Original citation:

Permanent WRAP url:
http://wrap.warwick.ac.uk/64465

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 3.0 (CC BY 3.0) license and may be reused according to the conditions of the license. For more details see: http://creativecommons.org/licenses/by/3.0/

A note on versions:
The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk
On Reducing the Effect of Covariate Factors in Gait Recognition: A Classifier Ensemble Method

Yu Guan, Chang-Tsun Li, Senior Member, IEEE, and Fabio Roli, Fellow, IEEE

Abstract—Robust human gait recognition is challenging because of the presence of covariate factors such as carrying condition, clothing, walking surface, etc. In this paper, we model the effect of covariates as an unknown partial feature corruption problem. Since the locations of corruptions may differ for different query gait videos, relevant features may become irrelevant when walking condition changes. In this case, it is difficult to train one fixed classifier that is robust to a large number of different covariates. To tackle this problem, we propose a classifier ensemble method based on the random subspace method (RSM) and majority voting (MV). Its theoretical basis suggests it is insensitive to locations of corrupted features, and thus can generalize well to a large number of covariates. We also extend this method by proposing two strategies, i.e., local enhancing (LE) and hybrid decision-level fusion (HDF) to suppress the ratio of false votes to true votes (before MV). The performance of our approach is competitive against the most challenging methods normally incur high computational costs.

Index Terms—Classifier ensemble, random subspace method, local enhancing, hybrid decision-level fusion, gait recognition, covariate factors, biometrics

1 INTRODUCTION

Compared with other biometric traits like fingerprint or iris, the most significant advantage of gait is that it can be used for remote human identification without subject cooperation. Gait analysis has contributed to convictions in criminal cases in some countries like Denmark [1] and UK [2]. However, the performance of automatic gait recognition can be affected by covariate factors such as carrying condition, camera viewpoint, walking surface, clothing, elapsed time, etc. Designing a robust system to address these problems is an acute challenge. Existing gait recognition methods can be roughly divided into two categories: model-based and appearance-based approaches. Model-based methods (e.g., [3]) employ the parameters of the body structure, while appearance-based approaches extract features directly from gait sequences regardless of the underlying structure. This work falls in the category of appearance-based methods, which can also work well on low-quality gait videos, when the parameters of the body structure are difficult to estimate precisely.

The average silhouette over one gait cycle, known as gait energy image (GEI), is a popular appearance-based representation [4]. The averaging operation encodes the information of the binary frames into a single grayscale image, which makes GEI less sensitive to segmentation errors [4]. Several GEI samples from the USF dataset [5] are shown in Fig. 1. Recently, Iwama et al. evaluated the effectiveness of several feature templates on a gait dataset consisting of more than 3,000 subjects. They found that good performance can be achieved by directly matching the GEI templates, when there are no covariates [6]. However, it is error-prone when covariates exist. Therefore, many researchers have been formulating various feature descriptors to capture discriminant information from GEIs to deal with different covariates. In [4], Han and Bhanu utilized principle component analysis (PCA) and linear discriminant analysis (LDA) to extract features from the concatenated GEIs. To extract gait descriptors, Li et al. proposed a discriminant locally linear embedding (DLEL) framework, which can preserve the local manifold structure [7]. By using two subspace learning methods, coupled subspace analysis (CSA) and discriminant analysis with tensor representation (DATR), Xu et al. extracted features directly from the GEIs [8]. They demonstrated that the matrix representation can yield much higher performance than the vector representation reported in [4]. In [9], after convolving a number of Gabor functions with the GEI representation, Tao et al. used the Gabor-filtered GEI as a new gait feature template. They also proposed the general tensor discriminant analysis (GTD) for feature extraction on the high-dimensional Gabor features [9]. To preserve the local manifold structure of the high-dimensional Gabor features, Chen et al. proposed a tensor-based Riemannian manifold distance-approximating projection (TRIMP) framework [10]. Since spatial misalignment may degrade the performance, based on Gabor representation, Image-to-Class distance was utilized in [11] to allow feature matching to be carried out within a spatial neighborhood. By using the techniques of universal background model (UBM) learning and maximum a posteriori (MAP) adaptation, Xu et al. proposed the Gabor-based patch distribution feature (Gabor-PDF) in [12], and the classification is performed based on locality-constrained group sparse representation (LGSR). Compared with GEI features (e.g., [4], [8]), Gabor features (e.g., [9], [11], [12]) tend to be more discriminant and can yield higher accuracies. However, due to the high dimension of Gabor features, these methods normally incur high computational costs.

