Gait Recognition Without Subject Cooperation

Khalid Bashir, Tao Xiang[1], and Shaogang Gong

School of Electronic Engineering and Computer Science

Queen Mary University of London, London E1 4NS, UK

Abstract

The strength of gait, compared to other biometrics, is that it does not require cooperative subjects. In previous work gait recognition approaches were evaluated using a gallery set consisting of gait sequences of people under similar covariate conditions (e.g. clothing, surface, carrying, and view conditions). This evaluation procedure, however, implies that the gait data are collected in a cooperative manner so that the covariate conditions are known a priori. In this work, gait recognition approaches are evaluated without the assumption on cooperative subjects, i.e. both the gallery and the probe sets consist of a mixture of gait sequences under different and unknown covariate conditions. The results indicate that the performance of the existing approaches would drop drastically under this more realistic experimental setup. We argue that selecting the most relevant gait features that are invariant to changes in gait covariate conditions is the key to developing a gait recognition system that works without subject cooperation. To this end, Gait Entropy Image (GEnI) is proposed to perform automatic feature selection on each pair of gallery and probe gait sequences. Moreover, an Adaptive Component and Discriminant Analysis (ACDA) is formulated which seamlessly integrates our feature selection method with subspace analysis for robust recognition, and importantly is computationally much more efficient compared to the conventional Component and Discriminant Analysis. Experiments are carried out on two comprehensive benchmarking databases: the CASIA database and the Southampton Human ID at a distance gait database (SOTON database). Our results demonstrate that the proposed approach significantly outperforms the existing techniques particularly when gait is captured with variable and unknown covariate conditions.

Index Terms

Biometrics, Gait Recognition, Gait Energy Image, Feature Selection, Adaptive Component and Discriminant Analysis.

1 Corresponding author. Tel:(+44)-(0)20-7882-8020; Fax: (+44)-(0)20-8980-6533; Email: txiang@dcs.qmul.ac.uk

I. Introduction

Gait is a behavioral biometric that measures the way people walk. Compared to physiological biometrics such as fingerprint, iris, and face, the advantage of gait is that it does not require subject cooperation and can operate without interrupting or interfering with the subject's activity. This makes gait ideal for situations where direct contact or cooperation with the subject is not possible (e.g. medium to long distance security and surveillance applications in public space).

Gait is sensitive to various covariate conditions, which are circumstantial and physical conditions that can affect either gait itself or the extracted gait features. Example of these conditions include clothing, surface, carrying condition (backpack, briefcase, handbag etc), view angle, speed, and shoe-wear type to name a few. The existing work on gait recognition uses a gallery set consisting of gait sequences of people [1] under similar covariate conditions and evaluate the performance of the proposed methods on probe sets of possibly different covariate conditions [6], [11], [13], [15], [25]–[27]. They therefore make the implicit assumption that the gallery data are collected in a cooperative manner so that the covariate conditions are known a priori. It is well known that given cooperative subjects, gait cannot compete with physiological biometrics in terms of recognition accuracy. It is thus necessary and crucial to evaluate the performance of the existing gait recognition approaches without the assumption on cooperative subjects, i.e. the gallery set is composed of a mixture of gait sequences under different and unknown covariate conditions. To the best of our knowledge, none of the existing work has done such an evaluation.

In this work, we evaluate the performance of existing gait recognition approaches under the aforementioned realistic experimental setup. The results show that the existing approaches yield very unsatisfactory performance (a nearly 3-fold decrease in recognition rate in some experiments compared to the result obtained using gallery sequences of similar covariate conditions). We argue that the main reason for the poor performance is that the existing approaches rely on both static appearance features and dynamic gait features for person identification, i.e. the identification is not achieved using gait alone [18], [19]. More specifically, most existing approaches represent gait using features extracted from silhouettes. By extracting silhouettes, a large part of physical appearance features have been removed from the image representation of human. Nevertheless,

[1] In object identification, a gallery set corresponds to a set of images of objects whose identities are known, whilst a probe set contains images of objects whose identities are unknown and need to matched against the gallery set for identification.

a silhouette still contains information about the shape of human body that is vulnerable to changes caused by conditions such as clothing and carrying. Although recent studies suggest that static shape information is more important than kinematics for most of the silhouette-based gait recognition approaches [18], [19], including static appearance features in gait representation also makes the existing approaches vulnerable to the changes of covariate conditions. To overcome the problem, it is crucial to select the most relevant gait features that reflect the unique characteristics of gait as a behavioral biometric, and importantly are invariant to appearance variations caused by changes of covariate conditions.

To this end, we propose a novel gait feature selection method to automatically select covariate condition invariant features for gait recognition. Specifically, Gait Entropy Image (GEnI) is proposed to measure the relevance of gait features extracted from the Gait Energy Image (GEI) introduced in [6]. A GEI represents a gait sequence using a single image; it is thus a compact representation which is an ideal starting point for feature selection. Constructed by computing Shannon entropy for the silhouettes extracted from a gait sequence, a GEnI can be readily used to distinguish dynamic gait information and static shape information contained in a GEI with the former being selected as features that are invariant to appearance changes. Since in a realistic experimental setup, the covariate conditions for both the gallery and probe sets are unknown, we propose to select a set of features that are unique to each pair of gallery and probe sequences.

After feature selection using Gait Entropy Image (GEnI), the gallery and probe GEIs can be used as templates and the gait recognition problem can be solved by measuring the distance between the templates directly. However, direct template matching has been shown to be sensitive to noise and small silhouette distortions by previous studies [6], [12] as well as our experiments (see Section V). To overcome this problem, statistical feature learning based on subspace Component and Discriminant Analysis (CDA) can be employed to further reduce the feature dimensionality [7]. Nevertheless, since different probe gait sequences yield different GEnIs, a different set of features will be selected for each probe GEI. Consequently a conventional CDA based approach is computationally costly because different subspaces have to be constructed given different probe GEIs. This problem is solved in this work by embedding feature selection into the subspace analysis. Specifically, a novel Adaptive Component and Discriminant Analysis (ACDA) is formulated, which, instead of computing a new subspace for each probe GEI, adapts a base subspace towards each probe GEI according to the selected features, therefore significantly reducing the computational cost.

To evaluate the effectiveness of our proposed approach, extensive experiments are carried out on two comprehensive benchmarking databases: the CASIA database [24] and the Southampton Human ID at a distance gait database (SOTON database) [16]. The results demonstrate that our feature selection based gait recognition method significantly outperforms previous approaches, especially when the gallery set is composed of sequences under variable unknown gait covariate conditions. Our experiments also suggest that the proposed Adaptive Component and Discriminant Analysis (ACDA) is much more efficient than the conventional Component and Discriminant Analysis (CDA) whilst being able to achieve very similar recognition accuracy.

