A fusion framework to estimate plantar ground force distributions and ankle dynamics

Fani Deligianni[a], Charence Wong[a], Benny Lo[a], Guang-Zhong Yang[a,*]

[a]*Hamlyn Centre, Imperial College London, United Kingdom*

Abstract

Gait analysis plays an important role in several conditions, including the rehabilitation of patients with orthopaedic and the monitoring of neurological conditions, mental health problems and the well-being of elderly subjects. It also constitutes an index of good posture and thus it can be used to prevent injuries in athletes and monitor mental health in typical subjects. Usually, accurate gait analysis is based on the measurement of ankle dynamics and ground reaction forces . Therefore, it requires expensive multi-camera systems and pressure sensors, which cannot be easily employed in a free-living environment. We propose a fusion framework that uses an ear worn activity recognition (e-AR) sensor and a single video camera to estimate foot angle during key gait events. To this end we use canonical correlation analysis with a fused-lasso penalty in a two-steps approach that firstly learns a model of the timing distribution of ground reaction forces based on e-AR signal only and subsequently models the eversion/inversion as well as the dorsiflexion of the ankle based on the combined features of e-AR sensor and the video. The results show that incorporating invariant features of angular ankle information from the video recordings improves the estimation of the foot progression angle, substantially.

Keywords: Gait analysis, e-AR, fusion of sensors gait data and video

2010 MSC: 00-01, 99-00

*Corresponding Author

Email address: g.z.yang@imperial.ac.uk (Guang-Zhong Yang)

URL: http://www.imperial.ac.uk/hamlyn-centre/ (Guang-Zhong Yang)

1. Introduction

Gait analysis is a well-established method for analysing the biomechanics of gait, and a means to capture effective and quantitative assessments for orthopaedic and neurological rehabilitation. Motion capture with topical tracking systems, force plates, instrumented treadmills and pressure sensing insoles are instruments for measuring the heel, subtalar, ankle and knee joint angles, and analysing the force exerted on the ground for accurate analysis of the biomechanical indices of subjects. Although such instrumentations are widely available, the high cost and long set up times typically required have restricted the use of such instruments in major hospitals for routine measurement of certain patients.

Pathological gait is difficult to describe, since it involves atypical ankle kinematics. Nevertheless, it is characterized by the periodic movement of each leg from one position to the next and the corresponding ground reaction forces that support the motion of the body. The ankle is the lower joint and the first to respond to the impact of the foot with the ground. In particular, the subtalar joint, which is lateral to the ankle, is responsible for most of the inversion and eversion of the foot, which plays a significant role in the toe-off phase of the gait as it provides the propulsion to lift the foot. In other words, ground reaction forces along with ankle eversion/inversion and dorsiflexion play a key role in the biomechanical dynamics. Several recent studies have shown that certain gait characteristics can be related to abnormal posture, the development of osteoarthritis and sports related injuries [1, 2, 3, 4, 5]. For example, Kuhman et al. has shown that lower leg and foot dynamics are related to the development of injuries in runners [2]. Furthermore, greater rear-foot pronation has been associated with greater pressure on the medial portions of the plantar surface during walking and it has been observed in individuals with poor postural control [5].

To measure the lower limb kinematics accurately expensive multi-camera configuration systems are used to detect and track reflective skin markers. However, the confined spaces typically available in clinics or at home means these methods cannot easily be applied in these scenarios. The use of monocular vi-

sion has also been proposed for a number of gait analysis applications, such as biometric authentication [6], diagnosis of Parkinson's disease [7], and identification of abnormalities for assisted living [8, 9]. Some of these works map the 2D extracted trajectories to 3D word coordinate system based on deep neural networks and require several labeled training sets. Furthermore, they assume large distances between the subject and the camera and assumptions that the body mass is planar. This does not allow an accurate estimation of the ankle dynamics and the foot progression angle. Furthermore, estimation of ground reaction forces are also important to determine the health risks over time due to excessive joint loading rates and stress. Accurate measurements of ground reaction forces normally require pressure insoles, which are placed inside the shoes.

Recently, wearable wireless body worn sensors have been proposed for detailed gait analysis [10, 11]. Our previous work has shown the feasibility and accuracy of using the ear-worn activity recognition (e-AR) sensor for detailed gait analysis and activity recognition [12, 13]. This lightweight and miniaturized sensor, e-AR, enables pervasive and continuous monitoring of user with negligible distraction to their normal daily activities. In previous work, we have demonstrated the feasibility of using the e-AR sensor with a hierarchical Bayesian Network framework for estimation of GRFs for normal gait [14]. This hierarchical model allowed characterisation of the plantar force timing distribution based on e-AR measurements only. In a recent article, Clark et al. showed that it is possible to predict vertical ground reaction forces in runners based on the body mass, the contact time between steps and the swing time only [15]. [16, 17] compare the advantages of inertial and vision for gait analysis. Although the wearable sensor can estimate the temporal distribution accurately, other detailed gait parameters, such as subtalar joint angle are more difficult to be determined based on inertial sensors only.

In this paper, we propose a novel integrated approach of using the e-AR sensor together with a single video camera, and introduce a framework to fuse the sensing and visual features to reveal the interaction of ground reaction

forces and ankle dynamics during normal and abnormal walking. In particular, we utilize Canonical Correlation Analysis with a fused-lasso penalty (fCCA) to extract features across steps that reveal correlations between the e-AR signal and the timing distribution of key gait events. These events occur when ground reaction forces are maximized in the plantar foot areas, such as heel , mid-foot , front-foot and toes . In a two-steps approach, we use fCCA again to fuse the e-AR signal with features derived from the video analysis of a single camera that reflect an angular interaction between the two legs during walking. In this way, we are able to create a prediction framework of the dorsiflexion and inversion/aversion foot angles during heel, mid-foot, front-foot and toe contacts with the ground.

2. Methods

2.1. Data fusion framework

Both normal and pathological gait exhibit repetitive patterns of motion of the lower limbs. In this paper, we utilize this to construct a fusion framework that samples across steps of e-AR signal and video recordings to extract features that predict well ground reaction forces timing distributions and subsequently foot angles in key gait events. Therefore, the framework has two main components that are constructed independently but they interact to provide detailed gait characteristics. The proposed fusion framework requires time-series derived from e-AR sensor, insole sensors and video features to be segmented into gait steps, independently. This is also important as it alleviate the need for accurate synchronization between different modalities. An overview of the framework is presented in Figure 1.

We are interested in learning a relationship between the e-AR acceleration data and the plantar force timing distributions across steps. The e-AR measures acceleration in three axes that are aligned to the body: Medial-Lateral (ML) axis, Superior-Inferior (SI) axis and Anterior-Posterior (AP) axes. On the other hand, ground reaction forces can be measured with foot pressure insoles that

record the pressure between the planar surface of the foot and the sole of the shoes. In order to estimate the plantar force distributions, we hierarchically subdivide the foot into the Heel, Mid-foot, Front-foot and Toe regions as well as Medial and Lateral regions. This results in eight sub-regions similar to our previous work [14]. The insole data are pre-processed to detect gait steps based on the pressure difference between left and right foot. Subsequently, for each step the timings of the maximums of the sub-plantar force distributions are defined within each region. These timings represent key gait events and they are important in identifying abnormal gait.

