2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) | 978-1-7281-3079-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/FG47880.2020.00064

Using a Skeleton Gait Energy Image for Pathological Gait Classification

João Loureiro and Paulo Lobato Correia
Instituto de Telecomunicações, Instituto Superior Técnico, Lisbon, Portugal

Abstract— **Gait, or the way of walking, of a person can be affected by several pathologies. Automatic gait analysis methods have recently been proposed to identify and distinguish different types of pathological gait. Vision-based methods are often adopted, using some representation of the gait video sequence as input to feature extraction and classification algorithms. The Gait Energy Image (GEI), a mean image of the person's silhouettes along a gait cycle, is one of the most used gait representations for biometric recognition purposes and also for pathology identification. In this paper we propose an alternative compact gait representation, the Skeleton Energy Image (SEI), computed based on skeleton images instead of silhouettes, to remove from the representation some characteristics mostly useful for biometric recognition purposes, like the body shape, while preserving and highlighting the movement information. A method for feature extraction and classification is developed, and a new pathological gait dataset is acquired to test the system, which is made available to the research community. The proposed SEI representation allows obtaining better gait pathology classification results than the GEI. A method combining both representations further improves the performance of the system.**

I. INTRODUCTION

Gait is defined as the set of cyclic movements of the limbs that results in locomotion [1], [2], [3]. Certain neurodegenerative disorders affect these movements, leading to an impaired gait. Different disorders lead to different types of pathological gait, with distinct characteristics that can be explored for gait pathology classification. The common practice for gait assessment is based on a subjective analysis performed by medical specialists, after observing the gait of a subject. This approach lacks the precision and detail that instrumented methods can provide. On the other side, there are solutions from specialized companies that use a wide variety of sophisticated equipment in laboratory conditions, which acquire motion data with high accuracy [4]. This, however, uses costly technology, typically vision-based, requiring time-consuming setup and calibration procedures. As such, there is a need for simpler and reliable solutions for gait assessment.

To tackle the issue of gait analysis, new technologies and methods have been developed to obtain objective measures of a varied selection of gait indicators, with the goal of classifying between normal and different pathological gait types. These methods use a variety of sensors, wearable or not [5].

Wearable sensors are attached to parts of the body in order to measure gait kinetics and kinematics. These methods, while being precise and not restricting their usage to a controlled environment, can be considered intrusive, as they may cause discomfort to patients or even affect their gait, producing inaccurate measurements.

Non wearable sensors include floor sensors and, mostly, vision-based methods, such as stereo cameras, infrared thermography, structured light, time-of-flight systems, or just simple 2D cameras. Vision-based methods typically include modules for image preprocessing, feature extraction and classification algorithms, taking gait image sequences as input. Feature extraction can use handcrafted or learned features, if the gait features are computed making specified measurements on the images or are extracted automatically using deep learning techniques, respectively.

Handcrafted feature extraction methods can be further classified into appearance- or model-based. Appearance-based techniques analyze the shape of the subject, often using silhouettes extracted by background subtraction, to compute the selected gait indicators [6], [7]. Model-based techniques attempt to fit a model to the gait sequence images, such as a human skeleton or some model relying on the segmentation of different body parts. The works in [6], [8] present methods that estimate skeleton information from silhouette images.

Methods using deep learning algorithms use a set of automatically learned features. These methods often require some image preprocessing beyond background subtraction, such as the computation of the Gait Energy Image (GEI) [9], a compact representation of a complete gait cycle using a single image, with the advantage of attenuating the effect of possible individual silhouette's segmentation errors. A GEI can then be used as input into a Convolutional Neural Network (CNN) that performs feature extraction [10], [11], and classification can be done using the fully connected layers of the same network, or alternatively using more conventional machine learning techniques operating on the features extracted by the CNN. Other deep learning methods use alternative gait representations, such as skeleton data [12], [13] or optical flow images [14]. Another approach is to adopt other deep learning architectures, such as recurrent neural networks, like the Long Short-Term Memory (LSTM) [15], which process a sequence of gait images.

