Gender classification on 2D human skeleton

Paola Barra
Carmen Bisogni
Michele Nappi
Dipartimento di Informatica
University of Salerno, Salerno, Italy
{pbarra, cbisogni, mnappi}@unisa.it

David Freire-Obregón
Modesto Castrillón-Santana
SIANI
Universidad de Las Palmas de Gran Canaria (ULPGC)
Gran Canaria, Spain
{david.freire, modesto.castrillon}@ulpgc.es

***Abstract*—Soft bimetrics has become a trending research topic over the past decade. In last years, the increase of new technologies such as the wearable camera devices has introduced a new challenge into the gender classification problem. In this sense, the ability to classify the gender not by an image but by the 2D estimated skeleton points is considered in this paper. Our experiments show that the human gender can be classified just considering the pose information provided by the body pose information. The proposed method have shown a remarkable performance on a dataset where subjects and camera are in movement.**

***Index Terms*—gender classification, gait analysis, supervised learning**

I. Introduction

Gender classification (GC) is one of the most important visual tasks for the human being. However, no only the social interactions depend on the correct gender perception of the parties involved, also recommendation systems can use this trait to improve the quality in their decisions. Some studies have found out that common collaborative filtering algorithms differ in the gender distribution of their recommendation lists [1].

In the past, GC primarily needed efficient feature extraction in order to feed robust classifiers. Most traditional methods use features extracted by hand as input to the the classification process. Among these traditional approaches we can find a large variety that goes from the attempt to exploit raw pixels [5] to the use of Haar-like wavelets [3] or Gabor wavelets [4]. However, these traditional approaches required an accurate face extraction. Moreover, most of these works agree that the distinct characteristics of the human face introduced a significant noise (facial shadows, presence of hair, skin color, and so on) which may result in degrading the robustness of the GC methods.

For the last years, pose analysis has been a very promising biometric topic. It has been widely used because of several reasons such as non invasive and it can be easily acquired at a certain distance [2]. But not only the human pose has been considered, also the human gait is gaining a increasing interest from researchers because of the possibility of extracting meaningful information of the human movement and information of the individual parts of the body even without considering the clothing and any occlusions [21].

In this work, we proposed a method to classify the gender not from the face, but from the pose estimation points extracted during the gait. We address the problem from a geometric perspective, considering several body points from different video frames. Several algorithms can be used for the body points extraction [6], [7]. In this sense, one of the most relevant work was introduced by Cao et al [6]. They proposed a powerful tool known as OpenPose. We made use of this algorithm to estimate the 2D human skeleton, in particular the different points of the body pose as a preprocessing step. Then, these points are computed to find a meaningful relation between them in order to recognize the gender of the observed subject.

Our main contribution is the use of a novel video dataset under different illumination conditions to achieve a high GC rate, no matter the environment. The paper is organized as follows: in the next section we address the state of the art, comparing different works regarding the GC when the human pose is considered. Then, Section III describes the experimental setup. Firstly it introduces the considered dataset and their main features. Secondly the algorithm pipeline is fully detailed, from the human points selection to the chosen classification algorithm. Results are presented in Section IV. Finally, Section V presents our conclusions.

II. Related work

As we stated before, automatic GC can be considered from different points of view. Perhaps, GC using face images is preferred mostly not only because face is most acceptable biometric trait, but also because there is a large variety of datasets that provides facial images.

On [20] is performed a large scale empirical evaluation of facial gender classification algorithms, with participation from five commercial providers and one university, is an interesting work because it uses large operational datasets comprised of facial images from visas and law enforcement mugshots, leveraging a combined corpus of close to 1 million images.

A remarkable work was developed by Makinen et al. [8] presented a systematic study on GC with automatically detected and aligned faces. These authors experimented with 120 combinations of automatic face detection, face alignment, and GC considering support vector machines (SVM). One of the findings was that the automatic face alignment methods

did not increase the GC rates. However, manual alignment increased classification rates a little, which suggests that automatic alignment would be useful when the alignment methods are further improved. Another main concern is the ethnic of the subject for the GC accuracy. Gao et al. [9] considered an Active Shape Model (ASM) is used for face texture normalization to overcome the non-uniformity of the images. They captured faces in real situations (varying in pose, illumination and expression) to conduct their experiment. Their method achieved better accuracy and robustness images in a multiethnic environment.