Once both insole and e-AR data are segmented into steps, we normalize the e-AR signal at each step with respect to the time axis based on linear interpolation so that all steps are equally sampled. Note that we concatenate horizontally the combined SI and AP signal along with the ML signal. Subsequently, these vectors are concatenated vertically to form a matrix, $\mathbf{X}$, $m \times 2n$, where m is the number of steps and n is the number of time samples. On the other hand, the response data $\mathbf{Y}$ is a $m \times k$ matrix that reflects the timings of the peaks of the plantar force distribution estimated based on the insole data. k is the number of plantar sub-regions defined. fCCA is used to relate the e-AR waveform data for each step with the GRFs timing distributions obtained from insole data. Canonical correlation analysis is a powerful tool of modelling the correlation between multivariate variables. The projection of $\mathbf{X}$ and $\mathbf{Y}$ on the derived canonical vectors result in maximally linearly correlated variables. Thus, it allows bi-direction predictive modelling of the associated variables and it has been used in high-dimensional spaces of multi-view gait recognition and numerous other applications [18, 19, 20]. fCCA is a variant of canonical correlation analysis that applies a fused lasso penalty, which penalizes the L_1 norm of both the coefficients and their successive differences. This enforces both sparsity and smoothness, which is important since the fCCA variables are time-series segments and ordered variables [21]. The implementation of fCCA is based on a penalized matrix decomposition framework, which obeys the following criterion

[22, 23]:

$$\begin{aligned} &\text{maximise}_{\mathrm{u,v}} u^T \mathbf{X}^T \mathbf{Y} v \\ &\text{subject to :} \|u\|^2 \leq 1,\ \|v\|^2 \leq 1,\ f_1 \leq c_1,\ f_2(v) \leq c_2 \end{aligned} \tag{1}$$

Here, f_1, f_2 are convex penalty functions that both impose a fused lasso penalty:

$$\mathrm{f(w)} = \sum_j \|\mathrm{w_j}\| + \sum_j \|\mathrm{w_j} - \mathrm{w_{j-1}}\| \tag{2}$$

Note that with u fixed, the criterion in eq. 1 is convex in v , and with v fixed, it is convex in u . Therefore, the objective function of this biconvex criterion increases in each step of an iterative algorithm [23]:

$$\begin{aligned} u &\leftarrow \mathrm{argmax_u} u^T \mathbf{X}^T \mathbf{Y} v \ \text{ subject to : } \|u\|^2 \leq 1, f_1(u) \leq c_1 \\ v &\leftarrow \mathrm{argmax_v} u^T \mathbf{X}^T \mathbf{Y} v \ \text{ subject to : } \|v\|^2 \leq 1, f_2(v) \leq c_2 \end{aligned} \tag{3}$$

Once the fCCA model has been trained it can be used for prediction:

$$\hat{\mathbf{Y}}_{\mathbf{s}} = (u\mathbf{X_s})^+ \mathbf{D} v^+ \tag{4}$$

Where, $\mathbf{D}$ is the diagonal matrix with the canonical correlation scores and + denotes the pseudo-inverse.

The second major component of the fusion framework is the incorporation of video features derived from a single camera. This provides us with the ability not only to delineate important timing gait events but also estimate the angles between the foot and leg that reflect dorsiflexion and inversion/aversion in these key gait points independently of the camera view point. To this end fCCA is applied again to find a relationship between the combined data derived from e-AR and video features, $\mathbf{Z}$, and foot angles, $\mathbf{W}$ estimated in key gait events, such as when GRFs are maximized during heel, mid-foot, frontal-foot and toe contacts with the ground. Therefore, $\mathbf{Z}$ is an $m \times 2n$ matrix, where m is the number of steps and n reflects the number of time samples. To form $\mathbf{Z}$ we concatenate horizontally the sum of the AP and SI eAR signal and an index based on cross ratio estimated extracted from single video recordings. $\mathbf{W}$ is a $m \times 2k$ matrix that encodes information about the foot angles (inversion/aversion and

dorsiflexion) in both the left and right foot when the GRFs are maximized in the key foot subregions Heel (H), Mid-foot (M), Front-foot (F) and Toe (T), Figure 3b. fCCA takes similar form as in equation 1. However, we have replaced f_2 fussed lasso penalty with a lasso penalty to reflect the fact that $\mathbf{W}$ does not encode ordered variables.

2.2. Processing of video data

To complement the inertial motion data captured using the ear-worn e-AR sensor, video is used to determine gait features that cannot be obtained through wearable devices. To capture the characteristics of the subject's gait from video, visual tracking and image segmentation steps are performed. Here, we have recorded the front and back views of the subjects. Our main assumption is that the camera has up-right orientation with respect to the ground.

Tracking. To ensure that the subjects' gait is evaluated consistently, a state-of-the-art tracking-by-detection method, kernelized correlation filter (Joao F Henriques, Caseiro, Martins, Batista, 2012; Joo F Henriques, Caseiro, Martins, Batista, 2015), is used to locate the lower limbs in each video frame. Histogram of oriented gradients and colour-space features are used to perform multi-scale tracking of the subject.

Image Segmentation. A clustering algorithm, K-means clustering, is then employed to segment the tracked region of the image into separate classes of foreground and background. The foreground clusters of interest in this work are the lower parts of the legs and feet. To improve the robustness of the method against non-uniform colours in the subject's footwear, GrabCut [24] is used to further refine the contours of the segmented body parts. GrabCut is based on the 'Graph Cut' algorithm, which uses a k-Gaussian mixture models to segment the target object from the background [25]. In an indirected graph, the Graph Cut aims to find a subset of edges C such that the two terminal nodes are separated in the induced graph. The algorithm minimizes an energy cost function $\mathbf{E}$

that measures how well a colour distribution model h fits the data and imposes smoothness constraints.

$$\mathbf{E}(\alpha, \theta, \mathbf{z}) = \sum_{n} -\log h(z_n; \alpha_n) + \gamma \sum_{(m,n \in C)} dis(m,n)^{-1} [\alpha_n \neq \alpha_m] \exp -\beta (z_m - z_n)^2 \tag{5}$$

where α is the pixels' label, $\mathbf{z}$ encodes the colour information and θ are the model's parameters. For each pair of neighbour pixels that do not have the same label, the energy function is increased according to the parameter β.

Grabcut requires a pre-estimation of a rectangular box or mask that surrounds the feet. We pre-estimate a rectangular area based on the position of the legs and subsequently we refine it iteratively based on the results of GrabCut. Figure 2 shows an example of the contours obtained for the legs and feet as well as the results for lower limb tracking for the four different types of gait considered in this paper.