In summary, this work has the following main contributions: 1) Gait recognition approaches are evaluated without assuming subject cooperation; 2) a novel feature selection method based on Gait Entropy Image (GEnI) is proposed for selecting the most relevant and informative gait features that are invariant to various covariate conditions. 3) a novel Adaptive Component and Discriminant Analysis (ACDA) is developed for fast gait recognition.

The rest of the paper is structured as follows: in Section II, related work is reviewed and the contribution and technical novelties of this paper are highlighted. The proposed gait feature selection method is detailed in Section III where the Gait Entropy Image (GEnI) is formulated and our pair-wise feature selection method is described. Section IV is focused on our gait recognition algorithm. In particular, the Adaptive Component and Discriminant Analysis (ACDA) is derived and its approximation accuracy and computational efficiency are analysed. Experimental results are presented in Section V. The paper concludes in Section VI.

II. Related Work

Existing gait recognition techniques mainly fall into two broad categories namely model based [13], [26], [27] and model free approaches [6], [11], [15], [25]. Model free approaches use motion information directly extracted from silhouettes, whilst model based approaches fit a model to human body and represent gait using the parameters of the model which are updated over time.

Model based approaches such as the Layered deformable models [13] tend to be more complex and computationally more expensive than model free approaches. For instance, Zhao et. al propose to perform 3D gait recognition using multiple cameras [27]. However, registration of gait images across camera view is non-trivial even in a well-controlled environment with clean background and little noise. A relatively simpler five link biped model based approach is used in [26] which introduces the idea of Sagittal plane (plane bisecting the human body) and claims that most gait movements are carried out on this plane. More recently the problem of gait recognition has been approached from a control systems perspective [5]. Nevertheless, existing model based approaches generally require good quality images to correctly extract the model parameters from a gait sequence, which may not be available in a real world application scenario such as CCTV surveillance in public space.

Recent trends in gait recognition research seem to favour model free approaches since they are computationally less intensive, more robust to noise, and have a comparable or better performance compared with the model based ones on benchmarking datasets [6]. BenAbdelkader et. al [3] propose to represent gait using image self similarity which measures the similarity between pairs of silhouettes in a gait sequence. It is claimed that the self similarity representation of gait encodes a projection of gait dynamics and is resistant to noise. An alternative representation is the Gait Energy Image (GEI) [6] which represents gait over a complete cycle as a single grey scale image by averaging the silhouettes extracted over the complete gait cycle. GEI has gained much popularity recently and has been employed in a number of state of the art gait recognition algorithms [17], [21], [22]. Model free gait recognition has also been studied in the transform domain [4] and using frequency based techniques [20], [23], [25].

Most model free approaches represent gait based on silhouettes extracted from the gait video sequences [15], [24]. This is because gait, as a behavioural biometric, is different from physical biometrics such as face in that gait mainly captures the dynamic aspect of human activity instead of the static appearance of human. By extracting silhouettes, a large part of physical appearance features have been removed from the image representation of human. Nevertheless, a silhouette still contains information about the shape and stance of human body. Recent studies suggest that static shape information is more important than kinematics for most of the silhouette-based gait recognition approaches [18], [19]. However, the inclusion of shape information in gait features can also introduce variations that will hinder the recognition performance especially in challenging cases where the same person wears different clothes and has different carrying conditions in the gallery and probe sequences [6], [24]. This problem becomes more acute when gait recognition is performed without subject cooperation, i.e. the gallery sequences also have variable and unknown covariate conditions. This has motivated the work presented in this paper which selects covariate condition invariant features for each pair of gallery and probe sequences using Gait Entropy Image (GEnI).

The idea of feature selection for gait recognition without subject cooperation was first exploited in our previous work [2]. Apart from performing more extensive analysis on additional datasets, this paper differs significantly from [2] in that a new feature selection method is formulated based on the novel Gait Entropy Image (GEnI). Better performance is achieved on the same benchmarking datasets, and importantly, the proposed feature selection method has only one free parameter and thus requires much less parameter tuning.

III. Feature Selection using Gait Entropy Image

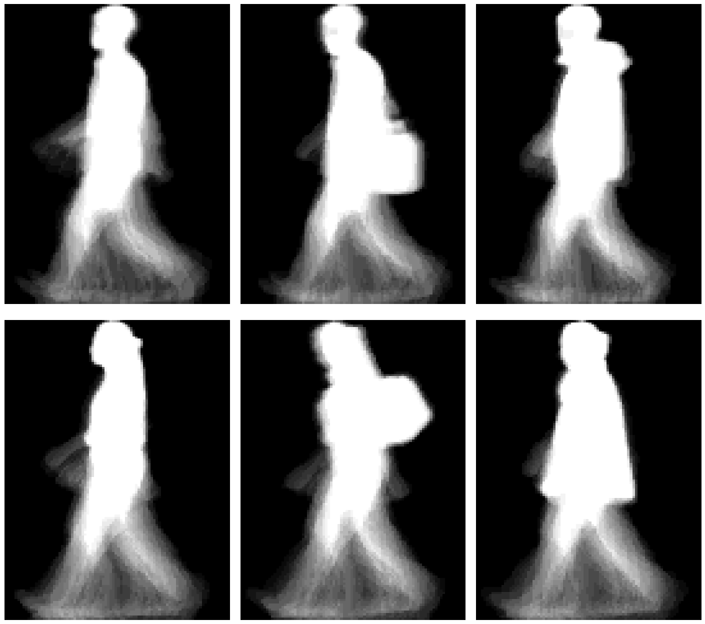
A. Gait Representation using Gait Energy Image

Given a human walking sequence, a human silhouette is extracted from each frame using the method in [15]. After applying size normalization and horizontal alignment to each extracted silhouette image, gait cycles are segmented by estimating gait frequency using a maximum entropy estimation technique presented in [15]. Gait Energy Image (GEI) is then computed as

$$GEI = G(x,y) = \frac{1}{T}\sum_{t=1}^{T} I(x,y,t), \tag{1}$$

where T is the number of frames in a complete gait cycle, I is a silhouette image whose pixel coordinates are given by x and y, and t is the frame number in the gait cycle.

