Maximization of Mutual Information for Multi-view Gait-based Gender Classification Using Gabor Feature

Maodi Hu, Yunhong Wang, and Zhaoxiang Zhang
Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing, 100191, China
Londeehu@gmail.com, {yhwang,zxzhang}@buaa.edu.cn

Abstract

Soft biometric classification has attracted more and more attention with its increasing potential applications. We address the issue of gender classification based on multi-view gait sequences, which are defined by two basic problems. One is one-to-one case while the other is many-to-one case. For the first problem, there exists only one view in the gallery set. We employ Gabor filter and Maximization of Mutual Information (MMI) method to extract low dimensional features, and apply the Bayes rules based on Hidden Markov Models (HMMs) to gender classification. For many-to-one problem, a hierarchical framework based on view-related population HMM and Bayes rule is proposed to deal with the multi-view gait sequences. Experimental results demonstrate the advantages of our solutions to existing methods.

1. Introduction

Gait, as an inoffensive biometric, has been widely studied in recent decades. Considering humans are able to visually recognize the gender, ethnicity and age of a person from the manner of walking, a variety of approaches including computer vision and statistical technologies have been proposed to realize the automatic classification [12, 6, 9, 18, 17]. The knowledge about gender, ethnicity, age, or other traits such as height, weight, dimensions of limbs, skin color, hair color, etc., can be termed as soft biometrics, but only a limited number of traits can be gathered from a given sensor [13].

Zhang et al. [18] inspected the effects of view angles when classifying gender with gait patterns, which provides the first investigation on multi-view gait-based gender classification. The common limitation of other related studies [12, 6, 9, 17] is that only the side view is referred in their experiments. Our approach extends the evaluation by integrated definitions.

For feature description of gait patterns, there are two choices in general: model-based method and model-free (appearance-based) method. Although model-based features such as trajectories of joint positions [4, 7] are more convenient to be view-normalized, some important information of shape variation is abandoned. In terms of view angle, soft biometric classification is different from identification. That is because we can always capture the gait motion of specific gender, age or other category in given views, but it can not be applied to the case that one subject requires to be identified. View transformation is a hot topic in gait identification. However, adequate samples of both gallery and probe views are required by appearance-based view transformation [8]. And relatively low recognition rate is received by joint trajectories [4, 7]. Since the soft biometric features of various views can be extracted from existing databases, it is inefficient to regulate the view if the samples in the probe view are adequate. Therefore, view transform is not considered in the application of multi-view gait-based gender classification.

There are also two main categories of HMM-based gait analysis. The first one is from feature to model, which classifies the probe data by selecting the model with maximum likelihood (or posterior). Cheng et al. [3] utilized Gaussian Process Latent Variable Model (GPLVM) to transform the gait silhouette into low dimensional embedding, and learned the temporal dynamics via HMM in the corresponding embedding space. The second category is from model to feature, and these works were aimed to extract the templates and stances without traditional period detection based on similarity. Liu et al. [11] proposed population Hidden Markov Model (pHMM), in which the hidden states represent gait stances over one gait cycle, and Viterbi decoding are used to arrive at one dynamics-normalized, averaged gait cycle of fixed length. To maximize the similarity between persons, shape distances were further computed in the Linear Discriminant Analysis (LDA) space. However, LDA is unsuitable to reduce the feature dimension for soft biometric classification. For example, the value of the upper
concatenate the serialized vectors to serve as the original
of several instances are displayed in Fig. 3. And then we
pass filtering with standard deviation 3. The filter output
are further downsampled to 15-by-10 after Gaussian low-
employed. The complex modulus of each response image
curs now and then. To enhance the representation, a bank
for data collection, incomplete silhouette segmentation oc-
nette cropping is implemented by background modeling. But
are required for feature extraction given probe data-cases.
Gabor responses to feature space. Note that only grey ones
feature extraction is shown in Fig. 2. In this figure, the white
classification. The whole procedure of moment-based gait
ratio, this database is sufficient for the investigation on gender
(Dataset B) [17]. In terms of view variation and gender ra-
ton 3 introduces the key issues for multi-view gender clas-
sion 2, we briefly describe the Gabor-MMI gait feature. Sec-
proposed HMM-based hierarchical framework accurately
models the granularity of view and gender, and achieves
high classification performance by Bayes decision.