Contour analysis. Contour analysis on the extracted body parts is then performed to find the orientation of the lower limbs and feet. Localization of individual body parts allows the distances between the left and right foot to be estimated in an invariant way with respect to the distance of the person from the camera. To calculate a view-invariant measure of the lower pose, a cross-ratio between the directions of the legs and feet is used, Fig. 1e):

$$CR = \frac{\cos(L_L, F_L) * \cos(L_R, F_R)}{\cos(L_L, L_L) * \cos(F_R, F_R)} \tag{6}$$

L_L F_L,L_R F_R correspond to fitted lines at the left leg, left foot, right leg and right foot, respectively, Figure 2.

2.3. Segmentation of multi-modal time-series data into steps

Both e-AR acceleration data and visual features are segmented into steps based on singular spectrum analysis and peak detection. The preprocessing of e-AR with singular spectrum analysis is based on the acceleration in the SI and AP axis, whereas the segmentation of the visual features is based on the normalized distance between the left and right ankle as it reflected on the

vertical direction of the image plane. Singular spectrum analysis is mainly used to denoise the signal and improve the detection of peaks that reflect foot contacts with the ground [12, 13]. If $\mathbf{s}$ is a time-series, a trajectory matrix is created based on an embedding dimension that reflects the window length of the sub-sampled time-series grouped together:

$$\mathbf{S} = [\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_k}] = \begin{pmatrix} s_0 & s_1 & \cdots & s_{k-1} \\ s_1 & s_2 & \cdots & s_k \\ \vdots & \vdots & \ddots & \vdots \\ s_{l-1} & s_l & \cdots & s_{n-1} \end{pmatrix} \tag{7}$$

where $k = n - l + 1$, n is the length of the time-series and l is the embedding dimension. This is a Hankel matrix with constant skew-diagonal elements. The singular value decomposition of a Hankel matrix is related to the state-space realization of a Hidden Markov model and it is appropriate for the decomposition of non-stationary signals. The covariance matrix of $\mathbf{S}$, $\mathbf{C} = \mathbf{S}\mathbf{S}^\mathbf{T}$ is decomposed into the eigenvectors $\mathbf{u}_i$ and the corresponding eigenvalues λ_i (Jarchi et al., 2014). Therefore, the trajectory matrix is written as:

$$\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2 + \cdots + \mathbf{S}_d \tag{8}$$

where $d = argmax_i\{\lambda_i > 0\}$, $\mathbf{S}_i = \sqrt{\lambda_i}\mathbf{u}_i\mathbf{v}_i^T$ and $\mathbf{v}_i = \mathbf{S}^\mathbf{T}\mathbf{u}_i/\sqrt{\lambda_i}$. Finally, the signal is reconstructed based on the diagonal averaging of a subset of the group elementary matrices derived from the decomposition.

3. Results

3.1. Data acquisition

We acquired simultaneous recordings of ear-sensor data and insole data (PAROTEC, Paromed, Germany) from seven healthy volunteers that performed four different styles of walking patterns: normal walking (normal), imitating limping based on unequal time steps (limping), imitating inversion injury (pronation) and imitating eversion injury (supination). For each condition we

recorded four sessions of each subject (total 16 sessions across all conditions) where the subject was asked to walk back and forth in a corridor for a total of approximately 11 meters. Visual information of each subject's gait is simultaneously captured using a 2D camera with a resolution of 1080p at 30 frames per second.

3.2. Training and validation

Figure 3a shows the average e-AR signal across all subjects and e-AR steps for normal walking, limping, walking with exaggerate pronation, and walking with exaggerated supination, respectively. The e-AR signal along IS and AP and ML axis have been processed based on singular spectrum analysis . Firstly, singular spectrum analysis is applied in IS, AP and ML axis independently. Subsequently, IS and AP are added together and the singular spectrum analysis is re-applied. The result is concatenated with the ML e-AR signal to form the predictive variables X. Here, we note that different styles of walking result in a phase shift of the e-AR signal when steps are normalized to have the same sample length. Figure 3b depicts a diagram of the insole sensors that shows how the foot plantar area has been segmented to each region. Firstly, with respect to the AP axis the foot is segmented to the Toe, Front-foot, Mid-foot and Heel regions. With respect to the ML axis, the foot is segmented to Medial and Lateral regions. Subsequently, sub-regions are defined as Heel-Lateral, Heel-Medial, Midfoot-Lateral, Midfoot-Medial, Front-foot-Lateral, Front-foot-Medial, Toe-Lateral, Toe-Medial. The predictive/response variables Y are shown in Figure 4. This shows the plantar GRFs timing distributions across all subjects for each condition. Gait events have been identified based on insole data and they have been normalized with respect to each insole step. Gait events have been sorted based on their mean value during normal walking.

Due to the difference in shoe size, foot pressure insoles are expensive. With the aim of providing a low cost approach for gait analysis, we are interested in devising a model that encodes the relationship between e-AR and insole data. Furthermore, we would like to be able to get accurate measurements of the

events timing in abnormal walking patterns. Therefore, we train and test our model based on leave-one-out cross validation in the following scenarios:

- Within subjects and within conditions: To form the training dataset, the data across three sessions are concatenated in each walking condition/pattern and each subject independently. The fourth session is used for testing. The mean errors and mean standard deviations are averaged across subjects and cross-validation rounds.

- Across subjects and within conditions: To form the training dataset, the data across all subjects and within each condition are concatenated. Subsequently, leave-one-out cross validation is used to estimate the mean error and mean standard deviation, Figure 5a.

- Across subjects and across conditions: Data across subjects and across conditions are concatenated and leave-one-out cross validation is used to estimate the mean error and mean standard deviation, Figure 5b.

We note a significant error reduction when we train the model across subjects and conditions. The results show that training the fCCA model with data acquired with normal walking as well as pathological variations enhances the prediction performance. Perhaps, this reflects that the model is more robust to outliers. Outliers can originate from both false positive and false negative detection of foot contacts with the ground. The results also show that the identification of heel contact is relatively the most accurate across all conditions and training scenarios. This is expected since heel contacts in normal walking is the first foot contact with the ground and therefore the change in the acceleration is rapid.

In order to evaluate the proposed measure of angular variation based on a cross ratio, CR, we estimate the correlation between several extracted video features and the normalized distance between left and right ankle in the vertical direction of the image plane . Figure 6a shows an example of the extracted features from the analysis of the video when the subject walk towards the camera.

These features include the dot products between the two legs , the two feet , the right leg and right foot and the left leg and left foot . Figure 6b shows that CR correlates well to the normalised distance between left and right ankle in the vertical direction of the image plane across all subjects and conditions.

