Gait Recognition Using 2D Poses

Daniel Ricardo dos Santos Jangua[1], Aparecido Nilceu Marana[1]

[1] Department of Computing
Faculty of Sciences
São Paulo State University (UNESP)
17033-360 – Bauru – São Paulo – Brazil

{daniel.jangua, nilceu.marana}@unesp.br

Abstract. *Over the last decades, the field of biometrics has become an important ally for human identification, mainly used for fraud prevention and access control in restricted areas, with the final purpose of increasing the security of the individuals in society. Nowadays, the most common biometric systems are those based in features like fingerprints, face and iris. Despite the great performance of state-of-art methods that use these traits, an important challenge remains, which is the automatic human identification in low-resolution videos, at a distance and without the need for subject cooperation. In this context, the usual biometric systems do not meet the expected performance, and using gait features to identify individuals may be the only viable option. The goal of this work is to propose a new method for gait recognition using gait information extracted from 2D poses estimated over video sequences with high accuracy and low computational cost when compared to other state-of-art methods. In order to estimate the 2D poses, we use OpenPose, an open-source and robust pose estimator. The proposed new method was assessed in two public gait datasets, CASIA Gait Dataset-A and CASIA Gait Dataset-B, and obtained recognition rates comparable with state-of-the-art results, but using smaller feature vectors.*

1. Introduction

Nowadays, biometrics, that consists in the statistical study of physical or behavioral characteristics [Jain et al. 2011], has become an essential tool to ensure the security of individuals in society in the most diverse scenarios, being used for human identification. The most usual biometric systems are based on fingerprint, face or iris traits, which are physical characteristics, and, thus, harder to imitate than gait, for instance, a behavioral characteristic. However, those physical traits may not perform well in scenarios where the biometric data is captured in low-resolution, at a distance, and the biometric system must be operated in covert mode. In such cases, the use of gait characteristics can be the only choice.

Gait can be defined as integrated and repetitive movements performed by the human body during the act of walking that form a pattern called "gait cycle" [Arantes and Gonzaga 2011]. According to researches conducted in the last decades, each person has a distinct way of walking [Nixon and Carter 2006]. In this context, the use of gait recognition for human identification is more advantageous than the classical biometrics methods since: (i) it can be performed at a distance, (ii) it presents a good performance in low-resolution images, and (iii) it does not depend on subject's cooperation, allowing covert identification [Wan et al. 2018].

The biometric methods based on gait analysis can be classified in two groups, according to the approach that is used for feature representation. The first group is composed by non-model based methods, that use the silhouette analysis to extract gait information, working mostly with the static aspect of gait. The second group is formed by model based methods, that use a spatio-temporal model to represent the gait. Model based methods are more computationally costly, but as they work with dynamic aspects of gait, they present more robustness and higher classification performances [Arantes and Gonzaga 2011].

The goal of this work is to propose a method for gait recognition based on 2D model of the human body. In our work, such models are estimated by using OpenPose [Cao et al. 2018], that outputs, for each frame of a video of a walking person, the positions of the body joints and other important key points like neck and nose. The proposed method uses the coordinates of the body joints to build two signals, one for angles and other for distances, that represent the movement of each limb part over the video frames. Histograms of angles and distances are built for each limb part, representing its behavior during the gait cycle. In order to assess the proposed method, experiments were conducted utilizing CASIA Gait Dataset-A [Liang Wang et al. 2003] and CASIA Gait Dataset-B [Yu, S. et al. 2006], two public gait datasets commonly used for assessment of gait recognition methods.

The rest of this paper is organized as follows: Section 2 discusses human pose estimation, focusing on OpenPose method; Section 3 introduces the main concepts of gait recognition; Section 4 presents some related works; Section 5 explains the proposed method in details; Section 6 presents the experimental results and, finally, Section 7 draws some conclusions obtained from the experimental results.

2. Human Pose Estimation

Human pose estimation is an important tool for computers to understand people in images and it can be described as the task of estimating key points of human bodies in a scene. Given an image, the pose estimator must return the coordinates of each key point of the human body, and by connecting those key points correctly it is possible to find the body limbs, and then calculate different features from the limbs related to their movements in a frame sequence.

