

A New Method for Gait Recognition Using 2D Poses

Daniel Ricardo dos Santos Jangua
UNESP - São Paulo State University
Bauru, Brazil
daniel.jangua@unesp.br

Aparecido Nilceu Marana
UNESP - São Paulo State University
Bauru, Brazil
nilceu.marana@unesp.br

Abstract—Over the last decades, biometrics has become an important way for human identification in many areas, since it can avoid frauds and increase the security of individuals in society. Nowadays, most popular biometric systems are based on fingerprint and face features. Despite the great development observed in Biometrics, an important challenge lasts, which is the automatic people identification in low-resolution videos captured in unconstrained scenarios, at a distance, in a covert and non-invasive way, with little or none subject cooperation. In these cases, gait biometrics can be the only choice. The goal of this work is to propose a new method for gait recognition using information extracted from 2D poses estimated over video sequences. For 2D pose estimation, our method uses OpenPose, an open-source robust pose estimator, capable of real-time multi-person detection and pose estimation with high accuracy and a good computational performance. In order to assess the new proposed method, we used two public gait datasets, CASIA Gait Dataset-A and CASIA Gait Dataset-B. Both datasets have videos of a number of people walking in different directions and conditions. In our new method, the classification is carried out by a 1-NN classifier. The best results were obtained by using the chi-square distance function, which obtained 95.00% of rank-1 recognition rate on CASIA Gait Dataset-A and 94.22% of rank-1 recognition rate on CASIA Gait Dataset-B, which are comparable to state-of-the-art results.

Index Terms—biometric, gait recognition, pose estimation, human identification.

I. INTRODUCTION

Over the last decades, Biometrics, that consists in the statistical study of physical or behavioral characteristics [1], has become an important tool for human identification in many areas since it can avoid frauds and increase the security of individuals in society. The most usual biometrics systems that have been deployed are based on fingerprint or face traits, which are biological characteristics, harder to imitate than the behavioral characteristic like voice and gait.

However, despite the great development observed in Biometrics, an important challenge lasts, which is the automatic people identification in low-resolution videos captured in unconstrained scenarios, at a distance, in a covert and non-invasive way, with little or none subject cooperation. In these cases, gait biometric characteristics can be the only choice.

Gait can be defined as motor behaviors composed by repetitive and integrated movements of the human body that form a pattern of corporal movements that repeat in each cycle [2]. Researches conducted in the last decades show that each

individual has a special and distinct way of walking [3]. In that context, gait recognition gains relevance due to its advantages in comparison with classical biometrics methods: (i) it can be executed at a distance; (ii) it presents a good classification performance even with low-resolution images; (iii) it does not depend on subject's cooperation; (iv) occlusions do not interfere so much on performance [4].

In gait analysis there are two different approaches for characteristics representation. The first one is based on the silhouette analysis, working mostly with static aspects. The second one uses a spatio-temporal model. Despite demanding more computational effort, the model-based methods present higher reliability and better classification performance because they work with dynamic aspects of gait [2]. For a model-based approach, it is necessary a robust pose estimator that can be utilized as a part of feature extraction, capable of estimate the position of the individual's skeleton joints in video with a good reliability. For that, one can use the OpenPose [5] algorithm that utilizes Part Affinity Fields (PAFs) to learn how to associate parts with individuals that are detected within an image.

The goal of our work is to propose a new method for modeling the human gait over the frames of a video, analyzing, at each frame of the video, the angles and distances of the individual's body parts to the neck position and building a signal that represents how each body part behave during the gait cycles. For this analysis, we utilize the body parts extracted from 2D human pose estimated by OpenPose [5]. Experiments were made utilizing the public datasets CASIA Gait Dataset-A [6] composed by 20 subjects with 12 video sequences each (4 sequences for each camera position: frontal, lateral and oblique) and CASIA Gait Dataset-B [7] composed by 124 subjects walking in three different conditions (normal, wearing a coat and carrying a bag) with 11 view angles each. The results obtained showed that the proposed method is promising since they approached state-of-the-art results.

The rest of this paper is organized as follows: in Section II some related works are briefly presented. Section III discusses Human Pose Estimation, focusing on OpenPose. Section IV gives a brief introduction to gait. Section V describes the proposed approach. Section VI shows the carried out experiments in detail and Section VII draws some conclusions obtained from the results.

