

Gender Recognition Based on Fusion of Face and Multi-view Gait

De Zhang and Yunhong Wang

Intelligent Recognition and Image Processing Laboratory,
School of Computer Science and Engineering, Beihang University,
Beijing 100191, China
zhangde@cse.buaa.edu.cn, yhwang@buaa.edu.cn

Abstract. In this paper, we consider the problem of gender recognition based on face and multi-view gait cues in the same walking sequence. The gait cues are derived from multiple simultaneous camera views. Meanwhile, the face cues are captured by a camera at front view. According to this setup, we build a database including 32 male subjects and 28 female subjects. Then, for face, we normalize the frame images decomposed from videos and introduce PCA to reduce image dimension. For gait, we extract silhouettes from videos and employ an improved spatio-temporal representation on the silhouettes to obtain gait features. SVM is then used to classify gender with face features and gait features from each view respectively. We employ three fusion approaches involving voting rule, weighted voting rule and Bayes combination rule at the decision level. The effectiveness of various approaches is evaluated on our database. The experimental results of integrating face and multi-view gait show an obvious improvement on the accuracy of gender recognition.

1 Introduction

It is known to all that human gender is a typical social category. Therefore, gender identification is an important requirement for different tasks including access control, vital statistics and commercial analysis etc. Gender recognition has received much attention in the computer vision literature. 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]. If a person is far away from the camera, it is hard to get face information at a high enough resolution for recognition tasks. However, when available, it is very powerful for recognition. The modality which can be detected and measured when the subject is far away from the camera is human gait or the style of walking. It will yield optimal performance if the system use as many cues as possible and combine them in meaningful ways.

There have been some attempts on fusion of face and gait for human identification. In [10], Shakhnarovich et al. compute an image based visual hull from a set of monocular views which is then used to render virtual canonical views for tracking and gait recognition. Also, this method is used to render frontal face images. They studied

the decision fusion of face and gait cues for a multi-camera indoor environment. Later, they further investigate different approaches of classifier combination for face and gait recognition in [11]. Zhou and Bhanu [12] present an approach to the fusion of gait and face which uses side view of non cooperating subjects, making use of PCA and MDA for gait and face features and combining them at the score level. In addition, Shan et al. applied the fusion of face and gait on gender recognition in [14]. They explored Canonical Correlation Analysis (CCA) to fuse the two modalities at feature level on CASIA Gait Database [15]. In their experiments, they used video sequences from two views: front view for face and side view for gait. They extracted facial part from the gait image sequences at front view. Therefore, these low-resolution facial images may do not contain enough information.

In this paper, we are interested in the further study of recognizing human gender by combining face and gait cues. Considering in surveillance environment it is possible to obtain a person's walking videos recorded by far-placed cameras at several views and the person's face images captured by near-placed cameras or the far-placed cameras when he/she comes near, we investigated the fusion of face and multi-view gait information from the same walking sequence. Our first work is to build a database consisting of gait data and face data captured simultaneously for each subject. Second, we carried out the task of gender recognition based on face or gait of each single view separately. Then, for the purpose of fusion, we tried with three schemes: voting rule, weighted voting rule and Bayes combination rule. The Bayes combination rule derives from the Bayes formula and for our case we designed a calculation method to obtain the fusion results.

The remainder of this paper is organized as follows. Section 2 describes our self-building database, and Section 3 briefly introduces the feature extraction for face and gait. The details of fusion strategies are presented in Section 4. Experiments and results are shown in Section 5. Section 6 concludes the paper.

2 Data Collection

We intend to investigate the effectiveness of a two-modality fusion of face and gait information acquired simultaneously for gender recognition. Especially, the gait information is expected to be captured from multiple views. Hence, we need to build such a database which not only contains face and gait data recorded simultaneously but also includes the similar number of males and females for the task of gender recognition.