The rest of this paper is organized as follows. In Sec-
section 4. Section 5 concludes the paper.

2. Moment-based Descriptor

Following most of gait analysis studies, we take silhou-
ettes as the starting point for our soft biometric classification.
Fig. 1 displays the multi-view silhouettes ranging from
0° to 180°, which are drawn from CASIA Gait Database
(Dataset B) [17]. In terms of view variation and gender ra-
tio, this database is sufficient for the investigation on gender
classification. The whole procedure of moment-based gait
feature extraction is shown in Fig. 2. In this figure, the white
cases indicate the training process of linear mapping from
Gabor responses to feature space. Note that only grey ones
are required for feature extraction given probe data-cases.

As studied in the related works, generally the silhou-
ette cropping is implemented by background modeling. But
there is no single universal method suitable for various envi-
rnments yet. Even with the controlled indoor environment
for data collection, incomplete silhouette segmentation oc-
curs now and then. To enhance the representation, a bank
of 18 Gabor filters spanning 3 scales and 6 orientations are
employed. The complex modulus of each response image
are further downsampled to 15-by-10 after Gaussian low-
pass filtering with standard deviation 3. The filter output
of several instances are displayed in Fig. 3. And then we
concatenate the serialized vectors to serve as the original
boundary dimension is one for gender classification, due to
the insufficient rank of inter-class scatter matrix.

On account of less computational complexity in test pro-
cedure, we prefer to learn the multi-view classifiers on
linear-transformed sequential features. The combination
of Gabor filter and Maximization of Mutual Information
(MMI) [5] provides an effective way to fuse nonlinear en-
hancement and linear dimension reduction. Moreover, our
proposed HMM-based hierarchical framework accurately
models the granularity of view and gender, and achieves
high classification performance by Bayes decision.

The low dimensional discriminative feature γ is similar
to fisherface [2] features. To learn the linear mapping W
efficiently, we randomly select a few samples from each
sequence. The training process of W includes two steps.
In the first PCA step, the eigenvectors corresponding
to the D largest eigenvalues of covariance matrix are con-
catenated to W_{PCA}, and the D dimensional intermediate
variable θ is obtained by θ = W_{PCA}x, which is also
known as Gabor-PCA feature [10]. The next step is to find
the optimal discriminative transformation W_{MMI} satisfying
γ = W_{MMI}θ. MMI is a linear mapping technique [15]
that maximizes quadratic divergence measure between soft
biometric label g (for gender, g ∈ {0, 1}) and low dimen-
sional feature γ = {γi}. The quadratic mutual information
measure is as following.

\[ I_T(g, γ) = \sum_g \int [p(g, γ)]^2 + [p(γ)]^2 [p(γ)]^2 \\ -2p(g, γ)p(γ)p(γ)dγ, \]  

where, the probability terms p(g), p(γ), p(g, γ) and their
corresponding derivatives are estimated by non-parametric
estimation. The detailed solution to this optimization prob-
lem can be found in [15]. The optimal moment-based linear
mapping W constructs a low dimensional discriminative s-
pace for classification. However, gait process is not only
a collection of stance representations but also the individual
motion characteristic. We model the temporal relationships
of extracted sequential features using HMM.

3. Multi-view Gender Classification

The issues of multi-view gait-based soft biometric clas-
sification are discussed in this section. The temporal rela-
tionship of extracted features are modeled by HMM, in
which each state is fitted by a Gaussian Mixture Model (GMM). Moreover, as depicted in Fig. 4(b), supervised learning makes the likelihoods of different categories more reliable by the virtue of discriminative subspace feature. The basic problems are defined as one-to-one and many-to-one soft biometric classifications. Other cases can be converted to either of them. Notice that, view angle and soft biometric category are assumed as different levels of decision granularity.

3.1. Solution to One-to-one Case

**Definition 1** Given the probe data from an arbitrary view \( v_r \), one-to-one evaluation is to assess the classification performance when there is only a specific view \( v_r \) in the gallery set.