In Figure 7, we demonstrate the performance of fCCA in fusing the e-AR signal with the derived video features, namely CR, to model the foot angle progression. Figure 7a show the error in radians based on leave-one-out cross validation of the proposed method. The fCCA model estimates the dorsiflexion and the aversion /inversion of the ankle when ground reaction forces at the heel , mid-foot , front-foot and toes are maximized. Therefore, we show the mean errors and standard deviation of estimating DF and EI at Heel, Mid-foot, Front-foot and Toes contact with the ground. Note that the results from left/right foot have been averaged accordingly. Ground truth data have been measured based on the identification of the knee, ankle and toes manually in each frame and subsequently estimating dorsiflexion as the angle of the foot with the vertical image plane axis and inversion/eversion as the angle of the foot with the horizontal image plane axis. Figure 7b-7e shows the difference of angular error between a model that uses only e-AR to predict foot angle progression and the proposed method for each of the conditions: normal walking, limping, pronation and supination, respectively. Although, we show the results summarized independently in each condition, the fCCA model has been trained based on cross-validation across all conditions at once. The proposed fusion method outperforms the model based on e-AR data only in most of the cases by a significant level. In fact, using the e-AR approach to estimate foot angle information during limping is inaccurate, possibly, due to the asymmetry between right and left gait steps. Nevertheless, combining e-AR signal and video features improves the performance in most other walking patterns too. The reduction in the performance of the proposed method during pronation may result from inaccuracies in segmenting the video features into steps. This could reflect the fact that the vertical distance between the right and left ankle is smaller in pronation.

4. Discussion

A number of approaches based on inertial sensors, cameras and computer vision algorithms have been proposed to measure gait characteristics, such as step length, foot angles and ground reaction forces. Most of them are based on system configurations that are difficult to be installed in a free-living environment. For example, these systems involve installing markers on the floor and or on the shoe [26]. On the other hand, advanced gait recognition systems use inertial sensors to assess gait as a biometric trait for security/surveillance applications [27, 28]. These systems are useful to identify individuals based on their gait characteristics but analysing gait for clinical assessment requires understanding the dynamics of the joint loading rates and stress, which are directly related to ground reaction forces and angles between the lower limp segments. These measurements are used to provide objective and reliable estimates of the progression of diseases in neurological conditions, stroke rehabilitation, ageing and orthopaedics [29].

The proposed framework exploits fCCA in a two-steps approach. Firstly, fCCA is used to extract the coefficients that relate the e-AR signal across steps to the timing distribution of ground reaction forces estimated based on insole sensors data. To verify that the proposed model is generalizable to new subjects, leave-one-out cross validation is used on the concatenated training sets across subjects and conditions. Subsequently, we use fCCA again to fuse the e-AR signal with visual features that reflect angular information of the lower limbs. This model learns a relationship between the fused information and the foot angle at key gait events, which reflect maximum ground reaction forces at heel, mid-foot, front-foot and toes, respectively. fCCA is based on a penalized decomposition method that imposes a fused lasso penalty on the coefficients that results in smoothness. We adapted this method because it can handle high dimensional data and it takes into account that the fCCA variables are ordered/time-series data. The extracted coefficients are sparse and they can reveal further which variables play a critical role in the prediction. Other approaches like deep neural

networks may be an interesting alternative of fusing information across different sensors and/or video. However, they require an abundance of training datasets that include insole recordings and angle measurements across patients groups.

Our method requires that both the e-AR signal and the extracted visual features are segmented into steps. There are several advantages with this approach. Firstly, the model is less sensitive to over-fitting even with a small number of subjects, since training samples are across steps. Furthermore, it does not require accurate synchronization between the multi-modal signals. However, the accuracy of the approach depends on the segmentation of multivariate time-series data, which is not a trivial problem and in abstract biological applications it involves high computational complexity [30]. Our previous work has shown that based on singular spectrum analysis, the e-AR signal can be reliably processed in a fraction of a second to derive clinically relevant gait parameters such as timing of initial foot contact, step time, swing time and stance time [12, 13]. Here, we have used singular spectrum analysis and peak detection to segment both the e-AR signal and the signal derived from the analysis of the video.

It is worth noting that our framework only requires to match steps across modalities, which is a significant simplification of the synchronisation problem [31, 32]. Steps are detected independently in each modality based on peak detection that correspond to initial foot contacts with the ground. For example, the peaks of the e-AR signal normally correspond to the initial foot contacts and they also correspond to the peaks of the ground reaction forces reflected in the insole data. The video data was processed to extract the normalised distance between left and right ankle in the vertical direction of the image plane and the cross-ratio. All these signals are periodic in nature with peaks that reflect left and right foot contact. However, the performance of the proposed algorithm would be affected when peaks are not consistently detected within each modality. In our approach is not important whether the peaks reflect exactly the same gait event across modalities. Furthermore, a large portion of errors in peak detection can be identified and filtered out based on the step time duration.

For the analysis of the video state-of-the art tracking technology is used and combined with advanced image segmentation approaches to identify the lower limb parts. Tracking of each limb segment independently is challenging even under controlled conditions. For this reason, we used tracking just to identify the area of legs and subsequently we utilize image segmentation approaches to segment background from foreground pixels. Subsequently, we applied Grabcut iteratively to segment the area of foot. Our results are promising and they can correctly segment the feet even when the subject is at the far end of the corridor.

To our knowledge there is very little work on gait analysis based on a single RGB camera. Accurate estimation of the angles between lower limbs, such as dorsiflexion and inversion/aversion angles based on marker-less motion capture is discussed by Sandau et al. [33]. In clinical settings, lower limbs angle estimation requires a laboratory environment for the placement of markers and multi-camera tracking, which is the gold standards for gait analysis. Marker-less systems have been developed but they also require multi-camera setup and even then, their accuracy is significantly compromised [33]. RGB-D sensors are able to acquire much more information on the scene with just one infrared camera. Out-of-the-box algorithms that come with devices such as Kinect extract the joints of the whole human body in real time. However, there is considerable jitter on the measured 3D location of the joints and their accuracy depends on the view angle. These factors compromise the use of these devices in estimating accurate angular measurements in clinical scenarios [34, 35, 36]. In fact, hip and knee angle correlations between estimated and multi-camera ground-truth data were lower than 0.3 and 0.8, respectively[34]. To deal with these problems Ye et al. [36]used markers and an RGB-D system to extract gait characteristics. Nevertheless, the error in the knee angle estimation can get up to 10 degrees during key gait events. Sundau et al. [33]used a marker-less motion capture system based on eight cameras to recover the joint angles with average accuracy of -0.7 ± 1.8 and 0.5 ± 2.9 degrees for dorsal/plantar flexion and eversion/inversion, respectively. However, the mean error in key gait events can go up to -1.0 ± 2.7 and 3.9 ± 2.8 degrees for dorsal/plantar flexion and ever-

sion/inversion, respectively. Our results based only on one camera and the e-AR sensor compare well with dorsal/plantar flexion average error around 4.5 ± 2.43 degrees and eversion-inversion mean error 2.29 ± 1.14 degrees at heel strike.