In our method, OpenPose [Cao et al. 2018] is utilized to estimate the 2D poses from each frame of an input video of a walking person. OpenPose is an open source real-time 2D pose estimator capable of perform multi-person detection with a high reliability and good computational performance and includes body, feet, hands and face key points. In contrast to most pose estimators, OpenPose takes a bottom-up approach that improves the performance of pose detection by treating the image globally, detecting all body parts candidates in the input image and associating them to form valid poses. Figure 1 shows the pipeline of the OpenPose method.

The method uses an approach named Part Affinity Fields (PAF) that maps the position and orientation of each body part in the image domain using 2D vector set along with Part Confidence Maps that represent the probability of a body part existence for each image pixel. And using this encoded global information the computational complexity is reduced by adopting a greedy approach to associate the joint points and form the 2D poses [Cao et al. 2018]. During the detection process, the PAFs are iteratively improved

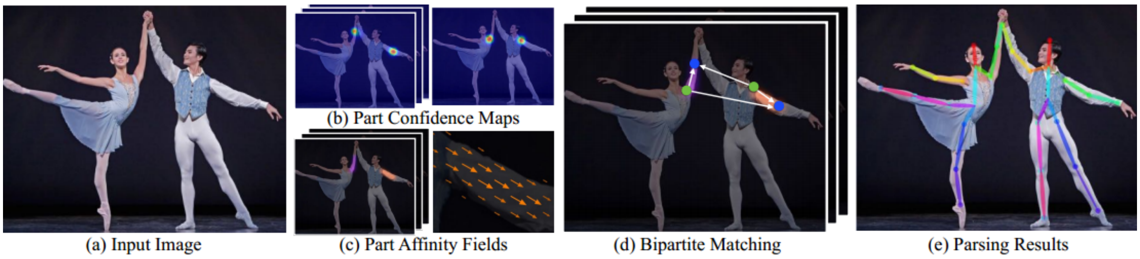

Figure 1. Pipeline of the OpenPose method [Cao et al. 2018].

in each stage $t \in \{1, ..., T\}$ along with the confidence maps through two interconnected convolutional neural networks (CNN). Figure 2 shows the OpenPose CNN architecture, the blue CNN predicting the PAFs and the beige CNN predicting the confidence maps.

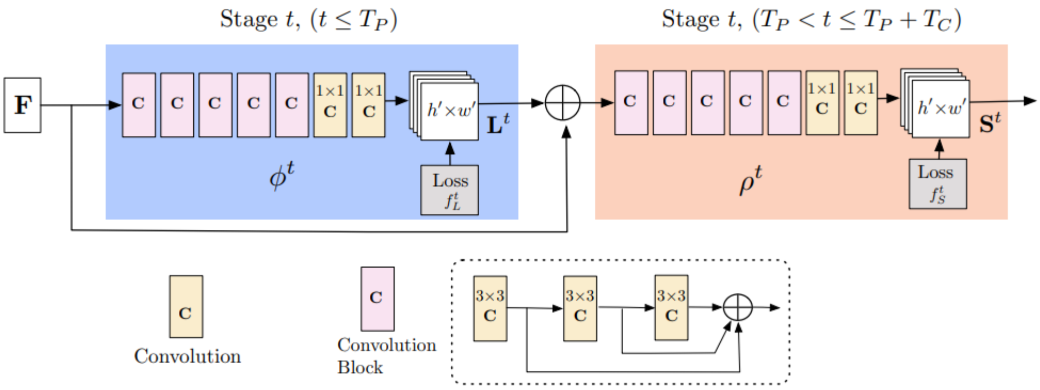

Figure 2. OpenPose CNN architecture [Cao et al. 2018].

3. Gait Recognition

Gait can be described as the periodic behavior of each body part movement in the act of walking. Early studies in Medicine and Psychology have shown that the use of gait to identify humans is feasible, since it has some components that are unique for every human being due to the unique muscular and skeletal structure [Wan et al. 2018].

Due to the periodic aspect of gait, it is possible to identify a pattern of movements repeated into a gait cycle. The gait cycle is the time interval between successive instances of initial foot-to-floor contact and can be divided in two periods or phases for each leg. The first phase, the stance phase, occurs when the foot is in contact with the floor supporting the body. The second phase, the swing phase, occurs when the foot is off the ground moving forward to the next step [Nixon, M. S. et al. 2006]. During each gait cycle the inferior and superior body members realize a pendulum movement varying their angulation in relation to the horizontal axis forming a pattern that can be used to extract gait information and perform gait recognition. Figure 3 illustrates the gait cycle with the stance and swing phases for each leg, starting from left to right, resumed in four frames from the dataset CASIA Gait Dataset-A [Liang Wang et al. 2003].