Through the “cutting and fitting” scheme, synthetic GEIs can be generated to simulate the walking surface effect [4]. In [13], the population hidden Markov model was used for dynamic-normalized gait recognition (DNGR). Both methods can yield encouraging performance against the walking surface covariate. In [14], Wang et al. proposed chrono-gait image (CGI), with the temporal information of the gait sequence encoded. Although CGI has similar performance to GEI in most cases [6], [14], it outperforms GEI in tackling the carrying condition covariate [14]. In [15], through fusing gait features from different cameras, it was found that short elapsed time does not affect the performance significantly in a controlled environment. Clothing was instead deemed as the most challenging covariate [15]. Based on a gait dataset with 32 different clothes combinations, Hossain et al. proposed an adaptive scheme for weighting different body parts to reduce the effect of clothing [16]. Yet it requires an additional training set that covers “all” possible clothes types, which is less practical in real-world applications.

Most of the previous works have satisfactory performance against some covariates like carrying condition, shoe type, (small changes in) viewpoint, etc. However, so far it still remains an open question to tackle covariates like walking surface, clothing, and elapsed time in less controlled environments. As such, our aim is not only to build a general framework that can generalize to a large number of covariates in unseen walking conditions, but also to address these open issues. Based on GEIs, we model the effect caused by various covariates as an unknown partial feature corruption problem and propose a weak classifier ensemble method to reduce such effect. Each weak classifier is generated by random sampling in the full feature space, so they may generalize to the unselected features in different directions [17]. This concept, named random subspace method (RSM), was initially proposed by Ho [17] for constructing decision forests. It was successfully applied to face recognition (e.g., [18]). In our previous works [19],
[0x0]Modelling of gait recognition challenges.

2 argmin d be the 2DPCA transformation matrix consisting of the that can suppress g 2½...

(C138)
t...

non-zero principle components such that (¼½¼½/C0 T¼ for different walking conditions.

Classifier ensemble solution and its extensions.

2½^ROBLEM indicates the corrupted gait fea-

2½^projection directions from f¼ 1 argmin T1;tt F/C138 for different walking conditions.

non-

1 corresponding to differ-

2 d...;t...should only extract the relevant features 2 different walking conditions. In (3), since

2¼M 2P

mentally evaluated and compared with other algorithms in Fig. 2. Several examples of ¼^I corresponding to differ-

t walking conditions are also illustrated in Fig. 3. For an unknown walking condition i, ¼^I indicates the corrupted gait fea-

tures with unknown locations. Before matching, we need to find a feature extractor T* that can suppress ¼^I, i.e.,

T* = argmin ||¼^I,T||2 , i 2 [1,F],

(2)

where T can extract features in the column direction of ¼^I. However, the locations of corruptions may differ for differ-

t walking conditions, as shown in Fig. 3. Given such non-deterministic nature of ¼^I, from (2) we can see that it is difficult to find a fixed T* that can extract effective features that can general-

eize to a large number of different covariates. In light of this, an effective T* should only extract the relevant features dynamically for different walking conditions.

2.2 Problem Formulation

For high dimensional gait feature templates, PCA is usually used for dimension reduction (e.g., [4], [14]). In this work, we instead use the matrix-based two-dimensional PCA (2DPCA) [24] because of its lower computational costs and higher performance.

Let T be the 2DPCA transformation matrix consisting of the leading d non-zero principle components such that T = [t1,t2,...,td]. Since [t1,t2,...,td] are pairwise orthogonal vectors, (2) can be written as:

T* = argmin d j=1 ||¼^I,tj||2 , i 2 [1,F].

(3)

It is difficult for traditional 2DPCA with a fixed T* = [t1,t2,...,td] to reduce the effect of F different walking conditions. In (3), since the terms ||¼^I,tj||2, j 2 [1,d] are non-negative, it is possible to reduce the effect of covariates by selecting a subset of [t1,t2,...,td] to form a new feature extractor.