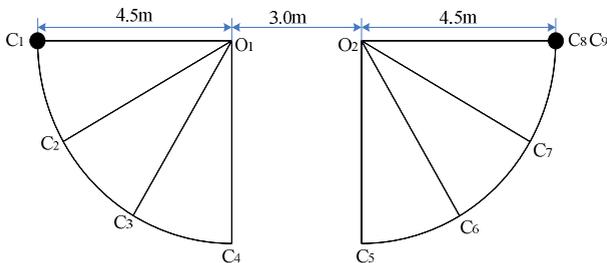


Fig. 1. Cameras setup for data collection. Cameras from C_1 to C_8 are used to record human gait at different views. Camera C_9 records human face.

Our data are captured in an indoor laboratory scenario. Eight cameras are placed at different angles recording the walking of a person. These cameras are divided into two groups each of which comprises four cameras and forms a 1/4 circle. The face of the person is captured by another camera from the front view. The setup of these nine cameras is illustrated in Fig. 1.

In the collecting process, there are 60 volunteers in all including 32 male subjects and 28 female subjects aged between 22–28. During the course of data collection, every participant was asked to walk along the straight line between camera C_1 and C_8 , which are denoted by two black points in Fig. 1, from left to right and then return, repeating five times. Thus, every camera recorded five left-to-right and five right-to-left walking video sequences for each person. Meanwhile, we label camera C_8 with the 0° view, C_7 with the 30° view, till C_1 with the 180° view. Camera C_4 and C_5 have the same view angle.

It is not enough in terms of the number of total subjects. We will continue to enlarge our database. Now we name it as IRIP-GF Database 1.0.

3 Gender Recognition

3.1 Face-Based Gender Recognition

For the face data in our database, we crop the facial part from original 24-bit frame images decomposed from face videos. Then, these facial images are normalized to 64-by-64 pixels. Finally, histogram equalization is used to generate a standard 8-bit gray image. Fig. 2 shows an example of a frame image and its preprocessing result.



Fig. 2. The left image is an original frame and the right is the result of preprocessing

Principle Component Analysis (PCA) has long been used for the face-based human recognition. PCA can reduce the dimension of feature space while retaining the most discriminatory information. Let $X \in R^N$ denote a vector representing an image, where N is the dimensionality. The covariance matrix of X is defined as $\Sigma_x = E([x - E(x)][x - E(x)]^T)$, where $E(\cdot)$ is the expectation operator and T

denotes the transpose operation. The covariance matrix Σ_x can be factorized into the following form:

$$\Sigma_x = \Phi \Lambda \Phi \quad (1)$$

where $\Phi = [\Phi_1 \Phi_2 \dots \Phi_N] \in R^{N \times N}$ is the orthogonal eigenvector matrix of Σ_x ; $\Lambda = \{\lambda_1 \lambda_2 \dots \lambda_N\} \in R^{N \times N}$ is the diagonal eigenvalue matrix of Σ_x with diagonal elements in descending order ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$). A useful property of PCA is the minimum mean square error (MSE) of signal reconstruction with using only a subset of principal components. The application of this property is the dimensionality reduction:

$$Y = P_{pca}^T X \quad (2)$$

where $P_{pca} = [\Phi_1 \Phi_2 \dots \Phi_m]$, $m < N$. The lower dimensional vector $Y \in R^m$ contains the most expressive features of the original data X .

Our 8-bit gray facial images are projected into a low dimensional feature space by using PCA. Then, we select Support Vector Machine (SVM) to classify gender.

3.2 Gait-Based Gender Recognition

For the independence of cloth color or texture information, silhouettes of walking humans are extracted by background subtraction firstly. Then, we need a gait representation that is able to be applied on different views. The spatio-temporal gait representation proposed in [13] for side-view is sensitive to silhouette deformations and robust to spurious pixels. So we take this representation into account and extend it to other views except the side-view. Additionally, we make an improvement to reduce data redundancies.