4. Related Works

In this section some works related to the proposed method are briefly presented. Such works propose different approaches for human identification using model based and non-

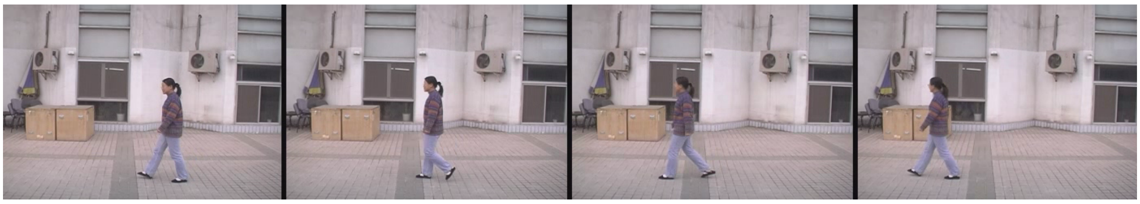

Figure 3. The gait cycle resumed in 4 frames from CASIA Gait Dataset-A [Liang Wang et al. 2003].

model based methods for gait analysis.

In [Wang, L. et al. 2003], the authors propose a method for gait recognition based on silhouette analysis. The silhouette of the walking person is extracted from the background in each frame of the video by a background subtraction procedure. Then the changes in the silhouette shape over time are represented using an associated sequence of complex vector configuration that is analyzed using the *Procrustes* method, creating a mean silhouette that encodes the gait information. This method achieved 90% of rank-1 accuracy on CASIA Gait Dataset-A.

In [Yu, S. et al. 2007], the proposed method utilizes a dynamic time warping (DTW) based contour similarity measure to perform the gait recognition based on silhouette shape analysis. This method aimed to reduce the influence of noise on classification process. This method obtained 83.5% of rank-1 accuracy on CASIA Gait Dataset-B.

In [Chen, C. et al. 2009], a representation named frame difference energy image (FDEI) is utilized to work with obstructed or incomplete silhouettes. In this method, the gait cycle is divided into clusters represented by the dominant energy image (DEI), which is utilized to build the FDEI representation. This method obtained 91.1% of rank-1 accuracy on CASIA Gait Dataset-B.

In [Liu, D. et al. 2016], the authors proposed a memory-based approach inspired by the brain sequential processing mechanism. The 2D joint points coordinates are extracted using the migratory articulated human detection, followed by a memory-based gait recognition (MGR) neural network that performs the gait recognition. This method achieved 95% of rank-1 accuracy on CASIA Gait Dataset-A.

In [Lima and Schwartz 2019], the gait recognition is performed utilizing a model-based approach. The positions of the subject's body joints are extracted by a 2D pose estimation method. Then, from this information, signals and movement histograms are built to be used as feature descriptor. This method utilizes a 1-NN classifier with Euclidean distance. This method achieved 97.5% of rank-1 accuracy on CASIA Gait Dataset-A and 98% on CASIA Gait Dataset-B.

5. Proposed Method

In this work we propose a new method for human identification based on gait recognition and 2D poses, likewise [Lima and Schwartz 2019].

First, the OpenPose [Cao et al. 2018] method is utilized to extract the 2D poses of individuals from the input video. After that, in each frame, the coordinate of the subject's body joint points are utilized to calculate the angulation of each limb part in relation to the