To see the reduced effect of covariates, we form different feature extractors by randomly selecting 2 projection directions from [t1,t2,...,td]. In Fig. 4, we report the corresponding normalized
Euclidean distances from the gallery sample in Fig. 1a to the probe samples in Figs. 1b, 1c, 1d, 1e, 1f, 1g, respectively. We can see the effect of covariates in Fig. 1b/1e can be reduced by using $[t_2, t_{35}]$, while the effect in Fig. 1g can be reduced based on $[t_{15}, t_{41}]$. These observations empirically validate that it is possible to select subsets of $[t_1, t_2, \ldots, t_d]$ as feature extractors to reduce the effect of covariates. However, the optimal subset may vary for different walking conditions due to the non-deterministic nature of $I$.

Therefore, we formulate the problem of reducing the effect of unknown covariates as a dynamic feature selection problem. For an unknown walking condition, we assume the irrelevant features (mainly corresponding to the corrupted pixels, as indicated by $I$) exist in $m(0 < m < d)$ unknown non-negative terms of (3). We aim to dynamically select $N \in [1, d]$ relevant features (mainly corresponding to the uncropped pixels) for classification. It is obvious that the probability of selecting $N$ relevant features is 0 when $N > d - m$. When $N \in [1, d - m]$ is randomly selecting $N$ features out of $d$ features, the probability of not choosing the $m$ irrelevant features $P(N)$ follows the hypergeometric distribution [25]. This can be expressed as the ratio of binomial coefficients:

$$
P(N) = \binom{(d-m)}{N} \binom{d}{N} / \binom{d}{N}, \quad N \in [1, d - m].$$

$P(N)$ is a generalization measure that is related in someway to the performance of the matching algorithm. Classifiers with greater $P(N)$ tend to generalize well to unseen covariates, and vice versa.

**Lemma 1.** Let $d, m, N$ be the numbers of the total features, irrelevant features (with $0 < m < d$), and randomly selected features, respectively, then $P(N)$ in (4) is inversely proportional to $N$ in the range $N \in [1, d - m]$.

**Proof.** It can be proved using mathematical induction. From (4), it is easy to see that $P(N + 1)/P(N) = 1 - m/(d - N) < 1$ for $N \in [1, d - m]$. This completes the proof of the lemma. □

It is preferable to set $N$ to smaller values, which can yield greater $P(N)$, according to Lemma 1. Specifically, when $N \ll d$, the hypergeometric distribution can be deemed as a binomial distribution [25], we have

$$
P(N) = \binom{N}{N} p^N (1-p)^{N-N} = p^N, \quad N \ll d,$$

where $p$ is the probability of randomly selecting one relevant feature. Since $p = 1 - m/d$, (5) can be written as:

$$
P(N) = \left(1 - \frac{m}{d}\right)^N, \quad N \ll d,$$

which clearly reflects the simple relationship between $P(N)$ and the other parameters when we set $N \ll d$.

### 2.3 Classifier Ensemble Solution and its Extensions

We use a classifier ensemble strategy for covariate-invariant gait recognition. After repeating the random feature selection process $L$ times ($L$ is a large number), $L$ base classifiers with generalization ability $P(N)$ (see (6)) are generated. Classification can then be performed by majority voting (MV) [26] among the labeling results of the base classifiers.

Given a query gait and the gallery consisting of $c$ classes, we define the true votes $V_{true}$ as the number of classifiers with correct prediction, while the false votes $\{V_{false}^{i}\}_{i=1}^{c}$ as the incorrectly predicted classifier number distribution over the other $c - 1$ classes. Given that, through MV, the final correct classification can be achieved only when $V_{true} > \max\{V_{false}^{i}\}_{i=1}^{c}$.

#### Ideal cases analysis

We make two assumptions in ideal cases:

1. When irrelevant features are not assigned to a classifier (referred to as unaffected classifier), correct classification should be achieved.
2. When irrelevant features are assigned to a classifier (referred to as affected classifier), the output label should be a random guess from the $c$ classes, with a probability of $1/c$.

