

Accepted Manuscript

Estimation and Validation of Temporal Gait Features using a Markerless 2D Video System

Tanmay T. Verlekar , Henri De Vroey , Kurt Claeys , Hans Hallez ,
Luís D. Soares , Paulo L. Correia

PII: S0169-2607(18)31040-X
DOI: <https://doi.org/10.1016/j.cmpb.2019.04.002>
Reference: COMM 4879



To appear in: *Computer Methods and Programs in Biomedicine*

Received date: 16 July 2018
Revised date: 12 March 2019
Accepted date: 1 April 2019

Please cite this article as: Tanmay T. Verlekar , Henri De Vroey , Kurt Claeys , Hans Hallez , Luís D. Soares , Paulo L. Correia , Estimation and Validation of Temporal Gait Features using a Markerless 2D Video System, *Computer Methods and Programs in Biomedicine* (2019), doi: <https://doi.org/10.1016/j.cmpb.2019.04.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- The proposed system estimates the biomedical gait indicators and the temporal gait features with a high level of accuracy, outperforming the current state-of-the-art markerless vision systems
- The system is robust to imperfections of the segmented silhouettes.
- The results of the proposed system are validated by comparing them with the “gold standard” optoelectronic motion capture system.
- The high correlation with the results of the optoelectronic motion capture system suggests that the proposed system can be an alternative to marker based systems in non-laboratory environments

ACCEPTED MANUSCRIPT

Index Terms—Biomedical Analysis, Biomedical Gait Indicators, Gait Analysis, Temporal Gait Features, Computer

Estimation and Validation of Temporal Gait Features using a Markerless 2D Video System

Tanmay T. Verlekar^a, Henri De Vroey^b, Kurt Claeys^b, Hans Hallez^b, Luís D. Soares^c, Paulo L. Correia^a

^a Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, Portugal.

^b KU Leuven, Campus Brugge, Belgium.

^c Instituto de Telecomunicações, Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal. (correspondence e-mail: tanmay.verlekar@lx.it.pt).

Abstract—

Background and Objective: Estimation of temporal gait features, such as stance time, swing time and gait cycle time, can be used for clinical evaluations of various patient groups having gait pathologies, such as Parkinson's diseases, neuropathy, hemiplegia and diplegia. Most clinical laboratories employ an optoelectronic motion capture system to acquire such features. However, the operation of these systems requires specially trained operators, a controlled environment and attaching reflective markers to the patient's body. To allow the estimation of the same features in a daily life setting, this paper presents a novel vision based system whose operation does not require the presence of skilled technicians or markers and uses a single 2D camera.

Method: The proposed system takes as input a 2D video, computes the silhouettes of the walking person, and then estimates key biomedical gait indicators, such as the initial foot contact with the ground and the toe off instants, from which several other temporal gait features can be derived.

Results: The proposed system is tested on two datasets: (i) a public gait dataset made available by CASIA, which contains 20 users, with 4 sequences per user; and (ii) a dataset acquired simultaneously by a marker-based optoelectronic motion capture system and a simple 2D video camera, containing 10 users, with 5 sequences per user. For the CASIA gait dataset A the relevant temporal biomedical gait indicators were manually annotated, and the proposed automated video analysis system achieved an accuracy of 99% on their identification. It was able to obtain accurate estimations even on segmented silhouettes where, the state-of-the-art markerless 2D video based systems fail. For the second database, the temporal features obtained by the proposed system achieved an average intra-class correlation coefficient of 0.86, when compared to the "gold standard" optoelectronic motion capture system.

Conclusions: The proposed markerless 2D video based system can be used to evaluate patients' gait without requiring the usage of complex laboratory settings and without the need for physical attachment of sensors/markers to the patients. The good accuracy of the results obtained suggests that the proposed system can be used as an alternative to the optoelectronic motion capture system in non-laboratory environments, which can be enable more regular clinical evaluations.

Vision, Video Processing.

I. INTRODUCTION

GAIT analysis is widely used in the functional evaluation of different patient groups. These patient groups vary from people suffering from local lower limb injuries (e.g., knee injuries) to patients with severe systemic disorders [1]. To perform such evaluation, key instants during the user's gait cycle must be identified, which are denoted as biomedical gait indicators (BGIs). A gait cycle typically starts with a BGI called the "initial contact", which indicates the initial contact of the heel of the evaluated leg with the floor. The initial contact also marks the end of the previous gait cycle. A gait cycle can be further subdivided into stance and swing phases, which are separated by another BGI called the "toe off" of the evaluated leg. The succession of recurrent left leg and right leg movements enables the computation of temporal gait features for each leg, such as stance time, swing time, cycle time and other gait cycle related timings.

Clinical gait analysis using temporal gait features, such as those mentioned above, has gained popularity for instance for the evaluation of rehabilitation progression in patients after a surgical procedure (e.g., after a knee replacement) [2], or in the prediction of fall risks in the elderly population [3]. However, the acquisition and evaluation of the features from which BGIs are derived is usually performed in dedicated laboratories, often resulting in an expensive and a time-consuming task. Therefore, a novel system for acquisition and evaluation of gait features in a daily life setting, such as a senior residence, hospital or home, is needed.