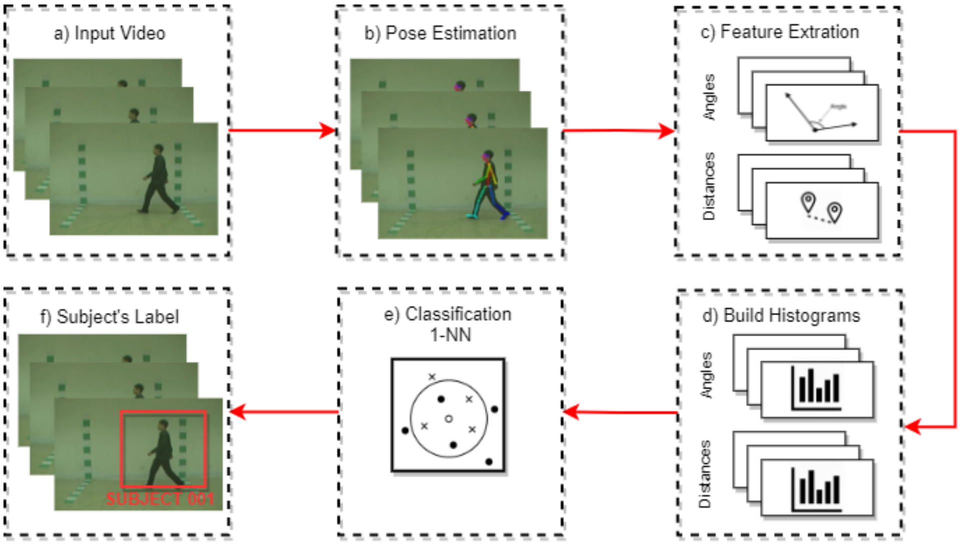

Figure 4. Block diagram of the proposed method for gait recognition using 2D poses.

horizontal axis and the distance between the line defined by the two points that formed the current limb part and the point that represents the neck. In the next step, these two data are utilized to build two histograms (angles and distances) for each limb part, that represent the gait signature over all frames. Finally, both histograms are used as the feature vector by a 1-NN classifier, with a predefined distance function, in order to assign the identity to the individual whose 2D poses were estimated in the input video. Figure 4 shows the block diagram of the proposed method.

5.1. Pose Estimation

The first step of the method is the pose estimation. In this step, OpenPose [Cao et al. 2018] is utilized to extract the coordinates of the individual's body joints in each frame of the video input. The vertical and horizontal coordinates are organized in a JSON text file according to the selected type of skeleton detection. In this work the model used was the BODY_25 format, which has 25 key points, as illustrated in Figure 5.

5.2. Feature Extraction

After the pose estimation, the key points that form the upper and lower limbs are selected to be used in the feature extraction process. According to the skeleton in Figure 5, the selected points are: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13 and 14. Given two selected key points $P_1 = (x_{P1}, y_{P1})$ and $P_2 = (x_{P2}, y_{P2})$, the limb part can be represented as a 2D vector $w = (x_1, y_1)$ where $(x_1, y_1) = (x_{P1} - x_{P2}, y_{P1} - y_{P2})$. For each limb part, we build a sequence formed by the angle φ between the limb part and the horizontal axis ($(x_2, y_2) = (1, 0)$) in each video frame, that can be calculated utilizing the Equation 1.

$$\varphi = \arccos \frac{x_1 * x_2 + y_1 * y_2}{\sqrt{x_1^2 + y_1^2} * \sqrt{x_2^2 + y_2^2}} \tag{1}$$

Analogously, the distance sequence is formed by the distance d between the line that the limb part is contained and the point that represents the neck (point 1 in Figure

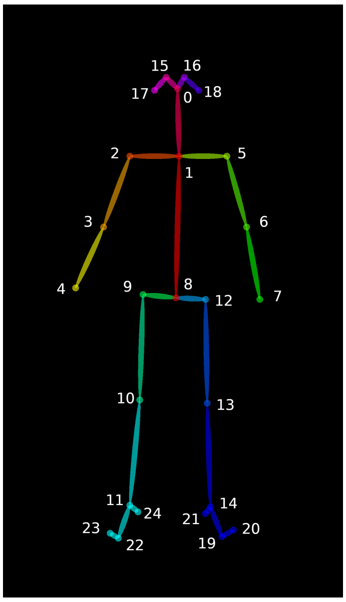

Figure 5. Pose output format of BODY_25 [Cao et al. 2018].

5). Considering the vector $v = P_{neck} - P_2$, in which P_{neck} is the neck point and P_2 is a key point that forms the member in question, we can use equations 2 and 3 to do this calculation:

$$Proj_w v = \left(\frac{v \cdot w}{\|w\|^2} \right) * w \tag{2}$$

$$d = \|v - Proj_w v\| \tag{3}$$

5.3. Histograms Building

In this step, for each limb part, we use the two sequences (angles, distances) to build one histogram for each of them. For all histograms, we used 16 bins (value found empirically), and as our method uses eight limb parts (left and right forearms, arms, legs and thighs), we end up with 8 histograms of angles and 8 of distances, with 16 bins each. The angle histograms are defined in the interval $[0, \pi]$, because the possible angle variation. The distance histograms are built applying the base 2 logarithmic function (log_2) in the distances, so the distance histograms are defined in the interval $[0, log_2(max\_dist)]$, in which max_dist is the longest calculated distance. The use of the log_2 function improves the performance of the method, as it maps the distances so that the difference between shortest distances (most recurring) is accentuated and the largest are grouped. At the end of this step, all 8 histograms of each type are concatenated to form only one histogram for each sequence. So we end up with two histograms (angles and distances) with 128 bins each (8 histograms x 16 bins).