Given $L$ classifiers, in the ideal cases the true votes $V_{true}$ can be approximated as the sum of the number of unaffected classifiers $P(N)/L$ (based on the law of large numbers [27]) and the number of affected classifiers:

$$
V_{true} \approx \text{round} \left( P(N) \frac{L}{c} \right).
$$

Similarly, in the ideal cases the false votes should correspond to the number of affected classifiers for the other $c - 1$ classes. Based on assumption 2), its maximum value, $\max\{V_{false}^{i}\}_{i=1}^{c-1}$, can be roughly expressed as the corresponding mean value:

$$
\max\{V_{false}^{i}\}_{i=1}^{c-1} \approx \text{mean} \left( V_{false}^{i} \right) = \text{round} \left( \frac{1}{c} P(N) L \right)
$$

Correct classification can be achieved when $V_{true} > \max\{V_{false}^{i}\}_{i=1}^{c-1}$. From (7) and (8), it is obvious that this condition can be met when $P(N) > \varepsilon$ ($\varepsilon$ is a small positive number).

From ideal cases analysis, it can be seen that we only care about the number of unaffected classifiers, instead of which ones are. According to (6), by setting $N$ to a small number to make $P(N) > \varepsilon$, the corresponding classifier ensemble method is insensitive to the locations of corrupted features (since $V_{true} > \max\{V_{false}^{i}\}_{i=1}^{c-1}$ as supported by (7) and (8)). Therefore, it is robust to a large number of covariates as long as the irrelevant feature number is not extremely large (i.e., $m < d$ in (6)).

#### Real cases analysis

Here the two assumptions made in the ideal cases analysis do not hold precisely: 1) unaffected classifiers do not always make the correct classification; 2) although the affected classifiers would result in label assigning in a relatively random manner, it is difficult to estimate $\max\{V_{false}^{i}\}_{i=1}^{c-1}$ precisely. However, in real cases we can deem $V_{true}$ and $\max\{V_{false}^{i}\}_{i=1}^{c-1}$ from ideal cases analysis as the upper bound and lower bound, respectively, i.e.,

$$
V_{true} \leq V_{true}, \quad \max\{V_{false}^{i}\}_{i=1}^{c-1} \geq \max\{V_{false}^{i}\}_{i=1}^{c-1}.
$$

Since correct classification can be achieved only when $V_{true} > \max\{V_{false}^{i}\}_{i=1}^{c-1}$, our objective is to increase $V_{true}$ and decrease $\max\{V_{false}^{i}\}_{i=1}^{c-1}$. Since it is difficult to estimate $\max\{V_{false}^{i}\}_{i=1}^{c-1}$ precisely, we relax the objective by decreasing $\sum_{i=1}^{c-1} V_{false}$. Then we can reformulate our objective as a problem of suppressing $\Gamma$ defined as

$$
\Gamma = \sum_{i=1}^{c-1} V_{false} / V_{true} + \varepsilon,
$$

where $\varepsilon$ is a small positive number to avoid trivial results. To suppress $\Gamma$ in (10), we extend the classifier ensemble method by proposing two strategies:

1) **Local enhancing.** If we use the output labels of all the $L$ classifiers as valid votes such that $L = V_{true} + \sum_{i=1}^{c-1} V_{false}$ and then according to (10) we can get $\Gamma \propto \sqrt{L}/(V_{true} + \varepsilon)$. Since $L$ is a constant, the first strategy is to increase $V_{true}$. To realize it, based on supervised learning, for each subspace we extract more discriminant features in order to enhance the classification performance. We name the second-stage supervised feature extraction (after the first-stage random feature selection) as local enhancing. Details of LE used in this work are presented in Section 4.
2) Hybrid decision-level fusion. The second strategy is to decrease \( \sum_{i=1}^{c-1} V_{\text{false}}^{i} \) by dynamically eliminating classifiers corresponding to the irrelevant features, before MV. Irrelevant features would lead to label assignment in a relatively random manner. Based on the same irrelevant features, a classifier pair corresponding to two different LE methods would produce two random labels, which are unlikely to be the same. Based on the “AND” rule, classifier pairs with different output labels are deemed as invalid votes and simply discarded. Although this scheme may also decrease \( V_{\text{true}} \) to some extent, it significantly reduces the value of \( \sum_{i=1}^{c-1} V_{\text{false}}^{i} \) to suppress \( \Gamma \) in (10). Details of HDF are presented in Section 5.