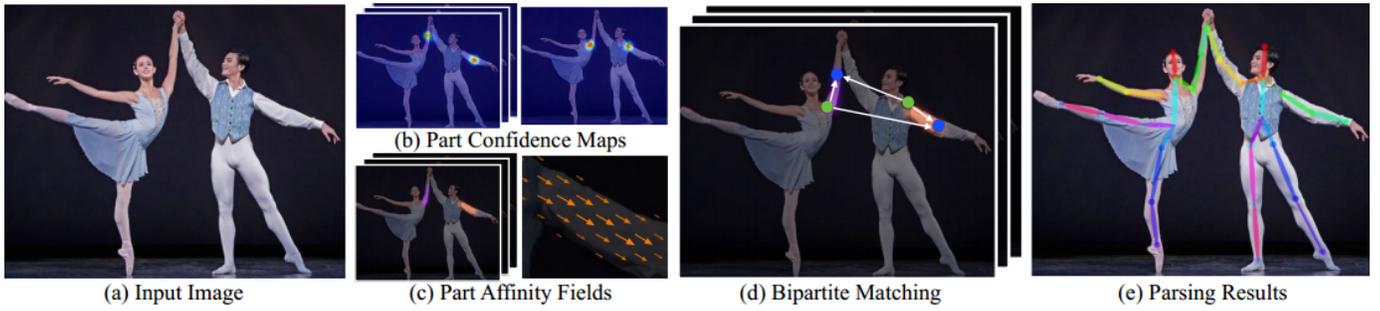


Fig. 1: Pipeline of the OpenPose method presented in [5]. The method uses the Part Confidence Maps (b) to detect the joints of human bodies in the input image and associate them using the Part Affinity Fields (c) by Bipartite Matching (d) forming poses of each individual in the image (e).

II. RELATED WORKS

In this section, some works that are related to the current proposition are briefly presented. All of them are focused on gait recognition and present results obtained on public available gait datasets.

In Wang *et al.* [8], the authors utilize a method based on silhouette analysis. The silhouette of a walking person is segmented from the video frames by a background subtraction procedure. Then, the changes of the detected silhouettes over time are represented using an associated sequence of complex vector configuration which is analyzed using the *Procrustes* shape analysis method in order to obtain a mean shape that describes all the gait sequence. On CASIA Gait Dataset-A this method achieved 90% of rank-1 accuracy in the best case and 88.75% in the worst.

In Yu *et al.* [9], a dynamic time warping (DTW) based contour similarity measure is proposed to be used in gait recognition based on silhouette analysis aiming to reduce the effect of noise on classification. On CASIA Gait Dataset-B this method achieved 83.5% of rank-1 accuracy.

In Chen *et al.* [10], the authors propose a dynamic gait representation scheme called frame difference energy image (FDEI) to work with human silhouettes even when they are incomplete. A gait cycle is divided into clusters. The FDEI of each frame is constructed using the dominant energy image (DEI) that represents a cluster. The FDEI representation can preserve the kinetic and static gait information of each frame even on incomplete silhouettes. On CASIA Gait Dataset-B the method achieved 91.1% of rank-1 accuracy.

In Liu *et al.* [11], the authors propose the use of a memory mechanism inspired by the mechanism of brain sequence processing. The 2D position of human joints are extracted using the migratory articulated human detection. Then, this information is used as input for the memory-based gait recognition (MGR) network which achieves the process of memory and identification of the gait sequence. On CASIA Gait Dataset-A this method achieved 95% of rank-1 accuracy in the best case and 85% in the worst.

In De Lima and Schwartz [12], a model-based approach is used to extract the position of the subject's joints from

each video frame utilizing a pose estimation algorithm. Then, this information is transformed into signals and movement histograms to be used as feature descriptors and the subject is classified using a 1-NN classifier, with Euclidean distance. On CASIA Gait Dataset-A this method achieved 97.5% of rank-1 accuracy in the best case and 92.5% in the worst. On CASIA Gait Dataset-B this method achieved 98% of rank-1 accuracy.