As baseline for the work presented here, a system was developed using the well-known GEI representation to classify a gait sequence as normal or impaired by one of four types of pathologies.

The main contribution of the paper is the usage of a new gait representation, the Skeleton Energy Image (SEI), computed from a sequence of skeletons, which is not affected by the body shape or the clothes being used. This representation decouples the subject's appearance from its

This work is funded by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020

978-1-7281-3079-8/20/$31.00 ©2020 IEEE
DOI 10.1109/FG47880.2020.00064

motion, in contrast to silhouette based representations as the GEI, allowing a better classification of gait pathologies. Early work on this direction includes the presentation, in 2014, of a Skeleton Variance Image (SVIM) [16], or the very recent work developed independently of a Skeleton Gait Energy Image (SGEI) [17], [18], both in the context of gait recognition applications to increase robustness against clothing changes, and using skeleton information in a different way from the proposed SEI.

Since the number of publicly available datasets for testing a pathological gait classification system is very limited, this paper also proposes a new pathological gait dataset, with 10 subjects walking normally as well as simulating four types of gait pathologies, each with two levels of severity. This dataset is publicly available to the research community at: http://www.img.lx.it.pt/GAIT-IST/.

The rest of the paper is organized as follows: section **II** describes the new GAIT-IST dataset. Section **III** presents the proposed SEI gait representation and section **IV** describes the adopted pathological gait classification solution. In section **V** results are presented and discussed. Final conclusions are drawn in section **VI** and future work directions are presented.

II. The GAIT-IST Dataset

In the context of gait recognition many datasets are available, with sequences of subjects walking normally, considering different observation angles as well as covariates such as clothing variations, carrying of items or different walking speeds. However, due to the difficulty of obtaining gait sequences from real patients, and also due to privacy and ethical issues for sharing those sequences, there are not many gait databases publicly available dedicated to the study of pathological gait. Moreover, the few publicly available datasets are composed of simulated gait impairments and have a reduced number of samples, which is not optimal. The most notable pathological gait datasets available are the following:

- The INIT Gait Dataset [19], with samples from 10 subjects, simulating impairments in the arms and legs.
- The DAI Gait Dataset [20], with samples from 5 subjects performing normal and random abnormal gait.
- The DAI Gait Dataset 2 (DAI2) [21] that includes sequences of 5 subjects performing normal and four types of abnormal gait simulations.

Given the above listed limitations, a new pathological gait dataset, called GAIT-IST, has been acquired to be used in this work, in addition to the existing ones. The GAIT-IST dataset includes silhouette and skeleton sequences, as well as the corresponding GEI and SEI representations, captured from 10 subjects walking normally and simulating four types of gait pathologies, each with two levels of severity.

The types of gait featured in the GAIT-IST dataset, besides Normal gait, are Diplegic, Hemiplegic, Neuropathic and Parkinsonian gait, as they are among the most common pathological gait types observed [1]. These are also the gait types simulated in the DAI2 dataset, making it possible to perform cross-database tests on gait pathology classification solutions.

A total of 10 subjects (8 males and 2 females) in the age range of 20-50 years old have participated voluntarily on the dataset acquisition and signed a consent form to allow usage of the images for research purposes. The GAIT-IST dataset includes normal walking and 4 pathological gait sequences, with 2 severity levels, 2 directions of walking and 2 repetitions, per participant, in a total of $10 \times (4 \times 2 \times 2 \times 2 + 2 \times 2) = 360$ gait sequences. Acquisitions considered the subject walking parallel to the camera plane, with a minimum of 2 complete gait cycles captured in each direction.