Examples of GEIs are shown in Fig. 1. Note that pixels with high intensity values in a GEI correspond to body parts that move little during a walking cycle (e.g. head, torso), while pixels with low intensity values correspond to body parts that move constantly (e.g. lower parts of legs and arms). The former mainly contain information about body shape and stance, whilst the later tells us more about how people move during walking. We call the former static areas of a GEI and the latter dynamic areas of a GEI. The dynamic areas are insensitive to human appearance changes caused by common covariate conditions such as carrying condition and clothing; they are thus the most informative part of the GEI representation for human identification given variable covariate conditions. The static areas of a GEI also contain useful information for identification (e.g. one's hair style). However, since they mainly contain body shape information, they are sensitive to changes in various covariate conditions. For instance, in each row of Fig. 1, three



(a) Normal (b) Carrying a bag (c) Wearing a coat

Fig. 1. Gait Energy Images of people under different carrying and clothing conditions. Top Row: a subject from the CASIA database [24]; bottom row: a subject from the SOTON database [16]. Compared to (b) and (c), the subjects in (a) did not carry a bag or wear a bulk coat.

GEIs are computed from three sequences of the same person walking under different conditions. The dynamic areas of the GEI suggest that they are the same person but the static areas suggest otherwise.

Based on this observation, an automatic feature selection method is developed to select the most informative gait features from a GEI. This is achieved using a novel Gait Entropy Image (GEnI) based approach formulated below.

B. Gait Entropy Image

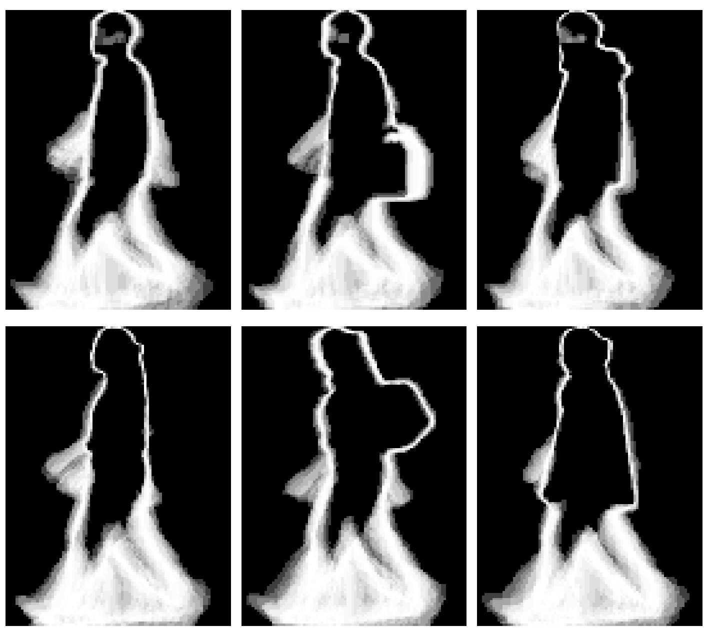
We propose to distinguish the dynamic and static areas of a GEI by measuring Shannon entropy at each pixel location in the GEI. More specifically, A gait cycle consists of a sequence of human silhouettes (T silhouettes); consider the intensity value of the silhouettes at a fixed pixel location as a discrete random variable, Shannon entropy measures the uncertainty associated with the random variable over a complete gait cycle and can be computed as

$$GEnI = H(x, y) = -\sum_{k=1}^{K} p_k(x, y) \log_2 p_k(x, y) \quad (2)$$

where x, y are the pixel coordinates and $p_k(x, y)$ is the probability that the pixel takes on the k^{th} value. In our case the silhouettes are binary images and we thus have $K = 2$, $p_1(x, y) = \frac{1}{T}\sum_{t=1}^{T} I(x, y, t)$ (i.e. the GEI) and $p_0(x, y) = 1 - p_1(x, y)$. Note that there is a close link between GEnI and a GEI. Specifically, let $z = p_1(x, y)$ representing a GEI, by expanding Eqn. 2 we have

$$GEnI = -z * \log_2 z - (1 - z) * \log_2(1 - z). \quad (3)$$

GEnI gives us an insight into the information content of the gait sequence as the intensity value at pixel location (x, y) is proportional to its entropy value $H(x, y)$. Fig. 2 shows the Gait Entropy Images (GEnI) obtained from the GEIs depicted in Fig. 1. It is evident from Fig. 2 that the dynamic areas in a GEI are featured with high intensity values in its corresponding GEnI whilst the static areas have low values. This is not surprising because silhouette pixel values in dynamic areas are more uncertain and thus more informative leading to higher entropy values.



(a) Normal (b) Carrying a bag (c) Wearing a coat

Fig. 2. Examples of Gait Entropy Images. Their corresponding Gait Energy Images are shown in Fig. 1.

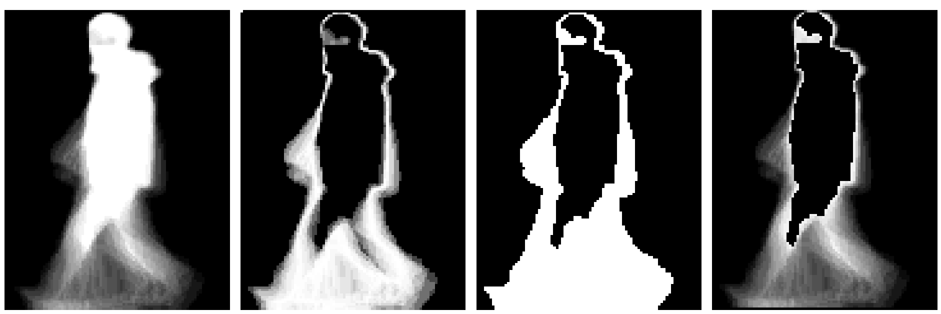
GEnI can be used directly for selecting informative gait features from GEI. However, Fig. 2 also suggests that the body shape changes caused by varying covariate conditions such as carrying and clothing are still visible in the GEnI. Consequently, the selected features from a GEI will contain information that is irrelevant to gait. To overcome this problem, a feature selection method is proposed which selects a unique set of features for each pair of gallery and probe GEI.

C. Pairwise Feature Selection using Gait Entropy Image

For each pair of gallery and probe gait sequences and their corresponding GEIs, we aim to select a set of gait features that are invariant to covariate condition changes and unique to the pair. To this end, first a binary feature selection mask [2] $M_G(x,y)$ is generated for each GEI using its corresponding GEnI $H(x,y)$.

$$M_G(x,y) = \begin{cases} 1, & \text{if } H(x,y) > \theta \\ 0, & \text{Otherwise} \end{cases} \tag{4}$$

where θ is a threshold. Examples of feature selection mask generated using GEnI or the GEI (both generate same feature selection mask) is shown in Fig. 3.