3 Random Subspace Construction

We construct the feature space using 2DPCA, from which we can generate \( L \) feature extractors. Given \( n \) gait templates (i.e., GEs) \( \{ I_{i} \in \mathbb{R}^{N_{i} \times N_{i}} \}_{i=1}^{n} \) in the training set (i.e., the gallery), we can compute the scatter matrix \( S^{i} = \frac{1}{n} \sum_{i=1}^{n}(I_{i} - \mu)^{T}(I_{i} - \mu) \), where \( \mu = \frac{1}{n} \sum_{i=1}^{n} I_{i} \). The eigenvectors of \( S^{i} \) can be calculated, and the leading \( d \) eigenvectors associated with non-zero eigenvalues are retained as candidates \( T = [t_{1}, t_{2}, \ldots, t_{d}] \in \mathbb{R}^{N_{i} \times d} \) to construct the random subspaces. By repeating \( L \) times the process of randomly selecting subsets of \( T \) (with size \( N < d \)), the random subspaces \( \{ R^{i} \in \mathbb{R}^{N_{i} \times N_{i}} \}_{i=1}^{L} \) are generated and can be used as random feature extractors. Then a gait template \( I \in \mathbb{R}^{N_{i} \times N_{i}} \) can be represented as a set of random features \( \{ X^{i} \in \mathbb{R}^{N_{i} \times N_{i}} \}_{i=1}^{L} \) such that

\[
X^{i} = IR^{i}, \quad l = 1, 2, \ldots, L. \tag{11}
\]

Note the random feature extraction is only performed in the column direction of \( I \), with the dimension reduced to \( \mathbb{R}^{N_{i} \times N_{i}} \).

4 Local Enhancing

The random features can be used for classification directly. However, these features may be redundant and less discriminant since 1) the random feature extraction process based on (11) is performed only in the column direction and 2) the random feature extractors are trained in an unsupervised manner, without using the label information. As a result, these features may lead to high computational costs and low recognition accuracies. Local enhancing is used to address this issue, as discussed in Section 2.3. In each subspace, we further extract more discriminant features based on two different supervised learning methods, i.e., two-dimensional LDA (2DLDA) [28], and IDR/QR [29]. In this paper, these two types of feature extractors are referred to as local enhancers (i.e., LE1 and LE2), and the process of training 2DLDA-based LE1 and IDR/QR-based LE2 are summarized in Algorithm 1 and Algorithm 2, respectively. More details about 2DLDA and IDR/QR can be found in [28],[29].

In the \( l \)th subspace, a gait template \( I \) with extracted random feature matrix \( X^{I} \) (or the corresponding concatenated vector \( \tilde{X}^{I} \)) can be enhanced by \( W^{I} \) (the output of Algorithm 1) or \( V^{I} \) (the output of Algorithm 2) through

\[
\tilde{x}^{I} = (W^{I})^{T}X^{I}, \quad l \in [1, L], \tag{12}
\]

or

\[
\tilde{x}^{I} = (V^{I})^{T}\tilde{X}^{I}, \quad l \in [1, L]. \tag{13}
\]

\( \tilde{x}^{I} \) (resp. \( \tilde{z}^{I} \)) are the enhanced random features by LE1 (resp. LE2), and they can be used for classification in the \( l \)th subspace.

5 Hybrid Decision-Level Fusion

Given the random feature extractors \( \{ R^{i} \}_{i=1}^{L} \), the corresponding LE1-based \( \{ W^{i} \}_{i=1}^{L} \), and LE2-based \( \{ V^{i} \}_{i=1}^{L} \), we can extract the sets of LE1-based (resp. LE2-based) features from a gait template using (11), and (12) (resp. (13)). Based on the new gait descriptors, for the query sequence \( P \) (including \( n_{p} \) gait cycles) with the corresponding descriptors \( [p_{1}, p_{2}, \ldots, p_{n_{p}}] \), the distance between \( P \) and the \( j \)th class centroid \( \phi_{j} \) is defined as:

\[
\delta(P, \phi_{j}) = \sum_{i=1}^{n_{p}}\| p_{i} - \phi_{j} \|^{2}, \quad j \in [1, c].
\]