To demonstrate the possibility of performing automatic gait analysis in conditions easy to replicate in rehabilitation clinics, or even at home, gait sequences were captured using a smartphone camera, with a resolution of 720x1280 pixels, fixed on a tripod, at a height of 1,5 meters above the ground. The distance between the camera and the walking subjects was about 4 meters. Illumination was fairly constant, mostly coming from the room's artificial illumination, and the background was constant and uniform, consisting in a white wall and a gray floor, with some shadows and reflections present, as illustrated in Fig. 1 (left). The GAIT-IST dataset is freely available to the research community.

III. Proposed Gait Representation

The goal of this paper is presenting a system able to, starting with videos of subjects depicting normal gait and different types of pathological gait, classify the gait sequences into one of those categories.

The gait representation typically used by state-of-the-art systems is the Gait Energy Image (GEI). A GEI can be computed by performing background subtraction on the original images, to obtain a sequence of silhouettes, which are then cropped and resized to a normalized height. If the goal is to use a convolutional neural network for feature extraction the silhouettes' width can be padded with zeros to obtain the desired aspect ratio, e.g. 224x224 pixels. After detecting the complete gait cycles in the sequence, e.g. by analyzing the variation of the silhouettes' width in the feet area along time, the GEI for a gait cycle corresponding to a set of N images, $I_t(x, y)$, can be computed by averaging those images:

$$GEI(x, y) = \frac{1}{N} \sum_{t=1}^{N} I_t(x, y) \tag{1}$$

However, even if a GEI captures the motion along a gait cycle, it also captures the physical characteristics and

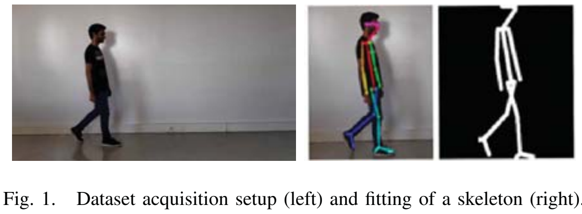

Fig. 1. Dataset acquisition setup (left) and fitting of a skeleton (right).

clothing or styling accessories, represented in the silhouette of the walking person. To obtain a gait representation that is only focused on the walking characteristics, this paper proposes using a new representation, the Skeleton Energy Image (SEI).

The first step to compute the SEI is obtaining a skeleton for each image of the walking person. By applying the OpenPose [13] algorithm a set of 24 pairs of 2D coordinates are obtained, corresponding to different parts of the human body: nose, neck, shoulders, elbows, wrists, middle hip, right and left hip, knees, ankles, eyes, ears, big toes, small toes, and heels. Notice that the SGEI model [17] uses only 13 keypoints, located in the limbs and the body torso, not including the tip of the feet or the head. As a consequence, the proposed SEI can better model impaired gait postures.

The SEI skeleton image is obtained by drawing lines connecting the 24 estimated coordinates using OpenCV [22], obtaining representations similar to the one illustrated in Fig. 1 (right). Then a resizing and averaging process is applied, as described for the GEI computation, to obtain the SEI.

Examples of GEI and SEI gait representations are included in Fig. 2.

IV. Proposed Gait Classification System

Having selected a gait representation, the classification of gait pathologies relies on the extraction of a set of features followed by a classification module.

In the proposed implementation feature extraction is performed using the VGG-19 architecture [23], a very deep CNN that is easy to implement and train as it converges quickly, and which has often been selected for similar tasks [10], [11]. However, as the available datasets for training are very small, transfer learning is used to limit overfitting, starting from a VGG-19 model pre-trained using the ImageNet database [24]. As the initial convolutional layers are responsible for detecting generic features, such as shapes in images, the weights of those layers can be frozen during training. The remaining VGG layers are fine-tuned using the training subset of the GAIT-IST dataset, to learn features that best represent our problem.

Two alternatives were considered for the classification step: (i) using the fully connected layers of the VGG-19; (ii) taking the first fully connected layer output as a feature vector, and applying Principal Component Analysis (PCA) for dimensionality reduction, followed by classification using Linear Discriminant Analysis (LDA) or a Support Vector Machine (SVM). LDA and SVM were selected after initial tests with a large amount of classifiers.