III. HUMAN POSE ESTIMATION

Human pose estimation can be described as the detection of joint points in the human body in a given image [13]. With this information it is possible to find the human limbs by connecting the joint points and, after, calculating different features from the person's limbs movement over time when the pose estimation is applied on all frames that compose a video. In our work the OpenPose [5] method was utilized to detect 2D poses.

A. OpenPose

OpenPose, proposed in [5], is a real-time method for multi-person 2D pose detection on images capable of performing detection with high accuracy and good computational performance. It is the first open-source method for real-time 2D pose detection that includes body, feet, hands and face key-points. Unlike the most common approaches that detects each subject in the input image and estimate their poses individually, OpenPose takes a bottom-up approach that treats the image globally, detecting all body parts in the input image and associating them forming each individual's pose. Figure 1 presents the pipeline of the OpenPose method.

This method gains accuracy and performance using an approach named Part Affinity Fields (PAF) that maps the position and orientation of body parts present in the image domain using 2D vectors set along with the Part Confidence Maps that represent the probability of the existence of a body part in a given pixel. By using this encoded global information it is possible to adopt a greedy approach of detection and association that allows to reduce the computational complexity without losing the confidence of the results [5].

In the pose detection process, the PAFs are iteratively improved together with the confidence maps through two

interconnected convolutional neural networks (CNN), one for the PAF and other for the confidence maps. Then, the parts are associated based on the most likely matches, forming the poses [5]. Figure 2 shows an example of 2D pose estimation using OpenPose.

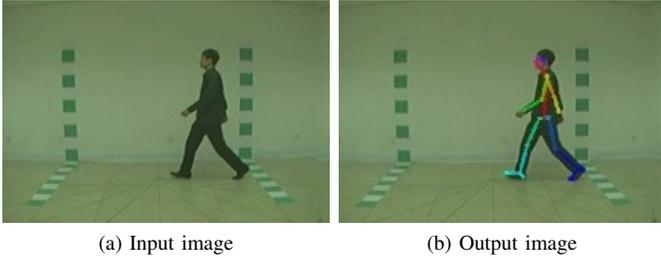


Fig. 2: Example of 2D pose estimation by using the OpenPose method. (a) Frame image of a video from the CASIA Gait Dataset-B; (b) 2D pose (each colored line represents a limb part of the individual in the image).

IV. GAIT RECOGNITION

Early studies in Medicine and Psychology have shown that human gait has some components that could be used to identify an individual [4] and indicated that every human being has a unique muscular and skeletal structure, indicating that human gait recognition is feasible. According to [4], gait has some unique properties that other biometric approaches do not have: (i) it can be captured far away and at low resolution, (ii) it can be done with simple instrumentation (*e.g.* a camera or an accelerometer), (iii) it does not need the subject cooperation, (iv) it is hard to impersonate; and (v) it works well even with partial occlusion of parts of the body.

Many studies show that gait is a periodic movement that repeats a pattern into a cycle. According to [14], a gait cycle is the time interval between successive instances of initial foot-to-floor contact and each leg has two periods, a stance phase, when the foot is in contact with the floor and a swing phase when the foot is off the ground moving forward to the next step. Figure 3 shows a gait cycle resumed in four frames from a CASIA Gait Dataset-A [6] video sequence. Inside each gait cycle the superior and inferior member of the human body realize a movement similar to a pendulum, varying its angulation in relation to the horizontal (or vertical) axis forming a pattern of angle variation.

The main hypothesis of our work is based on the results obtained by the works presented in Section II, mainly in [12], which shows that with the information of how the body members behave during the gait cycle, it is possible to determine a gait signature based on the angular variation of each limb part and this signature would keep sufficient spatio-temporal information about the gait for performing biometric identification.

