GENDER RECOGNITION BASED ON FUSION OF FACE AND GAIT INFORMATION

DE ZHANG, YUN-HONG WANG

Intelligent Recognition and Image Processing Laboratory, School of Computer Science and Engineering, Beihang University, Beijing 100083, China
E-MAIL: zhangde@cse.buaa.edu.cn, yhwang@buaa.edu.cn

Abstract:
This paper considers the combination of face and gait biometrics from the same walking sequence to carry out gender recognition. A camera is capturing the side view of a person, while another camera is placed to record the face of the same person at the front view. After these videos are acquired, we extract the silhouette images from the gait videos and normalized frame images decomposed from the face videos. Then, for face classification, we introduce PCA to reduce the image dimension and SVM to classify gender, for gait classification, we divide the silhouette into seven parts and extract features from each and also employ SVM to classify gender. On the decision level, the sum rule is applied to implement the fusion of these two classification results. The final fusion results show an improvement on correct classification rate.

Keywords:
Fusion; Silhouette; Face; Gender recognition; Sum rule

1. Introduction

With the development of Visual Surveillance and Human-Computer Interaction technology, computer vision systems are playing more and more important roles in our life. Gender recognition, as a part of the whole computer vision system, has received much attention in recent decades. In [4, 5, 6], we can find some gender identification systems based on the voice of speaker. In [1, 2, 3], the problem of classifying gender from facial images is described. Furthermore, gender can be recognized from human walking, as shown in [7, 8, 9]. It is true that other biometrics are also able to be applied in gender classification.

So far, there have been some attempts on combining multiple biometrics. In [19], Shakhnarovich et al. developed a view-normalization approach to integrate face and gait in human identification from multiple views. Otherwise, we intend to apply the fusion in gender recognition. In this paper, we designed a system for multi-modal gender recognition from face and human gait obtained in the same walking sequence. In visual surveillance, it is possible to capture the face images and human movements at the same time. Therefore, the fusion that we have conceived will be valuable in many surveillance applications.

Our system mainly consists of five parts: data collection, preprocessing, face classification, gait classification and fusion. In the data collection subsystem, eight cameras are used to record human walking from different views and one camera is responsible for face recording in front view (Figure 1). In the preprocessing phase, silhouettes of a walking person are extracted from video sequences and face images decomposed from the videos are normalized and grayed. In the part of face classification, PCA is used to extract features from gray facial images and SVM is used to classify gender. In the part of gait classification, we adopt a simple representation of gait appearance and also use SVM to implement gender classification. In the last part, we demonstrate better performance in recognizing gender which is achieved by fusion of these two modal classification results.

The reminder of this paper is organized as follows. In section 2, we will review some of previous work related to gender recognition from face or individual gait. Then in section 3 we will introduce the data collection part and preprocessing part in detail. The most important parts of our work will be described in section 4, including face classification, gait classification and fusion. We will present the results of our experiments involving 60 subjects in section 5. Finally, we conclude with a summary of the implications of these results and suggest future work to improve the fusion performance in section 6.

2. Related Work

The research work on gender recognition was started by psychologists. It has been an active area in
psychological literature [10, 11, 12, 13]. Computational algorithms for discriminating female from male have been explored in the past decade. In this section, we briefly review the previous work in visual gender recognition.

2.1. Learning gender from face

There has been much work to classify gender from human faces. In early 1990s, various neural network techniques were employed for gender classification from a frontal face. In paper [14] Golomb et al. trained a fully connected two-layer neural network, SEXNET, to identify gender from face images. Brunelli and Poggio [15] developed HyperBF networks for gender classification in which two competing networks, one for male and the other for female, are trained using 16 geometric features. To sum up, some of these techniques are appearance-based methods and others are based on geometric features.