Taking two-category classification as examples, the solution is presented as follows. Firstly, Two HMMs (denoted by \( \lambda_1 \) and \( \lambda_0 \)) are trained independently by the samples of the two categories. Note that, a category indicates one single gender group, and multi-classification can be defined in a similar way. An instance of modeling results is illustrated in Fig. 4(b), in which only diagonal covariance and one component are presumed. It can be clearly seen that the moment-based information is mapped to the discriminative Gabor-MMI space through supervised learning, and markov modeling of temporal relationship enhances the representation of movement distribution. Considering the purpose of evaluation, only the test process is presented. For the details of Baum-Welch training, please refer to [14].

Let \( g \in \{0, 1\} \) denote the category label. Given the Gabor-MMI feature sequence \( \Gamma_T = \{\gamma_i\}_{t=1}^T \) and \( P(\Gamma_T|\lambda_g) \) can be solved by forward algorithm. Similar to [14], we define \( \alpha_t(i) \) to facilitate the following descriptions.

\[
\alpha_t(i) = \pi_i b_{ij},
\]

\[
\alpha_{t+1}(j) = \sum_{i=1}^{n} \alpha_t(i) a_{ij} b_{jz},
\]

where \( a_{ij} \) denotes the probability of being in state \( i \) given that the previous state was \( j \), and \( b_{jz} \) denotes the probability of observed feature \( z \) in state \( j \). Let \( \alpha_{g,T(i)} \) indicate \( \alpha_T(i) \) in the model \( \lambda_g \), then the forward algorithm provides an online-test way for likelihood calculation, which is

\[
P(\Gamma_T|\lambda_g) = \sum_{i=1}^{n} \alpha_{g,T(i)}.\]

Considering the negative influence of occasional abnormal actions and part occlusions, small gap between model likelihoods can not be interpreted as the result of comparison. Intuitively, we design to optimize the frame length by maximizing the probability gap of the different models.

Given a probe feature sequence of length \( k \), the following error elimination method is presented to determine the optimal length for classification.

\[
\varphi(k, \Gamma, \{\lambda_g\}) = \begin{cases} 
1 & \text{if } k = 1 \\
\frac{k}{\varphi(k-1, \Gamma, \{\lambda_g\})} & \text{otherwise} \\
\end{cases}
\]

\[
\rho(k, \Gamma, \{\lambda_g\}) = \| \log P(\Gamma_k|\lambda_0) - \log P(\Gamma_k|\lambda_1) \|,\]

\[
\text{th} = \frac{1}{n} \sum_{i=k-n}^{k-1} \rho(i, \Gamma, \{\lambda_g\}).
\]

\( \text{th} \) serves as a dynamic threshold to limit the instantaneous deviation while keeping the sequence as long as possible. It is implemented by a First-In First-Out (FIFO) queue, and the value of \( n \) is 15 in our experiments. We write Equation (6) as \( \varphi(k) \) for short hereinafter. Assuming uniform prior probabilities for different models,

\[
P(\lambda_0) = P(\lambda_1),
\]

Maximum A Posterior (MAP) rule produces the the same decision Maximum Likelihood Estimate (MLE) one de-
duces to. Let \( r^e \in \{0, 1\} \) denote the estimated category label. Given a new data-case \( \Gamma_{v}(T) \), the Bayes decision is

\[
\hat{r}^e = \arg\max_g P(g | \Gamma_{v}(T)) = \arg\max_g P(\Gamma_{v}(T) | \lambda_g).
\]

In most benchmark databases, there are the equal amounts of data-cases in different views. Let \( U = \{u_i\}_{i=1}^{N_u} \) denote the probe subjects, \( V = \{v_i\}_{i=1}^{N_v} \) denote the view angles, \( v_{tr} \) be the specific view angle of the gallery set, \( r^e_{u_i} \) be the true label of \( u_i \), \( r_{v_{tr}, u_i, v_i} \) be the estimated one. The CCR of probe subject \( u_i \) in the view \( v_i \) is

\[
CCR_{v_{tr}, u_i, v_i} = \delta(r^e_{u_i}, r_{v_{tr}, u_i, v_i}),
\]

where \( \delta \) is Kronecker delta. And the one-to-one \((v_{tr} \to v_i)\) CCR can be derived as

\[
CCR_{v_{tr}, u_i, v_i} = \frac{1}{N_u} \sum_{u_i} CCR_{v_{tr}, u_i, v_i}.
\]