A. State-of-the-art

Devices used to acquire biomedical gait indicators of a user can be classified into sensor based and vision based systems [4]. Sensor based systems use devices that directly record signals representing the motion of the user. Such devices can be setup on the floor, such as force plates [5], or attached to the user for acquisition of signals. They allow the acquisition

of features such as velocity, cadence, step length and step time, to evaluate the user's gait. Such features can also be acquired using body worn sensors, as gyroscopes [6] and accelerometers [7]. Among them, wearable sensors have become popular in non-laboratory environments, due to their ease of usage [7]. However, signal acquisition can only be performed after the user is setup with the sensors, which in some cases are time consuming and may require operation by trained professionals, thus limiting their usage in daily life settings.

Vision based systems on the other hand rely on the use of video cameras to acquire gait related features from the user. Such systems can be further divided into model based and appearance based, depending on the process of acquisition and the type of information used [8]. Model based systems typically rely on the use of multiple 2D cameras, operating in the visible or infra-red spectra, possibly also in combination with depth sensing cameras, along with additional information about the scene geometry to model the gait of the user.

Among the model based systems, a large number rely on the use of depth sensing cameras to acquire the user's gait. Such systems typically compute a model of the user's skeletal structure to derive features such as leg length, normalized average stride length, and gait velocity [9] [10], or motion history [11], to classify the gait as either normal or abnormal. In the absence of depth information, calibrated cameras can be used to obtain features, such as the height of the user and the distance between the feet [12], or joint angles [13], to differentiate between normal and abnormal gait. Some systems use a combination of calibrated cameras and depth sensing cameras [14] to compute stride to stride variations and predict fall risks. It is worth mentioning that some depth sensing cameras have a limited range of operation, typically between 80 cm and 4 m, which can significantly restrict the conditions for gait acquisition.

The gold standard for clinical evaluations in laboratories corresponds to a model based optoelectronic motion capture system [15]. The system uses at least six calibrated infra-red (IR) cameras in combination with forty-four reflective markers that need to be set on the user, to obtain kinematic features from the joints and thus characterize the observed gait. The use of such systems is limited to laboratories, due to the required equipment, its setup, the need for calibration before use and professional knowledge to operate them.

On the other hand, appearance based systems typically rely on silhouettes obtained from a single 2D camera, from which spatiotemporal information about the user's gait is gathered. A significant amount of work using appearance based systems has been performed in the context of biometric recognition in unconstrained environments, such as those reported in [16], [17], as they do not require the use of reflective markers.

According to [18], major articulations during a gait cycle occur in the sagittal plane. Thus, some appearance based systems explore the lateral view of the user for the identification of different gait related pathologies. These

systems acquire features, such as stride length, leg angles, gait cycle time, speed [19], or the gait energy image (GEI) [20], to perform classification of pathologies, such as Parkinson disease, neuropathy, hemiplegia and diplegia. Although the lateral view is popular, the frontal view can also be used to extract features, such as stride length and leg angles [21], or homeomorphisms between 2D lattices and binary frontal view silhouettes [22] which can be used to perform abnormal gait detection.

Some appearance based systems estimate BGIs, such as initial contact and the toe off [23], or the stance feet/flat feet [24], using information acquired from the feet position. These BGIs can then be used as a starting point to compute additional features, such as step and stride length, cadence, or the duration of single and double support phases. Features such as the orientation of the user's body along with ramp angle [25], axial ratio, change in velocity and distance between legs [26] may be used to detect posture instabilities, such as a hunched back.

The detection of abnormalities in a user's gait in most appearance based systems is performed using a single 2D camera, making these systems easier to install and operate, which is an advantage that can be exploited in daily life settings. However, most state-of-the-art appearance based systems focus on classification of pathologies and not on the accuracy of the acquired features. For clinical evaluations, the accuracy of the features is extremely important. Thus, the results from such systems have to be validated, before accepting them for clinical use.

B. Motivation and contribution

Most medical professionals that analyze a user's gait acquire BGIs either using optoelectronic motion capture systems [15], or force plate based systems [5]. These systems are complex to handle and require trained personnel for operation. They are also expensive and inaccessible to most users. Moreover, processing data derived from such systems can also be a challenging and time-consuming task. In addition, the optoelectronic system [15] involves the application of reflective markers on to the skin of the user. This process requires an accurate application of the markers, as the system is very sensitive to noise and lighting conditions, and it can be tiring and uncomfortable for the users. Consequently, the use of such systems for the assessment of the user's gait on a regular basis may not be the easiest undertaking [6].

The use of inertial sensors can be used to tackle the identified shortcomings of the optoelectronic motion capture system [15], as many studies have demonstrated their validity in the estimation and assessment of gait features [3]. However, these sensors also need to be mounted on distinct and specific locations of the users' body when one wishes to monitor their gait. Thus, they too are limited in their usage in daily life settings.

In the state-of-the-art review, many systems that use video from a single 2D camera, without reflective markers to identify abnormal gait were presented. These systems can possibly yield a solution to the aforementioned limitations.

However, most of these systems focus on classification of pathologies, and only a few systems among them explore temporal gait features. The accuracy of the temporal gait features acquired by such systems is affected by the limitations in moving silhouettes' segmentation. Since clinical evaluations require accurate estimation of gait features, a novel system robust to such conditions is needed.

This paper presents a novel markerless 2D video based system to estimate two BGIs: initial contact and toe off. These estimates can be used to compute temporal gait features, such as the time intervals of the various phases of the user's gait, with a higher accuracy than the state-of-the-art. To validate the results, and emphasize the possibility to use them in clinical evaluations, the paper presents a comparison between the estimation of the gait features using the proposed system and the optoelectronic motion capture system [15]. That comparison can be made by computing the intra class correlation coefficient, to verify the level of agreement between the two systems [27]. The obtained intra class correlation results are in the range from 'good' to 'excellent', suggesting that the proposed system can be a valid alternative to the optoelectronic system [15] in non-laboratory environments.