5.4. Classification

For the classification process we use a 1-NN classifier. In order to decide which distance function should be used, we assessed two distance functions, the Euclidean and the chi-square. Results of these tests are presented in Section 6. Given a distance function, we calculate the distance d_1 between the angle histograms of the probe (query) and gallery (database) videos and d_2 between the distance histograms. The final distance between the probe and the gallery videos is $d_f = (d_1 + d_2)/2$. As both histograms, angles, distances and lengths, are normalized, there is no need to normalize the distances d_1 and d_2.

6. Experimental Results

In order to assess the proposed method, the experiments were carried out on CASIA Gait Dataset-A [Liang Wang et al. 2003] and CASIA Gait Dataset-B [Yu, S. et al. 2006], both datasets are briefly introduced in Section 1. For the first experiment, using the Dataset-A, we assessed the proposed method using two variations on classification step. One is using the Euclidean distance and the other is using the chi-square distance function in the classifier. The results were compared using the Cumulative Matching Characteristic (CMC) curve with the mean accuracy obtained for the three different camera views: lateral, oblique and frontal (totaling 240 walking sequences, 80 for each direction). Figure 6 shows the CMC curves obtained by each of the proposed method variations. It is possible to observe that the use of chi-square distance function increased the method performance, corroborating studies that indicate that this function is a good metric for histograms comparison [Marín-Reyes et al. 2016].

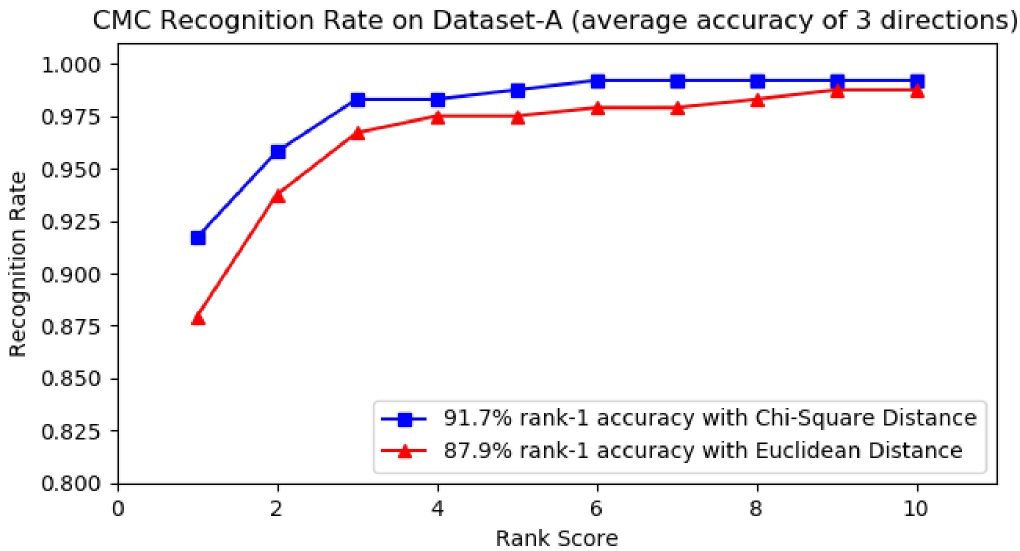

Figure 6. CMC curves obtained by the method variations on Dataset-A.

Table 1. Rank-1 Accuracy - CASIA Gait Dataset-A.