In Moghaddam and Yang’s paper [1], nonlinear SVM was investigated in gender classification for low-resolution thumbnail face (21-by-12 pixels) on 1,755 images from the FERET database. Shakhnarovich et al. developed a real-time face detection and demographic analysis (female/male and asian/non-asian) system [2]. The demographic classifier and the face detector are both based on the Adaboost algorithm.

Most of these studies suggest that PCA can encode face information in a psychologically plausible manner. Hence, we employ PCA to reduce the dimension of the facial image space in this paper.

2.2. Classify gender from gait

As for recognizing gender from human gait, researchers mainly focused on the recognition from moving light displays (MLDs) in early studies. In [16], a two-stage PCA framework was implemented to decompose male and female walking data into an Eigenspace, from which a linear classifier was used for gender recognition. The data consisted of three-dimensional motion-capture trajectories of 40 walkers (20 females, 20 males). Davis and Gao presented a three-mode expressive-feature model for recognizing gender from point-light displays of walking people [7]. The walking data used in their experiments was the same as above. They developed a tri-modal nature of female/male walkers (posture, time, and gender) in which an efficient three-mode PCA representation was employed. Then gender estimation could be constructed according to the result of three-mode factorization.

However, it is hard to find them applied in surveillance systems. So, video based approaches have been investigated in recent years. It is known that human identification from gait has been given much attention and a large amount of methods to implement human gait recognition have been developed. Up to now, some of them have been applied in gender classification from gait. Lee and Grimson developed a gait silhouette appearance representation by proportionally dividing the silhouette into 7 parts [8]. For each part, an ellipse was fitted and features were extracted from these seven ellipses. Finally, a SVM was used as gender classifier. In [9], Yoo et al. used a sequential set of 2D stick figure to represent the gait signature. SVM was also employed to carry out gender classification.

Likewise, SVM is used as the gender classifier in this paper.

3. Data collection and preprocessing

In this section, we will give a detailed description of the first two parts of our system.

This two modal fusion system that we intend to build should be based on a face database and gait database. Since there is no shared database, we collected data first. The face data and gait data of one sample that we need in this system ought to be captured simultaneously.

Preprocessing is to extract silhouettes of a walking person from a video sequence and to normalize facial images decomposed from a video. The result of preprocessing plays an important role in performance of gender classification.

3.1. Data collection

In any pattern recognition study, the database used for evaluation is necessary. Our data are recorded in an indoor laboratory scenario because the primary purpose of our database is to show basic practicality.

In our data collection subsystem, eight cameras are placed at different angles recording the movement of a person. These cameras are divided into two groups each of which consists of four cameras and forms a 1/4 circle. The face of the person is captured by another camera from the front view. The arrangement of these nine cameras is illustrated in Figure 1.
To increase the evaluation capacity on gender classification, we tried to make the same number of male and female participants. There were 60 volunteers in all including 32 male subjects and 28 female subjects aged between 22~28. At the beginning of collection, all of participants are asked to read, understand and sign an approved consent form. During the course of collection, each person walked along the direct line between two black points, as shown in Figure 1, from left to right and then return, repeating five times. Camera C9 is higher than C8, so that it can capture human face at front view in left-to-right turn.

3.2. Preprocessing

Frames are decomposed from videos first in this part. Additionally, the main aim of preprocessing gait data is to extract silhouettes from recorded videos and for face data we require normalized 8-bit facial images in experiments.

To get silhouettes, we employed a simple subtraction method to separate foreground and morphological filtering to reduce noise after binarization. Mean value of multi frames was used to do background subtraction. Suppose $f_i(x,y)$ is the i-th frame of a given video sequence and the mean value of first $k$ frames in the given sequence is $\bar{f}_i(x,y)$. Let $\mu_k$ denote the square deviation of the first $k$ frames. With respect to any other frame in the video sequence, we could extract the foreground using the following relation formula:

$$
\begin{align}
\| f_i(x,y) - \bar{f}_i(x,y) \| \leq \mu_k, \text{background} \\
\| f_i(x,y) - \bar{f}_i(x,y) \| > \mu_k, \text{foreground}
\end{align}
$$

To reduce noise in images, erosion and dilation are used to erase the small spots on the binarized image and to fix discontinuous point on the contour.