The remainder of the paper is organized as follows. Section II presents the proposed system, with the corresponding experimental results being reported in Section III. Section IV provides conclusions and directions for future work.

II. METHOD

The aim of the proposed system is to identify key BGIs, such as the initial contact and the toe off, by providing an estimate of the video frame number in which they occur. The state-of-the-art 2D vision based systems, such as those presented in [19] [23], typically perform poorly in the presence of imperfectly segmented silhouettes, often ignore the BGIs occurring at the start and end of a video. To address these limitations this paper presents a robust system that computes temporal gait features by estimating BGIs, following the architecture presented in Fig. 1.

The proposed system operates on silhouettes of the walking user to estimate the BGIs. The preprocessing module is used to obtain a set of silhouettes of the moving user from the 2D input video. The proposed system uses the mixture of Gaussians (MoG) foreground detection method presented in [28] to separate the dynamic part of the video sequence (user

from the static part (background), resulting in a foreground mask corresponding to the user silhouette. Morphological operators, such as dilation and erosion, are then applied to the silhouettes to fill holes and remove small isolated blobs.

Using the silhouettes obtained from the foreground detection, the proposed system performs BGI estimation in three blocks:

- **Flat feet identification** – The flat feet identification block selects the areas in the image where the user's feet are in complete contact with the floor. These regions are highlighted by computing an overlap between the silhouettes obtained from all the video frames, as illustrated in Fig. 2 (a). The locations of the flat feet are then obtained by applying a dynamic thresholding scheme to the foot regions of the overlapped image. The locations thus obtained can be used by the following block to estimate the BGIs.

- **Selection of candidate frames for initial contact and toe off** – The initial contact and toe off occur at specific points during a gait cycle. It is therefore possible to narrow down the search for the BGIs around these points, to prevent false positive detections. The selection step analyzes the width of the silhouettes belonging to a video sequence, selecting two sets of frames: (i) a set containing the candidate frames for initial contact, and (ii) a set containing candidate frames for toe off.

- **Initial contact and toe off estimation** – The third block uses the location where flat feet were detected and the sets of initial contact and toe off candidate frames, to estimate the BGIs. Since in each candidate frame the two feet are present, this step selects the appropriate foot at which the BGI occurs. The rate of change in the overlap between each flat foot and the (moving) candidate foot are analyzed to estimate the two BGIs. Gait features can then be computed using the BGIs.

The following subsections provide details for each of these blocks.

A. Flat feet identification

To estimate the initial contact and toe off, the proposed system needs to identify the locations where the feet are in complete contact with the floor. For this purpose, a "flat feet image" is created, capturing the corresponding locations in the video sequence.

The first step of the proposed system is the construction of a gait texture image (GTI), which is obtained by averaging the

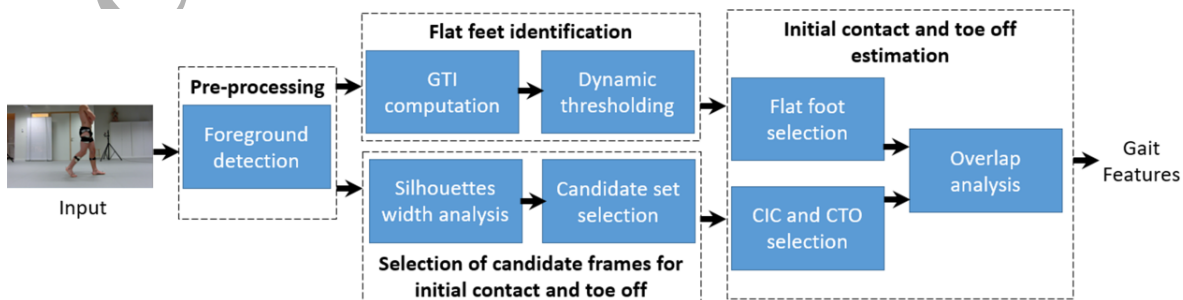


Fig. 1. System architecture for a 2D video based biomedical gait indicators estimation.

user silhouette images computed for each video frame, in their original positions, as illustrated in Fig. 2 (a). Thus, given T binary silhouette images, $I(x, y, t)$, where t represents the frame number, the $GTI(x, y)$ can be constructed by averaging the silhouettes according to Equation (1), where (x, y) represents the coordinates of a given pixel's position.

$$GTI(x, y) = \frac{1}{T} \sum_{t=1}^T I(x, y, t) \quad (1)$$

The GTI highlights the overlap between consecutive user silhouettes. Due to the nature of human gait, the highest overlap is observed at the locations that remain static for a longer period, i.e., where the user's feet are in full contact with the floor. These locations are represented by the highest intensity values in the GTI. The proposed system identifies these flat feet location by applying a dynamic thresholding to the GTI.