The nearest neighbour (NN) rule is employed for classification. Let \( \{ \omega_{j} \}_{j=1}^{c} \) be class labels of \( e \) subjects in the gallery, then the output label \( \bar{\Omega}^{I}(P^{I}) \) of the \( I \)th base classifier can be expressed as:

\[
\bar{\Omega}^{I}(P^{I}) = \arg\min_{j} \delta(P^{I}, \phi_{j}), \quad j \in [1, c].
\]
Then we use HDF to identify the query gait \( P = \{ P^l \}_{l=1}^L \). For simplicity reasons, let \( \Omega_{LE_1}(P^l), \Omega_{LE_2}(P^l) \) be the output labels of the classifier pairs corresponding to LE1 and LE2. HDF can be performed by assigning the class label to \( \Omega(P) \) with

\[
\Omega(P) = \arg \max_{\omega_j} \sum_{l=1}^L \Theta_j^{\omega_j}, \quad j \in [1, c],
\]

where

\[
\Theta_j^{\omega_j} = \begin{cases} 1, & \text{if } \Omega_{LE_1}(P^l) = \Omega_{LE_2}(P^l) = \omega_j; j \in [1, c], \\ 0, & \text{otherwise,} \end{cases}
\]

It can be seen that \( \sum_{j=1}^c \sum_{l=1}^L \Theta_j^{\omega_j} \leq L \). When there are a large number of irrelevant features for the query gait \( P \), most of the corresponding votes can be eliminated such that \( \sum_{j=1}^c \sum_{l=1}^L \Theta_j^{\omega_j} \ll L \). In this case, \( \sum_{j=1}^c \sum_{l=1}^L \Omega_{true} \) of (10) can be significantly reduced while \( V_{true} \) of (10) are less affected. As such, \( \Gamma \) of (10) can be effectively suppressed.

6 EXPERIMENTS

In this section, first we describe the two benchmark databases, namely USF [5] and OU-ISIR-B [16] datasets. Then we discuss the parameter settings and the time complexity of our method. Performance gain analysis by using the proposed LE and HDF strategies is also provided, followed by the algorithm comparison.

6.1 Datasets

The USF dataset [5] is a large outdoor gait database, consisting of 122 subjects. A number of covariates are presented: camera viewpoints (left/right), shoes (type A/type B), surface types (grass/concrete), carrying conditions (with/without a briefcase), elapsed time (May/Nov.), and clothing. Based on a gallery with 122 subjects, there are 12 pre-designed experiments for algorithm evaluation, which are summarized in Table 1. The second gait database, namely, the OU-ISIR-B dataset [16], was recently constructed for studying the effect of clothing. The evaluation set includes 48 subjects in standard clothes (i.e., type 9: regular pants + full shirt), whereas the probe set includes 856 gait sequences of the same 48 subjects in the other 31 clothes types. Several images from the USF and OU-ISIR-B datasets are shown in Fig. 5. For both datasets, the gallery is used for training. In this work, we employ two templates separately, namely, GEI (with the default size 128 x 88 pixels) and downsampled Gabor-filtered GEI (referred to as Gabor, with size 320 x 352 pixels).

We employ rank-1/rank-5 correct classification rate (CCR) to evaluate the performance. Considering the random nature of our method, the results of different runs may vary to some extent. We repeat all the experiments 10 times and report the statistics (mean, std, maxima and minima) in Tables 3 and 7 for both datasets. The small std values (listed in Tables 3 and 7) of the 10 runs indicate the stability of our method. Therefore, throughout the rest of the paper, we only report the mean values.