Our results show that the accuracy of the estimated ankle angle is affected by the walking pattern. For example, dorsal/plantar flexion and eversion/inversion have been identified more accurately during normal walking and pronation compared to limping and supination. The reason for this is that step identification is harder during limping and supination. It should be noted that step identification is independent in each modality. Nevertheless, our approach can be extended to minimise errors during asymmetric gait such as limping by explicitly accounting for left and right steps in the training procedure of the fCCA model. Another limitation is that we have identified 2D angles as opposed to 3D angles. To alleviate this issue we suggested a novel measure based on the cross-ratio between the foot and the leg, which is relatively invariant to the camera view. However, this measure is not yet adopted to clinical scenario and it more difficult to be interpreted. Finally, our method uses GrabCut and clustering approaches to segment the foot and leg. Therefore, its success depends on the contrast between for foreground and background pixels.

We asked healthy volunteers to walk normally as well as to imitate limping, pronation and supination. Therefore, our population variance and motion of the lower limbs may not represent accurately the variance observed in real pathological cases. We have used leave-one-out cross validation to determine the out-of-sample mean modelling error and standard deviation. The error is statistically significant in all cases. However, further work that would involve validation based on multi-camera tracking and confidence interval estimation is required for patient cohorts.

5. Acknowledgements

We would like to thank Dr. Daniele Ravi and acknowledge EPSRC as the funding body for this study: Smart Sensing for Surgery (EP/L014149/1).

References

[1] D. A. Bruening, T. E. Cooney, M. S. Ray, G. A. Daut, K. M. Cooney, S. M. Galey, Multisegment foot kinematic and kinetic compensations in level and uphill walking following tibiotalar arthrodesis, Foot and Ankle International 37 (10) (2016) 1119–1129. doi:10.1177/1071100716655205.
URL <GotoISI>://WOS:000385345300013

[2] D. J. Kuhman, M. R. Paquette, S. A. Peel, D. A. Melcher, Comparison of ankle kinematics and ground reaction forces between prospectively injured and uninjured collegiate cross country runners, Human Movement Science 47 (2016) 9–15. doi:10.1016/j.humov.2016.01.013.
URL <GotoISI>://WOS:000375167500002

[3] R. A. Resende, R. N. Kirkwood, K. J. Deluzio, E. A. Hassan, S. T. Fonseca, Ipsilateral and contralateral foot pronation affect lower limb and trunk biomechanics of individuals with knee osteoarthritis during gait, Clinical Biomechanics 34 (2016) 30–37. doi:10.1016/j.clinbiomech.2016.03.005.
URL <GotoISI>://WOS:000375808800006

[4] K. E. Roach, B. B. Wang, A. L. Kapron, N. M. Fiorentino, C. L. Saltzman, K. B. Foreman, A. E. Anderson, In vivo kinematics of the tibiotalar and subtalar joints in asymptomatic subjects: A high-speed dual fluoroscopy study, Journal of Biomechanical Engineering-Transactions of the Asme 138 (9). doi:Artn09100610.1115/1.4034263.
URL <GotoISI>://WOS:000383260200006

[5] D. D. Silva, F. H. Magalhaes, M. F. Pazzinatto, R. V. Briani, A. S. Ferreira, F. A. Aragao, F. M. de Azevedo, Contribution of altered hip, knee and foot kinematics to dynamic postural impairments in females with patellofemoral pain during stair ascent, Knee 23 (3) (2016) 376–381. doi:10.1016/j.knee.2016.01.014.
URL <GotoISI>://WOS:000378963600007

[6] G. Rogez, J. Rihan, J. J. Guerrero, C. Orrite, Monocular 3-d gait tracking in surveillance scenes, Ieee Transactions on Cybernetics 44 (6) (2014) 894–909.
URL <GotoISI>://WOS:000337960000014

[7] C. W. Cho, W. H. Chao, S. H. Lin, Y. Y. Chen, A vision-based analysis system for gait recognition in patients with parkinson's disease, Expert Systems with Applications 36 (3) (2009) 7033–7039.
URL <GotoISI>://WOS:000263817100156

[8] M. Nieto-Hidalgo, F. J. Ferrandez-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, J. M. Garcia-Chamizo, A vision based proposal for classification of normal and abnormal gait using rgb camera, Journal of Biomedical Informatics 63 (2016) 82–89.
URL <GotoISI>://WOS:000389557000009

[9] E. E. Stone, M. Skubic, Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing, 2011 Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc) (2011) 6491–6494.
URL <GotoISI>://WOS:000298810005008

[10] C. Wong, Z. Q. Zhang, B. Lo, G. Z. Yang, Wearable sensing for solid biomechanics: A review, Ieee Sensors Journal 15 (5) (2015) 2747–2760. doi:10.1109/Jsen.2015.2393883.
URL <GotoISI>://WOS:000352624500006

[11] N. Raveendranathan, S. Galzarano, V. Loseu, R. Gravina, R. Giannantonio, M. Sgroi, R. Jafari, G. Fortino, From modeling to implementation of virtual sensors in body sensor networks, IEEE Sensors Journal 12 (3) (2012) 583–593. doi:10.1109/JSEN.2011.2121059.

[12] D. Jarchi, C. Wong, R. M. Kwasnicki, B. Heller, G. A. Tew, G. Z. Yang, Gait parameter estimation from a miniaturized ear-worn sensor

using singular spectrum analysis and longest common subsequence, Ieee Transactions on Biomedical Engineering 61 (4) (2014) 1261–1273. doi:10.1109/Tbme.2014.2299772.
URL <GotoISI>://WOS:000337739300024

[13] D. Jarchi, B. Lo, C. Wong, E. Ieong, D. Nathwani, G. Z. Yang, Gait analysis from a single ear-worn sensor: Reliability and clinical evaluation for orthopaedic patients, Ieee Transactions on Neural Systems and Rehabilitation Engineering 24 (8) (2016) 882–892. doi:10.1109/Tnsre.2015.2477720.
URL <GotoISI>://WOS:000381496500007

[14] B. Lo, J. Pansiot, G. Z. Yang, Bayesian analysis of sub-plantar ground reaction force with bsn, Sixth International Workshop on Wearable and Implantable Body Sensor Networks, Proceedings (2009) 133–137doi:10.1109/P3644.37.
URL <GotoISI>://WOS:000275810500023

[15] K. P. Clark, L. J. Ryan, P. G. Weyand, A general relationship links gait mechanics and running ground reaction forces, Journal of Experimental Biology 220 (2) (2017) 247–258. doi:10.1242/jeb.138057.
URL <GotoISI>://WOS:000392154200018

[16] I. H. Lpez-Nava, I. Gonzlez, A. Muoz-Melndez, J. Bravo, Comparison of a Vision-Based System and a Wearable Inertial-Based System for a Quantitative Analysis and Calculation of Spatio-Temporal Parameters, Springer International Publishing, Cham, 2015, pp. 116–122. doi:10.1007/978-3-319-26508-7_12.
URL http://dx.doi.org/10.1007/978-3-319-26508-7_12

