Gait-Based Gender Classification in Unconstrained Environments

Jiwen Lu\textsuperscript{1}, Gang Wang\textsuperscript{1,2}, and Thomas S. Huang\textsuperscript{3}
\textsuperscript{1}Advanced Digital Sciences Center, Singapore
\textsuperscript{2}School of EEE, Nanyang Technological University, Singapore
\textsuperscript{3}Department of ECE, University of Illinois at Urbana-Champaign, IL USA
Email: jiwen.lu@adsc.com.sg; wanggang@ntu.edu.sg; huang@ifp.uiuc.edu

Abstract

This paper investigates the problem of gait-based gender classification in unconstrained environments. Different from existing human gait analysis and recognition methods which assume that humans walk in controlled environments, we aim to recognize human gender from uncontrolled gaits in which people can walk freely and the walking direction of human gaits may be time-varying in a single video clip. Given each gait sequence collected in an uncontrolled manner, we first obtain human silhouettes using background subtraction and cluster them into several groups. For each group, we compute the averaged gait image (AGI) as features. Then, we learn a distance metric under which the intraclass variations are minimized and the interclass variations are maximized, simultaneously, such that more discriminative information can be exploited for gender classification. Experimental results on our dataset demonstrate the efficacy of the proposed method.

1. Introduction

Over the past two decades, a number of human gait analysis and recognition methods have been proposed in the literature \cite{1, 5–8, 11}, and several benchmark datasets \cite{10, 12, 14} are also available for evaluation. However, to the best of our knowledge, all of the existing approaches (datasets) assume that people walk along a fixed direction in each video clip, which is impractical in many real applications because people usually walk freely and the walking direction may be time-varying. In this paper, we aim to classify human gender from uncontrolled gait, which means that people can walk freely along different directions, at different speeds, etc, in a single video sequence.

Classifying human gender from uncontrolled gait is difficult, partially because the view might change all the time and gait-based gender classification is known to be sensitive to the varying views \cite{14}. Our objective in this study is to classify human gender from such unconstrained gait sequences. We propose two methods to tackle such a challenge. First, for each video clip, we cluster the video frames into clusters. Each cluster is expected to gather video frames of similar views. Then, a robust averaged gait image (AGI) feature is extracted from each cluster. Since there are large intra-class variations among these AGIs extracted from the same gait sequence, it is hard to apply the point-to-point distance metric to measure the similarity. To address this, we propose to learn a discriminative metric, under which the intraclass and interclass variations described by the point-to-set distance are minimized and maximized, simultaneously.

We evaluate our approach on a newly collected dataset with 20 subjects, and our method can achieve 91.3\% correct classification rate using the leave one person out testing strategy, which is a promising performance number for such a hard problem. And the results also show our method significantly outperform several previous approaches.

2. Approach

Figure 1 shows the flow-chart of our approach. For each gait sequence, we first obtain human silhouettes using background subtraction and cluster them into several clusters. For each cluster, we compute the AGI as the gait feature. Then, we learn a distance metric under which the intraclass and interclass variations are minimized and maximized, simultaneously. The details are introduced in the following subsections.

2.1 Pre-processing

Given a gait sequence, the human silhouettes are extracted by background subtraction and thresholding in a way similar to the approach in \cite{10}. To make gait
representation insensitive to the distance between the camera and the subject, we resize each gait silhouette into $64 \times 44$ according to the centroid of each silhouette following the method in [10].

2.2 Clustering and Feature Extraction

The averaged gait image feature has been proven to be very powerful in representing human gaits because it is robust to preprocessing noises [4]. Since persons walk freely, human gait is no longer a periodical motion and we cannot detect the period in the gait sequence as the previous work [10]. Moreover, previous studies have shown that the AGI feature is sensitive to the varying views [14] and hence we cannot compute the AGI feature for the whole gait sequence directly due to the large view variation. To address this issue, we cluster each gait sequence into $K$ clusters. Each cluster is expected to gather human silhouettes of similar views. We simply apply the $K$-means algorithm and find good clustering results are achieved. More sophisticated clustering algorithms can also be used.