(a) GEI (b) GEnI (c) Mask Image (d) Masked GEI

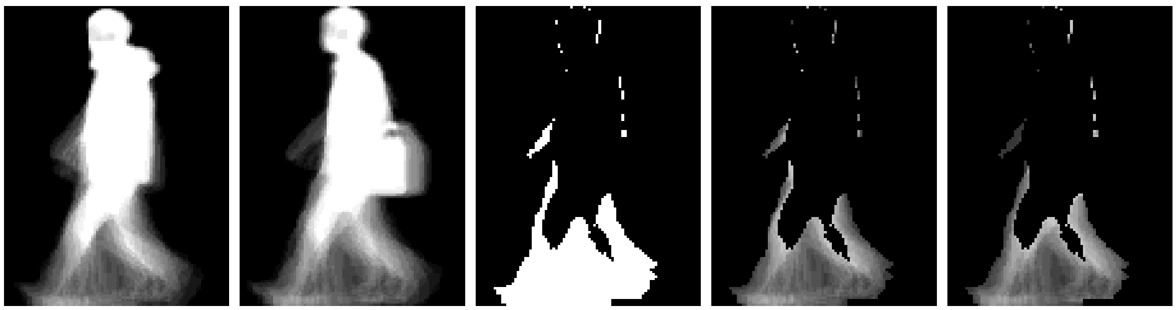
Fig. 3. An example of feature selection mask generated using GEnI.

Suppose the gallery set contains N GEIs belonging to C classes (subjects). For the ith gallery GEI, we generate a feature selection mask $M_G^i(x,y)$ using Eqn. (4); Similarly $M_G^j(x,y)$ is obtained for the jth probe GEI. Now to select features that are relevant for both GEIs, these two masks are to be combined. This is done using the simple binary 'AND' operation to select features specific to the probe/gallery pair in question.

[2] A feature selection mask determines whether features from a specific pixel location should be selected. It therefore has to be binary.

$$M_G^{ij}(x,y) = M_G^i(x,y) \&\& M_G^j(x,y) \tag{5}$$

where $\&\&$, is the binary 'AND' operator.



(a) (b) (c) (d) (e)

Fig. 4. Example of feature selection for a pair of gallery and probe GEIs. (a) gallery GEI; (b) probe GEI; (c) feature selection mask $M_G^{ij}(x,y)$; (d) gallery GEI with $M_G^{ij}(x,y)$ applied; (e) probe GEI with $M_G^{ij}(x,y)$ applied.

Fig. 4 shows an example of applying our feature selection method to a pair of gallery and probe GEIs under different covariate conditions. It is evident that after applying the pairwise feature selection mask generated using both GEIs, the effect of the changes in covariate conditions in the gallery and probe sequences is alleviated effectively. In particular, Fig. 4(d) and (e) give strong indication that the two images are captured from the same person, although both the carrying and clothing conditions are different as shown in Fig. 4(a) and (b).

IV. ADAPTIVE COMPONENT AND DISCRIMINANT ANALYSIS

After applying a feature selection mask $M_G^{ij}(x,y)$ to each pair of gallery and probe GEIs, gait recognition can be performed by matching a probe GEI to the gallery GEI that has the minimal distance between them. However, direct template matching has been shown to be sensitive to noise and small silhouette distortions [6], [12]. This is because that the dimensionality of the GEI feature space is high even after feature selection (typically in the order of thousands). To overcome this problem, subspace Component and Discriminant Analysis (CDA) based on Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) can be adopted which seeks to project the original features to a subspace of lower dimensionality so that the best data representation and class separability can be achieved simultaneously [7].

Suppose we have N d-dimensional gallery GEI templates $\{\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N\}$ belonging to C different classes (individuals), where each template is a column vector obtained by concatenating the rows of the corresponding GEI. To compute the distance between the ith gallery and the jth probe GEI, $M_G^{ij}(x, y)$ is applied to each gallery GEI, which gives us a new set of template $\{\mathbf{x}_1^{ij}, ..., \mathbf{x}_n^{ij}, ..., \mathbf{x}_N^{ij}\}$ of dimension d^{ij}. PCA is an orthogonal linear transformation that transforms the data to a subspace of dimensionality $\tilde{d^{ij}}$ (with $\tilde{d^{ij}} < d^{ij}$). The PCA subspace keeps the greatest variances by any projection of the data so that the reconstruction error defined below is minimized:

$$J_{\tilde{d^{ij}}} = \sum_{n=1}^{N} \left\| \left(\mathbf{m} + \sum_{k=1}^{\tilde{d^{ij}}} a_{nk} \mathbf{e}_k^{ij} \right) - \mathbf{x}_n^{ij} \right\|^2 \tag{6}$$

where $\mathbf{m}$ is the mean of the data, $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$ are a set of orthogonal unit vectors representing the new coordinate system of the subspace, a_{nk} is the projection of the nth data to $\mathbf{e}_k^{ij}$. $J_{\tilde{d^{ij}}}$ is minimised when $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$ are the $\tilde{d^{ij}}$ eigenvectors of the data covariance matrix with the largest eigenvalues (in decreasing order). Now the gallery template $\mathbf{x}_n^{ij}$ is represented as a $\tilde{d^{ij}}$-dimensional feature vector $\mathbf{y}_n^{ij}$ and we have

$$\mathbf{y}_n^{ij} = M_{pca}^{ij} \mathbf{x}_n^{ij} = [\mathbf{e}_1^{ij}, ..., \mathbf{e}_{\tilde{d^{ij}}}^{ij}]^T \mathbf{x}_n^{ij}. \tag{7}$$

PCA is followed by MDA which aims to find a subspace where data from different classes are best separated in a least square sense. Different from PCA, MDA is a supervised learning method which requires the gallery data to be labelled into classes. The MDA transformation matrix, W^{ij} maximizes

$$J(W^{ij}) = \frac{|{W^{ij}}^T S_B^{ij} W^{ij}|}{|{W^{ij}}^T S_W^{ij} W^{ij}|}$$

where S_B^{ij} is the between-class scatter matrix and S_W^{ij} the within-class scatter matrix of the gallery data in the PCA subspace $\{\mathbf{y}_1^{ij}, ..., \mathbf{y}_n^{ij}, ..., \mathbf{y}_N^{ij}\}$. $J(W^{ij})$ is maximized by setting the columns of W^{ij} to the generalized eigenvectors that correspond to the $C - 1$ nonzero eigenvalues in

$$S_B^{ij} \mathbf{w}_k^{ij} = \lambda_i^j S_W^{ij} \mathbf{w}_k^{ij}$$

where $\mathbf{w}_k^{ij}$ is the kth column of W^{ij} and C is the number of classes in the gallery data. Denoting these generalised eigenvectors as $\{\mathbf{v}_1^{ij}, \mathbf{v}_2^{ij}, ..., \mathbf{v}_{C-1}^{ij}\}$, a gallery template is represented in the MDA subspace as:

$$\mathbf{z}_n^{ij} = M_{mda}^{ij} \mathbf{y}_n^{ij} = [\mathbf{v}_1^{ij}, ..., \mathbf{v}_{C-1}^{ij}]^T \mathbf{y}_n^{ij}. \tag{8}$$

Note that the choice of $\tilde{d^{ij}}$ is affected by the dimensionality of the MDA subspace, i.e. $C-1$. In particular, S_W^{ij} becomes singular when $\tilde{d^{ij}} < C$ or $\tilde{d^{ij}} \gg C$. We therefore set $\tilde{d^{ij}} = 2C$ in this paper.

