The Research of Gait Recognition Based on Deep Learning: A Case Study of the Missing Elderly

Mingyang Liu
Shanghai Starriver Bilingual School, Shanghai, China
Hzhc_wjdi@163.com

Abstract. In our daily life, we often see some notices for the missing elderly. As we all know, the elderly people with Alzheimer's disease or memory impairment are more likely and more often to get lost. Therefore, I hope that I can make some contribution to help solve this problem. For this reason, the author a detection and tracking system with interactive interface was designed by the author based on gait recognition for the elderly people who are under the risk of getting lost. The walking characteristics of the elderly can be extracted from the gait image data of the elderly provided by their families. Combined with the video streaming information documented by the cameras of the public security organ, the detection and tracking system can identify the missing elderly at a long distance, in this way, it provides clues for searching the missing elderly. In this paper, all experiments were under the Windows 7 operating system. Python was used to invoke yolov3 model for character detection. The foreground images were obtained by applying background subtraction in the static background, and then gray processing, Gauss blurring, binarization and other image operations were done to extract the image features and obtain the GEI gait energy map with the help of gait feature extraction technology. The image data were stored in the database, and then pedestrian identity detection was carried out through an input video or by real-time cameras. In order to test conveniently, an interactive interface based on PyQt4 is designed, which can realize functions such as register pedestrian names on site, acquire pedestrian gait energy map and establish database, refresh background, detect and recognize pedestrians. Moreover, it supports arbitrary switching between real-time camera detection and input video detection.

1. Introduction
At present, the well-known biometric recognition technologies include iris identification technology, fingerprint identification technology, face identification technology, DNA identification technology and so on. Compared with other biometric recognition technologies which have been applied in the market, gait recognition technology is not yet mature and still has a bigger research space. In long-distance or complex scenes, gait is the only biometric feature that can be clearly imaged, which includes limb length, aspect ratio and so on. Even if when a person is of meters away, wearing a mask tens and facing a common camera with his back, gait recognition algorithm may also make an identity judgment, which can be applied to various resolutions, illumination and angles. Gait recognition technology can be
divided into the following steps: Gait data acquisition, image analysis, feature extraction, and data comparison.

Gait recognition can be applied in the fields of public safety, scenic spot construction and so on, with broad and high application prospects. At present, the widely used data sets are CASIA Gait, CMU Mobo, USF which use deep learning algorithm that requires complicated data acquisition process and excellent hardware support. Thus, the use of laptop camera to record data and gait energy graph algorithm to process data by the author greatly improved software speed and reduced hardware requirements.

There are many applications of gait recognition technologies, for example, the track of the lost elderly can be obtained through gait recognition. Some elderly people with Alzheimer's disease or memory deficits often go missing. The walking characteristics of the elderly can be extracted from the gait image data of the elderly provided by their families. The detection and tracking system can identify the missing elderly, by combining their walking features with the video streaming information documented by the cameras of the public security organ.

Gait recognition are still facing a verity of limitations: the difference among images caused by shooting angles will lead to recognition inaccuracy; certain requirements should be meet for dressing, etc., but it is still more accurate and faster than other biotechnologies in many applicable application scenarios.

2. Principles and Methods
The general idea of gait recognition is shown in Fig. 1.

![Figure 1. Principles of gait recognition](image)

After setting up the background model, the foreground image is obtained by the difference between the current image and the background model, and then the detected image is processed by image operations such as gray processing, Gauss blurring, binarization, after which the GEI image is obtained. The subject used Python language and designed an interactive interface based on pyqt4, which can realize the functions of on-site registration, acquisition of people's gait energy map, establishment of database, background refresh, and recognition and so on. It can support real-time camera detection or detect by inputting a video. The two methods can be switched freely to each other. When people click on the recognition button, they are able to select the test box and then the name of the person and the matching rate will be displayed.

Previously, we used to object detection built in opencv, but found that the effect of pedestrian detection fell flat, so we considered fusing some target detection algorithms with it. At present, the mainstream target detection algorithms are based on deep learning model, which can be divided into two categories.
- Two-stage detection algorithm, which divides the detection problem into two stages. Firstly, it produces candidate regions, and then classifies candidate regions which usually needs to be refined. The typical representatives of this kind of algorithm is R-CNN based on region proposal, such as R-CNN, Fast R-CNN, Faster R-CNN, etc.