V. PROPOSED METHOD

This work proposes a human identification method based on gait recognition. First, OpenPose [5] is utilized to extract the

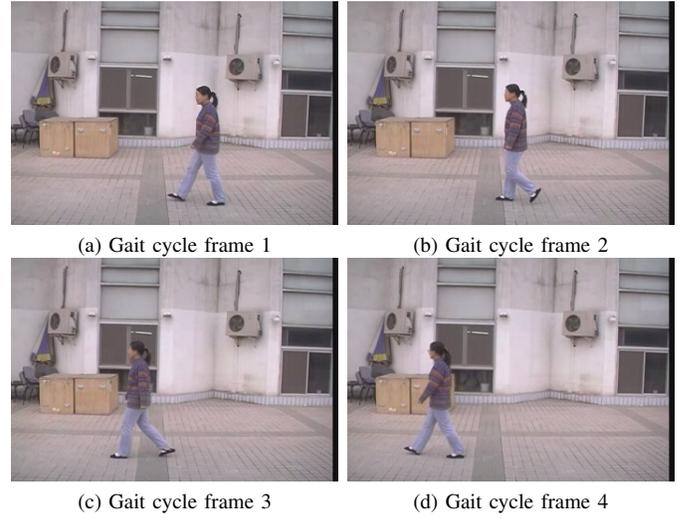


Fig. 3: Example of the gait cycle in a video sequence taken from CASIA Gait Dataset-A. The cycle begins and ends when the right heel touches the ground.

2D poses of individuals in all frames of the input video. After, for each frame, the coordinates of all joint points are utilized to calculate the angulation of each limb part in relation to the horizontal axis and the distance between the line defined by the two points representing the joints of a given limb part and the point that represents the neck. After that, these two information are utilized to build, for each limb part, over all video frames, two histograms (one for the angles and other for the distances) that are used as the gait feature vector. Finally, these feature vectors are used 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 our method.

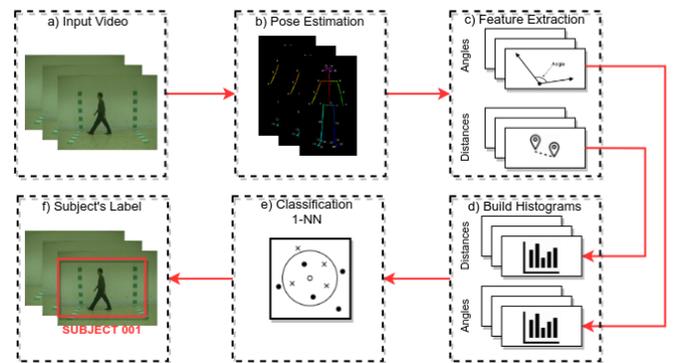


Fig. 4: Block diagram of the proposed method for gait recognition using 2D poses.

A. Pose Estimation

The pose estimation is the first step of the proposed method. In this step, the OpenPose [5] algorithm extracts, in each frame in the input video, the joints points of the person walking. The

output of this algorithm is a JSON file that contains the vertical and horizontal coordinates of each keypoint that composes the detected skeleton in each video frame. In OpenPose, it is possible to choose which skeleton type will be used in the pose estimation process. In our work, we used the BODY_25 format, which has 25 keypoints. Figure 5 shows the output of the BODY_25 format.

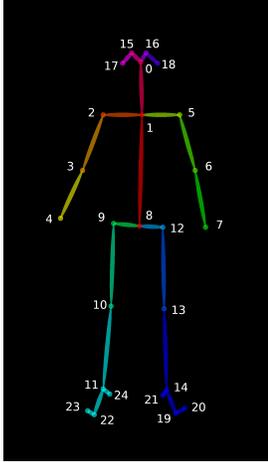


Fig. 5: Pose output format of BODY_25 [5].

B. Feature Extraction

After the pose estimation, we select the most important keypoints to be used in the feature extraction step, that is, the keypoints that encode more gait information: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13 and 14. These keypoints define the left and right arms, the forearms, the legs and the thighs, according to the BODY_25 format (Figure 5).

For each limb part, we build a sequence formed by the angles of that limb part in relation to the horizontal axis in each frame of the video sequence. For this calculation, given two keypoints $P_1 = (x_{P1}, y_{P1})$ and $P_2 = (x_{P2}, y_{P2})$ that form a limb part, we consider the member as a 2D vector $w = (x_1, y_1)$ where $(x_1, y_1) = (x_{P1} - x_{P2}, y_{P1} - y_{P2})$ and find the angle φ between it and the vector $(x_2, y_2) = (1, 0)$ utilizing the Equation 1.

