Gait Analysis for Gender Classification in Forensics

Paola Barra[1], Carmen Bisogni[1], Michele Nappi[1], David Freire-Obregón[2], and Modesto Castrillón-Santana[2]

[1] University of Salerno, Dipartimento di Informatica, Salerno, Italy 84084
[2] Universidad de Las Palmas de Gran Canaria (ULPGC), Gran Canaria, Spain 35000

Abstract. Gender Classification (GC) is a natural ability that belongs to the human beings. Recent improvements in computer vision provide the possibility to extract information for different classification/recognition purposes. Gender is a soft biometrics useful in video surveillance, especially in uncontrolled contexts such as low-light environments, with arbitrary poses, facial expressions, occlusions and motion blur. In this work we present a methodology for the construction of a gait analyzer. The methodology is divided into three major steps: (1) data extraction, where body keypoints are extracted from video sequences; (2) feature creation, where body features are constructed using body keypoints; and (3) classifier selection when such data are used to train four different classifiers in order to determine the one that best performs. The results are analyzed on the dataset Gotcha, characterized by user and camera either in motion.

Keywords: Gender Classification; Gait Analysis; Supervised Learning; SVC; Random Forest; AdaBoost; KNN

1 Introduction

Gender Classification (GC) is a natural ability that belongs to the human beings. In computer vision, GC can help increasing information when recognizing humans. Differently from other biometrics, soft biometrics take into account also behavioural traits during the recognition process; this allows us to use soft biometrics also with occlusions like clothes, backpacks, scarfs and others. Gender is a soft biometrics useful in video surveillance [3], especially in uncontrolled contexts such as low-light environments, with arbitrary poses, facial expressions, occlusions and motion blur.
A good method for feature extraction is important in computer vision. A lot of methods for feature extraction in computer vision has been used through the years, an example is the prediction on the facial landmark [1], raw pixels [8] the use of Haar-like wavelets [6] or Gabor wavelets [7]. Face Landmarks detectors were not preferred because in the contexts of video surveillance and forensics they often produce low quality data. In fact, unfavorable conditions are often observed for these data. Intentional and unintentional occlusions such as hair,

heavy makeup, masks or hands make feature extraction very difficult. In addition, low light conditions and shadows introduce noise and artifacts in images compromising the effectiveness of the method. On the other hand, the use of body information has shown to be most effective to this aim.
The human pose extracted from software like Openpose [5] also works properly with low camera resolution and long distance acquisitions. That kind of extraction is non-invasive and requires no collaboration from subjects. Furthermore, body features may be also acquired in the dark by means of infrared cameras or when the subject shows his back to the camera; in addition to the pose of the subject we are able to extract also soft biometrics like gait.
In this work we present a methodology for the construction of a gait analyzer. The methodology is detected in three major steps: (1) data extraction, where body keypoints are extracted from video sequences; (2) feature creation, where body features are extracted using body keypoints; and (3) classifier selection when such data are used to train four different classifiers in order to determine the best GC system. The problem has been faced using a geometric approach. As in [1], the algorithm uses skeleton information the algorithm uses the skeleton information for the extraction of the features of each frame of the video sequence. Then the features have been given in input to four classifiers, i.e. k-nearest neighbors, AdaBoost, support vector and random forest classifiers. Different subjects from different frames of video sequences has been considered.

The paper is organized as follows: in the following section we explore GC, when the human pose is considered. In section 3 we present we explain the particularity of Gotcha dataset on which the experiments were carried out. In section 4 we introduce the description of the method used. In Section 5 can be found experimental settings and experimental results. Finally, in Section 6 we present our conclusions.

2 Related work

GC using biometric traits has gained great interest in recent years. Divate and Ali [16] showed that finding a person's gender using biometrics like fingerprints, iris, face, gait, etc, is of interest in areas like forensics, video surveillance and security.
In the previous section we mentioned some limitations of a GC system based on face recognition for applications in security and foresics. In order to justify our choice of not using this biometric trait we will discuss some work on this field.
For example Ngan et al. in [17] presents a very rich work on gender classification based on face recognition, the experiments were carried out on several datasets under different conditions, i.e. ethnicity, age and "in the wild" context.
Castrillón-Santana in [18] perform GC on different datasets, in particular on GROUPS [20] a very challenging dataset as it contains cropped images, low resolutions and cluttered background; they analyze and compare different descriptors, selected the best and used the best configuration in order to obtain

better performances. Guo et al. [12] focus this problem with respect to the age of the subjects. Studying empirically GC using faces from a large database, they showed that the classification performs worse on children and elders. Recently, this has led some researchers to solve the problem by classifying children separately [13]. Another problem handled by researchers using faces regards ethnic aspects. Trying to mitigate the natural difference in images, [11] considered an Active Shape Model (ASM) for face texture normalization to overcome the non-uniformity. Their method achieved better accuracy and robustness with images in a multiethnic environment.
Classification algorithms can be divided in two main groups: features algorithms and Neural Networks (NN) algorithms. In the first area of research, the preprocessing of data plays a key role.