The proposed thresholding scheme prevents the selection of other body parts by considering only the bottom 10% of the user's body in the GTI, which contains the user's feet positions, according to the human anatomy [29] – see Fig. 2 (b). Since, the flat feet locations are represented by the highest intensity values in the GTI, threshold values are selected to highlight these locations. To select the threshold value for each flat foot location, the system selects the highest intensity values along the y-axis of the GTI (considering only the silhouette's bottom 10%) for every x-axis position– see Fig. 2 (c). As illustrated in Fig. 2 (c), the flat feet locations at the start and the end of the GTI may contain significantly lower intensity values, as in these areas of the GTI there is less silhouette overlap. The variation in intensity values may also be caused due to asymmetric walking patterns, where one foot of the user stays in contact with the floor significantly longer than the other, thus causing more silhouette overlaps at that foot positions in the GTI. To overcome this variability, the proposed system uses the valleys in the intensity plot to separate two neighboring flat feet positions, followed by the application of Otsu's thresholding [30] to each identified flat foot location. The resulting image contains the flat feet locations for a given video sequence and is called the “flat feet image” – see Fig. 2 (d).

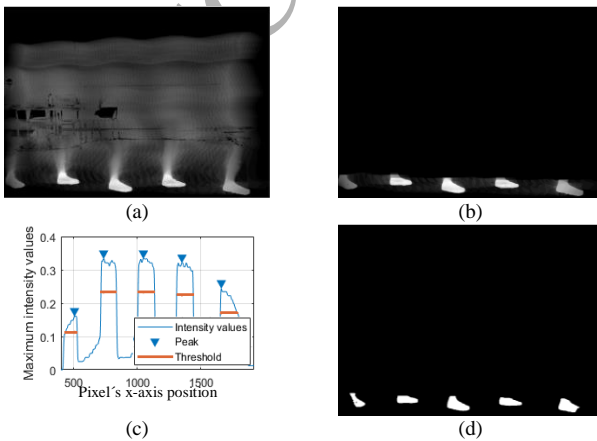


Fig. 2. Illustration of the proposed flat feet identification step: (a) GTI, (b) selected feet region of the GTI, (c) dynamic thresholding scheme, (d) resulting “flat feet image”.

B. Selection of candidate frames for initial contact and toe off

After flat feet detection, the proposed system can analyze the overlap between the flat feet image and every video frame to estimate the BGIs. However, the resulting overlaps can occur for both feet of the user, possibly causing several false positives. The number of false positives can be reduced by first identifying the set of frames in which each of the BGIs are most likely to occur. The proposed system reduces its search range by selecting two sets of frames near every double support phase, as candidates for initial contact and toe off, respectively. The selection is done by analyzing the distance between the feet of a user for the entire video sequence. Only the frames with the distance greater than the mean distance are selected as the candidates for the two sets. The two sets are then created using the midpoint of the double support phase.

The double support phase is the part of the gait cycle where both feet of the user are in contact with the floor. The initial contact BGI marks the start of double support phase, when a user's foot that is advancing, (i.e., the front foot) first touches the floor. The toe off BGI marks the end of the double support, when the user's standing foot, (i.e., the back foot) leaves the floor. In-between the two BGIs is the midpoint of the double support phase, corresponding to the point when the user's legs are the furthest apart from each other.

The proposed system identifies the midpoint of the double support phase by analyzing the distance between the feet, which can be approximated as the width of the silhouette's bounding box, which increases and decreases periodically during a gait cycle. However, the width of the bounding box may vary significantly as the user walks towards or away from the camera. Thus, the available width values must be normalized by subtracting them by their mean and dividing by their standard deviation. The points of maximum width then correspond to the local peaks of the normalized width, representing the midpoints of the double support phase. All frames with a positive normalized width occurring immediately before the midpoints of the double support phase (i.e., before the peak) are considered candidate frames for the initial contact. Similarly, the frames with a positive normalized width immediately after the peak are considered candidate frames for the toe off. These candidate frames are indicated in Fig. 3 by ‘*’ and ‘o’, respectively.

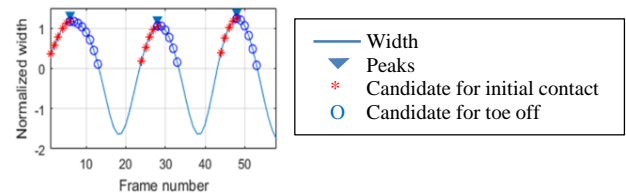


Fig. 3. Selection of candidate frames for the estimation of initial contact and toe off BGIs.

C. Initial contact and toe off estimation

The initial contact occurs at the user's front foot, while the

toe off occurs at the user's back foot. Thus, the proposed system selects only the front foot from every candidate frame of the initial contact set as a candidate for initial contact (CIC), and the back foot from every candidate frame of the toe off set as a candidate for toe off (CTO), as illustrated in Fig. 4 (a), (b).

As seen in Fig. 3, for a given video sequence, multiple peaks can be detected. For every peak i , a set of CICs (CIC_i^n) and CTOs (CTO_i^m) can be obtained, where n, m are the number of candidates in each set. The estimation of the frame numbers where a BGI occurs can be performed by analyzing the overlap between a foot from the flat feet image and every foot from the set of CICs and CTOs, respectively. For a set CTO_i^m there should be a toe off BGI that can be identified from the overlap of a CTO with the selected flat foot. Similarly, for a set CIC_i^n there should be an initial contact BGI that can be identified from the overlap of a CIC with the other selected flat foot – see Fig. 4 (d).