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

Analogously to the angle sequence, the distance sequence for each limb part is formed by the distance d between the straight line defined by the two limb part's keypoints and the keypoint that represents the neck (keypoint 1 in Figure 5), in each frame. Considering the vector $v = P_{neck} - P_2$, in which P_{neck} is the neck point and P_2 is a keypoint that forms the member in question, we can use equations 2 and 3 to do this calculation:

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

$$d = \|v - Proj_w v\| \quad (3)$$

C. Gait Histograms

With the sequences of angles and distances for each limb part in the video sequence, we build two histograms: one histogram for angles and other for distances. For both histograms, we use 16 bins, a parameter value found empirically. As our method considers eight limb parts, we have eight angle histograms and eight distance histograms, with 16 bins each.

The angle histograms are defined in the interval $[0, \pi]$, because the possible angle vary between 0 and π . 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.

Finally, we concatenate all angle histograms forming one 1-dimensional angle feature vector and do the same for the distance histograms. So, at the end, our gait descriptor is composed of two histograms (angles and distances) of 128 bins each.

D. 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 VI.

Given a distance function, we calculate the distance d_1 between the angle histograms of the probe (query) and gallery (database) videos. Then, we calculate the distance d_2 between the distance histograms of the probe and gallery videos. The final distance between the probe and the gallery videos is $d_d = (d_1 + d_2)/2$. As both histograms, angles and distances, are normalized, there is no need to normalize the distances d_1 and d_2 .

VI. EXPERIMENTAL RESULTS

In order to assess the proposed new method for gait recognition, we carried out experiments on two gait datasets, CASIA Gait Dataset-A [6] and CASIA Gait Dataset-B [7].

The CASIA Gait Dataset-A, created on 2001, includes 20 subjects, each one with 12 video sequences, 4 sequences for each of the three directions: 90, 45 and 0 degrees to the camera position, that represents the lateral, oblique and frontal view of the person walking, respectively.

The CASIA Gait Dataset-B, created in 2005, has 124 individuals walking in three different conditions: normal, wearing a coat and carrying a bag. Figure 6 shows an example of these variations. For each walking sequence there are 11 view angles varying from 0 to 180 degrees.

In our first experiment, utilizing the CASIA Gait Dataset-A, we applied our method of gait recognition with the Euclidean and chi-square distance functions and compared their performances using the Cumulative Matching Characteristic (CMC) curve using the mean accuracy obtained for the three different directions (totaling 240 walking sequences, 80 for



Fig. 6: Example of the walking condition variation on CASIA Gait Dataset-B [7] video sequences.

each direction). Figure 7 shows the CMC curves obtained in this experiment. One can observe that the chi-square distance function obtained a better performance. This result corroborates other studies that indicate that chi-square function is a good metric for histogram comparison [15].

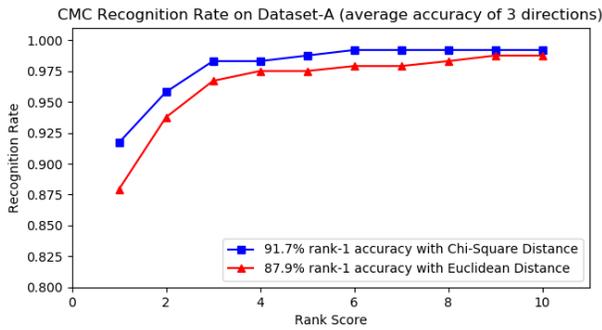


Fig. 7: CMC curves obtained by using the Euclidean and chi-square distance functions on CASIA Gait Dataset-A. Chi-square function obtained a better result.

TABLE I: Rank-1 Accuracy - CASIA Gait Dataset-A

Method	Lateral	Oblique	Frontal
Wang et al. [8]	88.75%	87.50%	90.00%
Liu et al. [11]	85.00%	87.50%	95.00%
De Lima and Schwartz [12]	92.50%	96.25%	97.50%
Our method (Euclidean)	80.00%	87.50%	96.25%
Our method (Chi-square)	87.50%	92.50%	95.00%

Table I shows the rank-1 accuracy values obtained by three methods presented in Section II ([4], [11], [12]), including two recent methods that can be considered state-of-the-art ([11] and [12]), and also by our method (using both distance functions) for each position of the camera in the CASIA Gait Dataset-A. One can observe that all methods showed better results when the person is in the front position to the camera. This should be because in this angle there is more information about the gait signature, mainly because there is no limb occlusions. One can also observe that our method, with the chi-square distance function, was superior to the method by Wang et al. [8] and was competitive with the state-of-the-art methods proposed by Liu et al. [11] and De Lima and Schwartz [12].