In this work we study GC from full body image sequences, according to the literature the first to address this problem were Cao et al. [21]; they face the first difficulties encountered with clothing, the difference between images taken from frontal or back views, the variety of different characteristics between different bodies, this approach has made it possible to classify the gender by considering all the features represented by the human body.

Wu and Guo in their work [19] analyze the advantages of classifying a person's gender from the bone structure of the body and in particular they make a systematic study of some critical issues in body-based GC, such as how informative each body part is, how many body parts are needed, what are good representations of body parts, and how accurate the GC system can achieve in challenging, unconstrained, real-world images. Connor and Ross [2] presented a survey on modern advanced techniques of human gait features extraction. They consider different methods by dividing temporal features in static and dynamics, model-based and model-free visual features; ground reaction force-based and finely resolved underfoot pressure features; wearable sensor features; and acoustic features. In [22] also silhouette bodies are extracted, but subsequently are divided into six regions. Features are extracted by a 2D wavelet transform and then classified by the K-Nearest Neighbor classifier.
Another use of the K-Nearest Neighbor classifier can be found in [23]; the authors use three different types of features; Spatio-Temporal Model, Leg Motion Detection, and Statistical Wavelet Model. Once the features are extracted, they use two different classification methods: the above-mentioned K-Nearest Neighbor classifier and the Support Vector Machine (SVM).
An alternative to SVM is presented in [24]; the authors treat every frame as a labeled instance, and replace SVM with linear discriminant analysis in conjunction with the Bayes rule. This method, however less accurate than NN methods, allows one to operate also with partially occluded gait cycles.
Some other approaches connect the GC task with the human identification. In [25] the authors use wavelet 5/3 lifting scheme to obtain the silhouettes and other simple features. Then the features are classified by using a C4.5 algorithm to generate a decision tree, so the latter is defined as a statistical classifier.

The method presented in [26] is suitable only for smartphones because it uses the built-in sensors like accelerometers and gyroscopes; the features obtained by means of these sensors are extracted by the histogram of gradient method and later classified by a bootstrap aggregating method that trains each classifier on a random redistribution of the training set.
One of the most preferred method in biometrics of recent years are the neural networks (NN). Neural networks have become increasingly important in the last decade in this field thanks to their ability to adapt to different environments and biometric traits. However, neural networks are not yet widely used in the specific field of gait analysis, probably due to the huge difference between human gait and other kinds of biometric traits, since some adaptation is required for NN to work properly.
An example represented by [27], in which a Convolutional Neural Network (CNN) has been modified for gait-based GC. The authors start from a pre-trained VGGNet-16, substituting the Softmax function in the last level of the network for the SVM. They developed also another model, called VGGNet-SVM using a hinge loss function using an L2 norm. They shown that the SVM perform better than Softmax in the last layer of a CNN suited for gender gait classification.
Another approach is presented in [4], using SVM and NN; the authors obtain a Gait Energy Image (GEI) normalizing and averaging all the silhouette images in one gait cycle for all the subjects and, after reducing its dimension, five spatio-temporal parameters are calculated and concatenated to the GEI image. The resulting features vectors are used to train and test the NN together with the SVM.

3 Gotcha Dataset

For training and experimentation purposes we used the Gotcha dataset. Gotcha dataset was firstly presented in [1]. This dataset contains different users behaviours, different environment settings and camera in motion. Related to the users behaviour, two subset can be distinguished: cooperative and non-cooperative behaviour. In cooperative video sequences, the subjects look at the camera during the acquisition and follow the camera lens during the motion, as in Figure 1-a left. In non-cooperative video sequences, the subjects try to avoid the camera during motion recording. This modality is clearly most challenging, as in Figure 1-a right. Related to the enviroments setting, we have two subsets of video: indoor with artificial lights and outdoor with natural lights. Indoor modality in turn is divided into videos with lights on and videos with lights off and flashlight. In conclusions, the choice of recording with a camera in motion makes the dataset very innovative, by simulating wearable cameras which the policeman are provided with [15]. Therefore, the features of this dataset are very suitable to simulate the video surveillance settings.
The acquisition device considered to capture videos of subjects was a Samsung

S9+. The subjects were free to wear any clothes and accessories they wants, in order to simulated as much as possible a real-world condition.
Summing up, the modalities are the following subset distinctions:

- (1) Cooperative mode.
- (2) Non-cooperative mode.
- (a) Indoor with artificial light.
- (b) Indoor without any lights but the camera flash, as in Figure 1-b .
- (c) Outdoor with sunlight.