Moreover, one-to-many CCR of gallery view \( v_{tr} \) is the average value of one-to-one CCRs among all the probe views, \( CCR_{v_{tr}} = \frac{1}{N_v} \sum_{v_i} CCR_{v_{tr}, u_i, v_i} \). One-to-one CCR and one-to-many CCR are two valuable benchmarks, especially when only one view is available in the gallery set. Most of prior works only focus on 90° to 90° CCR, which is one of the one-to-one CCRs. But more often, the gait patterns of unknown view angles are expected to be recognized. In this case, one-to-many CCR can reliably assess the robustness of a specific gallery view. In addition, the average of one-to-many CCRs in all the gallery views could be the evaluation of arbitrary single-view modeling.

### 3.2. Solution to Many-to-one Case

**Definition 2** Given the probe data from an arbitrary view \( v_i \), Many-to-one evaluation is to assess the classification performance when there are several views in the gallery set.

The only difference between the two basic problems is the number of view angles in the gallery set. Since multi-view gait sequences have been collected in various databases, soft biometric features in different views could be extracted with respect to the category labels. To solve the classification problem, many-to-one case is analogous with Multi-Instance Learning (MIL), in which each view is an instance of one category. But traditional classifier like MI-SVM [1] is not suitable for the sequential data. We apply a combination of Gabor-MMI, view-related pHMMs (vrpHMMs) and Bayes decision to the many-to-one classification.

Temporal relationship can be modeled by HMMs with respect to view angle and category label, in which two strategies are used. The first one is an intuitive method, which treats the views and soft biometric categories as the equivalent granules. Besides the uniform prior probabilities of different categories, we make the same assumption for view angles. Moreover, considering the independence between them, we can obtain

\[
P(v_i) = P(v_j), \ P(g, v_i) = P(g)P(v_i).
\]

The estimated category label \( r^e \) and view angle \( v^e \) are

\[
\{r^e, v^e\} = \arg\max_g \arg\max_{v_i} P(\Gamma_{v}(T) | \lambda_g, v_i),
\]

where, \( \lambda_{g, v_i} \) is trained by the feature sequences of category \( g \) and view \( v_i \). This equation requires \( N_v \times 2 \) times of forward tests, in which \( N_v \) is number of views in the gallery set.

Another intelligent hierarchical method we proposed here is to reckon the view first, and then to calculate the likelihood of each category with the estimated view. It can be interpreted as increasing the granularity by a coarse model and decreasing it by non-overlapped decisions. Inspired by pHMM [11], we propose the vrpHMM to represent the average dynamics of a set of individuals in each single view. The modeling process is similar to the description in one-to-one case, except the feature representation. Considering the feature function, Gabor-PCA features \( \Theta = \{\theta_i\}_{i=1}^{T} \) are used in the training and test processes of vrpHMM. Let \( \{\lambda_{v_i}\}_{i=1}^{N_v} \) represent the gait dynamics in their respective views. The hierarchical classification decision based on vrpHMMs can be written as

\[
v^e = \arg\max_{v_i} P(\theta_{v}(T) | \lambda_{v_i}), \quad r^e = \arg\max_g P(\Gamma_{v}(T) | \lambda_{g, v^e}),
\]

where, \( \lambda_{v_i} \) denotes the vrpHMM in the view \( v_i \), and \( v^e \) is the estimated view by vrpHMMs. Moreover, \( \rho(k, \theta, (\lambda_{v_i})) \) in Equation 6 is substituted as

\[
\hat{\rho}(k, \theta, (\lambda_{v_i})) = \sum_{v_i} \log P(\theta_k | \lambda_{v_i}) - \frac{1}{N_v} \sum_{v_j} \log P(\theta_k | \lambda_{v_j})
\]