But not only ethnic background can affect a robust GC, also ageing does. Guo et al. [10] showed that GC accuracy is affected significantly by the age of the person. Their empirical studies on a large face database of 8000 images with ages from 0 to 93 years showed that GC accuracy on adult faces can be a 10% higher than that on young or senior faces. Local binary pattern (LBP) and histograms of oriented gradients (HOG) methods were evaluated for gender characterization with age variation. More recently, Wang et al. [11] addressed the same ageing issue to apply GC on children. These authors developed a robust GC system considering principal components analysis and SVM techniques over the the FG-NET database.

Some of these works have shown a remarkable performance on the GC task. However, the performance of many of these state-of-the-art face classification proposals drop under the variation of lighting, pose and other factors. Furthermore, there is a large variability in facial appearance. The texture is subject to several other factors, including the facial pose, illumination or facial expression. When the GC must be done under an uncontrolled environment, then the task becomes quite harder.

This large variability problem along with the ethnic or the ageing issues can be omitted if the gender classifier does not consider the facial cues. Hull et al. [13] proposed a supervised modeling approach for gait-based GC. To accomplish this task, they made use of traditional temporal modeling methods to learn gender through male and female gait traits. Then, these authors considered several fusion techniques to boost the GC. Moreover, Yu et al. [12] presented a numerical analysis of the contributions of different human components, which shows that head and hair, back, chest and thigh are more discriminative than other components. More recently, Arai [15] classified the human gender in a spatial temporal reasoning using CASIA Gait Database [17]. This author used a SVM classifier, the accuracy was around 97.63%.

As can be seen, these state of the art approached produced accuracies up to 92-97% but are constrained by the necessity of requiring a complete gait cycle to function properly. We propose to remove this requirement by just considering the most relevant features of the pose keypoints extracted during the gait cycle. Ebenezer et al. [16] proposed a similar technique using a pose-based voting system, treating every frame as a labeled instance. Furthermore, Martínez et al. [14] also introduced an approach for gait-based GC where key biomechanical poses of a gait pattern were represented by partial Gait Energy Images (GEIs). Their classification was based on the weighted decision fusion of the pose-based GEIs.

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

(b) Indoor with the camera flash.

Fig. 1: Some Gotcha Dataset samples in cooperative and non cooperative modes.

However, many of these works present a view dependence and it is limited to classify test sequences taken from roughly the same viewing angle as the training sequences. As it will described in the next section, our evaluated dataset does not present this limitation, subject and camera are in movement.

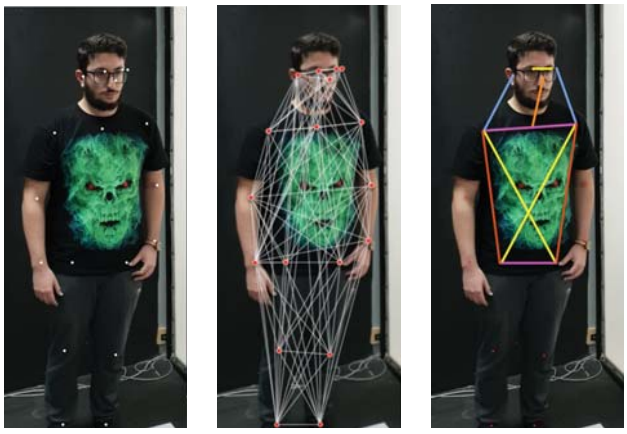
III. Experimental Setup

A. Gotcha Dataset

The considered dataset is aligned with the significant role that security plays in our society. Cameras are placed everywhere, from a nuclear reactor to the stairways in a housing block or as a wearable devices. The ”wearable” concept adds a new dimension of complexity. All the works described in Section II address the GC problem from the image based GC system. When the law enforcement uses a wearable acquisition device, there is not only one subject in movement but two (also the camera). The robustness of the classical approaches can be compromised due to this new configuration. For this reason, we have created a novel database known as Gotcha Dataset.

The acquisition device considered to capture videos of subjects was a Samsung S9+. The aim was to simulate the body-worn cameras [18]. To simulate real-world conditions, no accessories (clothes, hats or glasses) were controlled, they were left participant dependent.