However, as in real world, some problem can occur with particular illumination setting. For instance, the use of camera flash can generate blur frames in some sequences (see Figure 1-b). These kind of problems could affect facial recognition, are not a huge problem in gait analysis.
The dataset consists in 62 subjects, 15 females and 47 males. Each subject was therefore recorded in 6 different modalities for a total of 372 video sequences.
In this work we used all the 62 subjects for the classification.

4 Proposed method

In this section, we address the GC problem through the use of a geometric technique based on the poses extracted from the human gait. The pipeline is divided into three major steps: (a) data extraction, (b) feature creation and (c) classifier selection.

4.1 Data extraction

OpenPose is used to detect the body pose, it is defined as a real-time multi-person system to jointly detect human body keypoints on single images [9]. We make use of this system to extract the necessary data for the feature selection. As shown in Figure 2, all the 18 body keypoints are placed all over the subject body. These points provide the necessary data to generate our features.

4.2 Feature creation

Each examined video is composed of 200 frames; and from each frame 18 keypoints were extracted. Starting from the keypoints, the distance for each pair of points was calculated, resulting in 153 distances for each frame. If one or more keypoints were not present, the distance is first set to zero and then approximated with an interpolation with neighboring distances.

The gait for each video is described by 30,600 features (153 distances of keypoints * 200 frames).

(a) Outdoor cooperative and indoor non-cooperative samples with sunlight and artificial light respectively.

(b) Indoor cooperative with the camera flash, blurred frames.

Fig. 1: Some frames from Gotcha Dataset.

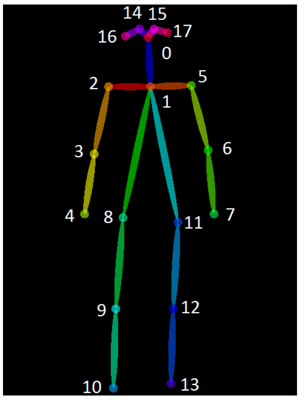

Fig. 2: Body keypoints extracted by OpenPose.

4.3 Classifier selection

The classifier is constructed by taking 30 subjects, 15 women and 15 men. For each of the 62 subjects a different classifier is thus constructed, the 15 women are always the same (as there are only 15 women in the dataset) and the 15 men are chosen randomly from time to time.

A seed = 1 was used which is increased by 1 at each iteration, so as to allow a different configuration for each classifier but at the same time repeatable.

For these experiments we used several known classifiers to study which of these is more efficient for the purpose, and with what parameters in order to allow repeatability.

Random Forest classifier. We considered a statistical classifier such as Random Forest (RF). Instead of relying on a single model, RF generates a collection of decision trees, and then the mode of the predictions from the trees is used as the model output. An important advantage is that RF does not tend to overfit if the maximum depth of the tree is limited, leaving the most important features on the top of the tree and the more specific features near the leaves. The maximum depth represents the depth of each tree in the forest. The deeper the tree, the more splits it has and it captures more information about the data.
Parameters:
- number of trees in the forest = 100;
- bootstrap samples are used when building trees;
- the function to measure the quality of a split is set entropy.

K-Nearest Neighbors classifier (KNN) is a supervised learning classification algorithm. The KNN algorithm classifies the data based on "how similar" they are. When a KNN algorithm receives an unclassified data, it measures its distance from classified data. It collect the K smaller distances and classifies the new

data as the class of the most frequent data having these K distances. Here the parameters chosen for the experimental settings:
- number of neighbors = 5;
- weight function used in prediction is *distance* : weight points by the inverse of their distance. In this case, closer neighbors of a query point will have a greater influence than neighbors which are further away;
- algorithm used to compute the nearest neighbors is *auto*: will attempt to decide the most appropriate algorithm based on the values passed to the fit method;
- power parameter for the Minkowski metric is the *euclidean distance*.

4.4 Support Vector classifier

Support Vector Classification (SVC) uses support vector machine, a known type of supervised machine learning classification algorithm.