Method	Lateral	Oblique	Frontal	Average
[Wang, L. et al. 2003]	88.75%	87.50%	90.00%	88.75%
[Liu, D. et al. 2016]	85.00%	87.50%	95.00%	89.17%
[Lima and Schwartz 2019]	92.50%	96.25%	97.50%	95.42%
Our method (Euclidean)	80.00%	87.50%	96.25%	87.92%
Our method (Chi-square)	87.50%	92.50%	95.00%	91.67%

Table 1 shows the rank-1 accuracy of our method (using both distance functions) with other methods presented in Section 4, including two recent state-of-art methods, [Lima and Schwartz 2019] and [Liu, D. et al. 2016] for each camera position on CASIA Gait Dataset-A [Liang Wang et al. 2003]. It is possible to observe that our method, with the chi-square distance function, was superior to the method by [Wang, L. et al. 2003] and was competitive with the state-of-the-art methods proposed by [Liu, D. et al. 2016] and [Lima and Schwartz 2019]. One can also observe that all methods presented lower accuracy on lateral view and higher accuracy on frontal view, this should be because in this angle there is more information about the gait signature, mainly because there is no limb occlusions.

In the other experiment, the two variation of the proposed method were assessed using the CMC curves on CASIA Gait Datatset-B [Yu, S. et al. 2006], utilizing only the walking sequences in the lateral direction (90 degrees to the camera position) and in normal walking condition (totaling 744 walking sequences). The result of this test is presented in Figure 7. It is possible to notice that, again, the chi-square distance function showed better results than Euclidean distance function.

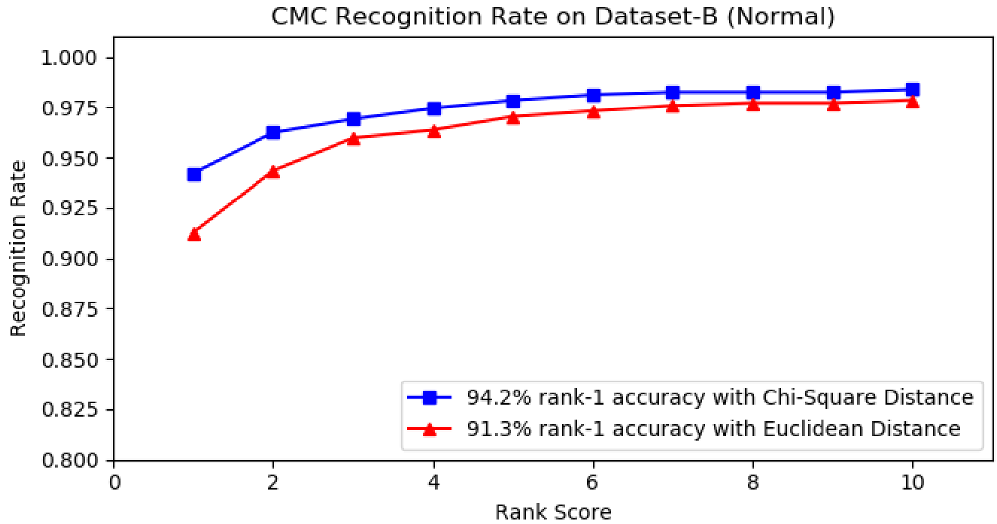

Figure 7. CMC curves obtained by the method variations on Dataset-B.

Table 2. Rank-1 Accuracy - CASIA Gait Dataset-B (Normal).

Method	Lateral
[Yu, S. et al. 2007]	83.50%
[Chen, C. et al. 2009]	91.10%
[Lima and Schwartz 2019]	98.00%
Our method (Euclidean)	91.26%
Our method (Chi-square)	94.22%

Table 2 shows the rank-1 accuracy of our method with other methods presented in Section 4, including one very recent method that can be considered state-of-the-art ([Lima and Schwartz 2019]), assessed on Dataset-B, using only the videos with lateral view. Although the proposed method did not get the highest rank-1 accuracy rate, it presented, with both distance functions (Euclidean and chi-square), higher results than the methods proposed by [Yu, S. et al. 2007] and [Chen, C. et al. 2009] and competitive results when compared with the [Lima and Schwartz 2019] method. In general, the tests in lateral view using the Dataset-B showed better results than the tests performed on Dataset-A. This probably happens because: (i) the walking sequences in Dataset-B are captured in an indoor environment and Dataset-A are captured in an outdoor environment, (ii) the walking sequences in Dataset-A have alternated directions in each sequence and in Dataset-B all walking sequences are right-to-left, (iii) the Dataset-B is significantly bigger than the Dataset-A.