In [13], for a sequence of silhouette images $b(x, y, t)$ indexed spatially by pixel location (x, y) and temporally by time t , the author forms two new 2D images $F_R(x, t)$ and $F_C(y, t)$. Considering the periodicity of gait, we form these two images only from a gait cycle of silhouettes instead of the whole sequence to reduce redundancies. This is our enhancing point for this method. Hence, the end value of t is the number of silhouettes included in one gait cycle, denoted by N_{gait} .

The silhouettes are normalized to be the same size and centered before feature acquisition. Define a silhouette as $s[i, j]$, $i = 0, 1, \dots, M - 1$, $j = 0, 1, \dots, N - 1$, where M and N denote the number of rows and columns of the silhouette, respectively. Let

$$s[i, j] = \begin{cases} 1 & \text{if } (i, j) \text{ belongs to the foreground} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

With the above definitions, the horizontal and vertical projection of silhouettes can be expressed as:

$$p_h[i] = \sum_{j=0}^{N-1} s[i, j], \quad i = 0, \dots, M-1 \tag{4}$$

$$p_v[j] = \sum_{i=0}^{M-1} s[i, j], \quad j = 0, \dots, N-1 \tag{5}$$

Consequently, two projection vectors can be defined as follows:

$$H = \{p_h[0], p_h[1], \dots, p_h[M-1]\} \tag{6}$$

$$V = \{p_v[0], p_v[1], \dots, p_v[N-1]\} \tag{7}$$

Here, one gait cycle has N_{gait} silhouettes. Therefore, the aforementioned image $F_R(x, t)$ is formed by N_{gait} H vectors arranged by columns. Likewise, N_{gait} V vectors form the image $F_C(y, t)$. Fig. 3 shows these images F_R and F_C at different view angles from a sample.

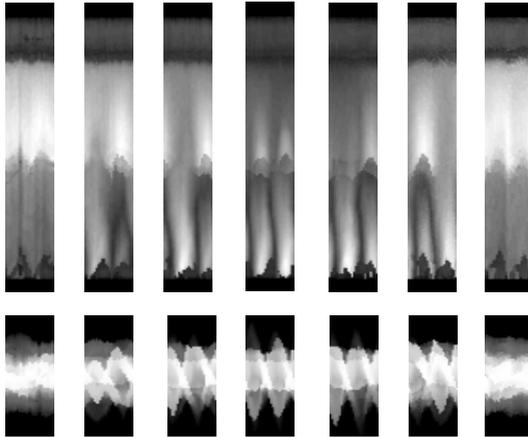


Fig. 3. An example of images F_R and F_C . Columns from left to right represent different view angles: $0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ, 180^\circ$. F_R images are listed on top row and their corresponding F_C images are shown on bottom row.

N_{gait} is calculated through the autocorrelation of the foreground sum signal. As different people may have different walking period, we resize all F_R and F_C images using bicubic interpolation to the same width (the end value of t i.e.). PCA is then used to retain only the important elements of F_R and F_C images. For each pair of F_R and F_C images, we combine their PCA results into one vector and use it as the feature vector which we intend to extract from a gait sequence. Finally, as the aforementioned face-based gender recognition, we also select Support Vector Machine (SVM) to classify gender.

4 Fusion Strategies

Generally, fusion can be done at two levels: feature level and decision level. Fusion on the feature level consists of combining the observations and mapping them into a single data point to be classified. As shown in [14], CCA is a powerful tool to relate two sets of measurements. Alternatively, one can treat all the input features as separate independent data and perform fusion on the decision level. In our case, each view of gait can be looked as an independent classifier. Adding the face modality, we have eight classifiers in all. Since our classifiers yield decision possibilities as results, the main interest is in fusion on decision level. As discussed in [16], we are interested in general fusion schemes. Particularly, we design a fusion strategy for our problem based on Bayesian inference. In the following paragraphs we will describe these methods.

