the Report for Biometrics coursework

Shurui Zhang
School of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom
sz2r20@soton.ac.uk

***Abstract*—The purpose of this coursework is to implement a Biometrics system to identify subjects according to their body shape. In this report, the shape of the person in the picture is used as the feature to classify this person. Semantic segmentation is used to obtain the subject mask. Hu Moments is used as the feature vectors of subjects. K-Nearest Neighbors Algorithm is used to classify people. Because each class in the training set contains few images, it is not possible to split the training set into the training set and the validation set. At the same time, since there is no label in the test set, the accuracy rate cannot be calculated. However, through visual inspection, it can be seen that the accuracy of this algorithm is low. This may be due to some reasons such as small a data set etc.**

***Index Terms*—Body Shape, Semantic Segmentation, K-Nearest Neighbors Algorithm, Classification, Hu Moments**

I. Introduction

Most Biometrics systems determine a person's identity based on facial features or fingerprints. This report will determine the identity of a person based on the body shape. The dataset used in this report is the subsets of the Large Southampton Gait Database. In this dataset, each subject has a front-on view and a side-on view. Figure 1 shows an example of these two views.

(a) Front-on view (b) Side-on view

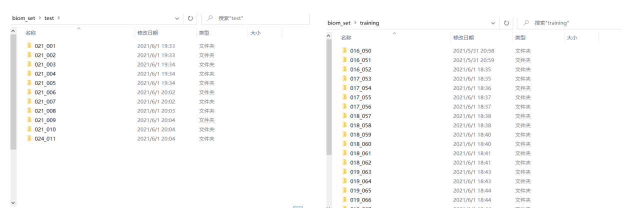
Fig. 1: Examples of both views

II. Method

The methods of reconstruction of this dataset, extraction of the shape of subjects, generation of binary graphs, extraction of feature vectors and classification will be introduced in this section.

A. Reconstruction of the given dataset

The algorithm uses the torchvision.datasets.ImageFolder to convert the given image data set to the data set that can be applied to the program. This class takes the name of a folder as the label for all images in that folder. Because this image data set given does not fit this directory structure, the directory structure of this image data set should be rebuilt. The new directory structure is shown in figure 2. Since each Subject has a front-on view and a side-on view, each folder in the training set contains the front-on view and side-on view of the subject and the name of this folder is the subject identification number. In the test set, each folder contains the front-on view and side-on view of the subject and the name of each folder is the subject's name views obtained by visual inspection. Although subject names are marked in the test set, these subject names are not used in the code to improve the algorithm's accuracy. They are only used to estimate the accuracy of this algorithm.



(a) Test set directory structure (b) Training set directory structure

Fig. 2: directory structure

This report uses semantic segmentation to extract the outline of a person in a picture. However, when torchvision.models.segmentation.deeplabv3_resnet101 directly be used for semantic segmentation, may be due to this class will consume a large amount of RAM resources, program will always be forced to interrupt (as shown in figure 3). Therefore, removebg is used to preprocess all the images in the image data set, that is, to remove their backgrounds. This can reduce the number of computation of torchvision.models.segmentation.deeplabv3_resnet101 , so that this program does not crash. Figure 4 shows the example image processed by removebg. These images processed by removebg become the images in the new image data set.

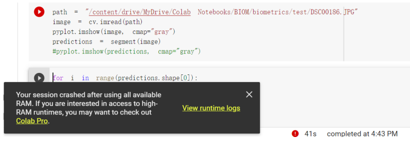


Fig. 3: The crashed session

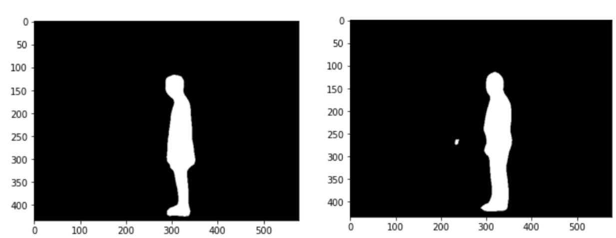
Fig. 4: Example image processed by removebg

B. Features

This algorithm will take Hu Moments as the feature vectors of images. Before calculating the Hu Moments, the subject mask needs to be obtained.

1) Subject mask: The image segmentation is used to obtain the subject mask from the image. Removebg can remove most of the background in the image, but it will retain the treadmill part in the image, so it is not possible to directly convert the image processed by the removebg into a binary image as a subject mask. The subject mask could be got by performing the following two steps:

1) The torchvision.models.segmentation.deeplabv3 _resnet101 is used for semantic segmentation. Most of the segmented images are shown in figure 5a. However, some of the segmented images will retain part of the treadmill in the image (as shown in figure 5b).



(a) (b)

Fig. 5: Examples of segmented image

2) The outline of the person in the image obtained by the he previous step needs to be moved to the center of the image and then the image needs to be cropped so that most of the abnormal segmented images‘ parts of the treadmill can be removed. Figure 6 shows the image processed by this step.

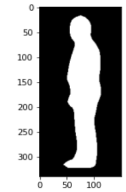


Fig. 6: Example of cropped image

2) Hu moments: Hu moments can be used to describe the shape of the subject mask. The shape of the subject mask of different subjects is different, and when the subject mask is translated, rotated and scaled, its Hu moments remain unchanged, so Hu moments can be used as feature vectors of a subject.

