Real-Time Gender Classification from Gait Features Using Convolutional Neural Networks

Kazuki Mizutani[1]

[1] a4525072@edu.gifu-u.ac.jp
[1] Gifu University (Japan)

Keywords: Gender Classification, Gait Silhouette, Average Gait Image, Convolutional Neural Network

ABSTRACT

I use walking silhouettes, which are not affected by wearing a mask, to classify gender. With conventional methods, only angle-dependent discriminators existed. However, by using deep learning, I have succeeded in creating a discriminator that does not depend on the angle and has the same accuracy.

1 Introduction

In recent years, there has been a lot of research on analyzing various features from the appearance of a person using image analysis techniques and deep learning. From a person's appearance, it is possible to obtain various information such as the person's gender, age, and emotional state.

However, the use of voice and face functions for gender classification is limited because it is difficult to obtain high quality voice and face images when the subject is far away from the sensor. Also, because of the recent coronavirus epidemic around the world, most people are wearing masks, making it difficult to acquire facial images. For this reason, I propose the gait function as a feature that can be used for gender classification even in this situation.

Gait features provide clues to solve recognition problems that occur at long distances. Compared to other biometric features, gait features can acquire gait information from a distance with an acceptable level of quality for specific tasks such as gait function and gender classification. It is also difficult to forge or falsify others because it is imbued with people's habits and characterizes their physical capabilities. Compared to other methods, it has the advantage of simplicity of equipment, as it can be easily acquired using only a camera.

Trung Dung Do [1] uses the Average Gait Image (AGI), which combines multiple gait silhouette frames into one, to create a gender classifier independent of each viewpoint. He uses a Support Vector Machine (SVM) for gender classification. The direction of the viewpoint is inferred from the feet, and the gender classifier for that viewpoint is used for classification. However, subjects move in various directions, which are common situations in real applications. A viewpoint-independent gender classifier decreases the accuracy if the appropriate viewpoint cannot be predicted. In this paper, I create a gender classifier that is robust to viewpoint using deep learning.

2 Experiment

2.1 Flow of the proposed method

Fig. 1. shows the flow of the proposed method. This paper proposes a method for gender classification from arbitrary viewpoint, and therefore, a dataset with multiple camera views is required. For this purpose, the CASIA gait Dataset B [2][3] as shown in Fig. 2 was utilized throughout this study for illustrative and experimental purposed.

2.2 Preprocessing

First, HOG (histogram of oriented gradient) [3] is used to identify a person from an image. Then, the background subtraction method [4] is used to extract only the silhouette of the person from the image including the background. Also, it is necessary to normalize the rectangular area of the person before training because the size of the person changes between frames. Finally, I padded the image with 0 and adjusted it.

2.3 AGI

A person can move in any direction in real situations. In this paper, I use AGI as shown in Fig. 3. to identify the features of gait. The AGI is created for each arbitrary direction α of the subject, and the model is trained by shuffling the AGI for all directions. The AGI is defined as follows.

$$AGI^{\alpha}(x,y) = \frac{1}{T}\sum_{t=1}^{N} S_t^{\alpha}(x,y)$$

where T is the walking interval. If the frame rate of the video is f and the approximate time of the walking cycle is μ, I can define $T = \mu \cdot f$. According to [4][5] When the frame rate f is 25 frames / s, the value of the gait cycle time μ must be 0.6 seconds to capture the most informative gait feature. For this reason, I use $T = 0.6 * f$ in this paper. S_t^{α} is the silhouette image at time t, and α is the direction in which the subject is facing.

2.4 Convolutional Neural Network

The Convolutional Neural Network used for training in this paper is introduced using PyTorch, an open-source deep learning framework. Referring to the structure of VGG [6], I construct a multilayer neural network consisting of an input layer, six convolutional layers and three pooling layers, three Batch Normalization layers [7], a fully connected layer, and an output layer, for a total of

14 layers. The structure of this convolutional neural network is shown in Fig. 4. The input data is an 8-bit grayscale image of 80 x 40 pixels. Before learning, all pixel values are normalized to be within the range of 0.0~1.0. Adam is used for optimization.

2.5 Dataset

In this experiment, I use CASIA Dataset B, a walking image database from the Chinese Academy of Science. This database contains images of 124 people (93 males and 31 females) taken from 11 viewpoints (0°~180°) in three conditions: walking normally, carrying a backpack, and wearing a coat. Each subject is filmed for a total of 10 times: 6 times walking normally, 2 times walking with a backpack, and 2 times walking with a coat. The subjects are all West Asians.

2.6 Method of Experiment

CASIA Dataset B is used for both training and test data. For the training data, I randomly select 15 males and 15 females, and shuffle about 10000 AGI images of normal walking patterns to be trained by CNN. For the test data, I randomly select 5 people each and evaluate the accuracy of about 3000 images. I use an intel® Core™ i5 7200 CPU for the processing system.

2.7 Evaluation Method

The Correct Classification Rates (CCRs) are used as a method of evaluation. The CCR is defined as:

$$CCR = \frac{TP + TN}{N}$$

where TP is the rate at which images with male labels are predicted to be male, TN is the rate at which images with female labels were predicted to be female, and N is the number of all test data. In this experiment, I labeled males as 1 and females as 0.