The second experiment utilizing the CASIA Gait Dataset-B was carried out with the goal of analyzing the influence of clothing in gait recognition. For this experiment were utilized the video sequences in which the individuals walk in the lateral direction, in normal conditions and wearing a coat. Table 3 shows the result of rank-1 accuracy rate for the proposed method (with its two variations) and the [Lima and Schwartz 2019] method. It's possible to observe that the results were inferior to the results presented in Table 2, showing that clothing variation can affect the gait recognition performance.

Table 3. Rank-1 Accuracy - CASIA Gait Dataset-B (Normal+Wearing a Coat).

Method	Lateral
[Lima and Schwartz 2019]	95.16%
Our method (Euclidean)	86.29%
Our method (Chi-square)	89.72%

7. Conclusion and Future Work

The results showed in this work are preliminary and the proposed method still have room for improvements. However, they have already proved to be competitive with other state-of-art methods and indicate that the angular variation of the limbs in gait sequence combined with the distance to the neck point can encode sufficient information about the gait signature to obtain good results in gait recognition.

The method proposed in this work shares some ideas with the method proposed in [Lima and Schwartz 2019], that obtained the best results in all carried out experiments. Both methods utilize 2D poses and histograms as gait descriptors, however in the best results the method proposed in [Lima and Schwartz 2019] utilizes two histograms for each key point of the detected skeleton (one histogram for the horizontal coordinate and other for the vertical coordinate) totaling 24 histograms with 85 bins each (that results in a 2040-dimensional feature vector), while our method utilizes two histograms for each limb part totaling 16 histograms with 16 bins each (that results in two 128-dimensional feature vector - one for distances and other for angles). This way, the method proposed in this work leads to a significant dimensionality reduction, improving the computational performance without losing much accuracy.

All conducted experiments showed that choosing an adequate distance function is essential to the method performance and the limb obstructions occurred in pose estimation step directly affects the classification accuracy. In this context, for future work we aim to assess more distance functions and improve the pose estimation process by adding a pre-processing phase in order to reduce noise and occlusions effects.

Acknowledgements. This paper is a result of the research of the scientific initiation conducted with the support of the Institutional Program for Scientific and Technological Initiation Scholarships (PIBIC) sponsored by National Council for Scientific and Technological Development (CNPq). This work was also sponsored by Petrobras/Fundunesp (Process 2662/2017) and the São Paulo Research Foundation (FAPESP), process 2020/14420-5.

References

Arantes, M. and Gonzaga, A. (2011). Human gait recognition using extraction and fusion of global motion features. *Multimedia Tools and Applications*, pages 655–675.

Cao, Z., Hidalgo, G., Simon, T., Wei, S., and Sheikh, Y. (2018). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *CoRR*, abs/1812.08008.

Chen, C., Liang, J., Zhao, H., Hu, H., and Tian, J. (2009). Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*.

Jain, A. K., Ross, A. A., and Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media.

Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu (2003). Silhoutte analysis based gait recognition for human identification. *IEEE trans Pattern Analysis and Machine Intelligence(PAMI)*, 25(12):1505–1518.

Lima, V. C. d. and Schwartz, R. (2019). Gait recognition using pose estimation and signal processing. *Iberoamerican on Pattern Recognition (CIARP 2019)*, 11896:719–728.

Liu, D., Ye, M., Li, X., Zhang, F., and Lin, L. (2016). Memory-based gait recognition. *British Machine Vision Conference 2016 (BMVC 2016)*, pages 82.1–82.12.

Marín-Reyes, P. A., Lorenzo-Navarro, J., and Castrillón-Santana, M. (2016). Comparative study of histogram distance measures for re-identification. *arXiv preprint arXiv:1611.08134*.

Nixon, M. S. and Carter, J. N. (2006). Automatic recognition by gait. *Proceedings of the IEEE*, 94(11):2013–2024.

Nixon, M. S., Tan, T. N., and Chellappa, R. (2006). *Human Identification Based on Gait*. Springer Science & Business Media.

Wan, C., Li, W., and Phoha, V. V. (2018). A survey on gait recognition. *ACM Computing Surveys*, 51(89):5–40.

Wang, L., Tan, T., Hu, W., and Ning, H. (2003). Automatic gait recognition based on statistical shape analysis. *IEEE transactions on pattern analysis and machine intelligence*, 12(9):1120–1131.

Yu, S., Tan, D., Huang, K., and Tan, T. (2007). Reducing the effect of noise on human contour in gait recognition. *International Conference on Biometrics*, pages 338–346.

Yu, S., Tan, D., and Tan, T. (2006). A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 4, pages 441–444.