4.1 Voting Rule

This strategy is motivated by the way humans are making decisions, especially when there is a group of people involved in the decision process. Each classifier is in the position of a human expert with one vote. The resulting class is determined by the majority of votes. Let $v(n)$ denote the number of classifiers with class n on the first rank:

$$v(n) = \sum_{n=n_i} C_i, n_i \in R_{C_i} \quad (8)$$

where n_i is the first rank of the i -th classifier C_i .

4.2 Weighted Voting Rule

The weighted voting method is no more than a variant of voting strategy, where each expert has its individual weight.

$$v(n) = \sum_{n=n_i} w C_i, n_i \in R_{C_i} \quad (9)$$

The weights w need to be defined before the fusion process is applied. The recognition rates of each classifier can be used as weights.

4.3 Bayes Combination Rule

The Bayesian theory is known to all. Given m possible classes $\omega_1, \dots, \omega_m$, we assume that both the prior probabilities $P(\omega_j)$ and the conditional densities $p(x|\omega_j)$ for an observation x are known. Bayes formula shows how to calculate the posteriori probability:

$$P(\omega_j | x) = \frac{p(x | \omega_j)P(\omega_j)}{\sum_{k=1}^m p(x | \omega_k)P(\omega_k)} \quad (10)$$

This theorem can be used for the inference of the joint probability of the input classifiers.

In our case, we intend to combine the results from multiple classifiers on decision level. For a person, there are five walking sequences at one view in our database. We use the voting rule to decide the person's gender, assigning one vote to each sequence. Let $P_m(v_{ij})$ be the possibility of the i -th person being classified as male from the j -th classifier. $P_m(v_{ij})$ can be defined as:

$$P_m(v_{ij}) = \frac{N_m}{N_i} \quad (11)$$

where N_m denotes the number of votes on male and N_i is the number of total votes which is 5 in our database. Likewise, $P_f(v_{ij})$ denotes the possibility being classified as female. Then, the fusion can result in the possibility $p_m(w_i)$ of recognizing a walker as a male with the following expression:

$$p_m(w_i) = \frac{P(m) \sum_j P_m(v_{ij})}{P(m) \sum_j P_m(v_{ij}) + P(f) \sum_j P_f(v_{ij})} \quad (12)$$

where $P(m)$ is the percent that males take up in our database and $P(f)$ is the percent of females. They can be looked as the prior probabilities of the two classes.

5 Experiments

The first part of our experiments is to carry out gender recognition with our face data and gait data. There are seven different views available for gait in our database. Because camera C_4 captured the walking videos in the same view as camera C_5 , we leave these videos from C_4 for later experiments. Now, every subject has five walking sequences from one view angle and five corresponding face sequences. We trained and tested SVM on face features and gait features of each view separately. Hence, there are eight gender classifiers comprising front-view face and seven different views' gait. For making the best of our samples, we used the leave-one-out validation method. 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. The correct recognition rates(CRR) are listed in Table 1. We ran SVM with different kernels among which linear kernel performed the best.

Table 1. Results of gender recognition using SVM with linear kernel

modality	gait							face
	0 ⁰	30 ⁰	60 ⁰	90 ⁰	120 ⁰	150 ⁰	180 ⁰	
CRR	78.7%	90.3%	91.7%	90.3%	88.3%	84.0%	82.3%	90.0%

The second part is our fusion experiments. The three combination strategies mentioned in above section lead to an obvious improvement on CRR, as shown in Table 2. The proposed Bayes combination rule achieved the same results compared with the other schemes. Additionally, we wondered whether the face cue gave its contribution in this combination. So, another set of fusion experiments of integrating only the seven different views of gait were done and the results can also be found in Table 2. The results show the fact that there is an obvious increase when face cues are added in the fusion.