Adaptive Boosting classifier (AdaBoost) is an algorithm that try to solve the classification problem converting a set of weak classifiers in a strong one. Starting from the minimum error in the weak classifier, AdaBoost assigns to each classifier a positive or negative weight, depending on the correct or incorrect classification of the selected classifier. At every step, misclassified cases would be updated with larger weights after an iteration and vice versa. More recently, it has been found that AdaBoost was minimizing an exponential loss function, thus this algorithm can be seen as a stagewise estimation procedure for fitting an additive logistic regression model [28].

5 Experiments

Regarding the conducted experiments, we have split our dataset into 70% of the validation/training set and 30% of the test set. The gender balance was not affected by this distribution.

Since the difficulty of the Gotcha dataset lies in the strong diversity of video acquisitions, we have divided the results into two tables. Table 1 summarizes the results related to cooperative videos, while Table 2 presents the results related to non cooperative videos. The accuracy is the percentage of videos correctly classified.

As we can see in the results, Random Forrest Classifier is the best performed approach for our purpose.
The results are very promising, even considering the difficulties of the Gotcha dataset.

6 Conclusions

The presented paper provides a gait analysis approach for Gender Classification. Our proposal achieves good results and prevents the overfitting by adjusting the

Classifier	***1a***	***1b***	***1c***	***tot.***
RF	**0.807**	**0.825**	**0.680**	**0.771**
KNN	0.691	0.741	0.631	0.688
AdaBoost	0.774	0.809	0.597	0.7271
SVC	0.741	0.777	0.639	0.719

Table 1: Results related to video cooperatives:
- column (1a) reports the accuracy results for cooperative indoor with artificial light videos;
- column (1b) reports the accuracy results for cooperative indoor videos without any lights but with camera flash;
- column (1c) reports the accuracy results the results for cooperative outdoor videos;
- column (tot.) reports the accuracy results for all cooperative videos.

Classifier	***2a***	***2b***	***2c***	***tot.***
RF	**0.755**	**0.779**	**0.681**	**0.738**
KNN	0.658	0.695	0.626	0.660
AdaBoost	0.720	0.774	0.602	0.699
SVC	0.666	0.694	0.627	0.662

Table 2: Results related to video non-cooperatives.
- column (2a) reports the accuracy results for non-cooperative indoor with artificial light videos;
- column (2b) reports the accuracy results for non-cooperative indoor videos without any lights but with camera flash;
- column (2c) reports the accuracy results for non-cooperative outdoor videos;
- column (tot.) reports the accuracy results for all non-cooperative videos.

RF depth. However, the proposed results prove that gait can be employed to classify gender at a distance with mobile devices. The information about gait analysis can be utilized also in forensic field and it can be acquired from non-frontal subjects, in the dark or with facial occlusions.

In combine this technique with other classification techniques such as neural networks and on other datasets known to the state of the art as CASIA Gait Database [14], OU-ISIR Gait Database, Multi-View Large Population Dataset (OU-MVLP) [29], The TUM-GAID Database.

References

1. Barra Paola, Bisogni Carmen, Nappi Michele, Freire-Obregn David Castrillon-Santana Modesto. (2019). Gender classification on 2D human skeleton. 1-4. 10.1109/BIOSMART.2019.8734198.
2. Connor, Patrick, Ross, Arun. (2018). Biometric Recognition By Gait: A Survey of Modalities and Features. Computer Vision and Image Understanding. 167. 10.1016/j.cviu.2018.01.007.
3. Neves, J., Narducci, F., Barra, S. et al. Artif Intell Rev (2016) 46: 515. https://doi.org/10.1007/s10462-016-9474-x
4. S. Choudhary, C. Prakash and R. Kumar, "A hybrid approach for gait based gender classification using gei and spatio temporal parameters," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1767-1771. doi: 10.1109/ICACCI.2017.8126100
5. arXiv:1812.08008
6. G. Shakhnarovich, P. Viola and B. Moghaddam. A unified learning framework for real time face detection and classification, in: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 16-19.
7. X. Leng and Y. Wang. Improving generalization for gender classification. International Conference on Image Processing, 2008, pp. 1656-1659.
8. B. Moghaddam and M. Yang. Learning gender with support faces. Pattern Analysis and Machine Intelligence, IEEE Transactions (24), 2002, pp. 707-711.
9. Z. Cao, G. Hidalgo, T. Simon, S. Wei and Y. Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. arXiv preprint arXiv:1812.08008. 2018.
10. T. Xiaohui, P. Xiaoyu, L. Liwen and X. Qing. Automatic human body feature extraction and personal size measurement. Journal of Visual Languages and Computing (46), 2018, 9–18 .
11. W. Gao and H. Ai. Face gender classification on consumer images in a multiethnic environment. Lecture Notes in Computer Science, 5558:169–178, 2009.
12. G.D. Guo, C. Dyer, Y. Fu,and T.S. Huang. Is gender recognition affected by age?. IEEE International Workshop on Human-Computer Interaction (HCI09), in conjunction with ICCV09, 2009.
13. Y. Wang, K. Ricanek, C. Chen, and Y. Chang. Gender classification from infants to seniors. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pages 1–6, 2010.
14. S. Zheng, J. Zhang, K. Huang, R. He and T. Tan. Robust View Transformation Model for Gait Recognition. Proceedings of the IEEE International Conference on Image Processing, 2011.

