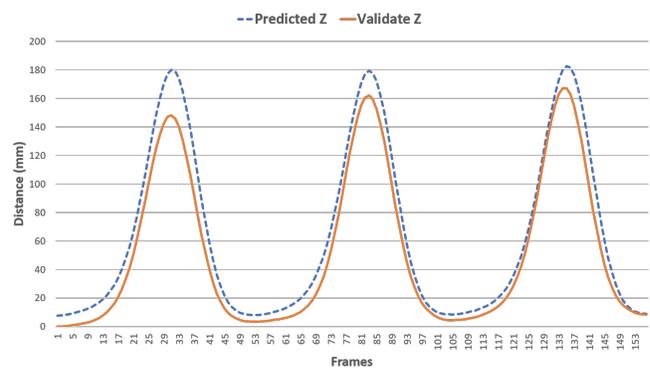


(c)

Fig. 8: Example of predicted stride compared to ground truth. (a) X-direction from side view with Vicon X-direction, (b) X-direction from frontal view with Vicon Y-direction, and (c) Y-direction from frontal view with Vicon Z-direction.



(a)



(b)

Fig. 9: Example of predicted and ground truth movements in sequence of strides in (a) X-direction and (b) Z-direction

of the marker that is placed on the heel, while the video extracted 2D joint does not represent the heel accurately as shown in Fig. 10 (a).

The walk base measurements show an average of 4.9cm

for the predicted gait movements from video compared to an average of 6cm for the Vicon data. Walk base is relatively small. The slight differences between estimated joints from video and Vicon markers as shown in Fig. 10 (b) and (c)

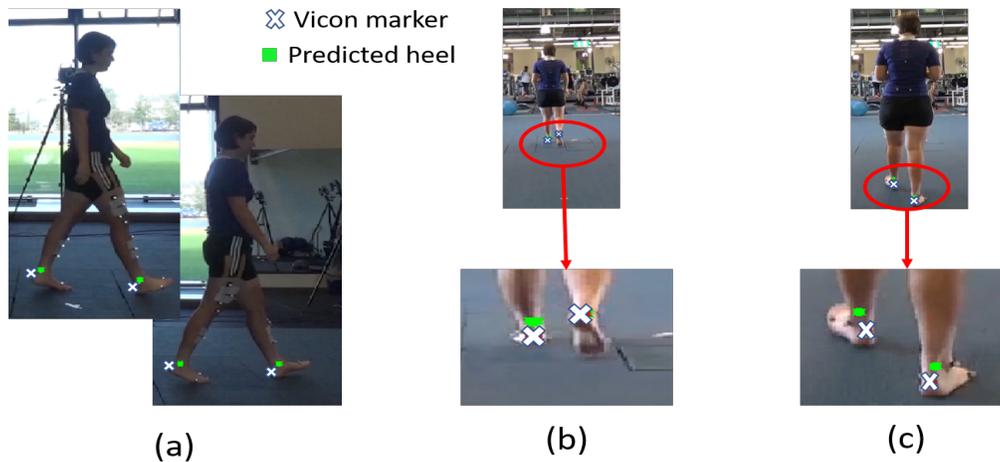


Fig. 10: Example of a foot’s predicted joints (square) and Vicon markers (x). (a) side view, (b) accurately estimated joints, and (c) inaccurately estimated joints for heels.

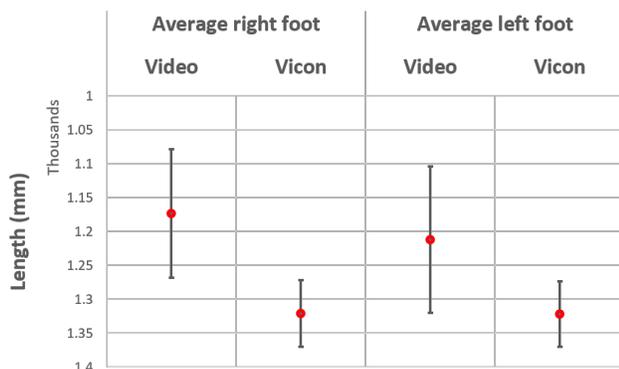


Fig. 11: The average strides from the estimated joints movement from video and Vicon for both feet with 95% confidence interval

explain the difference in the walk base measurements.

## VI. CONCLUSIONS AND FUTURE WORK

Experiments in this study demonstrate the possibility to measure and analyse gait from estimated gait movements from video – the approach is 90% correlated to the ground truth Vicon data. However, the frame-to-frame human pose estimation generates some errors in the predicted joint locations, which affects the output signals that represent gait movements, and thus also the derived parameters. Our work could be extended by utilising a 3D construction method to analyse gait from monocular view, and including more parameters for gait analysis from video, such as foot and hip angles for more precise gait analysis.

## REFERENCES

[1] Whittle, M., *Gait Analysis: An Introduction*, Butterworth-Heinemann, 2007  
 [2] Han, Ju and Bhanu, Bir, Statistical feature fusion for gait-based human recognition, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004*, IEEE, II-842, 2

[3] Mubashir, Muhammad and Shao, Ling and Seed, Luke, A survey on fall detection: Principles and approaches, *Neurocomputing*, 2013, 144–152, 100, Elsevier  
 [4] Toshev, Alexander and Szegedy, Christian, DeepPose: Human Pose Estimation via Deep Neural Networks, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 2014  
 [5] Baukhage, Christian and Tsotsos, John K and Bunn, Frank E, Automatic detection of abnormal gait, *Image and Vision Computing*, 2009, 108–115, 27, Elsevier  
 [6] Wang, Liang, Abnormal walking gait analysis using silhouette-masked flow histograms, *18th International Conference on Pattern Recognition (ICPR 2006)*, 2006, IEEE, 473–476, 3  
 [7] Nieto-Hidalgo, Mario and Ferrández-Pastor, Francisco Javier and Valdivieso-Sarabia, Rafael J and Mora-Pascual, Jerónimo and García-Chamizo, Juan Manuel, A vision based proposal for classification of normal and abnormal gait using RGB camera, *Journal of Biomedical Informatics*, 2016, 82–89, 63, Elsevier  
 [8] Wang, Jin and She, Mary and Nahavandi, Saeid and Kouzani, Abbas, A review of vision-based gait recognition methods for human identification, *2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2010, IEEE, 320–327  
 [9] Hollman, John H and McDade, Eric M and Petersen, Ronald C, Normative spatiotemporal gait parameters in older adults, *Gait & Posture*, 2011, 1, 111–118, 34.  
 [10] Muro-de-la-Herran, Alvaro and Garcia-Zapirain, Begonya and Mendez-Zorrilla, Amaia, *Gait analysis methods: an overview of wearable and Non-wearable systems, highlighting clinical applications*, *Sensors*, 2014, 2, 3362–3394, 14, Multidisciplinary Digital Publishing Institute,  
 [11] Butterworth, Stephen, On the theory of filter amplifiers, *Wireless Engineer*, 1930, 6, 536–541, 7  
 [12] Graves, Alex, Navdeep Jaitly, and Abdel-rahman Mohamed, Hybrid speech recognition with deep bidirectional LSTM, *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 2013.  
 [13] H. Ismail and I. Radwan and H. Suominen and G. Waddington and R. Goecke, "Human Postural Sway Estimation from Noisy Observations", *In 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 454-461.