As illustrated in Fig. 4 (a), (c) for k peaks, there exist k flat feet in the flat feet image. Thus, the selection of the flat foot and the corresponding set of CICs and CTOs, to estimate the BGIs can be done as follows. For CIC_i^n and CTO_i^m corresponding to peak i , the flat foot i from the flat feet image is selected for toe off, and flat foot $i + 1$ is selected for initial contact. As an example, considering Fig. 4, for CIC_2^n and CTO_2^m corresponding to the peak 2, flat foot 2 is selected for toe off and flat foot 3 is selected for the initial contact. The overlap between the flat foot 3 and CIC_2^1, CIC_2^6 and CIC_2^{10} and the flat foot 2 and CTO_2^1, CTO_2^5 and CTO_2^9 are illustrated in Fig. 4 (d), highlighting the instants where the BGIs occur. Also, in Fig. 4 (d) white represents the amount of overlap, while the gray areas represent the remaining parts of the two feet.

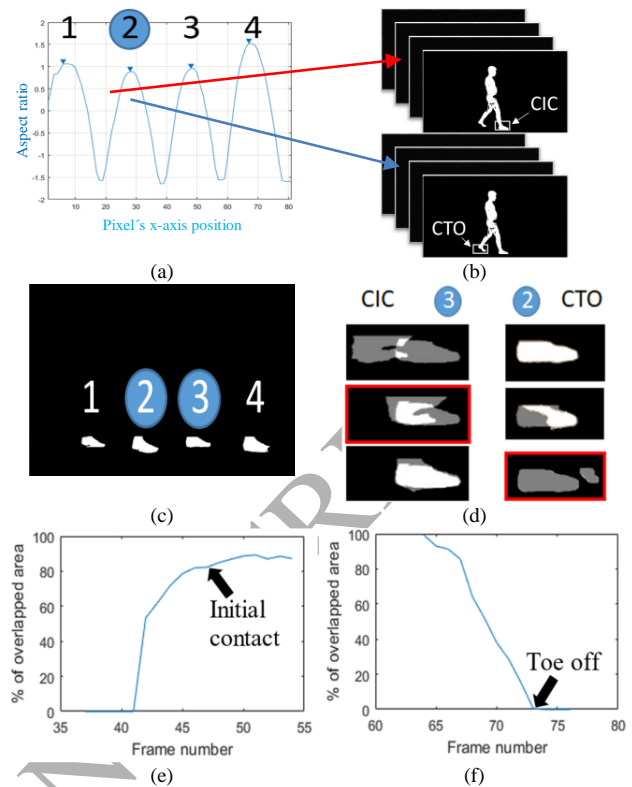


Fig. 4. Examples of intermediate steps in BGIs estimation: (a), (b) selection of candidate foot for initial contact (CIC) and candidate for toe off (CTO), (c) flat feet selection, (d) overlap between flat foot and CIC/CTO feet, estimation of: (e) Initial contact (f) toe off.

For the selected set of CICs/CTOs and the corresponding flat foot, the overlap can be measured as the percentage of CIC/CTO foot covered by the selected flat foot. As illustrated in Fig. 4 (e), for the initial contact, the percentage of overlapped area increases rapidly at the start and then slows down at the end. Thus, the initial contact can be estimated as the point where the increase in the overlapped area slows down significantly. It is observed that such slow down occurs when almost 80% of the CIC is overlapped. Thus, the proposed system estimates the initial contact as the first frame for which the overlap of the CIC and the flat foot exceeds 80%. Similarly, the toe off can be estimated by analyzing the overlap between the feet from the set of CTOs and the selected flat foot. The toe off can be estimated as the first frame for which the overlap between the CTO and the selected flat foot becomes zero – see Fig. 4 (f).

D. Computation of gait features

The result from the previous step is a sequence of frame numbers indicating the estimated initial contact and the toe off BGIs. To have a better understanding of the gait of the user, frame numbers can be converted into timestamps using the frame rate of the video sequence. The proposed system uses the timestamps to estimate six different gait features. The left and right stance times can be computed as the time from the initial contact to the toe off instant of the same leg, following Equation (2). The left and right swing times can be computed as the time from the toe off instant to the initial contact of the same leg, following Equation (3). Finally, the left and right

gait cycle times can be computed as the time interval from one initial contact to next the initial contact of the same leg, following Equation (4). These features can provide an insight about the symmetry of the user's gait.

$$\text{Stance time} = TO_i - IC_i \quad (2)$$

$$\text{Swing time} = IC_{i+1} - TO_i \quad (3)$$

$$\text{Gait cycle time} = IC_{i+1} - IC_i \quad (4)$$

III. RESULTS

The proposed system is tested using the CASIA gait dataset A, collected by the Institute of Automation of the Chinese Academy of Sciences [31], and the KU Leuven (KUL) gait dataset collected by the Department of Rehabilitation Sciences, KU Leuven, Belgium.

The CASIA dataset A consists of twenty users captured in three walking directions (0° , 45° , 90°) with respect to the camera, with each user having been recorded four times in each direction.

The KUL dataset consists of ten users, each recorded five times, using a single Casio Exilim EX-ZR100 camera. Each user is recorded in a lateral walking direction with respect to the camera, which records with a resolution of 1920×1080 and a frame rate of 30 fps. Each user is simultaneously recorded using a camera based optoelectronic motion capture system, consisting of six infrared Optitrack Flex 13 cameras with a resolution of 1280×1024 and a maximal frame rate of 120 fps. The user's gait is captured using forty-four reflective markers secured at different locations of the user's body, according to the lower limb and trunk model [15]. The six synchronized infrared cameras are synchronized, calibrated and operated using Motive Tracker v.01.90 to track reflective markers in a predefined 3D space. A pre-trial calibration is performed using a standardized procedure. A calibration is considered acceptable if the overall reconstruction errors of the marker trajectories are kept below 0.1 mm. Post-processing of marker data was performed in a dedicated software called V3D, by C-motion Inc [32]. Body segments are defined as rigid bodies, interconnected by joints with predefined degrees of freedom.