6.2 Parameter Settings and Time Complexity Analysis

There are only three parameters in our method, namely, the dimension of random subspace \( N \), the base classifier number \( L \), and the number of projection directions \( M \) for LE1/LE2. As discussed in Section 2.2, we should set \( N \) to a small number for great generalization ability \( P(N) \). The classifier number \( L \) should be large, since our classifier ensemble solution is based on the law of large numbers [27] (see Section 2.3). Like most subspace learning methods, the performance should not be sensitive to the number of projection directions \( M \) for LE1/LE2 (unless it is extremely small). On the USF dataset (Exp. A-L) we check the average performance sensitivity to \( N, M \), and \( L \), based on Gabor and GEI templates, respectively. By empirically setting \( L = 1,000 \) and \( M = 20 \), we run our methods with \( N \) set within the range \( [2, 6] \). The results in Fig. 6a indicate that the performance is not sensitive to \( N \). Based on \( L = 1,000 \), and \( N = 2 \), we conduct experiments with \( M = [20, 40, 60, 80, 100] \). The results in Fig. 6b suggest that the performance is stable for \( M \) with different values. By setting \( N = 2 \) and \( M = 20 \), we can also see from Fig. 6c that the performance is not decreasing with respect to the increasing number of classifiers. These observations are consistent with our expectation that the performance is not sensitive to these three parameters. For the rest of this paper, we only report the results based on \( N = 2, M = 20 \) and \( L = 1,000 \).
We analyze the time complexity for training the L LE1/LE2-based classifiers. For a LE1-based classifier in Algorithm 1, it takes \(O(NN_l^2)\) (resp. \(O(cNN_l^2)\)) for \(S_u\) (resp. \(S_l\)) and \(O(N_l^2)\) for eigenvalue decomposition. \(N_l\) is the template’s row number (i.e., \(N_l = 128\) for GEI, \(N_l = 320\) for Gabor), while \(n\) and \(c\) are the number of training samples and classes, respectively. Since, in our case, \(n > c\) and \(n > N_l\), the time complexity for generating \(L\) LE1-based classifiers can be written as \(O(NNN_l^2)\). For a LE2-based classifier in Algorithm 2, it takes \(O(C_l^2S)\) for the QR decomposition, where \(S = NN_l\). Calculating \(S_u\) and \(S_l\) requires \(O(c^2)\) and \(O(c^2n)\), while calculating \(Z\) and \(Y\) requires \(O(ncS)\) and \(O(Sc^2)\). Solving the eigenvalue decomposition problem of \((S_u^T)^{-1}S_u^T\) takes \(O(c^3)\) and the final solution \(V^i\) is obtained by matrix multiplication, which takes \(O(cSM)\). Since in our case \(n > c\) and \(S > c\), the time complexity for generating \(L\) LE2-based classifiers is \(O(LSc^3)\), which can also be written as \(O(LNN_l^2c)\). We run the matlab code of our method on a PC with an Intel Core i5 3.10 GHz processor and 16GB RAM. For the USF dataset, we report the training/query time (with \(L = 1,000\)) in Table 4. Note the classifiers are trained in an offline manner and can be used to identify probe sets with various covariates. It is clear that LE2 is very efficient when the dimension is large. For example, based on Gabor templates, LE2 only takes about 1/10 of LE1’s training time.

### 6.3 Performance Gain Analysis and Algorithms Comparison

Based on GEI and Gabor templates, we evaluate the effect of suppressing the ratio of false votes \(\sum_{i=1}^{K} V_{false}^i\) to true votes \(V_{true}\), as defined in (10) by using LE and HDF. For a probe set with \(K\) gait sequences, according to (10) we define \(\hat{\Gamma}\) as

\[
\hat{\Gamma} = \text{median} \left\{ \sum_{i=1}^{K} V_{false}^i / V_{true} + \epsilon \right\}_k^{k},
\]

where \(\epsilon = 1\) to avoid the trivial results. Over the 12 probe sets (A-L) on the USF dataset, the distribution of \(\Gamma\) and the performance are reported in Fig. 7 and Table 5. We can observe that, LE1/LE2 can reduce \(\Gamma\) to some extent, and RSM(LE1)/RSM(LE2) is less sensitive to covariates such as viewpoint, shoe, and briefcase (A-C, H-J). On the other hand, RSM-HDF can significantly suppress \(\Gamma\) and yields competitive accuracies in tackling the hard problems caused by walking surface and elapsed time (D-G, K-L).