15. A. Cavallaro and A. Brutti. Audio-visual learning for body-worn cameras. Computer Vision and Pattern Recognition. 2019. pp. 103–119
16. Chandrakant P. Divate and Syed Zakir Ali. Study of Different Bio-Metric Based Gender Classification Systems. International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2018, DOI: 10.1109/ICIRCA.2018.8597340.
17. Mei L. Ngan, Patrick J. Grother. Face Recognition Vendor Test (FRVT). Performance of Automated Gender Classification Algorithms, 2015. dx.doi.org/10.6028/NIST.IR.8052
18. M. Castrilln-Santana, J. Lorenzo-Navarro, E. Ramn-Balmaseda. Descriptors and regions of interest fusion for in- and cross-database gender classification in the wild, 2017. Journal Image and Vision Computing. Volume 57 Issue C, January 2017. Pages 15-24 Butterworth-Heinemann Newton, MA, USA doi¿10.1016/j.imavis.2016.10.004
19. Qin Wu1 and Guodong Guo. Gender Recognition from Unconstrained and Articulated Human Body. The Scientific World Journal. Volume 2014, Article ID 513240, 12 pages http://dx.doi.org/10.1155/2014/513240
20. P. Dago-Casas, D. Gonzlez-Jimnez, Long Long Yu and J. L. Alba-Castro, "Single- and cross- database benchmarks for gender classification under unconstrained settings," 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, 2011, pp. 2152-2159. doi: 10.1109/ICCVW.2011.6130514
21. Liangliang Cao, Mert Dikmen, Yun Fu and Thomas S. Huang. Gender recognition from body. Proceeding MM '08 Proceedings of the 16th ACM international conference on Multimedia. Pages 725-728. doi¿10.1145/1459359.1459470
22. P.B. Shelke and P.R. Deshmukh. Gait Based Gender Identification Approach. 2015 Fifth International Conference on Advanced Computing Communication Technologies, IEEE, 2015. DOI: 10.1109/ACCT.2015.66
23. Azhin Sabir, Naseer Al-Jawad, Sabah Jassim and Abdulbasit Al-Talabani. Human gait gender classification based on fusing spatio-temporal and wavelet statistical features. 2013 5th Computer Science and Electronic Engineering Conference (CEEC), IEEE, 2013. DOI: 10.1109/CEEC.2013.6659461
24. Ebenezer R.H.P. Isaac, Susan Elias, Srinivasan Rajagopalan and K.S. Easwarakumar. Multiview gait-based gender classification through pose-based voting. Pattern Recognition Letters, 2018. https://doi.org/10.1016/j.patrec.2018.04.020
25. Omer Mohammed Salih Hassan, Adnan Mohsin Abdulazeez and Volkan Mjdat TRYAK. Gait-Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis. 2018 International Conference on Advanced Science and Engineering (ICOASE), IEEE,2018. DOI: 10.1109/ICOASE.2018.8548909
26. Ankita Jain and Vivek Kanhangad. Gender classification in smartphones using gait information. Expert Systems with Applications. Volume 93, 1 March 2018, Pages 257-266. https://doi.org/10.1016/j.eswa.2017.10.017
27. Taocheng Liu, Xiangbin Ye and Bei Sun. Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition. 2018 Chinese Automation Congress (CAC),IEEE, 2018. DOI: 10.1109/CAC.2018.8623118.
28. Jerome H. Friedman, Trevor Hastie and Robert Tibshirani. Additive Logistic Regression: A Statistical View of Boosting. The Annals of Statistics 28(2):337-407, 2000. DOI: 10.1214/aos/1016218223.
29. N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, " Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition ", IPSJ Trans. on Computer Vision and Applications, Vol. 10, No. 4, pp. 1-14, Feb. 2018.

30. Martin Hofmann, Jrgen Geiger, Sebastian Bachmann, Bjrn Schuller and Gerhard Rigoll. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. Journal of Visual Communication and Image Representation 25(1):195206, 2014. DOI: 10.1016/j.jvcir.2013.02.006