The second classification option decouples classification from the learning of features, which might reduce overfitting. When using it, PCA was configured to keep enough components to explain 95% of the original variance. Concerning LDA, using full covariance structure performed better than a diagonal structure, being the option selected. Regarding the SVM, linear, quadratic, cubic and Gaussian kernels were tested, with the linear kernel producing the most consistent and generally best results.

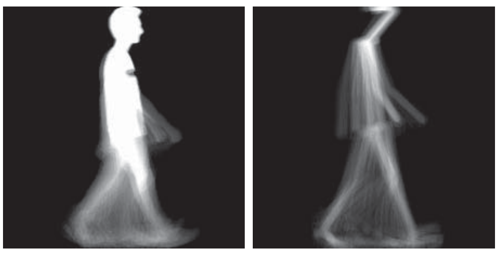

Fig. 2. Examples of GEI (left) and SEI (right) gait representations.

V. Experimental Results

To test the proposed system in the task of detecting and classifying different types of impaired gait, results were first obtained using the GAIT-IST dataset. Tests were conducted using 10-fold cross-validation, training with sequences of 9 subjects and using the remaining subject for validation. The reported results correspond to the mean accuracy of the 10-fold cross-validation tests.

Each test was repeated by varying the number of retrained VGG-19 layers, from retraining only the fully connected layers to one or more convolutional layers. It would be expected that retraining more layers would improve system performance, by better learning the image patterns present in the dataset being used.

A. Testing on the GAIT-IST Dataset

Results (mean accuracy of the 10-fold cross-validation) using gait representations based on GEI and SEI, for gait impairment classification using the full CNN, as well as LDA and SVM, are reported in Table I.

These results show that increasing the number of retrained layers tends to improve performance. It can be also noted that there is no improvement when using a model such as LDA or SVM for classification, as the best results are obtained when using the full CNN for classification. This may nevertheless be indicative of some overfitting, as the same dataset is being used for training and validation, despite using disjoint sets of sequences.

There is a significant overall increase of classification accuracy when using a gait representation based on the proposed SEI in comparison to the use of the GEI. This strongly suggests, as speculated before, that the SEI is a better representation to be used for pathological gait classification.

B. GEI and SEI Fusion Results

Since the computation of silhouettes and skeletons uses different algorithms, combining the results using gait representations based on GEI and SEI could further improve the performance of the system. As such, an experiment merging the classification results of the two representations was conducted. To do this, for each gait sequence, the CNN was used to perform classification, using both the GEI and

TABLE I

GAIT CLASSIFICATION RESULTS (%)

VGG-19 1st tuned block	Classifier CNN GEI : SEI	LDA GEI : SEI	SVM GEI : SEI	CNN GEI + SEI
FC	89.4 : 94.5	89.1 : 95.6	89.4 : 95.3	**96.2 ± 5.2**
CB5	93.3 : 96.7	88.7 : 95.9	87.4 : 95.5	**98.0 ± 2.8**
CB4	92.7 : 97.3	89.6 : 96.2	89.4 : 96.4	**97.8 ± 2.9**
CB3	94.5 : 97.1	89.3 : 96.1	90.4 : 95.6	**98.5 ± 2.3**
CB2	93.7 : 97.4	89.6 : 96.4	89.5 : 96.0	**97.6 ± 3.8**

FC - fully connected block; CB - convolutional block

the corresponding SEI, as the full CNN classifier obtained best results with the GAIT-IST dataset.

The resulting scores for each category, obtained before the softmax layer, represent the probability that the input observed belongs to that category, and the maximum between the values observed for the GEI and SEI was kept. A max-score fusion was therefore applied, selecting as the predicted category the one with the highest score corresponding to the usage of GEI or SEI. This can be interpreted as choosing the prediction of the model that performed the classification with higher certainty.