Now after three steps of dimensionality reduction (feature selection using $M_G^{ij}(x, y)$, PCA, and MDA), both the gallery and probe GEI feature vectors are represented in a $C-1$ dimensional subspace. This dimensionality reduction process is computationally expensive mainly due to the PCA step. This is because for each new gallery and probe GEI pair, a new mask $M_G^{ij}(x, y)$ is generated and we need to re-do the PCA which involves eigen-decomposition of a $N \times N$ matrix. To make our approach more computationally efficient, we develop an Adaptive Component and Discriminant Analysis (ACDA). More specifically, instead of applying each $M_G^{ij}(x, y)$ to the gallery templates and re-do the PCA on $\{\mathbf{x}_1^{ij}, ..., \mathbf{x}_n^{ij}, ..., \mathbf{x}_N^{ij}\}$, we compute PCA only once for the original gallery templates $\{\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N\}$, which results in a base PCA subspace. We then adapt the base PCA subspace towards each gallery and probe GEI pair by applying $M_G^{ij}(x, y)$ directly to the base principal components. Specifically, let $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{\tilde{d}}\}$ be the base components, each component can be treated as an eigenGEI, similar to eigenface for face recognition. The adapted components $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$ are then obtained by applying $M_G^{ij}(x, y)$ to the eigenGEIs. Now Eqn. (7) can be re-written as

$$\mathbf{y}_n^{ij} = M_{pca}^{ij}\mathbf{x}_n^{ij} = [\mathbf{u}_1^{ij}, ..., \mathbf{u}_{\tilde{d^{ij}}}^{ij}]^T \mathbf{x}_n^{ij}. \tag{9}$$

The MDA step that follows will remain unchanged (see Eqn. (8)).

In our Adaptive Component and Discriminant Analysis (ACDA) we approximate $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$ using $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$ in order to reduce the computational cost. What price we have to pay for this improvement in computational efficiency will depend on the accuracy of the approximation. Intuitively, applying a binary mask $M_G^{ij}(x, y)$ to the gallery data collapses some of the original coordinate axes. $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$, as the subspace expressed in the original coordinate system, should also have the corresponding axes collapsed, which is exactly how $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$ are generated. Theoretically, it can be readily proved that the projection of $\{\mathbf{x}_1^{ij}, ..., \mathbf{x}_n^{ij}, ..., \mathbf{x}_N^{ij}\}$ to $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$ will have an identical diagonalised covariance matrix as their projection on $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{\tilde{d^{ij}}}^{ij}\}$ [8]. More importantly, we demonstrate through experiments in the next section that the approximation is extremely accurate in practice.

V. Experiments

Two experiments were carried out in this study. In the first experiment, the gallery set contains sequences of people walking under similar covariate conditions, i.e. the same experimental setup as the existing work. In the second experiment, the gallery set is composed of a mixture of gait sequences collected under different unknown covariate conditions. The experimental setting for the second experiments is designed to reflect a real world scenario where no subject cooperation is required and therefore the covariate conditions for the subjects in both the gallery and probe sets are different and unknown.

A. Datasets

The CASIA Gait Database [24] and the Southampton Human ID at a distance gait database (SOTON database) [16] were used to evaluate the performance of the proposed approach. The CASIA database comprises of 124 subjects. For each subject there are 10 walking sequences consisting of 6 normal walking sequences where the subject does not carry a bag or wear a bulky coat (CASIASetA), 2 carrying-bag sequences (CASIASetB) and 2 wearing-coat sequences (CASIASetC). Each sequence contains multiple gait cycles resulting in multiple GEIs and GEnIs. The original image size of the database is 320x240.

The SOTON database consists of two datasets: a large dataset (with over 100 subjects) and a small dataset (with 11 subjects). All subjects in the large dataset were captured under the normal and fixed covariate conditions. The small dataset, on the other hand, was designed for investigating the robustness of gait recognition techniques to imagery of the same subject in various common conditions (e.g. carrying items, clothing). It is thus more suitable for evaluting uncooperative gait recognition and is employed in our experiments. For each subject we used 2 normal sequences (SotonSetA), 4 carrying-bag sequences (SotonSetB) and 2 wearing-coat sequences (SotonSetC). The original image size of the database is 720x576.

Note that both datasets used in the experiments contain various varying covariate conditions. In prevous work where they were used, gait sequences in either probe or gallery sets have the same covariate condition, whilst in our experiments, both the gallery and the probe sets consist of a mixture of gait sequences under different and unknown covariate conditions. The two datasets are perfectly suitable for our uncooperative gait recognition experiments under an experimental setting that differs from those in previous work.

The extracted silhouettes from each of the databases were centred and normalized, and the size of the GEIs and GEnIs is 128x88 for both databases (i.e. the original feature space has a dimensionality of 11264). Sample GEI images and GEnI images from both databases are shown in Fig. 1 and Fig. 2 respectively. Note that the only free parameter in our approach is the threshold value θ used in Eqn. (4). In the rest of the section, we report the results obtained when θ was set to 0.75. The effect of θ on the recognition performance is analysed in Section V-D.

B. Gallery Sequences under Similar Covariate Conditions

In this experiment, the gallery set used for the CASIA dataset consists of the first 4 sequences of each subject in CASIASetA (CASIASetA1). The probe set is the rest of the sequences in CASIASetA (CASIASetA2), CASIASetB and CASIASetC. For each subject the gallery set for the SOTON dataset consists of one of the normal sequences from SotonSetA (SotonSetA1). The probe sets are SotonSetA2 including the other normal sequence from SotonSetA, SotonSetB and SotonSetC.

The results obtained using our approach, termed as $M_G^{ij}(x,y)$+ACDA, was compared with the results published in [24] which were obtained using direct template matching on the same databases and the approach in [6] which was based on the standard CDA without any feature selection on the GEIs. The approach in [6] is widely regarded as one of best gait recognition approach. The performance was measured using recognition rates and is presented in Table I.