In another set of tests, we used the CASIA Gait Dataset-B, that is significantly bigger than the CASIA Gait Dataset-A

and has walking sequences that presents variation on clothing and carrying conditions. For the first test with this dataset, we utilized only the walking sequences in the lateral direction (90 degrees to the camera position) and in normal walking condition (totaling 744 walking sequences), and calculated CMC curves for our method using both distance functions. The result of this test is presented in Figure 8. It is possible to notice that, again, the chi-square distance function showed better results than Euclidean distance function.

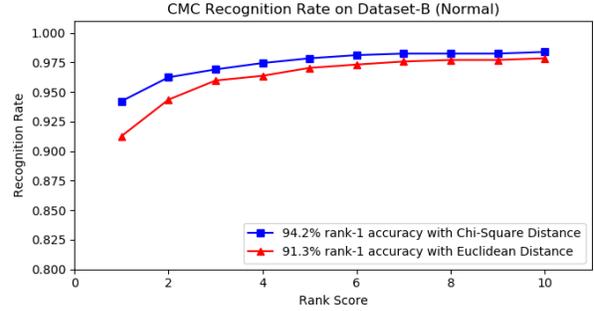


Fig. 8: CMC curve obtained by using the Euclidean and chi-square distance functions on CASIA Gait Dataset-B utilizing only the lateral view position and normal walking sequences. Chi-square function obtained a better result.

TABLE II: Rank-1 Accuracy - CASIA Gait Dataset-B (Normal)

Method	Lateral
Yu et al. [9]	83.50%
Chen et al. [10]	91.10%
De Lima and Schwartz [12]	98.00%
Our method (Euclidean)	91.26%
Our method (Chi-square)	94.22%

Table II shows the rank-1 accuracy values obtained by three methods presented in Section II ([9], [10], [12]), including one very recent method that can be considered state-of-the-art ([12]), and also by our method (using both distance functions) for the lateral position of the camera in the CASIA Gait Dataset-B. Again, although our method did not get the highest rank-1 accuracy rate, it obtained (with both distance functions) higher results than the methods proposed by Yu et al. [9] and Chen et al. [10]. In general, the tests with CASIA Gait Dataset-B showed better results considering the lateral view, this probably happens because: (i) the walking sequences in CASIA Gait Dataset-B are captured in an indoor environment and CASIA Gait Dataset-A are captured in an outdoor environment, (ii) the walking sequences in CASIA Gait Dataset-A have alternated directions in each sequence and in CASIA Gait Dataset-B all walking sequences are right-to-left, (iii) the CASIA Gait Dataset-B is significantly bigger than the CASIA Gait Dataset-A.

The second test utilizing the CASIA Gait Dataset-B was carried out with the goal of analyzing the influence of clothing

in gait recognition. We used the video sequences in which the individuals walk in the lateral direction, in normal conditions and wearing a coat. The rank-1 accuracy rates obtained by our method (with Euclidean and chi-square distance functions) and by De Lima and Schwartz’s method [12] are presented in Table III. One can observe that in this case, the three results were inferior to the results presented in Table II, indicating that variations in clothing may interfere in the gait recognition.

TABLE III: Rank-1 Accuracy - CASIA Gait Dataset-B (Normal+Wearing a Coat)

Method	Lateral
De Lima and Schwartz [12]	95.16%
Our method (Euclidean)	86.29%
Our method (Chi-square)	89.72%

We observe that our method share some ideas with the method proposed by De Lima and Schwartz [12], 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 by De Lima and Schwartz [12] utilizes two histograms for each keypoint 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). As the number of limb parts are lower, our method leads to a significant reduction in the dimensionality of the feature vectors and, consequently, improves the computational performance, while keeping comparable accuracy rates.