Using this method, an improvement of the overall classification accuracy was observed, when compared with the individual usage of GEI or SEI, as observed in the rightmost column of Table I. For the best configuration the average classification accuracy was 98.5 ± 2.3%.

C. Cross-database Results

With the objective of testing the generalization ability of the proposed system in the classification and feature extraction tasks, tests were made using a different gait dataset, namely the DAI Gait Dataset 2 (DAI2) [21]. This dataset is similar to the GAIT-IST, as it is composed of simulations of the same gait types, performed by 5 subjects. This makes it possible to use it for testing the models already trained with the GAIT-IST dataset, in a cross-dataset test, using the GEI as the gait representation. A similar test using the SEI representation was not possible, as the technique employed for skeleton computation requires access to the original images, which are not available on the DAI2 dataset.

The same classification methods were used, taking the CNN model already retrained using the GAIT-IST dataset and using it to perform feature extraction on the GEIs of the DAI2 dataset. This was only tested on the CNN retrained starting at the third convolutional block, as it was the configuration leading to the best results when using the GEIs. The extracted features were also used to train LDA and SVM models that were then tested using a similar cross-validation technique as before. The results are presented in Table II.

As expected, the results were significantly worse than when training and testing on samples of the same dataset, obtaining an overall accuracy of only 43.3%, using the CNN for classification. Many factors can contribute to the decrease in accuracy; for instance, the two datasets may have simulated each gait type by emphasizing different characteristics

TABLE II

CROSS-DATABASE RESULTS (%)

VGG-19 1st tuned block	Classifier CNN	LDA	SVM
CB3	43.3	**76.7**	68.8

CB3 - convolutional block 3

of the pathologies, and the GEIs may have been computed in slightly different ways. Besides, the silhouettes available from the DAI2 dataset have several segmentation errors, mainly in the form of holes in the silhouettes, which makes the classification task even harder.

In this case it is possible to observe a significant improvement of the classification accuracy when using different classification methods: an accuracy of 76.7% was obtained with LDA, and 68.8% with SVM. This shows that using the extracted features with a different classifier, that has been independently trained, improves the generalization ability of the system across different datasets.

VI. CONCLUSIONS

This paper presented a new gait dataset, GAIT-IST, with sequences acquired from 10 subjects simulating four different types of impaired gait, to be used in gait analysis and classification systems. This dataset is now publicly available to the research community.

A new gait representation, the Skeleton Energy Image (SEI), is adopted to better capture the motion of a walking person, while decoupling it from appearance characteristics, which may be useful for gait recognition, but are not of interest for gait impairment classification.

The gait impairment classification system proposed in this paper can use the well-known GEI representation as well as the proposed SEI. It was observed that the SEI representation performs better than the GEI, and that a fusion between GEI and SEI class probabilities can further increase performance.

For cross-dataset tests, training on GAIT-IST and testing on DAI2, it was observed that using a classifier trained independently of the feature extractor has achieved a better generalization performance.

An important topic for future work is the acquisition of larger datasets representing impaired and normal gait, as having access to large and representative datasets is of the utmost importance when dealing with gait pathology classification. Also acquiring longer gait sequences, with more than two gait cycles, can help overcoming any transient conditions that may lead to less representative data. Additional exploitation of the combined usage of SEI and GEI for classification might contribute to further improve the performance of the system. And using a recurrent neural network, such as the LSTM, could additionally be considered. Although the GEI and SEI are compact representations of the temporal information, the use of an LSTM could further emphasize the time aspect of the gait sequences. Input to the LSTM could be in the form of skeleton coordinates data or feature vectors extracted from the silhouettes composing each gait cycle.

References

[1] M. P. Chris Kirtley, “Clinical gait analysis: theory and practice”, Churchill Livingstone, 2006.

[2] P. Connor, A. Ross, ”Biometric recognition by gait: A survey of modalities and features”, Computer Vision and Image Understanding, 167, 1-27, 2018.