The hierarchical procedure for gender classification is shown in Fig. 5, in which the top chain denotes vrpHMM and the bottom one denotes the HMM for final classification. This algorithm requires only \( N_v + 2 \) times of likelihood calculations by forward algorithm, which significantly reduces the computational complexity. The Correct Classification Rate (CCR) of probe subject \( u_i \) in the view \( v_i \) is

\[
CCR_{u_i, v_i} = \delta(r^e_{u_i}, r_{v_{tr}, u_i, v_i}),
\]

and many-to-one CCR is

\[
CCR_{v_i} = \frac{1}{N_u} \sum_{u_i} CCR_{u_i, v_i},
\]

In case that gait patterns of several views are available in gallery set, the solution to multi-view problem no longer depends on view normalization. For example, we are able to extract the features covering various views from existing gait databases. Therefore, many-to-one and many-to-many
(\frac{1}{n} \sum_{v_i} CCR_{v_i}) CCRs are more important than the preceding ones. To be more specific, many-to-many CCR can assess the generalization ability in multi-view applications, whereas many-to-one ones appraise the results of specific probe view angle confronting various gallery views.

4. Experimental Results

In this section, we apply the proposed approach to multi-view gait-based gender classification. Note that, considering the focus of this work, Bayes decision on HMMs serves as the only classifier we used in the following experiments.

4.1. Gender Classification

Following most works on gait-based gender classification, we choose the CASIA Gait Database (Dataset B) [16] and IRIP Gait Database [18] for our experiments. CASIA (Dataset B) database captures the walking patterns from the 11 view angles ranging from 0° to 180°. Full length gait sequences of all the 31 females and 31 randomly selected males are covered in our experiments. Considering there are 6 sequences for each subject, the sum of their log-likelihoods is taken for comparison. Each of 31 subsets is composed by a male and a female. Each time we select one of the unchecked non-overlapping subsets as probe data, and use the remaining 60 gallery subjects to train the models.

As we described before, most of existing gender classification works focus on the CCR in the side view (90°), which is one of the fixed view one-to-one CCRs. The results of different algorithms on this database are shown in the Table 1.

Table 1. 90° view gender CCRs by different gender classification algorithms on CASIA Gait Database (Dataset B).

<table>
<thead>
<tr>
<th>Method</th>
<th>Dataset</th>
<th>CCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al. [12]</td>
<td>25 males and 25 females</td>
<td>85.0%</td>
</tr>
<tr>
<td>Huang et al. [6]</td>
<td>25 males and 25 females</td>
<td>85.0%</td>
</tr>
<tr>
<td>Li et al. [9]</td>
<td>31 males and 31 females</td>
<td>93.28%</td>
</tr>
<tr>
<td>Yu et al. [17]</td>
<td>31 males and 31 females</td>
<td>95.97%</td>
</tr>
<tr>
<td>Ours</td>
<td>31 males and 31 females</td>
<td>96.77%</td>
</tr>
</tbody>
</table>

Similar to multi-view identification work [4], Table 2 illustrates the one-to-one CCRs covering 6 conventional view angles, where the labels of rows and columns denote the view angles of gallery and probe sets respectively. It can serve as a baseline performance, since only 90° view gender classification on this database was considered in existing literatures. The variation trends of these CCRs coincide with common sense. Gait pattern in the 90° view performs better than others.

The CCRs of many-to-one case are shown in Figure 7, in which all the 11 views are used in gallery set. It can be clearly seen that our proposed viHMMs based hierarchical classification strategy achieves better CCRs than the intuitive one. Note that, the hierarchical approach also reduces the computational complexity. The granularity relationship of view and gender is correctly represented by the hierarchical models, which means the gap between different genders is much smaller than the gap between the population gait dynamics of adjacent (at intervals of 18°) views. We
suggest that the many-to-many CCR (91.79% in our work), which is also the average value of many-to-one CCRs of all the gallery views and all the probe views, should be the most notable benchmark for gait-based gender classification.

Table 2. One-to-one gender CCRs by Gabor-MMI feature on CASIA Gait Database (Dataset B), in which row (G) and column (P) denote gallery view and probe view respectively.