For face data, we first crop the facial part from original 24-bit frame images with the distance between eyes. Then, these facial images are normalized to 64-by-64 pixels. Finally, histogram equalization is used to generate a standard 8-bit gray image.

Figure 2 shows some sample frames and corresponding results of preprocessing.

4. Gender classification and fusion

In our system, the main focus is on the fusion of two kinds of biometrics to pursue better classifying performance. We have two options: early and late fusion. Early fusion, fusion on the sensor level, consists of combining the observations and mapping them into a single data point to be classified. Because of the large difference between silhouettes and face images, it is not easy to combine them at the sensor level. Late fusion, fusion on the decision level, can be performed on separate independent data [17]. Hence, gender classification from human face and gait are implemented respectively first.

![Figure 1. Cameras setup for data acquisition. Cameras from C1 to C8 are used to record human gait. Camera C9 records human face.](image)

Figure 1. Cameras setup for data acquisition. Cameras from C1 to C8 are used to record human gait. Camera C9 records human face.

![Figure 2. Sample images before preprocessing and after preprocessing. Top row comes from gait data. Next row demonstrates the face data from the same walking sequence as top row.](image)

Figure 2. Sample images before preprocessing and after preprocessing. Top row comes from gait data. Next row demonstrates the face data from the same walking sequence as top row.

4.1. Face-based gender classification

Given a set of $n$ observations on $p$ observed variables, principal component analysis (PCA) can be used to choose $r$ new variables which account for most of the variation in the $p$ original variables. Usually $r$ is much smaller than $p$.

We used PCA on our gray face images to extract $r$ features from 64-by-64 dimension space. Assume that there are $N$ face images in one sequence of a person. The early
fusion can be applied in a sequence before classification. We define the PCA result of the \(i\)-th image in one sequence as:

\[
V_i = (f_1, f_2, \cdots, f_r)
\]

In order to make use of all image information of one sequence, we combine the PCA results according to the idea of early fusion as following:

\[
S_i = (V_1, V_2, \cdots, V_N)
\]

Then all \(S_i\) are input into classifier. SVM, support vector machine, attempts to maximize the distance between the hyperplane and the closest training samples on either side of the hyperplane. It is a powerful technique for classification and in particular, for solving binary classification problems. In our system, we choose SVM to classify gender from face images.

If \(M\) denotes the number of total sequences of one person and \(K\) sequences of a female are recognized as male by SVM classifier, we define \(s=K/M\) as a discriminative score. When \(s\) is less than 0.5, this person is classified into female correctly. Otherwise, when \(s\) is bigger than or equal to 0.5, we deem the classification is false.

### 4.2. Gait-based gender classification

Gait, as a biometric, has some unique advantages in human identification. When applied in gender recognition, gait is also proved to be effective and convenient. We employed the method proposed by Lee and Grimson [8].

In our preliminary experiments, we found that the improvement of accuracy of gender classification by combining gait silhouettes from multiple views is quite limited. And the computation cost is high. So we only use the most accurate result from the side view in this paper.

![Silhouette divided into 7 parts](image)

Figure 3. An example of silhouette which is divided into 7 parts and ellipses are fitted to each region.

As described in [8], we divide the silhouette into seven parts as shown in Figure 3 (a). For each part, we fit an ellipse to the foreground in the region as shown in Figure 3 (b). The features extracted from each of these seven regions are the centroid \((\overline{X}, \overline{Y})\), aspect ratio \((L)\) of major and minor axis of the ellipse, and the orientation \((\alpha)\) of major axis of the ellipse. Then we have a vector \(R_i\) consisting of four parameters for the \(i\)-th region:

\[
R_i = (\overline{X}, \overline{Y}, L, \alpha)
\]

As a result, we can denote one frame with a 28 (7 regions × 4 parameters) dimension vector \(I_j\):

\[
I_j = (R_{i1}, R_{i2}, \cdots, R_{in})
\]

where \(j = 1, 2, \cdots, n\), \(n\) is the number of total frames in one sequence.