A. BGI estimation using CASIA dataset A

The first test is conducted using the CASIA dataset A to test the accuracy of the proposed system and compare it against the state-of-the-art. Following the setup presented in [19], the proposed system is tested using the 0° , lateral view sequences. Before the test, each frame of the dataset is checked manually to construct the ground truth. Considering the camera frame rate of 30 fps, and since normal gait cycles usually last for 1-2 seconds, it is observed that in several cases the exact point of initial contact or toe off are not recorded. Thus, the ground truth is constructed considering the frame just before or the frame just after the BGI. The decision is made by visually observing the BGI nearby frames. Therefore, for evaluation purposes, systems are allowed an error margin of ± 1 frame.

To analyze the significance of the error margin, a test is conducted by increasing the error margin for the proposed system from 0 to 2 in integer steps. With an error margin of 0, the proposed system performs a 72% correct estimation of initial contact and 88% correct estimation of toe off – see Fig. 5. In the results, the toe off estimation is significantly better than for initial contact, as it is very easy to detect zero overlap between the flat foot and the CTOs using the proposed system.

When allowing an error margin of ± 1 frame, both results improve significantly, to 99%. These results are significantly better than the state-of-the-art which adopts an error margin of ± 2 frames. Using that error margin for the proposed system, results are further improved to almost 100%, as shown in Table I. A root mean square error (RMSE) is also computed using the frame numbers estimated by the proposed system and the ground truth, with the proposed system performing significantly better than the state-of-the-art. It should also be noted that the methods presented in [19] and [23] ignore the BGIs occurring at the beginning and the end of the gait sequence due to incomplete/segmented silhouettes. The results presented in Table I report these ignored BGIs as a part of the failed detections. The proposed method, unlike the state-of-the-art, can operate even on such silhouettes, with a fail detection rate of 0.005 - see Table I.

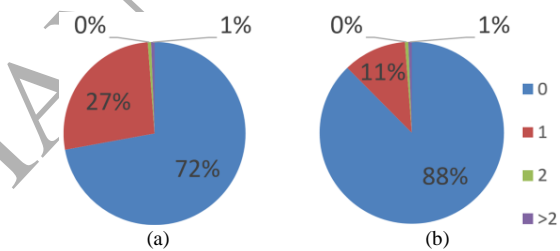


Fig. 5. Percentage of correctly estimated BGIs w.r.t. error margin for the proposed system: (a) Initial contact, (b) Toe off.

TABLE I
BGI ESTIMATION WITH AN ERROR MARGIN OF ± 2 FRAMES

	Method [23]		Method [19]		Proposed System	
	Initial contact	Toe off	Initial contact	Toe off	Initial contact	Toe off
# Correct Estimations	316	315	319	305	332	330
# False Estimations	8	5	9	8	0	1
# Failed Detections	10	12	6	19	2	1
RMSE	0.98	0.95	0.89	1.62	0.54	0.42

B. BGIs estimation using KU Leuven database

A second test is conducted to compare the performance of the proposed system against the gold standard optoelectronic motion capture system [15], using the KUL database. The left and right stance, swing and cycle times are derived for the optoelectronic system using the marker and coordinate based algorithm presented in [15]. The optoelectronic system results are considered the gold standard for clinical assessment of a user's gait. The same features are also acquired using the proposed system, from the 2D sequences included in the KUL database. It is observed that using the error margin of ± 1 frames, the

proposed system correctly estimates the BGIs with 99% accuracy. It should be noted that the proposed system attempts to perform the same type of assessment as the optoelectronic system [15], but without any calibrations or initial user setup and with a single camera. The single camera also operates at approximately one fourth the framerate of the optoelectronic system [15].

Table II reports the obtained results, where “L” and “R” represent the left and right legs, respectively. The agreement between the two methods is calculated using intra class correlation coefficients (ICC), which describes how strongly values in the two groups of results resemble each other. Notice that according to Fleiss and Cohen [27] an ICC in the ranges 0.00 – 0.39 is considered ‘bad’, 0.40 – 0.73 ‘moderate’, 0.74 – 0.90 ‘good’ and 0.91 – 1.00 ‘excellent’. Statistical analysis was performed using SPSS [33].

As seen in Table II, the ‘good’ to ‘excellent’ ICC values suggest a high level of agreement between the results of the proposed 2D video system and the optoelectronic system [15]. For the left and right gait cycle times, a small proportional bias (the difference between the means obtained by the two systems) of 0.02s is observed in Table II. This indicates that the proposed system can estimate initial contact with a high level of accuracy. A slightly higher bias is found when evaluating the left and right swing and stance times. However, a good correlation can still be observed between the two systems. The SPSS system also provides a p-value to indicate the reliability of the observed correlation. A score of less than 0.001 for the p-value of every entry in Table II also suggests a very strong evidence for the observed correlation.

From the obtained results it can be concluded that the proposed system is a viable alternative to the optoelectronic motion capture system [15] that can be considered for usage in non-laboratory environments, thus allowing a more frequent follow up of patients in between the more rigorous analysis done when coming to the laboratory where the optoelectronic motion capture is installed.