<table>
<thead>
<tr>
<th>G \ P</th>
<th>36°</th>
<th>54°</th>
<th>72°</th>
<th>90°</th>
<th>108°</th>
<th>126°</th>
</tr>
</thead>
<tbody>
<tr>
<td>36°</td>
<td><strong>91.94</strong></td>
<td>85.48</td>
<td>64.52</td>
<td>56.45</td>
<td>61.29</td>
<td>74.19</td>
</tr>
<tr>
<td>54°</td>
<td><strong>91.94</strong></td>
<td><strong>91.94</strong></td>
<td>79.03</td>
<td>75.81</td>
<td>75.81</td>
<td>83.87</td>
</tr>
<tr>
<td>72°</td>
<td>77.42</td>
<td>79.03</td>
<td><strong>95.16</strong></td>
<td>93.55</td>
<td>87.10</td>
<td>82.26</td>
</tr>
<tr>
<td>90°</td>
<td>62.90</td>
<td>72.58</td>
<td><strong>96.77</strong></td>
<td><strong>96.77</strong></td>
<td>95.16</td>
<td>91.94</td>
</tr>
<tr>
<td>108°</td>
<td>74.19</td>
<td>72.58</td>
<td>90.32</td>
<td><strong>93.55</strong></td>
<td><strong>95.16</strong></td>
<td>88.71</td>
</tr>
<tr>
<td>126°</td>
<td>69.35</td>
<td>85.48</td>
<td>91.94</td>
<td>88.71</td>
<td><strong>91.94</strong></td>
<td><strong>95.16</strong></td>
</tr>
</tbody>
</table>

Figure 7. Many-to-one gender CCRs by different strategies on CASIA Gait Database (Dataset B).

In addition, the experiments on IRIP Gait Database (32 males and 28 females) [18] have been further conducted. The previous work [18] provides baseline fixed-view one-to-one CCRs on this database by projection feature and kernel SVM, while none of work on CASIA one considers multi-view gender classification. With the same leave-one-out protocol in [18], the results of our approach are shown in Table 3. Obviously, the proposed method with Gabor-MMI feature and HMM-based decision achieves state-of-the-art performance in the two databases.

Table 3. Fixed view one-to-one gender CCRs on IRIP Gait Database.

<table>
<thead>
<tr>
<th>View Angle</th>
<th>linear [18]</th>
<th>polynomial [18]</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>78.7%</td>
<td>79.7%</td>
<td><strong>90.0%</strong></td>
</tr>
<tr>
<td>30</td>
<td>90.3%</td>
<td>91.0%</td>
<td><strong>91.7%</strong></td>
</tr>
<tr>
<td>60</td>
<td><strong>91.7%</strong></td>
<td>82.7%</td>
<td><strong>91.7%</strong></td>
</tr>
<tr>
<td>90</td>
<td>90.3%</td>
<td>78.0%</td>
<td><strong>93.6%</strong></td>
</tr>
<tr>
<td>120</td>
<td><strong>88.3%</strong></td>
<td>85.3%</td>
<td><strong>88.3%</strong></td>
</tr>
<tr>
<td>150</td>
<td>84.0%</td>
<td>83.0%</td>
<td><strong>86.7%</strong></td>
</tr>
<tr>
<td>180</td>
<td><strong>82.3%</strong></td>
<td>78.3%</td>
<td><strong>82.3%</strong></td>
</tr>
<tr>
<td>Average</td>
<td>86.5%</td>
<td>82.6%</td>
<td><strong>89.2%</strong></td>
</tr>
</tbody>
</table>

One can achieve acceptable CCRs of arbitrary views by the hierarchical strategy. The experimental results on two databases confirm the efficiency of our proposed method.

5. Conclusions

In this paper, we unify the definitions of soft biometric classification based on multi-view gait patterns, and propose efficient solutions for different gallery view cases. With adequate gait sequences captured from various views, one can achieve acceptable CCRs of arbitrary views by the hierarchical strategy. The experimental results on two databases confirm the efficiency of our proposed method.

Acknowledgement

This work is funded by the National Basic Research Program of China (No. 2010CB327902), the National Natural Science Foundation of China (No. 60873158, No. 61005016, No. 6106130560), the Fundamental Research Funds for the Central Universities, and the Innovation Foundation of BUAA for PhD Graduates.

References