By computing the mean and standard deviation of frame vector \(I_j\) across time in one sequence and considering mean height of centroid of each whole silhouette in the sequence, we get an average gait appearance feature vector of a sequence:

\[
Q = (\text{mean}(h), \text{mean}(I_j), \text{std}(I_j))
\]

\(Q\) is a 57-dimensional vector. Then all these sequence vectors are input into a SVM classifier. How to compute the final classification result is the same as face-based gender recognition.

### 4.3. Two-modal fusion

The results of face based and gait based gender classification are fused at decision level. In [18] Kittler et al. have provided a theoretical justification for a number of combination schemes. We tested all the combination rules in our experiments and our results showed that sum rule was the best.

We assigned an equal weight of 0.5 for face and gait classifiers since the correct classification rates of these two classifiers are equivalent in our experiments (c.f. Table (1)). Assume \(s_f\) is the score obtained from face classifier and \(s_g\) is the score from gait classifier. For the \(i\)-th person, both \(s_f\) and \(s_g\) denotes the possibility of recognizing the person as male.

According to sum rule, the fusion score is:

\[
s = 0.5 \times s_f + 0.5 \times s_g
\]

Final discriminative rule is as following:

\[
\begin{align*}
s < 0.5, & \quad P \in \text{female} \\
s > 0.5, & \quad P \in \text{male} \\
s = 0.5, & \quad \text{error}
\end{align*}
\]
5. Results

We tested our methods on our self-build database. In gait data, there are eighty sequences for one person because we use eight cameras to record left to right walking 5 times and right to left 5 times. In face data, there are only five sequences for one person because we use only one camera to capture face when a person walks from left to right.

In our experiments, we chose five sequences recorded by camera C4 (Figure 1) from one person’s gait sequences. All the five are left to right walking since face data are acquired during these walking courses.

We trained and tested SVM with linear kernel on PCA face features and gait appearance features respectively first. Leave-one-out method was used to partition gallery data and probe data. There are 60 subjects in all, 32 of which are male and the remainder is female. The five sequences of one person were chosen as probe data in turn and all the other sequences were used as gallery data.

In the end, we implemented the fusion of classification results from face and gait at the decision level. Table 1 lists our experimental results. These results show that the correct classification rate gains an obvious increase after fusion and the distinction between genders is consistent with a linear boundary.

Table 1. Summary of gender classification results

<table>
<thead>
<tr>
<th>modality</th>
<th>face</th>
<th>gait</th>
<th>face and gait</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>90%</td>
<td>90%</td>
<td>93.33%</td>
</tr>
<tr>
<td>classification rate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Conclusions and future work

We have developed a new system to evaluate the effect for gender recognition when combining visual cues from human face and gait in the same walking sequence. For face classification, PCA features manifested favorable results on SVM classifier. For gait classification, we only attempted the silhouettes acquired at the front-parallel view with a previously proposed method [8]. Leave-one-out test performed well on our database. In the experimental results, both gender classification accuracies achieved 90% and their combination obtained an improvement of more than 3% compared with single modality.

Interesting future work includes fusion of face and gait silhouettes of different views respectively or together, as well as introducing more complex combination rules. In this paper, decision-level fusion has been explored. With regard to sensor-level fusion, it seems a long way for such a bi-modal classification system with face and gait.

Acknowledgements

This work was supported by Program of New Century Excellent Talents in University, National Natural Science Foundation of China (No. 60575003, 60332010), Joint Project supported by National Science Foundation of China and Royal Society of UK (60710059), and Hi-Tech Research and Development Program of China (2006AA01Z133).

References