3 Results

3.1 Results of experiments

The experimental results are shown in Table. 1. I obtain a high accuracy of 99.7% for the normal walking state without any attachments. The accuracy for the two conditions with attachments is also obtained by using a gender classifier trained on normal walking conditions. I obtain 97.4% accuracy for the backpack state and 97.8% for the coat state, which is lower than that of the normal walking state, but still high accuracy. The reason for the decrease in accuracy is the change in silhouette caused by the backpack and coat.

3.2 Comparison with conventional method

The conventional methods [1] are all viewpoint independent gender classifiers, but in this paper, in order to create a viewpoint robust gender classifier, I use a Convolutional Neural Network to improve the learning performance and train it on training data with no known viewpoint. Overall accuracy is improved when compared to conventional methods. In the conventional method, the backpack and coat attachments are removed before the accuracy is achieved, but in this paper, I are able to achieve higher accuracy without this process. It is thought that the improved learning performance makes the method more robust against attachments.

4 Conclusions

I are able to extract the subject's walking silhouette from the video and efficiently perform gender recognition from it. Conventional methods only have viewpoint-independent gender classifiers, but in this paper, I have created a unique classifier using convolutional neural networks to enable viewpoint-robust gender classification. Although the silhouette changes with attachments and the accuracy decreases slightly, I are able to achieve a level of accuracy that is applicable to real world systems. However, the gender classification depends on the quality of the silhouette. As a future task, I would like to consider introducing an Auto Encoder that automatically completes extra or defective parts of the silhouette in order to make it of good quality.

References

[1] Trung dung Do, Hakil Kim, Van Huan Nguyen. (2019) Real-time and rubust multiple-view gender classification using gait features in video surveillance. In arXiv.org, https://arxiv.org/pdf/1905.01013.pdf.

[2] 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, pp. 441-444.

[3] YU, S., Tan, D., and Tan, T. (2006) Modelling the effect of view angle variation on appearance-based recognition. In Asian Conference on Computer Vision, Springer, pp.807-816.

[4] Yoo, J.-H., Hwang, D., and Nixon, M.S. (2005) Gender classification in human gait using support vector machine. In International Conference on Advanced Concepts for Inrelligent Vision Systems, Springer.

[5] Birch, I., Vernon, W., Burrow, G., et al. (2014) The effect of frame rate on the ability of experienced gait analysts to identify characteristics of gait from closed circuit television footage. In Science & Justice, 54(2), pp. 159-163

[6] Karen Simonyan, Andrew Zisserman. (2015) Very Deep Convolutinal Nerwork for Large-Scale Image Recognition. In arXiv.org, https://arxiv.org/pdf/1409.1556.pdf

[7] Sergey Ioffe, Christian Szegedy. (2015) Batch Normalization: Accelerating Deep Nerwork Training by Reducing Internal Covariate Shift. In arXiv.org, https://arxiv.org/pdf/1502.03167.pdf.

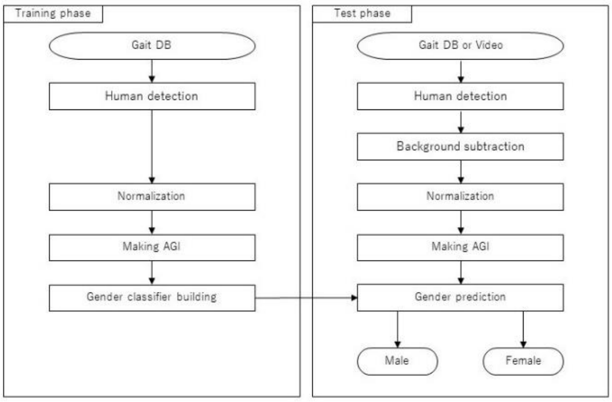

Fig. 1 Flowchart of the proposed method.

Fig. 2. Changes from the viewpoint of gait silhouette. Normal condition (top row), carrying an item (middle row), wearing a coat (bottom row)

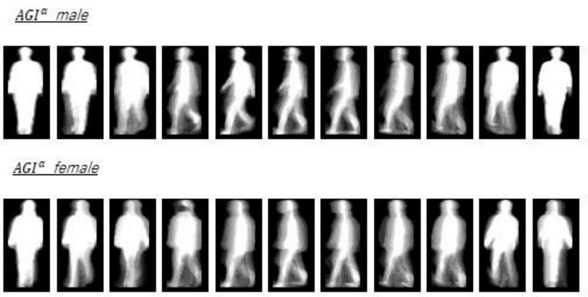

Fig. 3 Example of AGIs from different viewpoints.

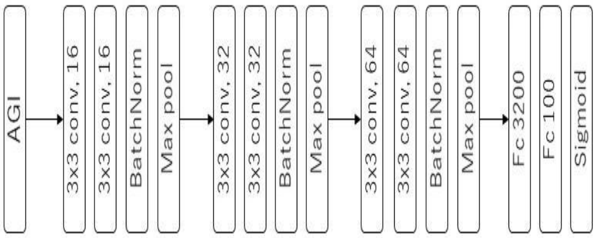

Fig. 4 6-layer Convolutional Neural Network used in this paper

	Proposed method	Previous method [2]
Normal walking	99.4	98.8
Carrying backpack	97.4	94.4
Wearing coat	97.8	93.5

Table. 1 CCRs (%) for the proposed method and the previous method