It can be seen from Table I that direct template matching gives the worst results. CDA based approach improves on template matching for most probe sets but there is still much room for further improvement, especially when the probe sets have different covariate conditions from the gallery sets. Table I shows that our approach ($M_G^{ij}(x,y)$+ACDA) significantly outperforms both template matching and CDA for all probe sets except for SotonSetB where people carry different bags (e.g. rucksack, laptop bag, and suitcases). The improvement is particularly substantial for the probe set with a different clothing condition (CASIASetC and SotonSetC), on which poor results were obtained without feature selection.

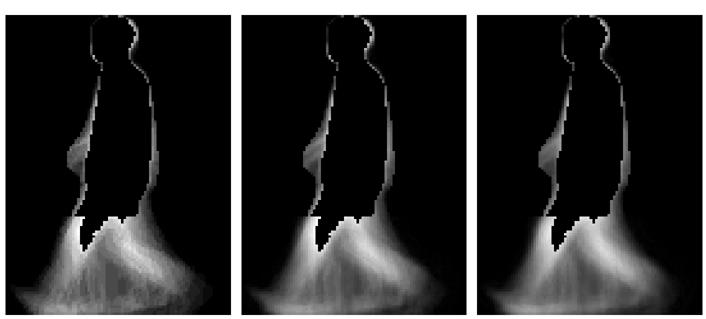
Our approach was also compared with an alternative approach which also performs features selection using GEnIs but use feature selection mask generated from the probe GEnI only (M_G^j+ACDA). These results will highlight the effect of having a pair-wise feature selection

Probe Set	TM	CDA	M_G^j+CDA	M_G^j+ACDA	$M_G^{ij}(x,y)$+CDA	$M_G^{ij}(x,y)$+ACDA
CASIASetA2	97.6%	99.4%	100%	99.1%	**100%**	100%
CASIASetB	52.0%	60.2%	70.5%	70.0%	**78.3%**	77.8%
CASIASetC	32.7%	30.0%	35.5%	35.1%	**44.0%**	43.1%
SotonSetA2	100%	100%	100%	100%	**100%**	100%
SotonSetB	54.5%	86.3%	86.3%	**86.3%**	81.8%	81.8%
SotonSetC	45.4%	72.7%	81.8%	81.8%	**83.3%**	81.8%

TABLE I

COMPARING DIFFERENT APPROACHES USING A GALLERY SET CONSISTING OF SEQUENCES UNDER SIMILAR COVARIATE CONDITIONS (WITHOUT CARRYING A BAG OR WEARING A COAT). TM: DIRECT GEI TEMPLATE MATCHING; CDA: METHOD IN [6] BASED ON CDA WITHOUT FEATURE SELECTION; M_G^j+CDA: FEATURE SELECTION USING MASKS GENERATED ONLY FROM THE PROBE GENI FOLLOWED BY CDA; M_G^j+ACDA: FEATURE SELECTION USING MASKS GENERATED ONLY FROM THE PROBE GENI FOLLOWED BY ACDA; $M_G^{ij}(x,y)$+CDA: THE PROPOSED APPROACH WITH FEATURE SELECTION USING MASKS GENERATED FROM EACH PAIR OF GALLERY AND PROBE GENIS FOLLOWED BY CDA; $M_G^{ij}(x,y)$+ACDA: THE PROPOSED APPROACH WITH FEATURE SELECTION USING MASKS GENERATED FROM EACH PAIR OF GALLERY AND PROBE GENIS FOLLOWED BY ACDA.

mask ($M_G^{ij}(x,y)$). Specifically, we generate a feature selection mask using Eqn. (4) for each probe GEI and apply it to all gallery GEIs. Table I shows that using the mask generate from the probe GEI only, the result is still much better compared to those of previous approaches without feature selection, albeit it is slightly worse than the result obtained using a mask generated for each gallery-probe GEI pair.



(a) (b) (c)

Fig. 5. (a): A GEI with $M_G^{ij}(x,y)$ applied; (b) The reconstructed GEI using $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{d^{ij}}^{ij}\}$; (c) The reconstructed GEI using $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$. The root-mean-square errors, which are $J_{\tilde{d^{ij}}}$ (see Eqn. (6)) normalized by the image size, was 0.0031 for (b) and 0.0060 for (c).

Table I also lists the results with feature selection using the mask from GEnIs follwed by CDA. The difference in the respective results obtained by using CDA (M_G^j+CDA, M_G^{ij}+CDA) and ACDA (M_G^j+ACDA, M_G^{ij}+ACDA) will indicate how much sacrifice in recognition accuracy needs to be made in exchange for lower computational cost for the proposed ACDA. It can be seen that our Adaptive Component and Discriminant Analysis (ACDA) method achieves almost identical results as the CDA approach. This suggests that our approximation of $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{d^{ij}}^{ij}\}$ using $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$ is accurate. Fig. 5(b) and (c) show examples of reconstructed GEIs using $\{\mathbf{e}_1^{ij}, \mathbf{e}_2^{ij}, ..., \mathbf{e}_{d^{ij}}^{ij}\}$ and $\{\mathbf{u}_1^{ij}, \mathbf{u}_2^{ij}, ..., \mathbf{u}_{\tilde{d}}^{ij}\}$ respectively. Both of them gave extremely small reconstruction errors. As for computational cost, as indicated by Table II, our ACDA is much more computationally efficient than CDA. The result was obtained using a platform with an Intel Dual Core 1.86GHz CPU and 2GB memory.

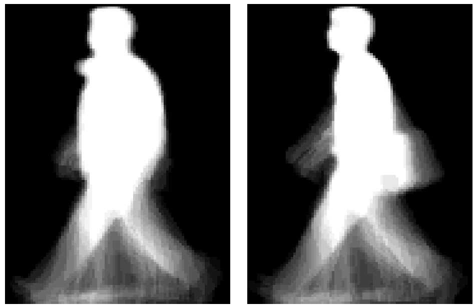
	CDA	ACDA
Computational cost	36.70 sec	0.18 sec

TABLE II

COMPARISON OF THE COMPUTATIONAL COST OF ACDA AND CDA FOR RECOGNISING A SINGLE GAIT SEQUENCE.