- One-stage detection algorithm, such as YOLO and SSD. With no need of the region proposal stage, it directly generates the class probability and position coordinate value of the object. The main performance indicators of target detection models are detection accuracy and speed. For accuracy, target detection should consider not only the classification accuracy, but also the accuracy of object location. Generally speaking, two-stage algorithm has plenty of advantages in accuracy, while one-stage algorithm has advantages in speed. However, with the development of research and technology, both kinds of algorithms will be improved in accuracy and speed. Because of the pyqt interface design added in our model, we hoped that the program could run faster with certain accuracy, therefore, we considered using Yolo algorithm for pedestrian detection. In the experiment, we used pre-recorded videos to test. Fig. 2 is the result of yolov3 pedestrian detection, it is obvious that the result has high accuracy.

![Figure 2. Pedestrian detection testing picture of yolov3](image)

2.1. Yolo algorithm

Instead of local convolution, we used fully convolutional which does not separate object location from category processing. In principle, they are quite different because the classification and location search are no longer separate. The central lattice of the object position window is responsible for classification, and each lattice outputs the prediction of the object position box of the prediction class. The final output dimension of the full connection layer of YOLO network is $S \times S \times (B \times 5 + C)$. The number of bounding boxes output by each lattice is $B$ (rectangular areas containing objects) and the number of objects’ probability information which indicates its belonging to a certain category output by each lattice is $C$.  

First of all, feature map of a certain size, such as $13 \times 13$, was obtained by inputting features to feature extraction network. Then the input image would be divided into $13 \times 13$ grid cells. If the central coordinate of an object in the ground truth fell in a grid cell, that grid cell would be responsible for predicting the object because each grid cell would predict a fixed number of bounding boxes $2$ in v1, $5$ in YOLO v2 and $3$ in YOLO v3). The initial sizes of these bounding boxes were different, and these bounding boxes were ultimately used in prediction combined with the largest bounding box of the IOU in ground truth. It can be seen that the predicted output feature map had two dimensions, such as $13 \times 13$, and one dimension (depth) is $B \times (5+C)$. Note: YOLO V1 is $B \times (5+C)$, where $B$ represented the number of bounding boxes predicted by each grid cell, such as 2 in YOLO v1, 5 in YOLO v2, 3 in YOLO v3, $C$ represented the number of categories of bounding boxes (There was no background class, so for VOC, the dataset was 20), and 5 represents four coordinate information and one objectivity score.

Yolov2 used Darknet-19, 19 layers of convolution, and yolov3 used Darknet-53, 53 layers of convolution, as shown in Fig. 3.
It can also be found from the papers of yolov3 algorithm that compared with other algorithms, the algorithm of yolov3 is faster, as shown in Fig. 4.

2.2. The principle of Binarization
Since the camera we used to record RGB tricolor images, and the final gait energy map is a binary black-and-white images, we need to process the original RGB images in a series of ways:
Figure 5. Comparison of original image, gray image and binary image

As shown in Fig. 5, the photos were taken by ourselves. The image in the middle was gray processed. Through the predefined cvtcolor function of OpenCV module, the three-dimensional image of RGB was transformed into one-dimensional gray image. The image on the right side was binarized. It was entered into the cv2. Threshold function of the OpenCV library. When the result exceeded the threshold value, it was converted to 255 (white) and when it was less than the threshold value, the result became 0 (black). The right image was obtained after binarization.

2.3. Gauss Fuzzy Principle
Gauss blurring is also a function defined by OpenCV module, which predefines a convolution kernel of Gauss probability distribution. The average value of the pixel blocks around the center is taken to achieve the blurring effect. By using the Gauss blurring method, it can effectively reduce the noise of the image and remove unnecessary interference which affects the accuracy. As can be seen from Fig. 6, the larger the size of the convolution core is, the fuzzier the output image becomes.

Figure 6. The left-most image is the original image and the following images are after Gauss Fuzzy.