### 6.3 Performance Gain Analysis and Algorithms Comparison

Based on GEI and Gabor templates, we evaluate the effect of suppressing the ratio of false votes \(\sum_{i=1}^{K} V_{false}^i\) to true votes \(V_{true}\), as defined in (10) by using LE and HDF. For a probe set with \(K\) gait sequences, according to (10) we define \(\hat{\Gamma}\) as

\[
\hat{\Gamma} = \text{median} \left\{ \sum_{i=1}^{K} V_{false}^i / V_{true} + \epsilon \right\}_k^{k},
\]

where \(\epsilon = 1\) to avoid the trivial results. Over the 12 probe sets (A-L) on the USF dataset, the distribution of \(\Gamma\) and the performance are reported in Fig. 7 and Table 5. We can observe that, LE1/LE2 can reduce \(\Gamma\) to some extent, and RSM(LE1)/RSM(LE2) is less sensitive to covariates such as viewpoint, shoe, and briefcase (A-C, H-J). On the other hand, RSM-HDF can significantly suppress \(\Gamma\) and yields competitive accuracies in tackling the hard problems caused by walking surface and elapsed time (D-G, K-L).

### On the performance sensitivity to the parameters on the USF dataset

#### (a) \(N\) is the dimension of the random subspace; \(M\) is the number of projection directions of LE1/LE2; \(L\) is the classifier number.

![Fig. 6](image1)

Fig. 6. On the performance sensitivity to the parameters on the USF dataset: (a) \(N\) is the dimension of the random subspace; (b) \(M\) is the number of projection directions of LE1/LE2; (c) \(L\) is the classifier number.

#### (b) \(N\) is the dimension of the random subspace; \(M\) is the number of projection directions of LE1/LE2; \(L\) is the classifier number.

![Fig. 7](image2)

Fig. 7. Over the 12 probe sets on the USF dataset, the distribution of \(\hat{\Gamma}\) (i.e., the general ratio of false votes to true votes).
notice that our method only has 42 percent rank-1 CCRs for probe sets K-L. In these cases, elapsed time is coupled with other covariates like walking surface, clothing, and shoe, as listed in Table 1. These walking conditions may significantly increase the number of irrelevant features, which would result in a lower $P(N)$ in (6). According to Section 2.3, a lower $P(N)$ would lead to a higher $G$ in (10), which contributes negatively to the performance. Nevertheless, experimental results suggest our method is robust to most covariates in the outdoor environment.

### 6.4 In Tackling the Clothing Challenges

Clothing was deemed as the most challenging covariate [15], and there are only a few works that have studied the effect of various clothes types. Recently, Hossain et al. built the OU-ISIR-B dataset [16] with 32 combinations of clothes types, as shown in Table 2. Based on an additional training set that covers all the possible clothes types, they proposed an adaptive part-based method [16] for clothing-invariant gait recognition. On this dataset, based on Gabor templates, we evaluate our methods RSM(LE1), RSM(LE2) and RSM-HDF. The statistics of our methods over 10 runs are reported in Table 7. Compared with the part-based method [16], Gabor+RSM-HDF can yield a much higher accuracy, as shown in Table 8. It is worth noting that different from [16], our method does not require the training set that covers all the possible clothes types and can generalize well to unseen clothes types.

We also study the effect of different clothes types, and the rank-1 CCRs for 31 probe clothes types are reported in Fig. 8. For most of the clothes types, our method can achieve more than 90 percent rank-1 accuracies. However, the performance can be affected with several clothes types that cover large parts of the human body.
7 Conclusion

In this paper, we model the effect of covariates as a partial feature corruption problem with unknown locations and propose a RSM-based classifier ensemble solution. The theoretical basis suggests that its insensitivity to a large number of covariates in ideal cases. To tackle the hard problems in realistic cases, we then propose two strategies, i.e., LE and HDF, to suppress the ratio of false votes to true votes before the majority voting. Experimental results suggest that our method is robust to clothing. It can be trained within minutes and perform real-time recognition in less than 1 second, which suggests that it is practical in real-world applications.

Acknowledgments

The authors would like to thank the support from the Royal Society’s International Exchange Programme (IE120092) and the constructive comments from the anonymous reviewers.

References