Table 2. Results of fusion

Fusion scheme	CRR by fusing face and multi-view gait	CCR by fusing of multi-view gait
Voting rule	98.3%	96.7%
Weighted voting rule	98.3%	96.7%
Bayes combination rule	98.3%	96.7%

In [14] Shan et al. exploit CCA to fuse face and gait at the feature level and their experiments achieve a superior recognition performance of 97.2% on CASIA Gait Database. However, this database doesn't include separate face data and Shan had to crop face images from front-view gait image sequences. Our self-building database can provide an advantage for the study of fusing face and gait and we will do a comprehensive investigation in this field.

6 Conclusions and Future Work

We have presented an effective two-modality fusion system for gender recognition. An important component of this system is a self-build database including front-view face data and gait data from seven different view. For every person in this database, these data are acquired in the same walking sequences. PCA is used to reduce facial image dimension and an improved spatio-temporal gait representation is introduced to extract the discriminating features from different gait view angles. Then, we employ SVM to implement gender classification based on the different modal features respectively. The final fusion is carried out according to three strategies: voting, weighted voting and Bayes combination. All of them reached the encouraging results.

In the future, we will apply different feature extraction methods to face and gait. In addition, more fusion schemes are waiting for us to make comparison with what we have done.

Acknowledgments

This work was supported by Program of New Century Excellent Talents in University, National Natural Science Foundation of China (No. 60575003, 60332010, 60873158), 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

1. Moghaddam, B., Yang, M.: Learning Gender with Support Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(5), 707–711 (2002)
2. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A Unified Learning Framework for Real Time Face Detection and Classification. In: *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 14–21 (2002)
3. Lu, H., Lin, H.: Gender Recognition using Adaboosted Feature. In: *The 3rd International Conference on Natural Computation*, vol. 2, pp. 646–650 (2002)
4. Parris, E.S., Carey, M.J.: Language Independent Gender Identification. In: *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 685–688 (1996)
5. Harb, H., Chen, L.: Gender Identification Using a General Audio Classifier. In: *Proc. of IEEE International Conference on Multimedia and Expo*, vol. 2, pp. 733–736 (2003)
6. Huang, T., Yang, Y., Wu, Z.: Combining MFCC and Pitch to Enhance the Performance of the Gender Recognition. In: *Proc. of IEEE International Conference on Signal Processing*, vol. 1, pp. 16–20 (2006)
7. Davis, J.W., Gao, H.: An Expressive Three-mode Principle Components Model for Gender Recognition. *Journal of Vision* 4, 362–377 (2004)
8. Lee, L., Grimson, W.: Gait Analysis for Recognition and Classification. In: *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 148–155 (2002)
9. Yoo, J., Hwang, D., Nixon, M.S.: Gender Classification in Human Gait with SVM. In: Blanc-Talon, J., Philips, W., Popescu, D.C., Scheunders, P. (eds.) *ACIVS 2005. LNCS*, vol. 3708, pp. 138–145. Springer, Heidelberg (2005)
10. Shakhnarovich, G., Lee, L., Darrell, T.: Integrated face and gait recognition from multiple views. In: *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 439–446 (2001)
11. Shakhnarovich, G., Darrell, T.: On probabilistic combination of face and gait cues for identification. In: *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 169–174 (2002)
12. Zhou, X., Bhanu, B.: Integrating face and gait for human recognition. In: *Proc. of IEEE workshop on Computer Vision and Pattern Recognition*, pp. 55–62 (2006)
13. Liu, Y., Collins, R., Tsin, Y.: Gait Sequence Analysis using Frieze Patterns. Technical Report CMU-RI-TR-01-38, Robotics Institute, Carnegie Mellon University (2001)
14. Shan, C., Gong, S., McOwan, P.W.: Learning Gender from Human Gaits and Faces. In: *IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 505–510 (2007)
15. Center for Biometrics and Security Research, CASIA, <http://www.cbsr.ia.ac.cn>
16. Bernard, A., Horst, B.: Combination of Classifiers on the Decision Level for Face Recognition. Tech. Rep. IAM-96-002, IAM, University Bern (1996)