VII. CONCLUSIONS AND FUTURE WORK

The results obtained by our method are preliminary and still have room for improvements. They 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 main advantage of our method is that compared with the method proposed by De Lima and Schwartz [12], for example, it presents a better computational performance because it defines a more compact gait signature information.

In all conducted experiments that confronted the Euclidean and chi-square distance functions, we could observe that the use of chi-square distance function improved the method’s accuracy. This probably happens because chi-square distance function seems to suit better for histogram comparisons. As future work, we intend to assess other distance functions, such as Bhattacharyya and Intersections [15], since the choice of a good distance functions matters. We also intend to investigate the best weights to be used for calculating the average distance between the angle and distance histograms.

From our experimental results we can also infer the importance of the pose estimation step to the method’s performance,

since we use its output coordinates to map the individual’s movements during the gait cycle. A pose estimator with less detection errors is of paramount importance for the robustness of the method. For future work, we intend to assess other algorithms for 2D pose estimations, such as PifPaf [16], and other skeleton formats.

For future work, we also intend to focus on improvements of the pose estimation step, mainly in error handling and noise attenuation that, according to the tests, seems to have the higher impact in the gait recognition performance, mostly when there are variations in clothing conditions, for instance.

VIII. ACKNOWLEDGMENTS

This article 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).

REFERENCES

- [1] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to biometrics*. Springer Science & Business Media, 2011.
- [2] M. Arantes and A. Gonzaga, “Human gait recognition using extraction and fusion of global motion features,” *Multimedia Tools and Applications*, pp. 655–675, 2011. [Online]. Available: <https://doi.org/10.1007/s11042-010-0587-y>
- [3] M. S. Nixon and J. N. Carter, “Automatic recognition by gait,” *Proceedings of the IEEE*, vol. 94, no. 11, pp. 2013–2024, 2006.
- [4] C. Wan, W. Li, and V. V. Phoha, “A survey on gait recognition,” *ACM Digital Library*, 2018. [Online]. Available: <https://dl.acm.org/doi/10.1145/3230633>
- [5] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, “Openpose: Realtime multi-person 2d pose estimation using part affinity fields,” *CoRR*, vol. abs/1812.08008, 2018. [Online]. Available: <http://arxiv.org/abs/1812.08008>
- [6] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu, “Silhouette analysis based gait recognition for human identification,” *IEEE trans Pattern Analysis and Machine Intelligence(PAMI)*, vol. 25, no. 12, pp. 1505–1518, 2003.
- [7] Shiqi Yu, Daoliang Tan, and Tieniu Tan, “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)*, vol. 4, Aug 2006, pp. 441–444.
- [8] Wang L., Tan T., Hu W., and Ning H., “Automatic gait recognition based on statistical shape analysis,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 12, no. 9, pp. 1120–1131, 2003.
- [9] Yu S., Tan D., Huang K., and Tan T., “Reducing the effect of noise on human contour in gait recognition,” *Internat. Conf. on Biometrics*, 2007.
- [10] Chen C., Liang J., Zhao H., Hu H., and Tian J., “Frame difference energy image for gait recognition with incomplete silhouettes,” *Pattern Recognition Letters*, 2009.
- [11] Liu D., Ye M., Li X., Zhang F., and Lin L., “Memory-based gait recognition,” *BMVC*, 2016.
- [12] V. C. de Lima and R. Schwartz, “Gait recognition using pose estimation and signal processing,” *Iberoamerican on Pattern Recognition - CIARP*, 2019.
- [13] A. Toshev and C. Szegedy, “DeepPose: Human pose estimation via deep neural networks,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1653–1660.
- [14] Mark S. Nixon, Tieniu N. Tan, and Rama Chellappa, *Human Identification Based on Gait*. Springer Science & Business Media, 2006.
- [15] P. A. Marín-Reyes, J. Lorenzo-Navarro, and M. Castrillón-Santana, “Comparative study of histogram distance measures for re-identification,” *arXiv preprint arXiv:1611.08134*, 2016.
- [16] S. Kreiss, L. Bertoni, and A. Alahi, “Pifpaf: Composite fields for human pose estimation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 977–11 986.