[17] I. Gonzalez, I. H. Lopez-Nava, J. Fontecha, A. Munoz-Melendez, A. I. Perez-SanPablo, I. Quinones-Uriostegui, Comparison between passive vision-based system and a wearable inertial-based system for estimating temporal gait parameters related to the gaitrite electronic walkway, Jour-

nal of Biomedical Informatics 62 (2016) 210–223.
URL <GotoISI>://WOS:000384703800020

[18] F. Deligianni, M. Centeno, D. W. Carmichael, J. D. Clayden, Relating resting-state fmri and eeg whole-brain connectomes across frequency bands, Frontiers in Neuroscience 8. doi:ARTN25810.3389/fnins.2014.00258.
URL <GotoISI>://WOS:000346512300001

[19] F. Deligianni, D. W. Carmichael, G. H. Zhang, C. A. Clark, J. D. Clayden, Noddi and tensor-based microstructural indices as predictors of functional connectivity, Plos One 11 (4). doi:ARTNe015340410.1371/journal.pone.0153404.
URL <GotoISI>://WOS:000374131700051

[20] X. L. Xing, K. J. Wang, T. Yan, Z. W. Lv, Complete canonical correlation analysis with application to multi-view gait recognition, Pattern Recognition 50 (2016) 107–117. doi:10.1016/j.patcog.2015.08.011.
URL <GotoISI>://WOS:000364893700008

[21] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, Journal of the Royal Statistical Society Series B-Statistical Methodology 67 (2005) 91–108.
URL <GotoISI>://WOS:000225686900006

[22] D. M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics 10 (3) (2009) 515–534.
URL <GotoISI>://WOS:000267213700010

[23] D. M. Witten, R. J. Tibshirani, Extensions of sparse canonical correlation analysis with applications to genomic data, Statistical Applications in Genetics and Molecular Biology 8 (1).
URL <GotoISI>://WOS:000267601500008

[24] C. Rother, V. Kolmogorov, A. Blake, "grabcut": interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. 23 (3) (2004) 309–314. doi:10.1145/1015706.1015720.

[25] Y. Y. Boykov, M. P. Jolly, Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images, Eighth IEEE International Conference on Computer Vision, Vol I, Proceedings (2001) 105–112.
URL <GotoISI>://WOS:000170817100014

[26] T. N. Do, Y. S. Suh, Gait analysis using floor markers and inertial sensors, Sensors 12 (2) (2012) 1594–1611. doi:10.3390/s120201594.
URL <GotoISI>://WOS:000300720300025

[27] G. Panahandeh, N. Mohammadiha, A. Leijon, P. Handel, Continuous hidden markov model for pedestrian activity classification and gait analysis, Ieee Transactions on Instrumentation and Measurement 62 (5) (2013) 1073–1083. doi:10.1109/Tim.2012.2236792.
URL <GotoISI>://WOS:000317361500023

[28] S. Sprager, M. B. Juric, Inertial sensor-based gait recognition: A review, Sensors 15 (9) (2015) 22089–22127. doi:10.3390/s150922089.
URL <GotoISI>://WOS:000362512200056

[29] A. Muro-de-la Herran, B. Garcia-Zapirain, A. Mendez-Zorrilla, Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications, Sensors 14 (2) (2014) 3362–3394. doi:10.3390/s140203362.
URL <GotoISI>://WOS:000335887900072

[30] N. Omranian, B. Mueller-Roeber, Z. Nikoloski, Segmentation of biological multivariate time-series data, Scientific Reports 5. doi:ARTN893710.1038/srep08937.
URL <GotoISI>://WOS:000351137000005

[31] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, R. Jafari, Enabling effective programming and flexible management of efficient body sensor network applications, IEEE Transactions on Human-Machine Systems 43 (1) (2013) 115–133. doi:10.1109/TSMCC.2012.2215852.

[32] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges, Information Fusion 35 (2017) 68 – 80. doi:https://doi.org/10.1016/j.inffus.2016.09.005.
URL http://www.sciencedirect.com/science/article/pii/S156625351630077X

[33] M. Sandau, H. Koblauch, T. B. Moeslund, H. Aanaes, T. Alkjaer, E. B. Simonsen, Markerless motion capture can provide reliable 3d gait kinematics in the sagittal and frontal plane, Medical Engineering and Physics 36 (9) (2014) 1168–1175. doi:http://dx.doi.org/10.1016/j.medengphy.2014.07.007.
URL http://www.sciencedirect.com/science/article/pii/S1350453314001775

[34] E. Auvinet, F. Multon, J. Meunier, New lower-limb gait asymmetry indices based on a depth camera, Sensors 15 (3) (2015) 4605–4623. doi:10.3390/s150304605.
URL http://www.mdpi.com/1424-8220/15/3/4605

[35] B. Muller, W. Ilg, M. A. Giese, N. Ludolph, Validation of enhanced kinect sensor based motion capturing for gait assessment, PLOS ONE 12 (4) (2017) 1–18. doi:10.1371/journal.pone.0175813.
URL https://doi.org/10.1371/journal.pone.0175813

[36] M. Ye, C. Yang, V. Stankovic, L. Stankovic, A. Kerr, A depth camera motion analysis framework for tele-rehabilitation: Motion capture and person-centric kinematics analysis, IEEE Journal of Selected Topics in Signal Processing 10 (5) (2016) 877–887. doi:10.1109/JSTSP.2016.2559446.

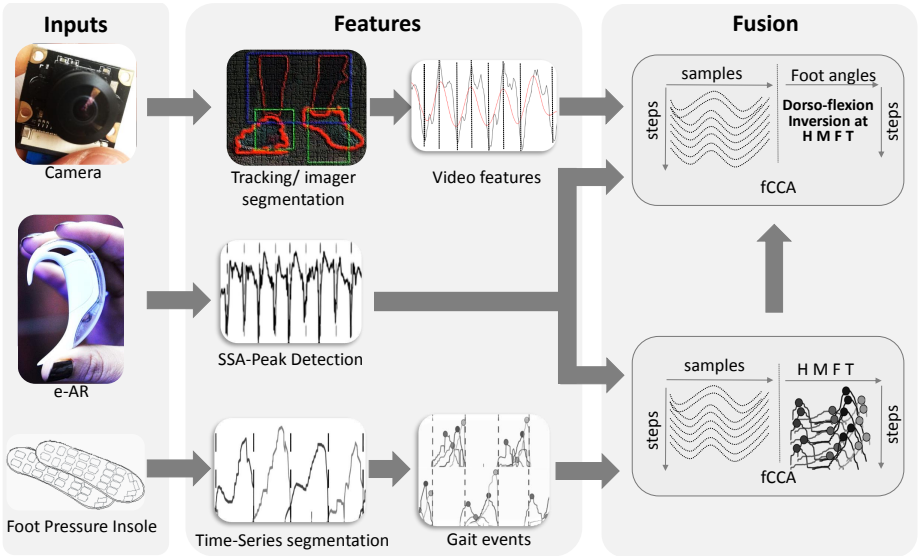


Figure 1. Data fusion framework for gait analysis. fCCA is used in a two steps approach to firstly model the timing distribution of GRFs that reflect key gait events such as when GRFs are maximized in plantar foot areas, such as the heel (H), midfoot (M), frontfoot (F) and toes (T). Subsequently, angular features derived from video tracking and image segmentation approaches are incorporated in another fCCA framework to model foot progression angle during these key gait events.