Assume there are $N_k$ frames for the $k$th cluster, $1 \leq k \leq K$, we compute the averaged gait image (AGI) similar to the gait energy image [4] feature for a cycle as the gait feature:

$$G_k(x, y) = \frac{1}{N_k} \sum_{p=1}^{N_k} l_{pk}(x, y)$$  \hspace{1cm} (1)

where $l_{pk}(x, y)$ is the $p$th human silhouette in the $k$ cluster, and $x$ and $y$ are 2-D image coordinates. Figure 1 shows twenty AGIs of two subjects’ gait sequences, where the top row shows AGIs of the first person (male) and the bottom row shows AGIs of the second person (female). We can see from this figure that there are large viewpoint variations within the gait sequence because the person walks in an uncontrolled manner.

2.3 Point-to-Set Metric Learning

Now, we have multiple AGI features for each gait sequence, and these AGIs describe human gaits from different views. Let $X = [x_1, x_2, \ldots, x_N]$ be a set of training AGI samples and $l = [l_1, l_2, \ldots, l_N]$ be the corresponding gender label, respectively, where $x_i \in R^d, l_i \in R^1, i = 1, 2, \cdots, N$. Assume there are $c$ classes and there are $N_j$ samples for the $j$th class, where $j = 1, 2, \cdots, c$. Hence, $N = \sum_{i=1}^{c} N_i$. Since gender classification is a binary classification problem, $c$ is 2 in our case. Conventional metric learning methods [2, 3, 13] usually employ the point-to-point distance to measure the pairwise similarity of two samples. However, such measure cannot effectively characterize the geometrical information of the AGIs, especially when they are captured from different views. To address this, we propose using the point-to-set distance to minimize the intraclass and maximize the interclass variations, respectively, by learning a distance metric $M$. We formulate the objective function as:

$$\max_{M} J = J_1 - J_2$$

$$= \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_1(\alpha))^{T} M (x_{ij} - D_1(\alpha))$$

$$= \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_2(\beta))^{T} M (x_{ij} - D_2(\beta))$$  \hspace{1cm} (2)

where $x_{ij}$ denotes the $j$th sample of the $i$th class, $D_1 \in R^{d \times k}$ and $D_2 \in R^{d \times k}$ are the $k$-nearest interclass and intraclass neighbors, respectively, $\alpha$ and $\beta$ are the corresponding reconstruction coefficients. The objective function of $J_1$ is to ensure that each $x_{ij}$ is linearly reconstructed by its $k$-nearest interclass samples and the reconstruction error is as large as possible. Similarly, the objective function of $J_1$ is to ensure each $x_{ij}$ is linearly reconstructed by its $k$-nearest intraclass samples and the reconstruction error is as small as possible under the distance metric $M$. $\alpha$ and $\beta$ can be easily obtained by solving the following constrained optimization problem as discussed in [9].

Since $M$ is symmetric and positive semidefinite,
where we can seek a nonsquare matrix \( W \) of size \( d \times l \), such that
\[
M = WW^T
\]  
(3)

Combining Eqs. (2) and (3), we simplify \( J \) to the following form
\[
J = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_1 \alpha)^T WW^T (x_{ij} - D_1 \alpha)
\]
\[
- \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_2 \beta)^T WW^T (x_{ij} - D_2 \beta)
\]
\[
= tr[W^T \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_1 \alpha)(x_{ij} - D_1 \alpha)^T W]
\]
\[
- tr[W^T \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_2 \beta)(x_{ij} - D_2 \beta)^T W]
\]
\[
= tr[W^T (H_1 - H_2) W]
\]  
(4)

where
\[
H_1 \triangleq \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_1 \alpha)(x_{ij} - D_1 \alpha)^T
\]  
(5)
\[
H_2 \triangleq \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_{ij} - D_2 \beta)(x_{ij} - D_2 \beta)^T
\]  
(6)

Now, we can formulate our point-to-set metric learning method as the following constrained optimization problem:
\[
\max_{W} J(W) = tr[W^T (H_1 - H_2) W]
\]  
(7)
subject to \( W^T W = I \).

where \( W^T W = I \) is a constraint to restrict the scale of \( W \) such that the optimization problem with respect to \( W \) is well-posed. Then, \( W \) can be obtained by solving the following eigenvalue problem
\[
(Z_1 - Z_2)w = \lambda w.
\]  
(8)

Let \( w_1, w_2, \ldots, w_l \) be the eigenvectors of Eq. (13) corresponding to the \( l \) largest eigenvalues ordered according to \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l \). A \( d \times l \) transformation matrix \( W = [w_1, w_2, \ldots, w_l] \) can be obtained to project each AGI feature sample \( x_{ij} \) into low-dimensional feature feature vectors \( y_{ij} \), as follows:
\[
y_{ij} = W^T x_{ij}.
\]  
(9)

where \( 1 \leq i \leq c \) and \( 1 \leq j \leq N_i \).