2.4. Principle of Contour Extraction
At the beginning of writing this paper, I used the cv2. FindContours() function of OpenCV module to crop and extract the contour of the pictures, and then obtained the contour size of moving objects (width, height): w, h. Then the contour area was calculated. If it was less than the threshold value, it was judged to be other objects, not people. When it was larger than the threshold value, whether the width and height of the object were larger than the set value should be determined. After satisfying all the conditions, the characters were selected, the coordinates of the current person were obtained, and then the image library was called to save the GEI image. At the same time, the portrait was framed by cv2. Rectangle(). Later, when data was added into the database, every person registered could get the current figure's gait map after clicking on the save button. In the later stages of the paper, after continuous testing, the human body detection effect of OpenCV was found to be not ideal, so the current popular deep learning for human body recognition came into my mind. I then used yolov3 algorithm, which can achieve a relatively high accuracy, frame travelers directly, and detection speed is faster than other algorithms.

2.5. Prospect Extraction Principle
Background differencing is a common and widely used sensing technology, which is mainly used to extract foreground in the case of fixed background. Its main principle is to subtract the current frame and background, and then use threshold to binarize to get the foreground mask. Fig. 7 is a sketch of background differencing.
Figure 7. The principle of Background differencing

It is a general method for motion segmentation of static scene. It calculates the difference between the currently acquired image frame and background image in order to get the gray image of the target moving area. The gray image was done thresholding to extract the moving area. In order to avoid the influence of environmental illumination, the background image was updated according to the current acquired image frame.

According to foreground detection, background maintenance and post-processing methods, there are several different background subtraction methods. If \( I_t \) and \( B_t \) are set to be the current frame and background frame images respectively, and \( T \) is the foreground gray threshold, the method adopted in this paper is as follows:

- Take the average of the previous frames as the initial background image \( B_t \);
- The current frame image and the background image are subtracted by gray scale and absolute values are taken; the formula is \( |I_t(x, y) - B_t(x, y)|\)
- For the pixel of the current frame \((x, y)\), if there is \( |I_t(x, y) - B_t(x, y)| > T \), the pixel is the foreground point;
- Morphological operations (corrosion, expansion, opening and closing operations) are performed on the foreground pixel images.
- Update the background images with the current frame images;

The background subtraction method was mainly adopted by us considering the simplicity of the algorithm. Besides, it can overcome the influence of ambient light to a certain extent. However, there were also some limitations, for example, for the cameras that could not be used in motion, it was difficult to update the background images in real time. When the cameras moved, the contour extracted would be difficult. When facing such problems, you need to click the \(<\text{Refresh Background}>\) button to re-select the background.

3. Result and Analysis

Figure 8. The illustration of gait energy of some pedestrians

During the experiment, different data were collected, including old people, children, men and women, as well as the same person wearing different cloth, with a total of more than 200 groups. Fig. 8 is an
example of gait energy diagrams of five people generated in the experiment. The final recognition accuracy of the experiment was about 70%, and it was found that with the increase of the amount of data collected from the same person, the recognition became more accurate.

4. Conclusion and Forecast
This paper can be used to generate its own database for the recognition of moving people in static background model. The accuracy can reach around 70%. However, a person's gait may change after illness, aging, injury, weight gain and loss, comfort in clothing and other factors, so when some extreme situations occur, it is very possible that gait may not be identified through the system. Therefore, we have to rely on a huge database, and need to collect different clothes and gaits of the same person for many times in order to greatly improve the accuracy. Moreover, the premise of gait energy map is static background, and if the tester are wearing clothes that are similar to the background color, the human image may be missed to some extent. Therefore, the research needs to be further improved. In the follow-up, we will combine in-depth learning to extract the gait energy map and solve the problem of dynamic background.

References
[1] Ju Han, Student Member, IEEE, and Bir Bhanu, Fellow, IEEE, "Individual Recognition Using Gait Energy Image" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 2, FEBRUARY 2006
[2] Joseph Redmon, and Ali Farhadi, University of Washington," YOLOv3: An Incremental Improvement", 2018
[3] Ahmad Puad Ismail ; Nooritawati Md TahirHuman, Cambridge, UK, IEEE, "Gait Silhouettes Extraction Using Haar Cascade Classifier on OpenCV", DOI:10.1109/UKSim.2017.25
[4] Davrondzhon Gafurov, Gjøvik University College, davrondzhon.gafurov@hig.no, "A Survey of Biometric Gait Recognition: Approaches, Security and Challenges".
[5] Perry Jacquelin M.D.; Slac Thoroughfare; Davids, Jon R. M.D. "Gait Analysis: Normal and Pathological Function", Journal of Pediatric Orthopaedics: November-December 1992-Volume 12 - Issue 6 - ppg 815