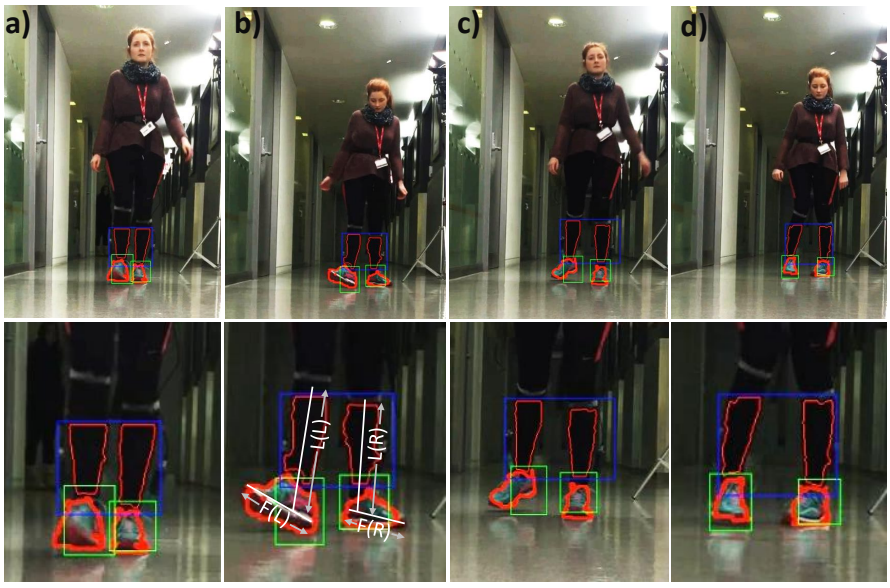
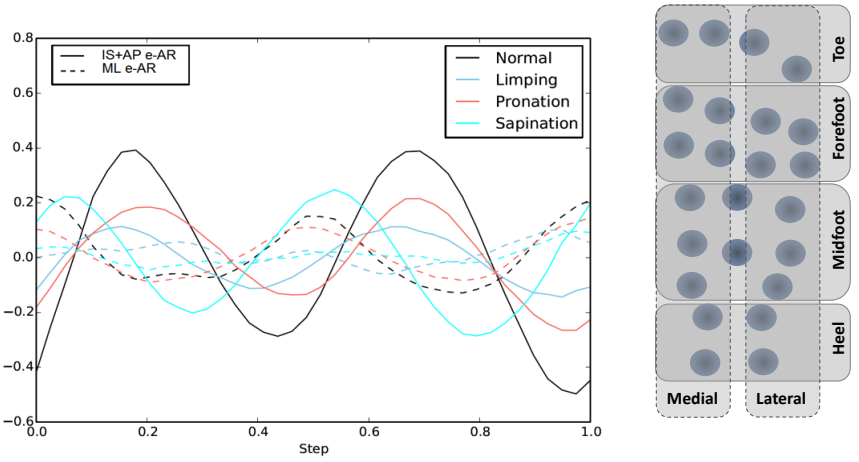


Figure 2. Examples of the lower limb visual tracking (blue), predicted foot regions (green), and leg and foot contours (red) obtained using K-means clustering and GrabCut are shown for the four types of gait considered in this paper; (a) normal, (b)supination, (c) limping and (d) pronation. A diagram of the fitted lines estimated based on image segmentation of the lower limbs has been overlayed on b) .



(a) e-AR training data (b) plantar foot subregions

Figure 3. a) Average e-AR signal across all subjects for normal walking, limping, walking with exaggerate pronation, and walking with exaggerated supination, respectively. The solid lines represent the summation of the e-AR signal along IS and AP after applying singular spectrum analysis . The dashed lines represent the e-AR signal in ML axis after the application of singular spectrum analysis. b) Plantar foot area subdivision into total eight areas. The plantar foot area is firstly subdivided into four regions across the anterios-posterior axis, namely Heel, Midffot, Frontfoot and Toes and in two regions across the medial-lateral axis, namely, Medial and Lateral.

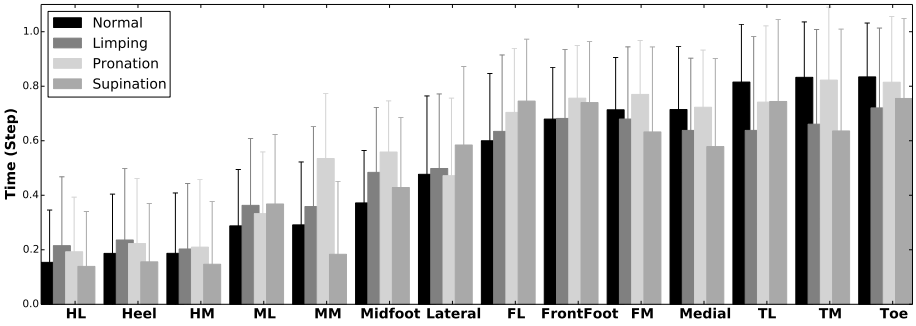
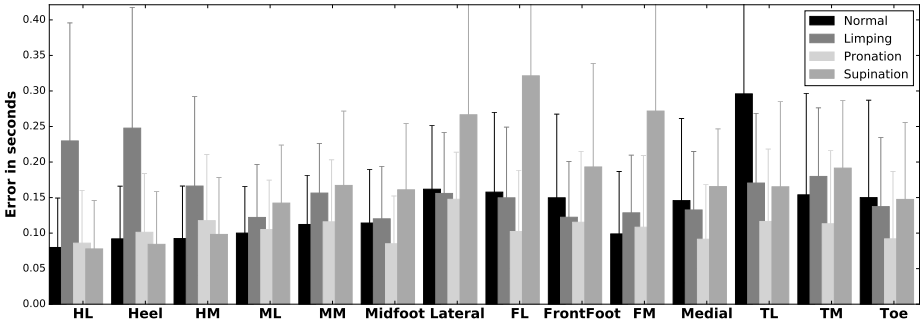
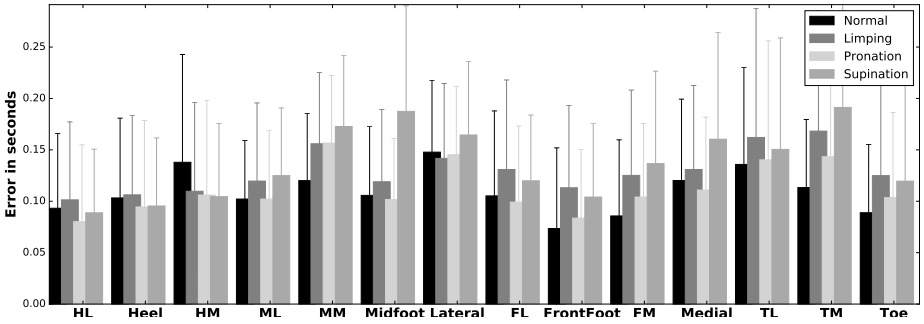