TABLE II
COMPARISON BETWEEN THE PROPOSED AND OPTOELECTRONIC SYSTEM

	Proposed System		Optoelectronic System [15]		ICC	Mean Diff	SD (sec)	RMSE (sec)
	Mean (sec)	SD (sec)	Mean (sec)	SD (sec)				
Stance Time L	0,77	0,08	0,72	0,07	0,85	0,05	0,06	0,06
Stance Time R	0,80	0,09	0,72	0,08	0,83	0,08	0,06	0,09
Swing Time L	0,42	0,05	0,47	0,05	0,81	-0,05	0,04	0,05
Swing Time R	0,45	0,05	0,47	0,06	0,83	-0,03	0,04	0,04
Cycle Time L	1,22	0,13	1,20	0,11	0,92	0,02	0,07	0,05
Cycle Time R	1,21	0,12	1,20	0,12	0,92	0,02	0,06	0,05

Overall p-value <0.001

IV. DISCUSSION

This paper presents a novel system to estimate temporal gait features using biomedical gait indicators, such as stance, swing and cycle times. The system achieved 99% correct BGI estimations, when considering an error margin of ± 1 frame. This contrasts with the currently available state-of-the art 2D vision-based systems [19] [23], which typically estimate the BGIs by performing gradient analysis on the difference between heel and toe positions in consecutive frames. To operate effectively, such systems require an accurate estimation of the user’s heel and toe positions. Therefore, imperfectly segmented silhouettes can significantly hamper their performance, leading to false and/or missed detections. Also, these systems often ignore the BGIs occurring at the start and end of a video, thus requiring long sequences, with multiple gait cycles, to perform a reliable BGI estimation.

The ‘good’ to ‘excellent’ ICCs found in the validation of the proposed markerless 2D video system in comparison to the gold standard optoelectronic motion capture system opens some new possibilities for the evaluation and management of different patient populations, in which the acquisition of spatiotemporal features of a user’s gait play an important role. The proposed system identifies the different temporal features of the user’s gait in a very reliable way using a single 2D camera, not requiring any special setup. This enables clinicians to acquire these features without the use of expensive laboratory equipment. The acquisition of the BGIs in a private practice or at home enables gait analysis in a more natural/unconstrained environment, which is preferred over a laboratory setting. The proposed system is easy to use and can be implemented even by professionals without technical background or education.

A limitation of the proposed system is that it currently estimates only temporal gait features. The optoelectronic and other sensor based systems can compute spatial and spatiotemporal features, such as stride length, step length, joint angles, joint angle velocity and acceleration. Such features can provide a greater insight into the user’s health. Thus, future work will include further improving the proposed system to also acquire spatiotemporal features. The proposed system is currently tested on normal gait. Some pathologies such as, Parkinson’s disease may cause self-occlusions, thus affecting the performance of the system. Thus, future work will also include testing and improving the system for the analysis of pathological gait. The BGIs acquired by the proposed system can also be used to analyze the fractal properties of step-to-step fluctuations [34], which can be related to higher nervous centers that control walking rhythm [35]. Since, the system is not bound to a laboratory environment, it allows collection of large amounts of BGIs, which is necessary to perform fractal analysis on gait dynamics.

ACKNOWLEDGEMENTS

This work was partially supported by Instituto de Telecomunicações under Fundação para a Ciência e

Tecnologia Grant UID/EEA/50008/2013 and by the European Regional Development Fund, project 1047.

The authors declare that this manuscript has not been published elsewhere; and do not have any conflicts of interest.