The original method to obtain the subject's feature vector is to obtain the subject's Gait Energy Image (GEI) first by using the subject's front-on view and side-on view's masks. Then this subject's GEI's Hu moment is calculated. However, by looking up literature, it can be found that most of the GEI were generated by the walking video of a subject in a continuous period. However, in this image data set, each subject has only two static images at two different time points, so this feature vector is not suitable to be obtained through GEI. Because of this, this algorithm will determine the identity of this subject through both of the front-on view and side-on view of the subject. Therefore, the Hu moments of a subject's front-on view and side-on view's masks are calculated respectively and pieced together to become the feature vector of this subject.

C. Classifier

K-Nearest Neighbors algorithm is used as the classifier of this algorithm. It mainly calculates the distance between one instance pulled from the test set and each instance in the training set. The distance here refers to the Euclidean Distance of these two instances' Hu moments. As there is only one front-on view and one side-on view for each subject in this image data set, the number of nearest neighbors (k) is set to 1. Therefore, the subject corresponding to a feature vector extracted from the test set is the subject corresponding to the feature vector which is from the training set nearest to it.

III. Results

Table I shows the results of that this model recognize the images in the test set. Based on the objects‘ names obtained by visual inspection, the accuracy of this model is low. There are 11 groups of images in the test set, and only one group is correctly identified. The model consumes a long time in semantic segmentation of all images in the image data set, and it consumes a short time in calculating Hu moments and using K-Nearest Neighbors algorithm to identify images in the test set.

TABLE I: The subject corresponding to images in test set

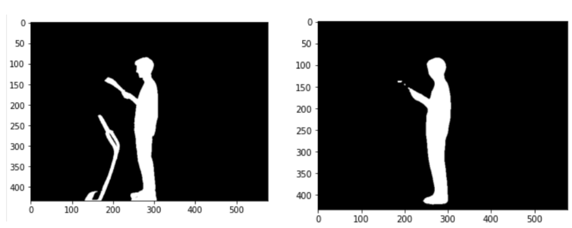
the name of image 1	the name of image 2	the subject's name
DSC00165.JPG	DSC00166.JPG	$018_0 58$
DSC00167.JPG	DSC00168.JPG	$023_0 81$
DSC00169.JPG	DSC00170.JPG	$021_0 03$
DSC00171.JPG	DSC00172.JPG	$020_0 70$
DSC00173.JPG	DSC00174.JPG	$020_0 70$
DSC00175.JPG	DSC00176.JPG	$023_0 81$
DSC00177.JPG	DSC00178.JPG	$018_0 58$
DSC00179.JPG	DSC00180.JPG	$019_0 65$
DSC00181.JPG	DSC00182.JPG	$020_0 70$
DSC00183.JPG	DSC00184.JPG	$016_0 51$
DSC00185.JPG	DSC00186.JPG	$018_0 57$

IV. Discussion

A. Performance

Although the recognition accuracy of this algorithm is low, it can be found that the body shapes of different subjects have great differences during the completion of this coursework, which indicates that the method of identifying people through their body shapes is feasible. The evaluation of each component of this algorithm is as follows:

1) Semantic segmentation: using resnet101 for semantic segmentation can indeed extract the outline of a person in a picture, but the accuracy of the extracted outline reduced to some extent. This conclusion can be reached by comparing these two images in Figure 7. Figure 7a is a binary image directly converted from the image processed by removebg. Figure 7b is the binary graph obtained by semantic segmentation. Figure 7a contains some detailed features of the shape of the human body, such as the nose, etc., but these detailed features have been lost in Figure 7b. The original scheme of the experiment was to use the binary image processed by removebg as a subject mask directly, but after the experiment, it was found that the treadmill part in the subject mask obtained in this way would have a great impact on the accuracy of this model.



(a) (b)

Fig. 7: Two kinds of subject mask

2) Feature vector: this algorithm calculates the Hu moments of the subject's side-on view and front-on view respectively, and then pieced these two Hu moments together to become the feature vector of this subject. Due to the low accuracy of this model, this feature may not fully represent the body shape. The original solution was not to combine the Hu moments of the two views as one feature vector, but to have two Hu Moments as two feature vectors for the corresponding view. Through experiments, it is found that the feature vectors obtained in this way will make the accuracy of this model lower. There are 22 pictures in the test set, and only 1 picture is classified correctly. This may be due to the fact that Hu moments cannot describe the details of the image, resulting in an incomplete description of the image. The shape of the clothes of the people in the test set is quite different from that of the people in the training set, which may be one of the reasons why the body shape of one person cannot be described more accurately by Hu moments.
3) K-nearest neighbors algorithm is suitable to solve this problem. This is because if the feature vector of the subject in the test set is similar to the feature vector in the training set, they can be determined to be the same subject. If the feature vector can more accurately represent the subject, k-nearest neighbors algorithm will be used for identification with higher accuracy

B. Improvements

In the future, some more accurate features can be extracted from the images. For example, a virtual skeleton can be fitted for the body shape, and features of the body shape such as the relative position of the head and the body can be extracted according to these virtual skeletons. Since the shape of the person's clothing has less impact on the fitting of the virtual skeleton, such feature vectors may be more accurate.

The semantic segmentation of this algorithm consumes a lot of time, because this algorithm is all run on the CPU. In the future, the semantic segmentation of this algorithm can be run on the GPU, which will greatly improve the running speed of this algorithm.

V. Conclusion

This report proposes a biometrics system that can determine a person's identity based on some still images of this person's body. This report also analyzes the performance of each part of this system, points out the factors that prevent the system from improving the recognition accuracy, and proposes some improvement measures. This report confirms the feasibility of identifying a person through their body shape.