About the procedure followed by each participant, there were two recording procedures: (1) a cooperative mode with the camera where the subject walks and collaborates with the camera watching it during the walk, and (2) a non cooperative



(a) OpenPose data (b) Full body features (c) Upper-body features

Fig. 2: Processed features.

mode where the same subject walks trying to avoid the camera (see Figure 1-a). In order to be able to create robust systems, three possible scenarios were considered for the previous described procedures: (a) indoor with artificial light, (b) indoor without any lights but the camera flash, and (c) outdoor with sunlight. Some real world problems can occur regarding the selected illumination settings. For instance, the use of camera flash can generate blur frames in some sequences (see Figure 1-b). These kind of problems are not easy to handle when the facial classification approaches are used.

Finally, from a demographic perspective, this work analyzes and classified the gender of 90 subjects perfectly balanced (half of them are men and half of them are women).

B. Proposed method

In this subsection, 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) the classifier selection.

OpenPose, is defined as a real-time multi-person system to jointly detect human body keypoints on single images [6]. We make use of this system to extract the necessary data for the feature selection. As can be appreciated in Figure 2-a, all these keypoints are placed all over the subjects body. This points provide the necessary data to generate our features. As we mentioned before, we have considered a subset of features based on the information provided by these keypoints.

We have conducted two different experiments. The first one considers the generated distance feature mesh, the full-body features as can be seen in Figure 2-b. However, not all the keypoints are always available, occlusions in the bottom part of the human body may happen (see Figure 1). For this reason we decided to also conduct a second experiment by generating a geometric approach focus just on the upper part of the human body, the Upper-body features as can be observed in Figure 2-c. This second experiment relies on the classical scientific assumptions about the anthropometric differences between men and women [19]. These assumptions stated that women generally have hips that are wider than the shoulders while is generally the opposite for males.

Finally, 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.

IV. Experiments

In the full-body feature experiments, 5.850 total features were extracted, each feature describes a frame. These features are divided: 3.970 (1.985 men and 1.985 women) in cooperative videos and 1.830 (915 men and 915 women) in non-cooperative videos.

In the upper-body feature experiments, 17.000 total features were extracted, each feature describes a frame. These features are divided: 3.600 (1.800 men and 1.800 women) in cooperative videos and 1.400 (700 men and 700 women) in non-cooperative videos.

There are few features if we consider the quantity of frames contained in the 90 videos taken into consideration. This number is low because only the frames containing all the body landmarks were selected and many videos frame only the upper body, so we have lost many features in full-body mode. We can also see that all frames captured in the non-cooperative mode are less than those in cooperative mode, because due to the shaking of the camera and the subject we have lost many features.

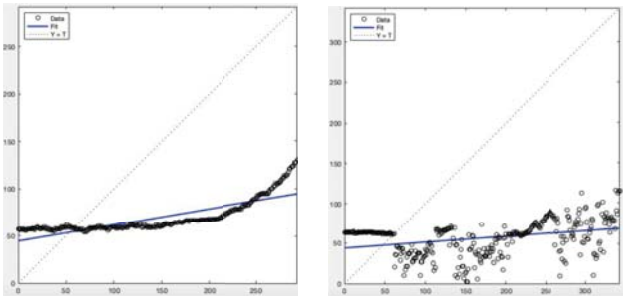
Regarding the conducted experiments, we have split our dataset into 70% of the validation/training set and 30% of the test set. The gender balance was not affected by this distribution. In the experiments we fit each decision tree with depths 4-8-12 or Not limited.

On the one hand, the results of the RF classifier with considering full-body features can be observed in Table I. On the other hand, the results of the RF classifier considering only the upper-body features can be observed in Table II.

The accuracy is the percentage of frames correctly classified on all frames.

TABLE I: Results considering full-body features. See Figure 2-b.

RF Depth	Full-Body Accuracy		
	Cooperative mode	*Non-cooperative mode*	*both*
4	0.965	0.537	0.834
8	0.979	0.543	0.797
12	0.983	0.554	0.787
Not limited	0.899	0.552	0.799



(a) Cooperative mode (b) Non-cooperative mode

Fig. 3: The measurement of the distance from the coordinate of the nose to that of the neck

In general, it can be appreciated that results are more accurated when all the distances of the body are considered. This fact can may be because even the movements of the body -such as the movement of the arms while walking- are different for men and women.