C. Gallery sequences under different covariate conditions

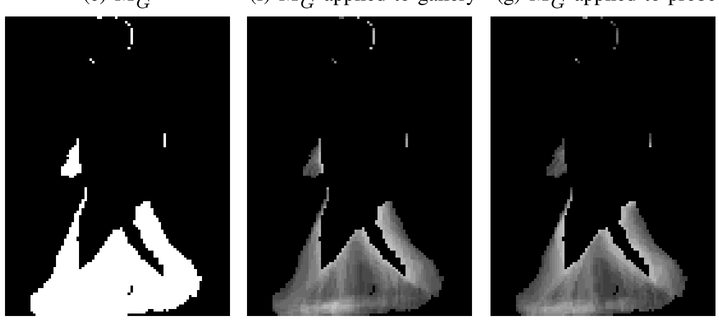
In this experiment, the gallery sets include a mixture of normal, carrying-bag, and wearing-coat sequences, which give us a challenging experimental setting closely representing the condition for gait recognition with uncooperative subjects. More specifically, for the CASIA dataset we selected the first one third of the sequences from CASIASetC, the second one third from CASIASetB and the last one third from CASIASetA. The probe sets consist of the rest of the dataset and are referred to as CASIASetA3, CASIASetB2 and CASIASetC2. For the SOTON dataset the mixed gallery set contains the first one third of the subjects from SotonSetA the second one third from SotonSetB and last one third from SotoSetC. The probe sets include the rest of the sequences and are termed as SotonSetA3, SotonSetB2 and SotonSetC2 respectively. The probe and the gallery sets are mutually exclusive.



(a) Gallery GEI (b) Probe GEI

(c) Gallery GEnI (d) Probe GEnI

(e) M_G^j (f) M_G^j applied to gallery (g) M_G^j applied to probe



(h) M_G^{ij} (i) M_G^{ij} applied to gallery (j) M_G^{ij} applied to probe

Fig. 6. Comparing the effectiveness of the feature selection mask M_G^j generated using probe GEI only, and M_G^{ij} generated using both the gallery and probe GEIs. The subject wears a coat in the gallery sequence and carries a bag in the proble sequence.

Probe Set	CDA	M_G^j+ACDA	$M_G^{ij}(x,y)$+ACDA
CASIASetA3	48.1%	58.2%	**69.1%**
CASIASetB2	31.9%	37.5%	**55.6%**
CASIASetC2	9.7%	23.6%	**34.7%**
TotalCASIA	32.6%	42.5%	**55.5%**
SotonSetA3	45.5%	45.5%	**63.6%**
SotonSetB2	31.82%	50.0%	**50.0%**
SotonSetC2	36.4%	36.3%	**54.6%**
TotalSOTON	36.3%	45.5%	**54.5%**

TABLE III

COMPARING DIFFERENT APPROACHES USING GALLERY SETS CONSISTING OF SEQUENCES UNDER DIFFERENT COVARIATE CONDITIONS.

The experimental results are presented in Tables III. The results indicate a drastic degradation in performance for the CDA based method without feature selection [6] and the approach with feature selection using the probe GEnI only. In comparison, our approach ($M_G^{ij}(x,y)$+ACDA) achieves much better result, especially for the probe sequences where people carry a bag or wear bulking clothes. This result suggests that under such a realistic experimental setup, feature selection based on each pair of gallery and probe gait sequences is critical for selecting the gait features that are invariant to covariate condition changes. This is evident from an example shown in Fig. 6. It can be seen from Fig. 6 that after applying the mask generated using both the gallery and probe GEnIs, the gallery and probe sequences can be correctly matched, whilst the mask generated using the probe sequence alone cannot deal with the variations in GEnIs caused by changes in covariate conditions resulting in an incorrect match.

D. The effect of the feature selection threshold θ

We have investigated the effect of the only free variable θ of our method on the recognition rate for the CASIA database without subject cooperation. The result is shown in Fig. 7. The value of θ ranges from 0 to 1 with smaller values corresponding to less features being selected. Fig. 7 show that similar recognition rate can be achieved when the value of θ is between 0 to

0.95. This suggests that our method is insensitive to the setting of θ.

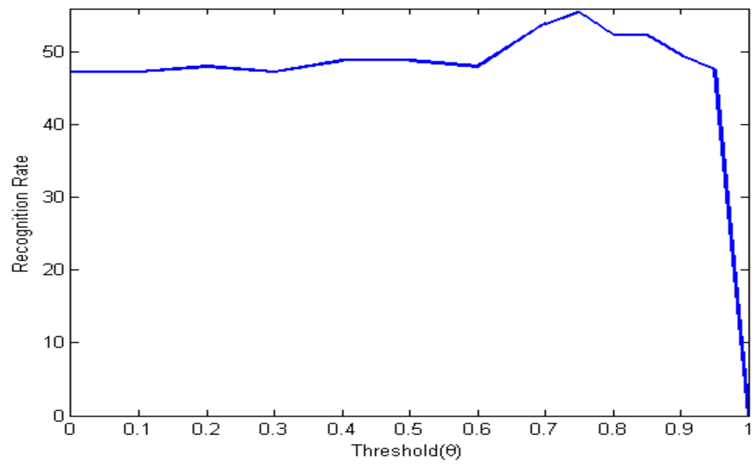


Fig. 7. The effect of θ on the recognition rate.

E. Discussions

The key findings of our experiments are summarised and discussed as below:

1) Our approach significantly outperforms direct template matching and Han and Bhanu's GEI+CDA approach [6] in both experimental conditions i.e. the traditional experimental setup and the new setup proposed in this work assuming no subject cooperation. This is mainly due to the novel Gait Entropy Image (GEnI) based feature selection method proposed in this paper. In particular, even with the same gait representation (GEI) and recognition algorithm (CDA), large improvement can be achieved by selecting features that are invariant to covariate condition changes.
2) When both the gallery and probe sets contain sequences of different and unknown covariate conditions, all gait recognition approaches, including ours, suffer from significant decrease in recognition performance. This is hardly surprising as human identification without cooperative subjects is the 'holy grail' of biometrics research and is widely regarded as the most challenging problem yet to be solved. But importantly our results show that under this challenging and more realistic experimental setting, performing feature selection, particularly selecting a unique set of features for each pair of gallery and probe sequences, is crucial and much more promising than alternative approaches in solving the problem.