\subsection{2.4 Classification}

Given a testing gait sequence \( T \), we also first obtain human silhouettes using background substraction and cluster them into \( K \) clusters. For each cluster, we calculate the averaged gait image (AGI) as the gait feature: \( T_1, T_2, \ldots, T_K \). For the \( k \)th AGI \( T_k \), we can recognize the gender label by using the conventional nearest neighbor (NN) classifier. Lastly, we assign a gender label to the whole testing gait sequence using the majority voting rule for the \( K \) AGIs.

\section{Experimental Results}

\subsection{3.1 Dataset}

To investigate the problem of gait-based gender classification in unconstrained environments, we collect a new dataset for our experiments. A Microsoft Kinect depth camera fixed on a tripod is used to capture gait sequences on two different days in two indoor rooms. There are 20 subjects (13 males and 7 females) in our dataset, and each subject walks freely in the room. For each subject, we collect his/her gait sequences for four times, and hence there are a total of 80 gait sequences in our database. These depth sequence images are captured at a rate of 30 frames per second and the original resolution is \( 320 \times 240 \). The length of each gait sequence varies from 300 to 800 frames.

\subsection{3.2 Results and Analysis}

In our experiments, we adopt the leave one person out strategy to conduct experiments. Specifically, we take four gait videos for each person as the testing set and the remaining as the training set. We repeat this 20 times and record the average classification rate as the final classification accuracy. The number of clusters \( K \) in each sequence, the feature dimension \( l \), and the neighborhood parameters \( k \) of our proposed metric learning method are set as 10, 100, and 7 respectively.

**Comparisons with Existing Metric Learning Algorithms:** We compare our proposed metric learning method with three existing metric learning algorithms: neighborhood component analysis (NCA) [3], large margin nearest neighbor (LMNN) [13], and information-theoretic metric learning (ITML) [2]. For these three methods, we empirically set the number of the nearest neighbors as 5. We apply principal component analysis (PCA) to reduce each AGI feature into 100 dimensions for all metric learning algorithms to improve the speed. Table 1 tabulates the classification rate of different methods. We can observe that our proposed metric learning method
Table 1. Correct classification rate (%) of different metric learning methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCA [3]</td>
<td>88.8</td>
</tr>
<tr>
<td>LMNN [13]</td>
<td>87.5</td>
</tr>
<tr>
<td>ITML [2]</td>
<td>86.3</td>
</tr>
<tr>
<td>Our approach</td>
<td>91.3</td>
</tr>
</tbody>
</table>

Table 2. Correct classification rate (%) of different clustering strategies.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correct classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal clustering</td>
<td>82.5</td>
</tr>
<tr>
<td>Global clustering</td>
<td>80.0</td>
</tr>
<tr>
<td>Our clustering approach</td>
<td>91.3</td>
</tr>
</tbody>
</table>

consistently outperforms the other compared methods in terms of the classification accuracy. That is because our method applies the point-to-set similarity measure to learn the distance metric while others use the point-to-point distance which may not effectively model the large intraclass view variation.

Comparisons with Different Clustering Strategies: We perform our approach with another two clustering strategies: 1) temporal clustering: we cluster each gait sequence into K groups according to the temporal information of these image frames. Specifically, for a gait sequence containing N frames, we equally divide it into K segments and compute the AGI feature for each segment. 2) global clustering: we cluster all the image frames of all sequences into K group and calculate the AGI feature for each gait sequence. Table 2 records the classification accuracy when different clustering methods were used. We can see that the clustering strategy used in our approach achieves much better performance than both the temporal clustering and global clustering methods. The reason is that human gait view in our dataset is time-varying. Grouping from the temporal information mixes different views together. The global clustering method results in much quantization error, which decreases the discriminative power of the extracted AGI features.

4 Conclusion

In this paper, we have proposed a point-to-set metric learning approach to gait-based gender classification in unconstrained environments. To our knowledge, this is the first attempt to formally address this challenging problem. Experimental results on our dataset have shown the efficacy of our approach. For future work, we are going to apply the proposed technique to real world applications.

Acknowledgement

This study is supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sciences Center from the Agency for Science, Technology and Research (A*STAR) of Singapore.

References