REFERENCES

- [1] J. Jellish, J. J. Abbas, T. M. Ingalls, P. Mahant, J. Samanta, M. C. Ospina, N. Krishnamurthi. (2015). A system for real-time feedback to improve gait and posture in Parkinson's disease. *IEEE journal of biomedical and health informatics*, 19(6), pp. 1809-1819.
- [2] W. Wang, D. Ackland, J. McClelland, K. Webster, S. Halgamuge. (2018). Assessment of Gait Characteristics in Total Knee Arthroplasty Patients Using a Hierarchical Partial Least Squares Method. *IEEE journal of biomedical and health informatics*, 22(1), pp. 205-214.
- [3] K. Wang, K. Delbaere, M. A. Brodie, N. H. Lovell, L. Kark, S. R. Lord, S. Redmond. (2017). Differences between gait on stairs and flat surfaces in relation to fall risk and future falls. *IEEE journal of biomedical and health informatics*, 21(6), pp. 1479-1486.
- [4] D. Gafurov. (2007). A survey of biometric gait recognition: Approaches, security and challenges. Presented at annual Norwegian computer science conference.
- [5] D. Slijepcevic, M. Zeppelzauer, A. Gorgas, C. Schwab, M. Schüller, A. Baca, B. Horsak. (2017). Automatic Classification of Functional Gait Disorders. *IEEE Journal of Biomedical and Health Informatics*.
- [6] B. Greene, D. McGrath, R. O'Neill, K. O'Donovan, A. Burns, B. Caulfield. (2010). An adaptive gyroscope-based algorithm for temporal gait analysis. *Medical & biological engineering & computing*, 48(12), pp. 1251-1260.
- [7] J. Wang, C. Lin, Y. Yang and Y. Ho. (2012). Walking pattern classification and walking distance estimation algorithms using gait phase information. *IEEE Transactions on Biomedical Engineering*, 59(10), pp. 2884-2892.
- [8] M. Yasushi, D. S. Matovski, M. S. Nixon, J. N. Carter and Y. Yagi. (2015). Gait recognition: Databases, representations, and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*.
- [9] O. Ťupa, A. Procházka, O. Vyšata, M. Schätz, J. Mareš, M. Vališ, V. Mařík. (2015). Motion tracking and gait feature estimation for recognising Parkinson's disease using MS Kinect. *BioMedical Engineering OnLine*, 14(97), pp. 1-20.
- [10] A. Procházka, O. Vyšata, M. Schätz, O. Ťupa, J. Mareš, M. Vališ, V. Mařík. (2015). Bayesian Classification and Analysis of Gait Disorders Using Image and Depth Sensors of Microsoft Kinect. *Elsevier: Digital Signal Processing*, 12(47), pp. 169-177.
- [11] A. Chaaoui, J. Padilla-López and F. Flórez-Revuelta. (2015, May). Abnormal gait detection with RGB-D devices using joint motion history features. Presented at IEEE International Conference and Workshops Automatic Face and Gesture Recognition.
- [12] K. Lin, S. Wang, P. Chung and C. Yang. (2013). A New View-Calibrated Approach for Abnormal Gait Detection. Presented at Advances in Intelligent Systems and Applications.
- [13] A. Leu, D. Ristić-Durrant and A. Gräser. (2011, May). A robust markerless vision-based human gait analysis system. Presented at IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI).
- [14] E. Stone and M. Skubic. (2011, August). Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing. Presented at Annual International Conference of the Engineering in Medicine and Biology Society, EMBC.
- [15] M. Robinson, and J. Vanrenterghem. (2012). An evaluation of anatomical and functional knee axis definition in the context of side-cutting. *Journal of biomechanics*, 45(11), pp. 1941-1946.
- [16] T. T. Verlekar, L. Correia and L. Soares. (2017). View-invariant gait recognition system using a gait energy image decomposition method. *IET Biometrics*.
- [17] T. T. Verlekar, L. Correia and L. Soares. (2016) View-Invariant Gait Recognition Exploiting Spatio-Temporal Information and a Dissimilarity Metric," Presented at International Conference of the IEEE Biometrics Special Interest Group (BIOSIG).
- [18] W. Stuberger, V. Colerick, D. Blanke and W. Bruce (1988) Comparison of a clinical gait analysis method using videography and temporal-distance measures with 16-mm cinematography. *Physical therapy*, 68(8), pp.1221-1225.
- [19] M. Nieto-Hidalgo, F. Ferrández-Pastor, R. Valdivieso-Sarabia, J. Mora-Pascual and J. García-Chamizo. (2016). A vision based proposal for classification of normal and abnormal gait using RGB camera. *Journal of biomedical informatics*, 63, pp.82-89.
- [20] M. Nieto-Hidalgo, and J. García-Chamizo. (2017, November). Classification of Pathologies Using a Vision Based Feature Extraction. Presented at International Conference on Ubiquitous Computing and Ambient Intelligence.
- [21] M. Nieto-Hidalgo, F. Ferrández-Pastor, R. Valdivieso-Sarabia, J. Mora-Pascual and J. García-Chamizo. (2016). Vision Based Gait Analysis for Frontal View Gait Sequences Using RGB Camera. In Ubiquitous Computing and Ambient Intelligence. Presented at 10th International Conference, UCAmI.
- [22] C. Bauckhage, J. Tsotsos and F. Bunn. (2009). Automatic detection of abnormal gait. *Image and Vision Computing*, 27(1), pp. 108-115.
- [23] M. Nieto-Hidalgo, F. Ferrández-Pastor, R. Valdivieso-Sarabia, J. Mora-Pascual and J. García-Chamizo. (2015, December). Vision based extraction of dynamic gait features focused on feet movement using RGB camera. Presented at Ambient Intelligence for Health.
- [24] M. Serrano, Y. Chen, A. Howard and P. Vela. (2016, August). Automated feet detection for clinical gait assessment. Presented at IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)
- [25] R. Krishnan, M. Sivarathinabala and S. Abirami. (2016). Abnormal gait detection using lean and ramp angle features. Presented at Computational Intelligence in Data Mining.
- [26] T. Nguyen, H. Huynh and J. Meunier. (2014, December). Extracting silhouette-based characteristics for human gait analysis using one camera. Presented at Fifth Symposium on Information and Communication Technology.
- [27] J. Fleiss and J. Cohen. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3), pp. 613-619.
- [28] Z. Zivkovic, and F. Van Der Heijden. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7), pp. 773-780.
- [29] S. Choudhury, and T. Tjahjadi. (2012). Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors. *Pattern Recognition*, 45(9), pp. 3414-3426.
- [30] N. Otsu. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), pp. 62-66.
- [31] "CASIA Gait Database". <http://www.sinobiometrics.com>.
- [32] <http://www2.c-motion.com/index.php>
- [33] IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.
- [34] P. Mukli, Z. Nagy, A. Eke. (2015). Multifractal formalism by enforcing the universal behavior of scaling functions. *Physica A: Statistical Mechanics and its Applications*, 417, pp. 150-167.
- [35] J. M. Hausdorff, P. L. Purdon, C. K. Peng, Z. V. Ladin, J. Y. Wei, and A. L. Goldberger. (1996). Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *Journal of applied physiology*, 80(5), pp. 1448-1457.