3) The improvement obtained by using the pair-wise feature selection strategy potentially comes with a price, that is, the computational cost can be very high which may hinder the implementation of our approach for real time applications. Fortunately, the proposed Adaptive Component and Discriminant Analysis (ACDA) provides a solution to this problem. ACDA performs approximation in the PCA subspace rather than re-computing the subspace for each pair of gallery and probe sequences. The experimental results suggest that this approximation is accurate in practice and the gain in speed is much greater than the loss in recognition accuracy.
4) It is noted that our approach did not achieve the same amount of improvement for the SOTON carrying-bag sequences (SotonSetB) compared with other probe sets (see Table I). The reason is that for these sequences subjects often carry suitcases which occlude the dynamic areas around the leg region. Our feature selection method can remove the static areas caused by the suitcases; but it will also remove some of the informative dynamic areas. Under this circumstance and with subject cooperation, the existing approach without feature selection may perform better as they utilise some of the shape information in the static areas located in the upper body region. However, because our objective is to develop a human identification method that is invariant to covariate condition changes and without relying on cooperative subjects, this is not considered as a drawback of the approach. Our results in Tables III clearly show that without subject cooperation, our approach performs much better than the alternatives even for the SOTON carrying-bag sequences.
5) There exist a number of methods [1], [10] which also learn a PCA space adaptively. However, there is a fundamental different between these methods and our ACDA method. Specifically, the existing adaptive PCA methods are designed for incremental updating of PCA space when a new data point become available. Consequently, the updated PCA space is fairly similar to the old one as it is learned mostly based on the unchanged old data. In our problem, given a new probe sequence, it does not introduce a new data point. Instead, it changes the whole dataset for learning PCA as a new feature selection mask is generated. Therefore the existing adaptive PCA methods are not suitable here because it is designed for a completely different problem.
6) It is worth pointing out that although our method can cope with most other covariate condition changes, it only works for moderate view angle changes. Particularly, the two

datasets used in our experiments contain mainly people walking from left to right (or vice versa) in front of a side-view camera. In a more realistic situation, the camera view angle could be arbitary. However, we note that a number existing approaches [9], [14] have been proposed recently to address the problem of view recognition and feature transformation across view. When the view angle is unknown and differs greatly from gallery and probe sets, these methods can be readily integrated with the method proposed in this paper.

VI. Conclusions

We have investigated the performance of state-of-the-art gait recognition approaches under a realistic experimental setup where no subject cooperation is required. Our experimental results suggest that the existing approaches are unable to cope with changes in gait covariate conditions in a gallery set, therefore are unsuitable for a truly uncooperative person identification challenges presented by the real world. To overcome this problem, we have proposed a novel gait recognition approach, which performs feature selection on each pair of gallery and probe gait sequences using Gait Entropy Image (GEnI), and seamlessly integrate feature selection with an Adaptive Component and Discriminant Analysis (ACDA) for fast recognition. Experiments are carried out to demonstrate that the proposed approach significantly outperforms the existing techniques. It is worth pointing out that the proposed feature selection method is designed mainly for mitigating the effect of changes in covariate conditions that affect gait feature extraction rather than gait itself. Our ongoing work includes further extending the proposed feature selection methods to deal with a wider range of covariate conditions that can affect gait itself including injury, mood, shoe-wear type, and elapsed time.

References

[1] Cynthia Archer and Todd K. Leen. Adaptive principal component analysis. Technical report, 2002.

[2] K. Bashir, T. Xiang, and S. Gong. Feature selection for gait recognition without subject cooperation. In *British Machine Vision Conference*, 2008.

[3] C. BenAbdelkader, R. Cutler, and L. Davis. Gait recognition using image self-similarity. *EURASIP J. Appl. Signal Process.*, 2004(1):572–585, 2004.

[4] N. Boulgouris and Z. X. Chi. Gait recognition using radon transform and linear discriminant analysis. *IEEE Transactions on Image Processing*, 16(3):731–740, 2007.

[5] T. Ding. A robust identification approach to gait recognition. In *CVPR08*, pages 1–8, 2008.

[6] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on PAMI*, 28(2):316–322, Feb 2006.

[7] P.S. Huang, C.J. Harris, and M.S. Nixon. Recognizing humans by gait via parametric canonical space. *Artificial Intelligence in Eng.*, 13:359–366, 1999.

[8] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag New York Inc., 2002.

[9] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *ICCV Workshop*, 2009.

[10] Dae Sung Lee, Jong Moon Park, and Peter A. Vanrolleghem. Adaptive multiscale principal component analysis for on-line monitoring of a sequencing batch reactor. *Journal of Biotechnology*, 116(2):195 – 210, 2005.

[11] Z. Liu and S. Sarkar. Improved gait recognition by gait dynamics normalization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(6):863–876, 2006.

[12] Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image Vision Comput.*, 25(6):817–832, 2007.

[13] H. Lu and P. Venetsanopoulos. A layered deformable model for gait analysis. In *7th Intl Conf, Automatic Face and Gesture Recognition*, pages 249–254, April 2006.

[14] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *ECCV (3)*, pages 151–163, 2006.

[15] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The humanID gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on PAMI*, 27(2):162–177, 2005.

[16] J. Shutler, M. Grant, M. Nixon, and J Carter. On a large sequence-based human gait database. In *4th International Conference on Recent Advances in Soft Computing*, pages 66–71, 2002.

[17] D. Tao, X. Li, X. Wu, and S.unhong Maybank. General tensor discriminant analysis and gabor features for gait recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(10):1700–1715, 2007.

[18] A. Veeraraghavan, A. Chowdhury, and R. Chellappa. Role of shape and kinematics in human movement analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 730–737, June 2004.

[19] G.V. Veres, L. Gordon, J.N. Carter, and M.S. Nixon. What image information is important in silhouette-based gait recognition? In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 776–782, June 2004.

[20] Y. Wang, S. Yu, Y. Wang, and T. Tan. Gait recognition based on fusion of multi-view gait sequences. In *ICB*, pages 605–611, 2006.

[21] D. Xu, S. Yan, D. Tao, S. Lin, and H. Zhang. Marginal fisher analysis and its variants for human gait recognition and content- based image retrieval. *IEEE Transactions on Image Processing*, 16(11):2811–2821, 2007.

[22] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. Zhang. Human gait recognition with matrix representation. *IEEE Trans. Circuits Syst. Video Techn.*, 16(7):896–903, 2006.

[23] S. Yu, D. Tan, K. Huang, and T. Tan. Reducing the effect of noise on human contour in gait recognition. In *ICB*, pages 338–346, 2007.

[24] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *18th International Conference on Pattern Recognition*, pages 441–444, 2006.

[25] S. Yu, L. Wang, W. Hu, and T. Tan. Gait analysis for human identification in frequency domain. In *Image and Graphics*, pages 282–285, Dec 2004.

[26] R. Zhang, C. Vogler, and D. Metaxas. Human gait recognition at sagittal plane. *Image Vision Comput.*, 25(3):321–330, 2007.

[27] G. Zhao, G. Liu, H. Li, and Pietikainen. 3D gait recognition using multiple cameras. In *7th Intl Conf, Automatic Face and Gesture Recognition*, pages 529–534, April 2006.