TABLE II: Results considering the upper-body features. See Figure 2-c.

RF Depth	Upper-body features Accuracy		
	Cooperative mode	*Non-cooperative mode*	*both*
4	0.833	0.598	0.737
8	0.786	0.574	0.721
12	0.736	0.566	0.714
Not limited	0.725	0.568	0.714

Furthermore, as can be seen in Table I and II, the GC accuracy depends on the collaborative mode of the subject. For instance, cooperative subjects provide a better accuracy than the non cooperative ones. This behaviour can be modeled from a mathematical perspective. The variance of the features for the non cooperative videos is higher than the cooperative ones. The Figure 3 shows the measurement of the distance from the nose to the neck coordinates over time: in Fig. 3-a shows this distance in a cooperative video and Fig. 3-b shows this distance in a non-cooperative video of the same subject. We can observe that the cooperative mode exhibits a linear behaviour, while the non cooperative mode behaves irregular. The difference in the regularity of the movements of the subjects in cooperative and non-cooperative videos has led us to carry out experiments on different methods to study the results.

V. CONCLUSIONS

Soft biometrics can boost the performance of hard biometrics systems. The presented paper provides an interesting gait analysis proposal for GC. Our proposal achieves good results and prevent the overfitting by adjusting the RF depth. We strongly believe that there is room for improvement if we combine this technique with other classification techniques such as convolutional neural networks or recurrent neural networks. However, the proposed results prove that gait can be employed to classify gender at a distance with mobile devices.

REFERENCES

[1] M. Ekstrand, M. Tian, M. Kazi, H. Mehrpouyan, Hoda and D. Kluver. Exploring Author Gender in Book Rating and Recommendation. Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 242–250.

[2] K. Takeichi, M. Ichikawa, R. Shinayama and T. Tagawa. A Mobile Application for Running Form Analysis Based On Pose Estimation Technique. 2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2018, pp. 1-4.

[3] 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.

[4] X. Leng and Y. Wang. Improving generalization for gender classification. International Conference on Image Processing, 2008, pp. 1656-1659.

[5] B. Moghaddam and M. Yang. Learning gender with support faces. Pattern Analysis and Machine Intelligence, IEEE Transactions (24), 2002, pp. 707-711.

[6] 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.

[7] 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 .

[8] E. Makinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. PAMI, 30(3):541547, March 2008.

[9] W. Gao and H. Ai. Face gender classification on consumer images in a multiethnic environment. Lecture Notes in Computer Science, 5558:169–178, 2009.

[10] 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.

[11] 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.

[12] S.Yu, T.Tan, K.Huang, K.Jia and X.Wu. A study on gait-based gender classification. Image Processing, IEEE Transactions on. 2009. 18(8):1905–1910.

[13] M. Hu, Y. Wang, Z. Zhang, and D. Zhang. Gait-based gender classification using mixed conditional random field. IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 41, no. 5. 2011. pp. 1429–1439.

[14] R. Martínez-Félez, R. Alberto-Mollineda and J. Salvador-Sánchez. Gender Classification from Pose-Based GEIs. Proceedings of the Computer Vision and Graphics (ICCVG). 2012. pp.501-508

[15] K. Arai. Human Gait Gender Classification in Spatial and Temporal Reasoning. International Journal of Advanced Research in Artificial Intelligence. 2012.

[16] I. Ebenezer, S. Elias, S. Rajagopalan and K. Easwarakumar. Multi-view gait-based gender classification through pose-based voting. Pattern Recognition Letters. 2018.

[17] 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.

[18] A. Cavallaro and A. Brutti. Audio-visual learning for body-worn cameras. Computer Vision and Pattern Recognition. 2019. pp. 103–119

[19] A. Rajivan. Measurement of Gender Differences Using Anthropometry. Economic and Political Weekly, no. 43 (31). 1996.

[20] Mei Ngan and Patrick Grother. Face recognition vendor test (FRVT) performance of automated gender classification algorithms. Technical Report NIST IR 8052, National Institute of Standars and Technology, April 2015.

[21] RIDA, Imad, JIANG, Xudong, et MARCIALIS, Gian Luca. Human body part selection by group lasso of motion for model-free gait recognition. IEEE Signal Processing Letters, 2016, vol. 23, no 1, p. 154-158.