Figure 4. Plantar ground reaction forces timing distributions across all subjects for each conditions. Gait events have been identified based on insole data and they have been normalized with respect to each insole step. Sub-regions are defined as Heel-Lateral (HL), Heel-Medial (HM), Midfoot-Lateral (ML), Midfoot-Medial (MM), Frontfoot-Lateral (FL), FrontFoot-Medial (FM), Toe-Lateral (TL), Toe-Medial (TM) based on the subdivision of the plantar foot area shown in Figure 3b.

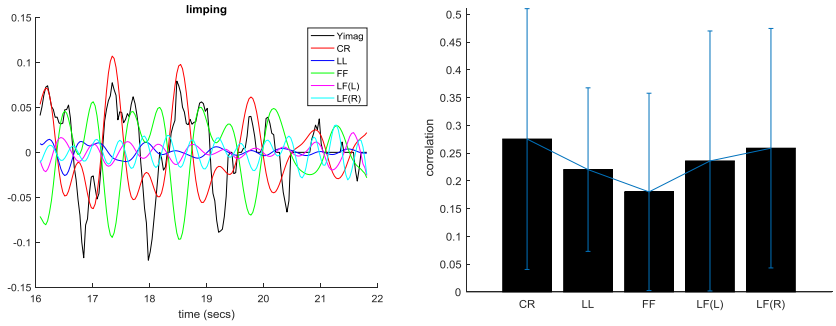


(a) Across subjects and within conditions



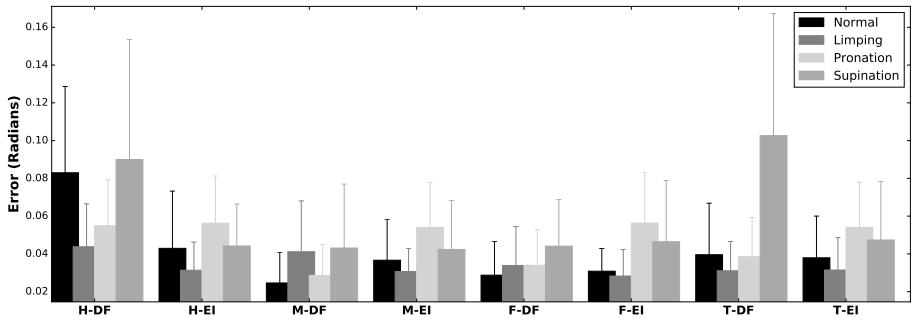
(b) Across subjects and across conditions

Figure 5. Leave-on-out cross validation results of the fCCA model in predicting the ground reaction forces timing distributions. a) Across subjects and within conditions: To form the training dataset, the data across all subjects and within each condition are concatenated. b) Across subjects and across conditions: Data across subjects and across conditions are concatenated.

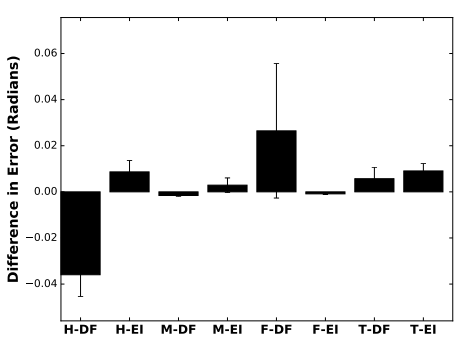


(a) Video features (b) Correlation

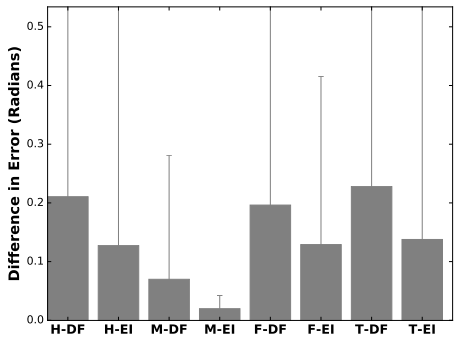
Figure 6. a) An example of the extracted features from the analysis of the video when the subject walks towards the camera. Yimag is the scale invariant distance in the vertical direction of the image plane between left and right ankle. LL is the dot product between the two legs, FF is the dot product between the feet, LF(R) is the dot product between the right leg and right foot and LF(L) is the dot product between the left leg and left foot. b)Absolute correlation values between Yimg, which reflects the distance between right and left feet, and the CR, LL, FF, LF(L) and LF(R), respectively.



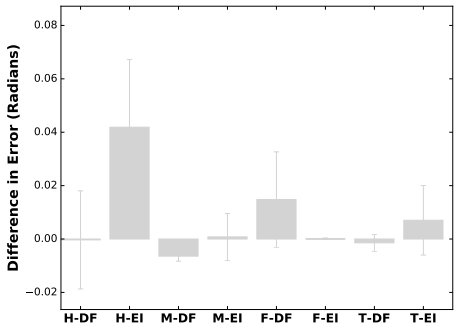
(a) e-AR & Visual Features prediction performance



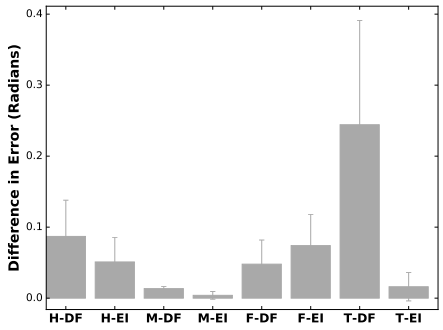
(b) Normal



(c) Limping



(d) Pronation



(e) Supination

Figure 7. The performance of fCCA in fusing the e-AR signal with the derived video features, namely CR, to model the foot angle progression. a) Error in radians based on leave-one-out cross validation of the proposed method. The fCCA model estimates the dorsiflexion (DF) and the aversion /inversion (EI) of the ankle when ground reaction forces at the Heel (H), Mid-foot (M), Front-foot (F) and Toes (T) are maximized. Therefore, we show the mean errors and standard deviation of estimating DF and EI at Heel, Mid-foot, Front-foot and Toes contact with the ground as H-DF, H-EI, M-DF, M-EI, F-DF, F-EI, T-DF, T-EI, respectively. b-e) Differences of angular error between a model that uses only e-AR to predict foot angle progression and the proposed method.