[3] I. Rida, N. Almaadeed, S. Almaadeed, ”Robust gait recognition: a comprehensive survey”, IET Biometrics, 8(1), 14-28, 2018

[4] J. Vanrenterghem, D. Gormley, M. Robinson, and A. Lees, ”Solutions for representing the whole-body centre of mass in side cutting manoeuvres based on data that is typically available for lower limb kinematics,” Gait & Posture, Vol. 31 (4), pp. 517-521, 2010

[5] A. Muro-de-la-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, ”Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications,” Sensors, Vol. 14 (2), pp. 3362-3394, 2014

[6] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo, “A vision based proposal for classification of normal and abnormal gait using RGB camera,” Journal of Biomedical Informatics, vol. 63, pp. 82–89, 2016.

[7] T. T. Verlekar, L. D. Soares, and P. L. Correia, “Automatic classification of gait impairments using a markerless 2D video-based system,” Sensors, Vol. 18 (9), 2018.

[8] A. Leu, D. Ristic-Durrant, and A. Graser, “A robust markerless vision-based human gait analysis system,” in IEEE Symposium on Applied Computational Intelligence and Informatics, pp. 415–420, 2011.

[9] J. Han and B. Bhanu, “Individual recognition using gait energy image,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28 (2), pp. 316–322, 2006.

[10] T. T. Verlekar, P. L. Correia, and L. D. Soares, “Using transfer learning for classification of gait pathologies,” in IEEE International Conference on Bioinformatics and Biomedicine, pp. 2376–2381, 2018.

[11] C. Li, X. Min, S. Sun, W. Lin, and Z. Tang, “DeepGait: a learning deep convolutional representation for view-invariant gait recognition using joint Bayesian,” Applied Sciences, Vol. 7 (3), 2017.

[12] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, “Convolutional pose machines,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-December, pp. 4724–4732, 2016.

[13] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 1302–1310, 2017.

[14] K. Simonyan and A. Zisserman, ”Two-stream convolutional networks for action recognition in videos,” in Advances in Neural Information Processing Systems (NIPS), pp. 568-576, 2014.

[15] R. Liao, C. Cao, and E. B. Garcia, “Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations,” in Chinese Conference on Biometric Recognition, China, Oct. 2017.

[16] T. Whytock, A. Belyaev, and N.M. Robertson, ”Dynamic distance-based shape features for gait recognition,” Journal of Mathematical Imaging and Vision, Vol. 50, pp. 314–326, 2014.

[17] L. Yao, W. Kusakunniran, Q. Wu, J. Zhang, and Z. Tang, ”Robust CNN-based gait verification and identification using skeleton gait energy image,” in International Conference on Digital Image Computing: Techniques and Applications, Dec. 2018.

[18] Q. Wu, J. Zhang, Z. Tang, and W. Yang, ”Robust gait recognition using hybrid descriptors based on skeleton gait energy image”, Pattern Recognition Letters, in press, 2019.

[19] J. Ortells and H. Ram, “Vision-based gait impairment analysis for aided diagnosis,” Medical and Biological Engineering and Computing, Vol. 56, pp. 1553–1564, 2018.

[20] M. Nieto-Hidalgo, F. J. Ferrández-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. García-Chamizo, “Vision based extraction of dynamic gait features focused on feet movement using RGB camera,” in Ambient Intelligence for Health, pp. 155–166, 2015.

[21] M. Nieto-Hidalgo and J. M. García-Chamizo, “Classification of pathologies using a vision based feature extraction,” in Ubiquitous Computing and Ambient Intelligence, 2017, pp. 265–274.

[22] G. Bradski, ”The OpenCV Library”, Dr. Dobb’s Journal of Software Tools, 2000

[23] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in International Conference on Learning Representations, 2015.

[24] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-fei, “ImageNet: a large-scale hierarchical image database,” in IEEE Computer Vision and Pattern Recognition, pp. 2–9, 2009.